

TIETOKONEELLA MIELI?

Ajattelevien koneiden teemat akateemisissa
filosofisissa joulaleissa Englannissa 1950–1975

Eero Kangas

Pro gradu -tutkielma

Turun yliopisto

Historian, kulttuurin ja taiteiden tutkimuksen laitos

Kulttuurituotannon ja maisemantutkimuksen tutkinto-ohjelma

Digitaalinen kulttuuri

Huhtikuu 2021

”Turun yliopiston laatu järjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.”

TURUN YLIOPISTO

Historian, kulttuurin ja taiteiden tutkimuksen laitos/

Humanistinen tiedekunta

KANGAS, EERO: Tietokoneella mieli? Ajattelevien koneiden teemat akateemisissa filosofisissa joulunaleissa Englannissa 1950–1975.

Pro gradu -tutkielma, 65 s.

Digitaalinen kulttuuri

Huhtikuu 2021

Tutkielman aihe on ajattelevien koneiden teemat filosofian alan merkittävässä akateemisissa joulunaleissa Englannissa 1950–1975. Sen tarkoituksena on selvittää ajattelevia koneita käsittelevää filosofista keskustelua vuosina 1950–1975, avata keskustelun teemojen taustoja sekä yhteyksiä muihin keskustelun teemoihin. Aineistolähtöisesti tutkimuskohteeseen asennoituvassa tutkielmassa laadullisen sisällönanalyysin ja teemoittelun avulla muodostetaan historiallista rekonstruktiota tutkimuskohteesta. Tutkimus on toteutettu tarkastelemalla merkittävässä englantilaisissa akateemisissa filosofian joulunaleissa julkaistuja artikkeleja vuosina 1950–1975 ja erittelemällä niissä esiintyneitä teemoja ja selvittämällä teemojen yhteyksiä toisiinsa sekä yleisempään akateemiseen historialliseen kontekstiin. Tutkielman merkittävimpiä tuloksina voidaan pitää tärkeimpien keskustelun teemojen – tavallisen kielen filosofian, ihmisen ja koneen kykyjen ja ominaisuuksien vertailun, matkimispelin, toiset mielet - ongelman sekä yliaistillisen havaitsemisen teemojen – erittelyä sekä kontekstualisointia. Tuloksista pääteltiin erityisesti yleisempien filosofian keskustelujen, kuten kielifilosofian ja mielenfilosofian ongelmien osaltaan siirtyneen myös ajattelevia koneita koskevaan keskusteluun ja ajattelevien koneiden ideaa ensisijaisesti käytetyn filosofian perinteisten ongelmien käsittelyn uutena tapana.

Avainsanat: ajattelevat koneet, filosofia, Englanti, joulunale, teemat.

SISÄLLYSLUETTELO

1. Johdanto.....	1
1.1. Tutkimuskysymykset ja teoreettinen tausta.....	2
1.2. Tutkimuksen tavoitteet ja merkitys.....	8
1.3. Tieteenhistoriallinen filosofian tutkimus.....	10
1.4. Aikaisempi tutkimus.....	14
1.5. Käsitteet ja termit.....	15
2. Aineisto ja menetelmät.....	17
2.1. Aineistonkeruu ja aineisto.....	17
2.2. Laadullinen sisällönanalyysi.....	19
3. Ajattelevien koneiden idean taustaa.....	21
3.1. Ajattelevien koneiden yleistä historiaa.....	21
3.2. Akateeminen filosofia Englannissa 1900-luvun puolivälissä.....	24
3.3. Alan Turing (1912–1954).....	25
3.4. Turingin testi ja kiinalaisen huoneen argumentti.....	26
4. Ajattelevat koneet.....	30
4.1. Matkimispeli.....	30
4.2. Ajattelevat koneet kielen filosofian ongelmana.....	33
4.3. Elävät koneet.....	39
4.4. Ajattelevat koneet mielen filosofian ongelmana: toiset mielet -ongelma.....	41
5. Turing ja yliaistillinen havaitseminen.....	44
5.1. Yliaistillisuuden argumentti	44
5.2. Yliaistillisuuden tutkimukset 1900-luvun alkupuolella.....	46
5.2.1. J. B. Rhine (1895–1980).....	46
5.2.2. S. G. Soal (1899–1975).....	47
5.2.3. C. D. Broad <i>Philosophy</i> ssä 1944.....	48
5.3. Turing ja yliaistillinen havaitseminen matkimispelissä.....	49
6. Lopuksi.....	55
Lähteet.....	61

