Arto Laitinen

Voiko kone ajatella?
Tieteellinen maailmankuva ja arkifenomenologia

Voiko kone ajatella? Voivatko koneet olla itsetietoisia tai autonomisia? Voivatko ne olla ihmisten kanssa vastavuoroisissa tunnustussuhteissa? Nämä näennäisen empiiriset kysymykset sekä pakottavat selventämään kysymyksissä esiintyviä käsitteitä että paljastavat näkökulmaeroja tieteellisen ja arkifenomenologisen maailmankuvan välillä. Wilfrid Sellarsin mukaan filosofian tehtävä on pyrkiä pelastamaan kaikki pelastamisen arvoinen molempien lähestymistapojen paljastamista maailmoista – tai pikemminkin yhden ja saman maailman aspekteista.

Filosofian tehtävä, tai ainakin yksi niistä, on yrittää sovittaa yhteen näkökulmia, jotka paljastavat maailmasta jotain olennaista[1]. Jos näkökulmia ei voi sovittaa yhteen, filosofia voi pyrkiä ottamaan kantaa siihen, miten näkökulmien perustavanlaatuisiin eroihin tulisi suhtautua. Tieteellisen näkökulman ja tässä ”arkifenomenologiseksi” kutsutun lähestymistavan välillä vallitsee tällainen perustavanlaatuinen jännite, joka käy ilmi myös keskustelussa koneiden ja ihmisten eroista. Aihepiirin keskeiset empiiriset kysymykset, kuten ”voiko kone ajatella?”, ymmärretään tieteellisestä ja arkifenomenologisesta näkökulmasta toisistaan poikkeavin tavoin, joista voi kartoittaa näkyviin systemaattisia merkityseroja. Merkitysero on systemaattinen silloin, kun samaa sanastoa (kuten ’ajattelu’, ’tietoisuus’) käytetään radikaalisti ja systemaattisesti eri merkityksissä ja kun eri osapuolet ymmärtävät erimielisyyden eri tavoin, kumpikin itselleen suotuisasti. Tieteellisestä näkökulmasta on houkuttelevaa erottaa koneiden ja ihmisten todellinen ajattelu ihmisajattelun erityislaatuisuutta paisuttelevasta myyttisestä Ajattelusta (ajattelusta isolla A:lla), joka on oikeastaan filosofinen fiktio. Arkifenomenologian näkökulmasta puolestaan on houkuttelevaa erottaa todellinen inhimillinen ajattelu pelkästä näennäisestä ”ajattelusta” (ajattelusta vain lainausmerkeissä), joka ei kirjaimellisesti ottaen ole ajattelua lainkaan. Samalla voidaan kieltää oletus, että arkifenomenologitkaan haikailisivat liioitellun Ajattelun perään.

Alan Turing kirjoitti vuonna 1950 seuraavasti:

> ”Kysymystä ’voivatko koneet ajatella’ pidän liian merkityksettömänä keskusteltavaksi. Uskon kuitenkin, että vuosisadan lopulla sanojen käyttö ja yleinen valistunut mielipide ovat kylliksi muuttuneet, jotta koneiden ajattelusta kyetään puhumaan odottamatta vastaväitteitä.”[2]

Näyttää siltä, että Turingin ennustus osui oikeaan. Sanojen käyttö on muuttunut niin paljon, että koneiden ajattelusta kyetään puhumaan. On kuitenkin syntynyt terminologista sekaannusta, jota voi selkeyttää suhteuttamalla eri määritelmät niiden taustalla oleviin lähestymistapoihin. Erimielisyyden paikantamista hankaloittaa se, että sanoja käytetään eri merkityksissä. Saattaa vaikuttaa, että erimielisyyttä ei olekaan: kaikki voivat olla samaa mieltä, että koneet eivät Ajattele mutta ”ajattelevat”. Kun kysymys tarkentuu muotoon ”osaavatko koneet nykyisin ajatella relevantisti samassa mielessä kuin ihmiset?”, lähestymistapojen erimielisyys tulee paremmin esiin.

Tieteellinen maailmankuva ja arkifenomenologia

Wilfrid Sellars erotti toisistaan tieteellisen maailmankuvan ja ”ilmikuvan” (*scientific and manifest image*)[3]. Tieteellinen maailmankuva sisältää vain ne entiteetit, joita paras tiede olettaa teorioissaan. Ilmikuva puolestaan kattaa ihmisten maailman kaikessa rikkaudessaan riippumatta siitä, ovatko kyseisen maailman sisältämät asiat myös tieteellisten teorioiden ainesta. Sellarsin mukaan ilmikuva voi kehittyä (esimerkiksi jotkin ennakkoluulot voivat osoittautua perusteettomiksi); kaikki kehitys ei ole tieteellisen tarkastelun ansiota. Ilmikuvaan luottavaa lähestymistapaa voidaan kutsua arkifenomenologiseksi. Toisin kuin fenomenologisen filosofian perinne (esimerkiksi Husserl), arkifenomenologia ei pyri sulkeistamaan luonnollista asennetta ja siinä tehtäviä oletuksia, vaan arkiset oletukset ovat lähtökohtaisesti osa tätä lähestymistapaa. Toisaalta kaikkiin ”maalaisjärjen” tai ”kansan näkemysten” oletuksiin ei dogmaattisesti tarvitse sitoutua, vaan ne voivat kokemuksen ja historian myötä osoittautua ongelmallisiksi osallistujien oman perspektiivin sisällä.

Vastaavan erottelun ovat tehneet monet filosofit. Charles Taylor erottaa toisistaan elämismaailmaan ”kiinnittyvän” (*engaged*) ja siitä ”irtautuvan” (*disengaged*) lähestymistavan. Thomas Nagelille ”näkökulma eimistään” (*view from nowhere*) ja John McDowelille ”si-

Ilari Hautamäki, *Explorer* (2018), akryyli ja öljy kankaalle, 180 × 140 cm

vuttainen näkymä" (*sideways-on picture*) eroavat elävän osallistujan näkökulmasta.[4]

Sellars puolustaa ajatusta, että filosofian päämääränä on "stereoskooppinen" kuva, jossa tieteellinen ja ilmikuva yhdistyvät. Sellars itse päätyy painottamaan tieteellistä kuvaa todellisuuden mittana. Elämismaailman ensisijaisuutta ovat puolestaan puolustaneet esimerkiksi edellä mainitut McDowell ja Taylor. Heidän mukaansa nykyisin vallitsevat tavat kysyä ja vastata filosofisiin kysymyksiin kietoutuvat poikkeuksetta yhtäältä yksipuolisen "objektivististen" (esimerkiksi fysikalismi) ja toisaalta yksipuolisen "subjektivististen" intuitioiden vahvaan asemaan modernissa filosofiassa (ja kulttuurissa laajemminkin).[5] He pyrkivät artikuloimaan niitä filosofisia (McDowell) ja käytännöllisiä (Taylor) motivaatioita, jotka johtavat kannattamaan fysikalistisia, subjektivistisia tai nämä yhdistäviä dualistisia teorioita. Heistä on tärkeätä selvittää, miksi nämä taustakuvat ja -motiivit "pitävät vankeinaan" sellaisiakin, jotka pyrkivät niistä "vapautumaan".[6] Elämismaailmallisen perspektiivin puolustaminen on tavallaan "ensimmäistä filosofiaa", jonka tarkoitus on poistaa vääränlaiset jännitteet filosofisten erityiskysymysten käsittelystä, vapauttaa meidät vääristä ennakko-oletuksista terapeuttisesti.