1. Johdanto

”Voivatko koneet ajatella?” Tätä kysymystä on pohdittu erityisesti sen jälkeen, kun Alan Turing esitti kuuluisan *matkimispelinsä*, jota sittemmin on myös *Turingin testiksi* kutsuttu, lokakuussa 1950. Koneiden ajattelukyvyistä voisi tehdä monenlaista tutkimusta, mutta tässä keskitytään siihen, millä tavoin filosofit lähestyivät aihetta keskustelun alkuaikoina, Turingin artikkelin ”Computing Machinery and Intelligence” jälkeen. Tätä pyritään selvittämään tutkimalla artikkeleita englantilaisissa filosofisissa jurnaaleissa 1950–1975. Mielenkiinnon kohteena on siis se, miten filosofit ajattelevista koneista ammattimaisesti ja julkaisukelpoisesti ajattelivat sekä minkälaisia teemoja heidän argumentoinnissaan välittyi. Tämä on siis ideoiden ja filosofian historiaa käsittelevä työ, joka keskittyy erityisesti filosofien toimintatapoihin ja pyrkii teemoittelun avulla rekonstruoimaan keskustelua ja esittämään tarkasteluvälin aikana käydyn keskustelun teemoja.

Tämänkaltainen historiallisia akateemisia julkaisuja, varsinkin filosofisia julkaisuja, historiallis-teemoittelevasta näkökulmasta lähestyvä tutkimus, joka pyrkii myös hieman selvittämään tutkimuskohteina olevien argumenttien sisäistä logiikkaa ja selittämään sitä ensisijaisesti kontekstissaan, on jossain määrin jäänyt vähäiseksi. Tämän vuoksi tämä työ pyrkii paikkaamaan tätä puutetta tutkimuskohteensa osalta, mutta toisaalta myös hieman muotoilemaan mahdollista metodologiaa tämänkaltaisen tutkimuksen tekemiseen myös tulevaisuudessa.

Tutkimuskohteena olevaa materiaalia on olemassa runsaasti ja varsinkin yliopiston käyttäjätunnusten kautta se on helposti saatavissa, mutta aineisto on hajanaista sekä historiallisen kontekstin muutoksen vuoksi mahdollisesti vaikeasti lähestyttävää. Tämä työ pyrkii teemoittelemaan keskustelua helpommin hahmotettavaksi kokonaisuudeksi, sekä selvittämään historiallista kontekstia ja sen avulla myös selittämään, miksi juuri kyseisistä teemoista puhuttiin sillä tavoin, kun niistä puhuttiin. Filosofian historia on usein jätetty filosofeille, mikä on ymmärrettävää, mutta usein filosofisissa filosofian historian töissä on vahva filosofinen asenne historiallisen sijaan, mikä saattaa kallistaa töiden mielenkiintoa enemmän argumenttien logiikkojen ymmärtämiseen kyseisten argumenttien historiallisten kontekstien ymmärtämisen ja sitä kautta argumenttia lähestyen, sijaan.

Tämä työ on jaettu viiteen varsinaiseen lukuun, joista kaksi ensimmäistä, johdanto sekä aineistot ja menetelmät -luku selvittävät tutkimuksen perustuksen, tutkimuskohteen sekä metodologian, tavoitteet sekä muut tutkimuksen toteuttamisen kannalta perustavanlaatuiset asiat, kuten tapana on. Tämän jälkeen, kolmannessa luvussa, selvitetään ajattelevien koneiden idean historiaa sekä tutkimuksen tarkasteluajankohdan tilannetta pyrkien luomaan hieman kontekstia, johon aineisto asettuu. Kolmannessa luvussa myös käyn lyhyesti läpi Alan Turingin henkilöhistoriaa ja Turingin testin asetelmaa, pyrkien näin ennen varsinaista teemojen esittelyä esittelemään, mistä keskustelu oikeastaan kumpuaa. Turing katsotaan merkittäväksi keskustelun alullepanijaksi, vaikka hän ei sitä tietenkään yksin tehnyt, vaan asiasta keskustelivat myös muut. Turing kuitenkin nähdään tässä merkittävimpanä, ja siksi häntä tarkastellaan tarkimmin.

Luvussa neljä esittelen merkittävimät teemat, jotka aineistosta nousevat esiin ajattelevia koneita koskien. Pyrin myös parhaani mukaan selvittämään kunkin teeman kohdalla sen merkitystä sekä sitä, minkä vuoksi se niin merkittäväksi ja usein toistuvaksi teemaksi on muodostunut. Luku viisi tarkastelee erityistä teemaa, nimittäin Turingin tapaa käsitellä ja esittää yliluonnollisen havaitsemisen ideaa artikkelissaan ”Computing Machinery and Intelligence.” Teema on melko heikosti kontekstiin sopiva, ja siitä on hyvin vähän Turing-tutkimuksen piirissä keskusteltu, joten tässä on nähty luontevaksi tarkastella sitä hieman muita teemoja laajemmin pyrkien selvittämään sen taustoja ja sopivuutta Turingin maailmankuvaan.

1.1 Tutkimuskysymykset ja teoreettinen tausta

Tässä työssä pyritään selvittämään, minkälaisia teemoja ajattelevia koneita käsitelleessä filosofisessa keskustelussa nousi esiin vuosina 1950–1975, varsinkin Ison-Britannian kontekstissa. Työn perustutkimuksellisen luonteen vuoksi on luontevampaa pitää aihe yksinkertaisena ja rajattuna ja pyrkiä välttämään tutkimuskysymysviidakkoa.

Päätutkimuskysymys kuuluu: *”Mitä teemoja ja argumentointitapoja esiintyi merkittävässä englantilaisissa akateemisissa filosofian alan jurnaaleissa vuosina 1950–1975 koskien ajattelevista koneista käytyä keskustelua ja miksi?”*

Alatutkimuskysymyksenä on *”miksi Turing piti yliluonnollista aistimista matkimispelinsä kannalta ongelmallisena?”*

Painotus ensimmäisenkin päätutkimuskysymyksen osalta on Turingin artikkelista lähteneessä keskustelussa esiintyneissä merkittävämmissä teemoissa ja argumentaatiostrategioissa. Tässä työssä pyritään siis ensisijaisesti teemoittelemaan keskustelua ajattelevista koneista ja selvittämään, miten filosofit käsillä olevaa ongelmaa lähestyvät ja miksi juuri siten. Työssä tarkastellaan merkittävässä englantilaisissa filosofian alan akateemisissa jurnaaleissa julkaistuja artikkeleita, jotka otsikkotasolla ilmaisevat käsittelevänsä ajattelevia koneita tai kommentoivansa Turingin aloittamaa keskustelua. Aineistoon valikoituvat siis luontevasti artikkelit, jotka selkeästi kuuluvat kyseiseen aihepiiriin, mutta ei esimerkiksi kirja-arvosteluja, vaikka niissä saatettaisiinkin kommentoida jotain ajattelevia koneita käsittelevää uutta kirjaa. Myös aineiston rajaamiseksi tämänkokoiseen työhön sopivaksi tarkastelun ulkopuolelle jätettiin jurnaaleissa julkaistut *Discussion*-osioiden tekstit, vaikka ne olisivat otsikkotasolla ilmaisseet käsittelevänsä aihepiiriä. Tämänkaltainen rajaus on tämän työn koon puitteissa realistinen tarkasteltava siten, että on mahdollista saada aikaan muutakin kuin monotoninen luettelo käsitellyistä teemoista. *Englantilainen filosofia* on selkeämmin rajattavissa kuin esimerkiksi *englanninkielinen filosofia* ja tämän vuoksi tässä keskitytään siihen. Englanninkielisellä filosofialla voitaisiin tarkoittaa maantieteellisesti myös varsinkin Yhdysvalloissa, Kanadassa ja Australiassa käytäviä keskusteluja, mutta myös esimerkiksi saksankielisessä julkaisussa julkaistavaa englanninkielistä tekstiä. Tämän vuoksi maantieteellinen rajaus Englannissa julkaistuihin merkittäviin jurnaaleihin on järkevää. Niissä tietenkin on julkaistu kirjoituksia laajemmin Ison-Britannian alueelta, esimerkiksi Edinburghin yliopiston filosofeilta, mutta myös muiden maiden filosofien tekstejä, esimerkiksi Yhdysvalloista ja Irlannista.