Modernin luonnontieteen edistysaskeleet luonnontapahtumien selittämisessä ja luonnonhallinnan mahdollistamisessa ovat toki Taylorin ja McDowellin mukaan kiistattomia. Moderni luonnontiede asettaa kuitenkin filosofisen tai maailmankatsomuksellisen haasteen: kuinka modernin luonnontieteen "taiasta vapautunut" maailmankuva voidaan sovittaa yhteen ihmisten kokeman todellisuuden kanssa?

Yritys vastata tähän kysymykseen on johtanut Taylorin ja McDowellin mukaan filosofisesta ylilyönnistä toiseen. *Objektivistinen* ylilyönti on esimerkiksi radikaali eliminativistinen fysikalismi, jonka mukaan modernin luonnontieteen ontologia on kaiken todellisuuden mitta: mikään sellainen, mitä ei löydy luonnontieteiden ontologisesta kehikosta, ei ole todellista. Ei ole olemassa pöytiä, on vain hiukkasia pöytämäisessä järjestyksessä; ei ole olemassa kokemuksia, on vain aivotapahtumia; ei ole olemassa normatiivisia tai institutionaalisia tosiasioita, on vain aivokemiaa, joka synnyttää vaikutelmia sellaisista. Aiheellinen pelko, että tämä johtaisi subjektien ja kaiken subjektiivisen todellisuuden kiistämiseen, voi motivoida siirtymää toiseen ääripäähän, *subjektiiviseen* idealismiin, jonka mukaan vain subjektit mielen sisältöineen ja *sense datoineen* ovat todellisia. Ei ole olemassa pöytiä tai instituutioita, on vain ihmismielen sisäisiä ajatuksia niistä. Subjektiivinen idealismi liitetään usein Berkeleyn nimeen, joskin hänen ajattelussaan Jumalan mielellä on keskeinen osansa, joten hän ei redusoi todellisuutta ainoastaan ihmismielen sisältöihin.

Objektivismiin on luontevaa liittää subjektiivinen idealismi suhtautumisessa kaikkeen muuhun kuin luonnontieteen tutkimuskohteisiin: kokonaisia pöytiä (ei vain hiukkasia), instituutioita, kokemuksia ja normatiivisuutta on vain subjektien mielissä. *Dualismia*, joka tuttuimmin yhdistetään Descartesiin, motivoikin näkemys, että objektivismi on oikeassa yhtäällä (ulottuvaisen ulkomaailman osalta) ja subjektivismi toisaalla (subjektien osalta). Modernien dualismien mukaan fysikalistisesti käsitetty luonto on riippumattomasti olemassa, mutta subjektit ovat jollain tavoin sen ulkopuolella niin ikään todellisia. Kartesiolainen ajatus luonnosta erillisestä "ajattelevasta substanssista" sekä kantilainen erottelu "vapauden kausaalisuuden" ja "luonnon kausaalisuuden" välillä puolustavat subjektia dualistisesti, erottamalla sen fysikalistisesti käsitetystä luonnosta. Kenties voidaan kuitenkin ajatella, että tämäntyyppinen dualismi vain yhdistää molemmat ylilyönnit ja on eräänlainen ojan ja allikon yhdistelmä.[7]

McDowellin ja Taylorin keskeinen sanoma on, että näiden ylilyöntien taustalla on sama perustava virhe: irrottautuminen elämismaailmallisesta perspektiivistä. Heidän mukaansa jaettu elämismaailmallinen todellisuus ei redusoidu sen enempää fysikalistiseen luontoon kuin subjektien mieliinkään. Filosofiassaan he pyrkivät terapeuttisesti vapauttamaan meidät rajoittuneista näkökannoista. Vaikka puhtaat objektivistiset ja subjektivistiset teoriat voivat olla harvinaisia, moniin kantoihin sisältyy tendenssejä ajatella niiden mukaisesti: esimerkiksi todellisuuden mittana voidaan pitää riippumattomuutta mielestä tai normatiivisuuden voidaan ajatella palautuvan subjektien käsityksiin normatiivisuudesta.

Elämismaailman ensisijaisuutta puolustavan leirin sisällä on kuitenkin varsin erilaisia käsityksiä siitä, kuinka vahvasti filosofia voi puolustaa synteettistä tai stereoskooppista käsitystä todellisuuden ykseydestä, jossa sekä luonnolla että subjekteilla ja elämismaailmalla on oma asemansa ja jonka paljastamisessa tieteellisellä maailmankuvalla on oma tehtävänsä.[8] Siinä missä Sellarsin stereoskooppinen kuva painottaa tieteellistä lähestymistapaa ja Taylorin ja McDowellin lähestymistapa arkifenomenologista lähestymistapaa, esimerkiksi Hegelin ja Ricœurin filosofiset systeemit pyrkivät kunnianhimoisemmin ottamaan molemmat kuvat huomioon.[9] Heihin verrattuna erityisesti McDowell tyytyy pelkkään filosofian "terapeuttiseen" tehtävään ja pidättäytyy rakentavien väitteiden esittämisestä.

Hegelin systeemissä kokevat subjektit ja heidän perspektiivinsä maailmaan ovat todellisia maailmankaikkeuden rakenneosasia. Kun subjektit kokevat, ymmärtävät ja filosofisella järjellä tavoittavat maailman käsitteellisen perusrakenteen, toteutuu tavallaan maailman itsetietoisuus. Hegelin filosofian kunnianhimoinen tavoite onkin, että maailman kategorinen perusrakenne hahmotetaan olennaisilta osiltaan. Näin filosofinen tutkimus on todellisuuden itsetietoisuuden asialla lähestyessään tätä tavoitetta.[10]

Paul Ricœurin filosofiassa oletetaan, että tämä ei koskaan onnistu: eri näkökulmien synteesi jää aina keskeneräiseksi ja näkökulmat yhteismitattomiksi. Silti eri näkökulmat, kuten modernin luonnontieteen näkökulma tai toimijuuden ohittava strukturalistinen teoria kielestä ja yhteiskunnasta (kuten Saussurella tai Lévi-

”Maailmoja on vain yksi, joten olisi helppoa ajatella eri teorioiden puhuvan tyystin eri asioista.”