Merkittävät filosofian jurnaalit taas katsotaan tässä työssä niiksi, joita julkaisivat Englannin merkittävimpien yliopistojen, Oxfordin sekä Cambridgen yliopistojen kirjapainot. Näitä kirjapainoja käyttivät julkaisukanavinaan myös yhdistykset, joten kyse ei ole aina Oxford University Pressin tai Cambridge University Pressin omista julkaisuista, mutta niiden merkittävyys akateemisina kustantamoina katsotaan sopiviksi rajaustavoiksi, sillä niiden logoilla varustetuilla kirjallisilla tuotoksilla on akateemisessa maailmassa enemmän arvoa kuin esimerkiksi yksittäisen pienen yhdistyksen omakustanteisella julkaisulla. Näin siis rajauksen ulkopuolelle jäävät pienten yhdistysten julkaisut. Lisäksi julkaisujen tulee olla digitaalisessa kirjastossa *JSTOR* koko tarkasteluvälin ajalta, vuosilta 1950–1975. Tällöin esimerkiksi

vasta myöhemmin tarkasteluajan kuluessa perustetut jornaalit tai sen aikana lopetetut – tai sen aikana taukoa pitäneet jornaalit – jäivät tarkastelun ulkopuolelle.

Tarkasteluväliksi muodostui 1950–1975, koska vuonna 1950 Alan Turing julkaisi merkittävän artikkelinsa ”Computing Machinery and Intelligence” ja monet sitä seuranneet ajattelevia koneita käsittelevät artikkelit vaikuttivat tavalla tai toisella olevan Turingin artikkelia kommentoivia. Tarkasteluväli loppuu vuoteen 1975, koska 1970-luvun puolivälin paikkeilla ainakin tekoälytutkimuksessa katsotaan alkaneen niin kutsuttu ”tekoälytalvi” (engl. AI-winter) ja tämä lienee sen vuoksi sopiva vuosiluku tarkasteluvälin lopuksi.¹ Tässä työssä siis tarkastellaan ajallisesti sitä ajanjaksoa, jolloin yleisesti ottaen ainakin tekoälyä käytännössä tutkineet olivat optimistisia sen mahdollisuuksista ja nopeasta edistyksestä. Filosofinen keskustelu ei tietenkään täysin seuraile muuta tieteellistä keskustelua, vaan etenee hieman omaa tahtiaan. Toisaalta vuosi 1975 sopii tarkasteluvälin päätteeksi myös käytännöllisemmistä syistä. Tällä 26 vuoden aikavälillä tarkasteltavissa jornaaleissa julkaistiin riittävä määrä ajattelevia koneita käsitteleviä artikkeleita, jotta teemoittelua voi tehdä, mutta ei kuitenkaan niin monia, että niiden analysoiminen olisi mahdotonta tämän kokoisessa tutkielmassa.

Tutkimuskohteena filosofian alan akateemiset jornaalit valikoituiivat aineistoksi sen vuoksi, että tutkimusongelmassa mielenkiintoisena nähdään juuri filosofien kirjoitukset ja ajattelu. Toki filosofian jornaaleissa voivat julkaista myös ei-filosofit, kuten esimerkiksi Turing, mutta tällöin artikkelin luonne on joka tapauksessa filosofinen ja voidaan laskea filosofiaksi.