Straussilla), lisäävät olennaisesti ymmärrystämme ja järkevää käsitystämme maailmasta, vaikka ne Ricœurin mukaan ovatkin arkifenomenologian näkökulmasta väärässä. Maailmoja on vain yksi, joten olisi helppoa ajatella eri teorioiden puhuvan tyystin eri asioista: ne esittävät ristiriitaisia väitteitä todellisuuden luonteesta, joskin ristiriitaa on hankala paikantaa näkökulmien yhteismitattomuuden vuoksi. Ricœur puolustaa ”epäsuoraa hermeneutiikkaa” ja kiertotietä, jonka nämä erilaiset yhteismitattomat näkökulmat tarjoavat. Suuren synteesin sijaan tavoitteena voi olla fragmentoitu kokonaiskuva. Ricœurlaisittain ajatellen suora arkifenomenologinen lähestymistapa on yksinään köyhempi kuin pluralistinen maailmankuva.[11]

Tieteellistä ja arkifenomenologista näkökulmaa voidaan siis koettaa yhdistää monella tapaa – joko alistaen toisen toiselle (kuten Sellars, Taylor tai McDowell) tai pyrkien tasapainoiseen synteesiin (kuten Hegel) tai mosaiikkimaisempaan pluralistiseen kokonaiskuvaan (kuten Ricœur). Miten yhdistäminen tehdäänkin ja mitä näillä binaariparin vastakkaisilla päillä tarkkaan ottaen tarkoitetaankin, kaksijakoisuus voi joka tapauksessa tuoda lisävalaistusta lukuisiin eri filosofian aihepiireihin ihmispersoonien tai mielen ontologiasta luonnon, yhteiskunnan, arkiesineiden tai moraalin metafysiikkaan. Kysymys koneiden ja ihmisten ajattelun ja tietoisuuden eroista ja yhtäläisyyksistä kuuluu myös niiden joukkoon.

Ihmisten ja tekoälyn ero arkifenomenologian näkökulmasta

Eräs kiertotie myös ihmisten ymmärtämiseksi voi olla ihmisen vertaaminen sellaisiin kognitiivisiin laitteisiin, joilla ei oletettavasti ole kokemuksia, tuntoisuutta, aitoa ymmärrystä, mielikuvitusta, kykyä normien rikkomiseen ja vapaan tahdon päätöksiin ja niin edelleen. Robotit, tekoälyjärjestelmät tai autonomiset älykoneet ovat tällainen vertailukohta. Arkifenomenologian näkökulmasta niiltä puuttuvat oleelliset inhimilliset kyvyt. Toisaalta voidaan löytää tieteellinen lähestymistapa, joka kyseenalaistaa ihmisten ja koneiden perustavanlaatuisen eron.

Tällaisen lähestymistavan tarjoaa kyberneettisen systeemiteorian pohjalle rakentuva kognitiotiede. Koska kognitiotiede on tieteenala, joka on parempi identifioida pikemminkin sen kysymien kysymysten kuin sen tarjoamien vastausten avulla, käytän tietynlaisesta sisällöllisestä kannasta tässä nimitystä ”kyberneettinen” lähestymistapa. Tarkoitus ei ole kuitenkaan sitoutua ”vanhaan kybernetiikkaan”, joka edelsi kognitiotiedettä, sen enempää kuin ”arkifenomenologia” sitoutuu fenomenologian klassikoiden kantoihin. Molemmat lähestymistavat ovat jossain määrin ideaalityyppisiä konstruktioita, joiden avulla pyrin havainnollistamaan todellisia näkemyksiä todellisissa debateissa.[12]

Kognitiotieteellisessä kuvassa (osana tieteellistä kuvaa) ihmisen ajatteluun kuuluu sekä tiedostamattomia informaation käsittelyprosesseja (tyypin 1 prosessit) että tiedostettuja prosesseja (tyypin 2 prosessit). Tietoinen ajattelu on rationaalista ja sensitiivistä perusteille, tiedostamaton on automaattista ja assosiatiivista. Tiedostamatonkin ajattelu tuottaa silti päätelmiä ja käyttäytymistä, jotka ovat rationaalisia ja joita ihmiset rationalisoivat perusteilla, mikäli heiltä kysytään toimintansa syitä jälkikäteen. Arkifenomenologisesta näkökulmasta voidaan hyväksyä, että ihmisillä on sekä tyypin 1 että tyypin 2 prosesseja, mutta koneilla vaikuttaisi voivan olla vain tyypin 1 prosesseja. Tietynlaisesta tieteellisestä näkökulmasta tarkasteltuna koneillakin voi kuitenkin olla myös jossain mielessä tyypin 2 prosesseja, kuten myöhemmin tuon esiin.

Arkifenomenologian mukaan elävillä olennoilla on elävä, eletty keho (*Leib*; *lived body*), joka ei palaudu pelkäksi fysikaaliseksi kappaleeksi tai ”ruumiiksi” (*Körper*; *physical body*). Niillä on myös ”elämä”, joka prosessina

”Ihmispersoonilla on lukuisia piirteitä, jotka erottavat ne esimerkiksi kivistä tai mekaanisista koneista.”

eroaa elottomien olentojen historiasta. Ihmispersoonilla on lisäksi lukuisia piirteitä, jotka erottavat ne esimerkiksi kivistä tai mekaanisista koneista. Nämä piirteet voivat olla toisistaan riippuvaisia, kuten kyky tuntea kipuja, aistimuksia ja tunteita, kyky haluta, tahtoa ja aikoa, kyky muodostaa uskomuksia, sitoumuksia ja ajatuksia, kyky välittää asioista ja toimia sekä kyky järkeillä ja kommunikoida toisten järkeilijöiden kanssa. Persoonat ovat tietoisia, itsetietoisia, kykeneviä itsemääräämiseen, vastuulliseen toimintaan ja vastuun ottamiseen. Persoonilla on myös moraalinen status, arvokkuus tai ”ihmisarvo”, ja ne kykenevät vastavuoroiseen kunnioittamiseen, jossa ne tunnustavat toistensa moraalisen statuksen.

Arkifenomenologian näkökulmasta nykyrobotit voivat kenties simuloida näitä piirteitä, mutta kirjaimellisesti niillä ei näitä piirteitä ole[13]. Ne eivät voi aikoa, välittää tai uskoa, mutta kenties ne voivat ”aikoa”, ”välittää” tai ”uskoa”. Arkifenomenologian näkökulmasta ei tarvitse ottaa kantaa kysymykseen siitä, mitä tietokoneet eivät koskaan tule osaamaan[14]. Sen sijaan kantana on, että jos koneet tulevat joskus omaksumaan olennaisesti ihmistä vastaavat kyvyt, on tapahtunut huima *kvalitatiivinen loikka.* Tuon loikan ottamiseen asti koneet vain simuloivat näitä piirteitä: kykyä ajatella, tietoisuutta, itsetietoisuutta, tuntoisuutta ja niin edelleen. On siis esitettävissä kokonainen luettelo piirteitä, joita ihmisillä on mutta koneilla ei ole.

Koska (tai jos) nämä piirteet liittyvät toisiinsa (kuten elävä keho, kyky toimia, kyky ajatella, kyky ottaa toiminnan ja ajattelun perusteet huomioon ja niin edelleen), roboteilla tai muillakaan tekoälyjärjestelmillä ei voi olla yhtä tämän listan piirteistä ennen kuin sillä on muutkin[15]. Roboteilla ei esimerkiksi voi olla haluja tai aikomuksia ilman uskomuksia: halut tai aikomukset eivät voi johtaa toimijaa kohti uutta maailmantilaa ilman uskomuksia nykyisestä maailmantilasta. Käytännöllisiä perusteita (”se aiheuttaisi kuoleman, joten vältä sitä”) ei voi ymmärtää ilman kykyä välittää asioista (”ennenaikainen kuolema on negatiivinen asia”), mikä puolestaan edellyttänee kykyä kokea tunteita ja tuntemuksia, niin sanottua sentienssiä. Jotta yksi listan piirteistä todella voidaan omaksua, täytyy jossain määrin kehittää myös joitain muita listan piirteitä. Niitä ei voi saada täysimääräisessä muodossaan yksi kerrallaan. On varmastikin monimutkainen empiirinen kysymys, mitkä kyvyt riippuvat mistäkin toisista kyvyistä ja millä tavoin, mutta arkifenomenologian näkökulmasta näiden piirteiden holistinen riippuvuus toisistaan vaikuttaa uskottavalta oletukselta. Ideaalityyppisesti voidaan olettaa yleinen keskinäinen riippuvuus: kaikki listan asiat riippuvat toisistaan. (Toki on mahdollista muotoilla myös atomistinen arkifenomenologinen kanta, jossa kykyjen riippuvuutta toisistaan ei oleteta.) Joka tapauksessa olennaista on, että arkifenomenologia käsittää kaikki listan termit omalla tavallaan ja kyberneettinen näkökulma omalla tavallaan.

Arkinäkökulmasta koneiden ja ihmisten ero vaikuttaa varsin selvältä, kuten seuraavissa koneiden ja ihmisten eroa käsittelevien artikkeleiden nimissä: ”Ei minuutta eikä kykyä pahantahtoisuuteen”, ”Koneet eivät ajattele ikuisuuskysymyksiä”, ”Itsetietoinen tekoäly? Ei tuhanteen vuoteen!”, ”Koneet eivät piittaa suhteista”[16]. Alan Turing tarkasteli (edustamatta sitä itse) kantaa, jonka mukaan koneet eivät ikinä kykene seuraaviin asioihin:

> ”Olemaan huomaavaisia, neuvokkaita, kauniita, ystävällisiä, aloitteellisia, huumorintajuisia, kykeneviä erottamaan oikean ja väärän, tekemään virheitä, rakastumaan, nauttimaan mansikoista ja kermasta, saada jotakuta rakastumaan itseensä, oppimaan kokemuksistaan, käyttämään sanoja oikein, olemaan omien ajatustensa kohde, osoittamaan yhtä paljon moninaisuutta käytöksessään kuin ihminen, tekemään jotain oikeasti uutta.”[17]

Kuten johdannossa jo tuli esiin, hän kirjoitti myös, että kysymys koneiden ajattelukyvystä on sellaisenaan merki-

”Tieteellinen näkökulma suhtautuu kansanpsykologian oletuksiin kuin alkeelliseen tieteelliseen teoriaan.”

tyksetön. Turing kuitenkin mainitsi mahdollisuuden, että vuosisadan vaihteeseen mennessä sanojen käyttö ja yleistieto saattaisivat muuttua niin, että kysymyksestä tulee mielekäs ja koneiden ajattelusta voi puhua vakavasti. Sanassa ’ajattelu’ voikin katsoa tapahtuneen Turingin mainitseman merkityssiirtymän. Arkifenomenologian rinnalle on kehittynyt tarkastelutapa, jonka mukaan koneet kirjaimellisesti ajattelevat. Samanlaisen merkityssiirtymän voinee ajatella tapahtuneen myös monissa muissa keskeisissä käsitteissä, kuten ’tietoisuudessa’, ’itsetietoisuudessa’ ja niin edelleen – *jossain mielessä* kaikki nämä käsitteet soveltuvat myös koneisiin. Arkifenomenologiankin näkökulmasta kaikille edellä listatuille piirteille voi antaa redusoidun, köyhtyneen merkityksen, jossa koneet kenties ”ajattelevat” tai ”ovat tietoisia” (lainausmerkeissä). Jossain merkityksessä koneilla on siis myös systeemin 2 prosesseja. On täten syntynyt sekava tilanne, jossa ei aina huomata selventää, puhutaanko ajattelusta tai tietoisuudesta vaativammassa vai köyhtyneemmässä merkityksessä – eikä havaita, että eri osapuolien näkemysero voikin palautua pelkkään kiistaan sanojen käytöstä.

Arkifenomenologian näkökulmasta voimme erottaa ajattelun ja robottien ”ajattelun”, ihmisten toiminnan ja robottien ”toiminnan” ja niin edelleen. Arkifenomenologian kannalta katsottuna kilpailevat lähestymistavat analysoivat jossain määrin pelkkää korviketta, eivät aitoa asiaa.

Ihmisten ja tekoälyn ero vastakkaisesta tieteellisestä näkökulmasta

Arkifenomenologian kanssa kilpaileva tieteellinen näkökulma irtautuu elämismaailmasta ja suhtautuu epäluuloisesti arkisen ”kansanpsykologian” olettamuksiin. Tämän näkemyksen mukaan ei tarvita laadullista hyppyä ennen kuin robotit voivat olla samalla viivalla ihmisten kanssa. Ihmiset ja robotit ovat jo yhtä lailla informaationkäsittelyjärjestelmiä, ja monelta osin tietokoneohjelmat, tekoälyjärjestelmät tai robotit suoriutuvat tiedonkäsittelystä ihmisiä paremmin. Sean Carrollin esseen otsikko ilmaisee tämän napakasti: ”Olemme kaikki ajattelevia koneita”[18].