Rajaus pelkälle otsikkotasolla ilmenevälle viittaukselle ajatteleviin koneisiin tai Turingin aloittamaan keskusteluun saattaa ensisilmäyksellä vaikuttaa riittämättömältä, ja esimerkiksi abstraktien lukeminen voisi tuntua luontevammalta. Varsinkin vanhemmissa filosofisissa artikkeleissa kuitenkin on usein tapana kehitellä käsiteltävää asiaa läpi tekstin vasta lopulta päätyen varsinaiseen argumenttiin. Niissä ei myöskään – ainakaan tarkastelemieni jornaalien tapauksissa – ole yleensä abstraktia. On siis joko luettava kaikki artikkelit tarkasteluajaväliltä tai valittava tarkasteluun ne tekstit, joissa yhteys tutkimuskohteeseen ilmenee otsikkotasolla. On luontevaa olettaa, että filosofiseen jornaaliin kirjoittava haluaa tietystä keskustelusta kiinnostuneen henkilön huomaavan oman kirjoituksensa ja siksi viittaus keskusteluun näkyy selvästi jo otsikkotasolla.

¹ Bostrom, 2014, 8.

Yliaistillisen havainnoimisen ajatuksen alkuperää selvittävä alatutkimuskysymys saattaa ensi silmäyksellä vaikuttaa päätutkimuskysymyksestä hieman irralliselta siinä mielessä, että se ei ilmene aineistossa merkittävänä yksittäisenä teemana. Kysymys on sivuutettu filosofisessa keskustelussa ja myöhemmässä Turing-tutkimuksessa lähes kokonaan, minkä takia sitä voi olla mielekästä tässä hieman käsitellä. Se on myös ainakin nykypäivän kontekstistamme hieman erikoinen väite ja siksikin tarpeellinen käsiteltäväksi. Yliaistillinen havaitseminen Turingin matkimispelin kontekstissa on siis enemmänkin ”anti-teema” teemojen joukossa, juuri siksi, että siitä ei ole keskusteltu.²

Tutkimuksen pyrkimyksenä on rekonstruoida käytyä filosofista keskustelua teemoittelemalla sitä ja luomalla yhteyksiä eri tekstien ja eri joulleissa julkaistujen tekstien välillä. Ei voi olettaa, että tietyssä joulleissa alkanut keskustelu pysyisi ainoastaan kyseisessä joulleissa, vaan varsinkin laajaa huomiota saavista ongelmista keskustellaan monissa eri joulleissa ja siksi on tärkeää pyrkiä luomaan yhteyksiä eri joulleissa julkaistujen tekstien välille. Näin menneisyydessä tapahtuneesta keskustelusta tulee selkeämpää ja mielekkäämpää myös nykypäivän kontekstista käsin. Ottaen myös huomioon, että koska keskustelu yhä jatkuu koko ajan laajentuen, tuo tämä tutkielma selkeyttä keskustelun alkuvuosien teemoihin.

Tässä työssä siis jossain määrin katsotaan sosiaalisten käytäntöjen luovan todellisuutta. Tällä tarkoitetaan tässä sitä, että keskustelun alkuhetket ovat oletuksen mukaan vaikuttaneet myöhempään keskusteluun ja siksikin niihin on tärkeää palata. On tosin toisen tutkimuksen aihe selvittää, miten vuosien 1950–1975 englantilainen keskustelu ajattelevista koneista näkyy esimerkiksi nykypäivän filosofisessa keskustelussa ajattelevista koneista tai tekoälystä.

Tämä tutkimus on melko aineistolähtöinen, mutta koska puhtaan aineistolähtöinen tutkimus on mahdotonta – tutkijan aiempi ymmärrys ja käsitykset tutkittavasta ilmiöstä tietenkin vaikuttavat aina tutkimukseen käytännössä – lienee paikallaan selvittää tutkimuksen teoriataustaa. Tutkimus on ensisijaisesti historiatutkimusta, ja kuten historioitsija Marjatta Hietala toteaa, historian tutkijan on usein vaikea kuvata tarkkaan metodologiaan tai tutkimuksen teoreettista viitekehystä.³ Tämän ei kuitenkaan tule antaa olla tekosyynä teoriataustan pohtimisen välttellemiselle. Tutkimuksessa on ensinnäkin selviä *hermeneuttisen tutkimuksen* piirteitä, lähtien jo

² Leavitt, 2017.