Tämä tieteellinen näkökulma suhtautuu kansanpsykologian oletuksiin kuin alkeelliseen tieteelliseen teoriaan. Jos kansanpsykologian mukaan tarvitaan fenomenaalisen tietoisuuden piirteet selittämään älykästä käyttäytymistä, onko teoria oikeassa? Eikö yhtä hyviä tai parempia kilpailevia selityksiä voida tarjota olettamatta fenomenaalista tietoisuutta tai mieltä?

Tämä suhtautumistapa näkyy esimerkiksi Daniel Dennettin vastauksissa John Searlelle keskustelussa tietoisuuden tutkimuksesta:

> ”John Searle ja minä olemme syvästi erimielisiä siitä, kuinka tutkia mieltä. Searlelle asiat ovat yksinkertaisia. Meillä kaikilla on ajan myötä hyväksi havaittuja perusintuitioita tietoisuudesta, ja jokainen teoria, joka haastaa niitä, on kertakaikkisen tolkuton. Minä puolestani ajattelen, että tietoisuuden ongelma säilyy mysteerinä, kunnes löydämme jonkin tuollaisen selvääkin selvemmän intuition ja osoitamme, että ensivaikutelmasta huolimatta se on väärässä!”[19]

Kun syvään juurtuneista intuitioista, jotka ovat tavallaan hypoteeseja siinä missä tieteellisetkin hypoteesit, on päästy eroon, voidaan tarkastella asiaa objektiivisesti. Objektiivisesti tarkasteltaessa alkaakin vaikuttaa siltä, että tekoälyjärjestelmä tai robotti voi *jossain mielessä* olla esimerkiksi tietoinen: koneella on representaatioita ympäristöstään, ja se kykenee operoimaan näiden representaatioiden avulla. Jossain mielessä myös intentiot ovat mahdollisia, sillä representaatiot lopputiloista, joihin pyritään, voivat ohjata robotin toimintaa. Roboteilla voi olla myös representaatioita omista tiloistaan, joten ainakin jonkinlainen itsetietoisuus vaikuttaa mahdolliselta. Näin siis myös koneilla, ei vain ihmisillä, on sekä systeemin 1 että systeemin 2 prosesseja.

Dennett esimerkiksi kirjoittaa, että robottien intentionaalisuus on ”yhtä todellista kuin mikä tahansa intentionaalisuus planeetallamme”[20]. Keinotekoisuus ei intentionaalisuutta pahenna. Dennettin mukaan nykyihmisten ja nykytekoälyn välillä ei ole laadullista hyppäystä. Voidaan kirjaimellisesti puhua vahvasta tekoälystä (joka todella ymmärtää, ei vain ”ymmärrä”, ja joka on intentionaalinen, ei vain ”intentionaalinen”), koska ihmistenkään intentionaalisuuden ja ymmärryksen kohdalla ei pidä luottaa arkipsykologian syvään juurtuneisiin ”intuitioihin”.

Tekoälysysteemit siis tavallaan osaavat erottaa itsensä ja ympäristönsä (eli niillä on jonkinlainen ”maailma”). Kun robotissa on virta päällä, sillä on jossain mielessä toimiva keho funktionaalisine ominaisuuksineen, jotka eivät palaudu pelkästään fysikaalisiin tiloihin (jotka tekevät mahdollisiksi nämä funktionaaliset tilat). Robotti voi tietää raajojensa sijainnin, se voi käyttää niitä saadakseen maailmassa aikaan muutoksia ja niin edelleen. Kenties se ei – isolla kirjaimella – Tiedä tai Toimi (jos Fenomenaalinen Tietoisuus on näiden edellytys), mutta entä sitten? Arkipsykologian oletukset Mielen tiloista (isoilla kirjaimilla) on syytä joka tapauksessa kyseenalaistaa. Miksi kaipailla Tietoisuutta, Itsetietoisuutta, Järkeä, Moraalia, ja miksi ei sen sijaan tyytyä sellaiseen tietoisuuteen, johon koneetkin kykenevät? Dennett pitää itse asiassa hyödyllisenä keskittyä sellaisiin kykyihin, joita roboteillakin on, jotta saadaan ”mysteerikot savustettua ulos”[21].

Siinä missä arkifenomenologia erotti ajattelun (johon robotit eivät kykene) ja pelkän simulaation eli ”ajattelun” (johon robotitkin kykenevät), kognitiotieteellinen tai kyberneettinen näkökulma erottaa ajattelun (johon robotitkin kykenevät) ja Ajattelun (jonka määrittelyssä nojataan ”kansanpsykologisiin” syvään juurtuneisiin intuitioihin, jotka tulee haastaa). Jälkimmäisestä näkökulmasta ihmisten ja robottien välille syntyy ero, mutta se on kognitiotieteen kyberneettiseltä kannalta yhdentekevää, koska koko näkökulmasta on syytä luopua. Dennett käyttää Sellarsin erottelua tieteellisen ja ilmikuvan välillä ja jättää (perinteisen) fenomenologian tehtäväksi ilmikuvan jäsennysten artikuloinnin, kun taas kognitiotiede voi hänen mielestään paljastaa arkinäkökulmasta epäintuitiivisen ontologian, joka kuitenkin selittää ihmisten kyvyt paremmin kuin ilmikuvan arkiontologia[22]. Tieteen tehtävä on ratkaista, missä määrin arkinen ilmikuva on virheellinen.

Arkifenomenologia kuitenkin kiistää esittävänsä liioitellun käsityksen Ajattelusta, vaikka siltä kyberneettisestä näkökulmasta vaikuttaakin (kukapa myöntäisi liioittelevansa). Omasta näkökulmastaan arkifenomenologia pyrkii erottamaan ajattelun ja ”ajattelun”. Toisaalta kyberneettinen näkökulma puolestaan kiistää, että käsittelisi pelkkää ”ajattelua” (kukapa myöntäisi näkemyksensä olevan köyhdytetty), vaikka siltä arkifenomenologisesta näkökulmasta näyttääkin.

Kohti stereoskooppista kuvaa?

Alan Turing siis ennusti, että vuosituhannen vaihteessa on mielekästä kysyä, voivatko koneet ajatella. Olen käyttänyt Dennettiä esimerkkinä kannasta, jonka mukaan tämä tosiaan on mielekästä ja vastauskin on myönteinen – vahvaa tekoälyä eli tekoälyä, joka kykenee kirjaimellisesti ajattelemaan, on Dennettin mielestä jo olemassa. Turingin mukaan kysymyksen mielekkyys kertoo siitä, että termien merkitykset ovat muuttuneet. (Myös sanan ’kone’ merkitykset ovat varmastikin eläneet.) Toisaalta olen käyttänyt Searlea esimerkkinä kannasta, jonka mukaan vastaus on kielteinen.

Arkifenomenologia ja kognitiotieteellinen näkökulma käsittävät näkökulmiensa eron eri tavalla. Informaationkäsittely, joka arkifenomenologiselta kannalta on pelkkää ”ajattelua”, on kognitiotieteellisestä näkökulmasta niin täysimääräistä ajattelua kuin sopii olettaakin. Ja kognitiotieteellisestä näkökulmasta paisutelluilta vaikuttavat käsitykset Ajattelusta, Tietoisuudesta, Itsetietoisuudesta (ja niin edelleen) ovat arkifenomenologisesta näkökulmasta osuva kuvaus ihmispersoonien ajattelusta, tietoisuudesta ja itsetietoisuudesta – siis laadullisesti enemmän kuin pelkkä ”ajattelu”, ”tietoisuus” tai ”itsetietoisuus”.

Jos halutaan välttää ohipuhuminen ja pelkästään semanttinen kiistely, on syytä huomata, että molemmat kiistan osapuolet omivat tavalliset termit (ajattelu, tietoisuus, itsetietoisuus) itselleen ja puhuvat pejoratiivisesti siitä, mistä heidän vastustajansa vaikuttaa puhuvan (”ajattelu” tai Ajattelu, ”tietoisuus” tai Tietoisuus, ”itsetietoisuus” tai Itsetietoisuus). Vaikka yhden kannan mukaan robotit eivät kykene ajattelemaan ja toisen kannan mukaan ne kykenevät ajattelemaan, voidaan huomata, että jos osapuolet suostuisivat käyttämään vastustajiensa termejä, löydettäisiin yksimielisyys siitä, että robotit eivät kykene Ajattelemaan mutta ne kykenevät ”ajattelemaan”. Erimielisyys ei siis ole pelkästään empiirinen vaan käsitteellinen, ja se heijastelee laajempaa eroa arkifenomenologisen ja kognitiotieteellisen lähestymistavan välillä. Aito erimielisyys kuitenkin koskee sitä, kykenevätkö nykyrobotit ajattelemaan relevantisti samassa mielessä kuin ihmiset.

Voiko näkökulmien eron kuroa umpeen? Kuten todettua, Sellars ja Dennett päätyvät stereoskooppisen kuvan luonnehdinnoissaan puolustamaan tieteellisen maailmankuvan ensisijaisuutta ja esimerkiksi Taylor ja McDowell elämismaailman ensisijaisuutta. Askel kohti tasapainoisempaa ”stereoskooppista kuvaa” otetaan, kun huomataan, että näillä lähestymistavoilla on omat vahvuutensa eri ”tiedonintresseissä”: toisen mukaan todellisena pitämisen mitta on hyödyllisyys tieteelliselle selittämiselle, toisen mukaan taas ensisijaista on välttämättömyys elämismaailmallisiin käytäntöihin osallistumiselle tai elämismaailmallisten käytäntöjen kritiikille.[23]

Dennett ja muut tieteellisen näkökulman edustajat lienevät oikeassa siinä, että arkipsykologian oletuksia ei tarvita kaikessa *selittämisessä.* ”Kansanpsykologian” oletukset eivät varmastikaan ole kilpailukykyisiä hypoteeseja selityksiä haettaessa. Niitä saatetaan kuitenkin tarvita elämismaailmaan osallistumisessa, elämisessä. Eletyt kokemukset eivät lähtökohtaisesti ole hypoteeseja, ja monet ilmiöt kokemuksista arkiesineisiin paljastuvat ennen kaikkea osallistujien arkifenomenologisesta perspektii-

Ilari Hautamäki, *Vapors* (2017), akryyli ja öljy kankaalle, 140 x 110 cm

vistä katsottuina. Saattaa siis olla, että kahden eri tiedonintressin (tieteellisen ja osallistuvan) vuoksi tarvitaan stereoskooppista kuvaa – Sellarsin tavoite sovittaa yhteen kaksi eri näkökulmaa on yhä perusteltu.

Joidenkin ilmiöiden tavoittamiseen arkifenomenologian näkökulma on välttämätön. Searle argumentoi, että näin on ainakin fenomenaalisen tietoisuuden kohdalla. Searlen mukaan tietoisuutta ei voi käsittää muuten kuin ensimmäisen persoonan näkökulmasta. Hän pyytää lukijoitaan nipistämään itseään kämmenselästä tuottaakseen pienen kiputuntemuksen. Kiputuntemukseen sisältyy tietynlainen laadullinen tuntemus, ja tällaiset laadulliset tuntemukset ovat keskeisiä sekä valve- että unitiloissamme. Tuntemuksilla on ensimmäisen persoonan subjektiivinen ontologinen status: ne ovat olemassa vain jonkin subjektin kokemina. Searlen mukaan Dennett kieltää niiden olemassaolon, joten hän ei näe eroa monimutkaisten zombien ja ihmisten välillä. (Filosofiassa zombeilla viitataan yleensä olioihin, jotka toimivat, mutta joilla ei ole lainkaan ensimmäisen persoonan kokemushorisonttia, tajunnanvirtaa, tietoisuutta.)[24]

Searlen mukaan (fenomenaalisessa) tietoisuudessa koko ilmiö on riippuvainen sitä koskevien kokemusten tai tuntemusten olemassaolosta. Vaikka neurobiologia voi osoittaa jotkin intuitiot vääriksi (kuten Searlen mukaan sen, että käsikipu sijaitsisi käden sijaintikohdassa), tietoisten tilojen olemassaolo ei samaan tapaan riipu intuitioista, jotka voisivat osoittautua vääriksi. Jotta käsitys voisi osoittautua vääräksi, pitäisi vallita ero sen välillä miltä asia näyttää ja miten asia on:

> ”Mutta mitä tulee tietoisten tilojen olemassaoloon, erottelua ilmenemisen ja todellisuuden välille ei voida tehdä, sillä todellista on juuri tämä ilmenemisen olemassaolo. Jos minusta tietoisesti vaikuttaa siltä, että olen tietoinen, niin olen tietoinen.”[25]

Searle myös kirjoittaa, nähdäkseni osuvasti, Dennettin kannan seuraavan kahdesta perusolettamasta: objektivistisesta tiedekäsityksestä, joka suosii kolmannen persoonan menetelmiä ja karsastaa ensimmäisen persoonan subjektiivista näkökulmaa, ja verifikationismista, jonka mukaan jokin on olemassa vain, jos se voidaan todentaa tieteen menetelmin olemassa olevaksi.