³ Hietala, 2001, 21.

aineiston hankinnasta, jossa aineiston tarjoaviksi joulunäköiksi valittujen artikkelien otsikot ja niiden sisältämät ajatteleviin koneisiin ja Turingin matkimispeliin viittaavat elementit muodostivat sen käsitteistön, jonka puitteissa aineiston muodostavat artikkelit valittiin. Tämä on osaltaan myös melko aineistolähtöinen lähestymistapa jo tällä aineiston keräämisen tasolla, mutta aineistolähtöisyys on luonnollisesti myös merkittävässä osassa aineiston analyysivaiheessa.

Hermeneutiikalla tutkimuksellisenä suuntauksena tarkoitetaan yleisesti ottaen merkityksiä sisältävien kokonaisuuksien tulkintaa sekä erityisesti ymmärtämistä. Siihen liitetään usein myös niin kutsuttu *hermeneuttinen kehä*. Sillä viitataan siihen ajatukseen, että yksityiskohdat ja niiden tulkinnat vaikuttavat kokonaistulkintaan ja siten tutkimuskohteen uudelleentulkinnan muodostavaan, jatkuvasti syvenevään ja laajenevaan tutkimuskohteen tulkintaan.⁴ Tämä toiminta muodostaa ”kehän”, tai ehkä tarkemmin spiraalin, sillä ”kehä” viittaisi alkuun palaamiseen ilman, että mitään muutosta olisi tapahtunut, mutta spiraalissa ikään kuin palataan alkuun, mutta on kuitenkin edetty alkupisteestä. Kuitenkin kehän analogia on terminologisesti vakiintunut tässä kontekstissa, joten tässä pitäydytään siinä, kuitenkin tiedostaen, että kehällä ei tarkoiteta varsinaista alkuun palaamista ja kehämäisyyttä.

Tässä työssä on myös tiettyjä *konstruktivistisen* suuntauksen elementtejä, erityisesti siinä, millaisena filosofinen keskustelu nähdään. Konstruktivismilla tutkimussuuntauksena tarkoitetaan sellaista, jossa tiedon nähdään muodostuvan tutkimusprosessin aikana ja tutkijan rakentavan tieteellistä tietoa. Radikaalimmissa konstruktivismin muodoissa saatetaan jopa ajatella tutkijan rakentavan totuutta ja relativistisesti ajateltavan, että maailmassa ei ole muuttumattomia totuuksia, joita tiede voisi lähestyä, vaan se kaikki on tutkijoiden rakentamaa.⁵ Tämänkaltaisella suuntauksella kuitenkin saattaa olla tiettyjä episteemisiä heikkouksia ja siksi tässä työssä pidättyydytään tästä radikaalimmasta muodosta ja lähestytään tutkimuskohdetta osaltaan maltillisemmän konstruktivistisen asenteen kannalta ja pidetään mahdollisena myös sitä, että on luultavasti olemassa tosiasioita, jotka ovat tutkijasta tai ihmisestä riippumattomasti tosia.

Konstruktivismin suuntaukseen liittyy läheisesti *sosiaalisen konstruktionismin* suuntaus, joka on tämän työn kannalta ehkä konstruktivismia olennaisempi käytännöllisessä mielessä. Siinä keskitytään mm. tiedon, rakenteiden ja

⁴ Lähdesmäki ym, 2009a.

⁵ Lähdesmäki ym, 2009c.

ilmiöiden esittämiseen ihmisen toiminnan tuloksina sen sijaan, että ne olisivat luonnollisesti muotoutuneet tai itsestään selviä.⁶ Filosofi John Searle esittelee realistisen sosiaalisen konstruktionismin ajatusta ja esittää tiettyjen asioiden olevan *raakoja tosiasioita* (engl. *brute facts*), jotka ovat maailmassa tosiasiasa (kuten vuoret ja luonnonlait), mutta samalla toisten asioiden olevan *institutionaalisia tosiasioita* (engl. *institutional facts*), kuten raha tai jossain määrin tieteelliset käytänteet.⁷ Tässä työssä tämä nähdään luontevana episteemisenä taustana tutkimukselle, sillä se myöntää tosiasian statuksen tietyille tosiasioille, kuten luonnonlaeille ja esimerkiksi ihmisen hermojärjestelmälle sekä tietokoneen osille, mutta samalla toteaa tiettyjen asioiden olevan ihmisten toiminnan tuloksia ja siitä riippuvaisia, esimerkkeinä yliopistot ja filosofiset jurnaalit sekä keskustelut jurnaaaleissa argumentteineen ja ajatuskokeineen.