Tietoisuus ei ole ainoa asia, jonka ontologisesta statuksesta arkifenomenologia ja tieteellinen näkökulma voivat olla erimielisiä. Myös esimerkiksi vastavuoroisten tunnustussuhteiden (persoonien välisen kunnioituksen, arvostuksen ja rakkauden) toteutuminen edellyttää osallistujilta arkifenomenologista lähestymistapaa: toisiin ei voi suhtautua persoonina ja yhteistyötahoina, jos pitää heitä vain atomiryppäinä. Toki toisia voi pitää *myös* atomiryppäinä (olisi outoa, jos kaikki muu luonto koostuisi atomeista, mutta ihmiset eivät), mutta tunnustussuhteiden syntymiseksi toisia tulee pitää (myös) intentionaalisina, persoonallisina olentoina. Samaa edellyttää kokemus tunnustuksen, kunnioituksen ja rakkauden vastaanottamisesta toisilta. Siinä määrin kuin suhtaudumme toisten käyttäytymiseen kliinisesti vain oireena atomitason tapahtumien lainalaisuuksista, emme suhtaudu näihin toisiin persoonina vaan esimerkiksi persoonattomien luonnonvoimien ilmentyminä. Vastavuoroiseen tunnustukseen sisältyvä persoonana pitäminen ei siis ensisijaisesti motivoidu selittävänä hypoteesina, jota koetellaan, vaan elämismaailmallisena varmuutena, jota arkinen kanssakäyminen edellyttää. Myös oikeudet, velvollisuudet, ihmisarvo tai kysymykset vastuusta voidaan sivuuttaa selityksiä tavoittelevasta näkökulmasta, mutta käytännölliseen elämismaailmaan ne kuuluvat.

Kaikkiaan olen tässä artikkelissa puolustanut kolmea väitettä. (1) Empiirinen kysymys ”voiko kone ajatella?” osoittautuu monitulkintaiseksi, käsitteellistä selvennystä vaativaksi kysymykseksi. (2) Jos kysymys tarkennetaan muotoon ”osaavatko koneet nykyisin ajatella relevantisti samassa mielessä kuin ihmiset?”, paljastuu seuraava erimielisyys: arkifenomenologian mukaan koneet vain ”ajattelevat” ja tulevaisuudessa vaaditaan laadullinen hyppäys teknologian kehityksessä, ennen kuin niiden voidaan katsoa ajattelevan ihmisten tapaan; elämismaailmasta irtautuvan kannan, kuten kyberneettisen kognitiotieteen mukaan koneet ajattelevat jo relevantisti samassa mielessä kuin ihmiset – kunhan hylätään arkipsykologiset hypoteesit siitä, että ihmiset Ajattelevat isolla A:lla. Sama erimielisyys paljastuu muistakin keskeisistä käsitteistä kuten tietoisuudesta, itsetietoisuudesta tai tuntoisuudesta. Yhdestä näkökulmasta koneiden ja ihmisten välillä on selvä laadullinen ero ja koneilta puuttuu jotain, toisesta näkökulmasta koneet ja ihmiset ovat jo nyt samalla viivalla. (3) Kun huomataan, että elämismaailmaan kiinnittyvää arkifenomenologista näkökulmaa tarvitaan ennen kaikkea elämiseen, eivätkä sen varmuudet ole oikeastaan tieteellisiä hypoteeseja vaan asioita, joiden todellisuutta ei arkinäkökulmasta ole vahvoja perusteita epäillä, voidaan ottaa askel kohti stereoskooppista kuvaa. (Joidenkin filosofien mukaan voidaan ottaa jatkoaskeleita ja saavuttaa synteettinen kokonaiskuva, toisten mukaan jäädään perspektiivien moneuteen.) Näin voidaan sallia näkemys, että vaikka tieteellisen näkökulman lähtökohdat voivat olla liian reduktiivisia ollakseen koko totuus, ne kuitenkin voivat paljastaa paljonkin asioita, jotka eivät arkinäkökulmasta avaudu ja joiden paljastaminen vaatii systemaattista tieteellistä tutkimusta. Tämä on laajemminkin relevantti erottelu, jonka havainnollistamisessa kysymys koneiden ja ihmisten ajattelun erilaisuudesta on vain yksi esimerkki.

Vastaus kysymykseen ”voiko kone ajatella relevantisti samassa mielessä kuin ihminen?” kuuluu siis seuraavasti: ei, jos tarkoitetaan inhimilliseen maailmassa olemiseen kuuluvaa ajattelua kaikessa rikkaudessaan, kuten elämismaailmaan kiinnittyvä arkifenomenologia tekee; ja kyllä, jos asiaa tarkastellaan rajatummin määriteltyjen kognitiivisten prosessien näkökulmasta, kuten erilaiset tieteelliset lähestymistavat, esimerkiksi kyberneettinen kognitiotiede, tekevät.[26]

Viitteet

1 Sellars 1963.
2 Turing 2004 (1950), 449.
3 Sellars 1963.
4 Taylor 1985; McDowell 1996; Nagel 1986.
5 Taylor 1985; McDowell 1996. Muita objektivismin muotoja voivat olla sellaiset subjektin illusorisuutta puolustavat systeemiteoreettiset, psykoanalyyttiset tai strukturalistiset mallit, joissa ontologinen perusta nähdään yhteiskunnan, kielen tai alitajuisten voimien tasolla, ja subjektius nähdään näihin redusoituvana. Ks. Ricœur 1970.
6 McDowell 1996**,** 4–7; Taylor 1985, ”Introduction”. McDowell pyrkii osoittamaan erilaiset kritisoimansa kannat (kuten annetun myytti tai koherentismi) filosofisesti motivoituneiksi. Taylor sen sijaan on vakuuttunut, että reduktiivisten ja eliminativististen kantojen houkuttelevuus perustuu osittain moraalisiin motiiveihin (ne voivat esimerkiksi liittyä tiettyihin vapausihanteisiin).
7 Taylor 1985; McDowell 1996.
8 Olen tarkastellut tätä tarkemmin teoksessa Laitinen 2009.
9 Hegel 1978; 1981; 2011; Ricœur 1992.
10 Hegel 1978; 1981; 2011.
11 Ricœur 1970; 1992.
12 Hans Jonas (1966) kutsuu näitä näkökulmia fenomenologiseksi ja kyberneettiseksi.
13 Ks. Seibt 2017.
14 Vrt. Dreyfus 1972; 1992.
15 Ks. esim. Ricœur 1992, 58 & *passim*; vrt. myös Ihde 1983.
16 Baumeister 2015; Chalupa 2015; Dobelli 2015; Enfield 2015.
17 Turing 2004 (1950), 453.
18 Carroll 2015.
19 Dennett 1995.
20 Sama.
21 Dennett 2007, 249.
22 Sama, 250–251.
23 Habermas 1971.
24 Searle 1995.
25 Sama.
26 Kiitokset Juho Rantalalle, Risto Koskensillalle, Jaakko Beltille sekä kahdelle anonyymille arvioijalle lukuisista tekstiä koskevista parannusehdotuksista ja kysymyksistä; samoin osallistujille tilaisuuksissa, joissa olen esittänyt versioita tässä esitellyistä ajatuksista.

Kirjallisuus

Artificial Intelligence: A Modern Approach. 3. p. Toim. Stuart Russell & Peter Norvig. Prentice Hall, Amsterdam 2010.

Baumeister, Roy, No ’I’ and no capacity for malice. Teoksessa *What to Think About Machines That Think: Today’s Leading Thinkers on the Age of Machine Intelligence.* Toim. John Brockman. Harper Perennial, New York 2015, 72–73.

Bickhard, Mark H., Robot Sociality: Genuine or Simulation? Teoksessa *Sociality and Normativity for Robots.* Toim. Johanna Seibt & Raul Hakli. Springer, Cham 2017, 41–66.

Carroll, Sean, We are all machines that think. Teoksessa *What to Think About Machines That Think: Today’s Leading Thinkers on the Age of Machine Intelligence.* Toim. John Brockman. Harper Perennial, New York 2015, 56–58.

Chalupa, Leo M., No machine thinks about the eternal questions. Teoksessa *What to Think About Machines That Think: Today’s Leading Thinkers on the Age of Machine Intelligence.* Toim. John Brockman. Harper Perennial, New York 2015, 83–84.

Dennett, Daniel, The Mystery of Consciousness: An Exchange [with John Searle]. *New York Review of Books*. December 21, 1995. Verkossa: https://www.nybooks.com/articles/1995/12/21/the-mystery-of-consciousness-an-exchange/

Dennett, Daniel, When HAL Kills, Who’s to Blame? Computer Ethics. Teoksessa in *HAL’s Legacy: 2001’s Computer as Dream and Reality*. Toim. D. G. Stork. MIT Press, Cambridge (MA) 1997.

Dennett, Daniel, Heterophenomenology Reconsidered. *Phenomenology and the Cognitive Sciences.* Vol. 6, No. 1–2, 2007, 247–270.

Dobelli, Ralf, Self-aware AI? Not in 1000 years. Teoksessa *What to Think About Machines That Think: Today’s Leading Thinkers on the Age of Machine Intelligence.* Toim. John Brockman. Harper Perennial, New York 2015, 98–101.

Dreyfus, Hubert, *What Computers Can’t Do.* MIT Press, New York 1972.

Dreyfus, Hubert, *What Computers Still Can’t Do.* MIT Press, New York 1992.

Enfield, N. J., Machines aren’t into relationships. Teoksessa *What to Think About Machines That Think: Today’s Leading Thinkers on the Age of Machine Intelligence.* Toim. John Brockman. Harper Perennial, New York 2015, 397–398.

Habermas, Jürgen, *Knowledge and Human Interests*. Käänt. Jeremy J. Shapiro. Beacon Press, Boston 1971.

Hegel, Georg Wilhelm Friedrich, *Wissenschaft der Logik. Erster Band. Objektive Logik* (1812/1813). *Gesammelte Werke*. Band 11. Felix Meiner Verlag, Hampuri 1978.

Hegel, Georg Wilhelm Friedrich, *Wissenschaft der Logik. Zweiter Band. Subjektive Logik* (1816). *Gesammelte Werke*. Band 12. Felix Meiner Verlag, Hampuri 1981.

Hegel, Georg Wilhelm Friedrich, *Logiikan tiede I*. Suom. Ilmari Jauhiainen. Summa, Helsinki 2011.

Ihde, Don, Why Do Humans Think They Are Machines? Teoksessa *Existential Technics*. SUNY Press, Albany 1983.

Jonas, Hans, *The Phenomenon of Life: Toward a Philosophical Biology*. Northwestern University Press, Evanston 1966.

Laitinen, Arto, *Itseään tulkitseva eläin.* Gaudeamus, Helsinki 2009.

Machine Ethics. Toim. Michael Anderson & Susan Leigh Anderson. Cambridge UP, Cambridge 2011.

McDowell, John, *Mind and World. With a New Introduction by the Author.* Harvard UP, Cambridge (MA) 1996.

McDowell, John, *Mind, Value and Reality*. Harvard UP, Cambridge (MA) 1998.

Merleau-Ponty, Maurice, *Phenomenology of Perception* (Phénoménologie de la perception, 1945). Käänt. Colin Smith. Routledge. London 1962.

Nagel, Thomas, What Is It Like to Be a Bat? *Philosophical Review*. Vol. 83, No. 4, 1974, 435–450.

Nagel, Thomas, *The View from Nowhere.* Oxford University Press, Oxford 1986.

Ricœur, Paul, *Freud and Philosophy. An Essay on Interpretation.* New Haven, Yale 1970.

Ricœur, Paul, *Oneself as Another*. The University of Chicago Press, Chicago 1992.

Searle, J. R., Minds, Brains, and Programs. *Behavioral and Brain Sciences*. Vol. 3, No. 3, 1980, 417–457.

Searle, J. R., The Mystery of Consciousness: Part II. *The New York Review of Books.* November 16, 1995.

Searle, J. R., The Mystery of Consciousness: An Exchange. A Reply [to Dennett]. *New York Review of Books*. December 21, 1995. Verkossa: www.nybooks.com/articles/1995/12/21/the-mystery-of-consciousness-an-exchange/

Seibt, Johanna, Towards an Ontology of Simulated Social Interaction: Varieties of the ”As If” for Robots and Humans. Teoksessa *Sociality and Normativity for Robots.* Toim. Johanna Seibt & Raul Hakli. Springer, Cham 2017.

Sellars, Wilfrid, Philosophy and the Scientific Image of Man (1963). Teoksessa *Science, Perception and Reality*. Ridgeview, Atascadero 1991, 7–43.

Smith, Nicholas H., *Charles Taylor. Meaning, Morals and Modernity.* Polity Press, Cambridge 2002.

Taylor, Charles, *Human Agency and Language: Philosophical Papers vol. 1.* Cambridge University Press, Cambridge 1985.

Turing, Alan, Intelligent Machinery (1948). Teoksessa *The Essential Turing: The Ideas That Gave Birth to the Computer Age*. Toim. Jack B. Copeland. Clarendon Press, Oxford 2004, 410–432.

Turing, Alan, Computer Machinery and Intelligence (1950). Teoksessa *The Essential Turing. The Ideas That Gave Birth to the Computer Age*. Toim. Jack B. Copeland. Clarendon Press, Oxford 2004, 441–464.