Tutkimuksen historiatutkimuksellinen tutkimusstrategia pyrkii näitä työn tieteenfilosofisia suuntauksia käytännöllisemmin lähestymään tutkittavaa kohdetta. Historiatutkimus nähdään tässä tutkimusstrategiana, jolla pyritään selvittämään menneisyyden ilmiöiden kehityskulkuja, merkitystä tai ominaisuuksia.⁸ Historian tutkimisessa usein saatetaan pyrkiä tarkastelemaan ilmiön syy-seuraussuhteita ja selittää siten jotain menneisyyden ilmiötä. Vaikka tässä pyritään menneisyyden ilmiön ymmärtämiseen ja selittämiseen, selittäminen tapahtuu enemmän ilmiön sisäisesti, tässä tapauksessa teemoittelemalla filosofisen keskustelun teemoja ja siten pyrkien ymmärtämään ajattelevien koneiden keskustelun ilmiötä filosofisessa keskustelussa. Tietenkin myös jonkin verran selvitetään esimerkiksi esiin nousevien teemojen taustoja ja pyritään ne asettamaan kontekstiinsa. Tätä perinteistä syy-seuraussuhteita selvittävää strategiaa sovelletaan vahvemmin Alan Turingin ja yliluonnollisen aistimisen (anti-) teemaa käsittelevässä luvussa, jossa pyritään selvittämään, miksi kyseinen teema ilmenee Turingin kirjoittamassa artikkelissa. Tämä teema nostetaan tämänlaiseen erityisasemaan sen huomiota herättävän erilaisuuden vuoksi: siinä missä Turingin artikkelin muut osat ovat hyvin behavioristisia ja loogisen positivismin asenteiden mukaisia, niin artikkelin yliaistillista havainnoimista käsittelevä lyhyt tekstinpätkä irtoaa selkeästi tästä perinteestä ja pitää jopa yliaistillisia ilmiöitä kuten telepatiaa mahdollisena. Tätä ilmiötä käsittelevässä luvussa pyritään siis myös selvittämään ilmiön syytä tarkastelemalla hieman laajemmin aiemmin julkaistuja tieteellisinä pidettyjä tekstejä ja hakemalla niistä yhtäläisyyksiä Turingin tekstiin.

⁶ Lähdesmäki ym, 2009d.

⁷ Searle, 1995, 2.

⁸ Lähdesmäki ym, 2009b.

Jonkinlaisena hypoteesina voidaan pitää, että aineistosta nousee joitain yhteneviä teemoja tai argumentaatiotapoja ja ne ovat jollain tavalla yhteydessä historialliseen tilanteeseen filosofian alalla Isossa-Britanniassa. Kuitenkin, koska tässä työssä painotetaan aineistolähtöisyyttä, ei liene tarpeen muodostaa kovin tarkkoja hypoteeseja esimerkiksi jo olemassa olevien teorioiden pohjalta.

Tämä työ ei ole varsinaista tekoälyn tutkimusta siinä mielessä, että se pyrkisi suoraan edistämään tekoälyn kehittämistä tai tutkimista. Kuitenkin osaltaan se auttaa tässäkin tarjoamalla historiallista kontekstia ja merkitysten selvittämistä kyseisessä kontekstissa. Tässä työssä ei myöskään pyritä muodostamaan mitään selkeää filosofista argumenttia tekoälystä tai ajattelevista koneista, vaan tutkitaan, millaisia ne olivat valitussa kontekstissa valittuna ajanjaksona. Näin se osaltaan auttaa kuitenkin ymmärtämään, miksi filosofiassa argumentoitiin siten, kun argumentoitiin.

1.2 Tutkimuksen tavoitteet ja merkitys

Tämän tutkielman tavoitteena on valottaa akateemisen tekoälykeskustelun alkuvuosia erityisesti filosofian alalla Englannissa. Pyrin muodostamaan kuvaa siitä, minkälaista keskustelu oli ja erityisesti mihin teemoihin keskustelu keskittyi, varsinkin ajattelevia koneita käsittelevän keskustelun osalta.

Tutkimusta argumenteista liittyen Turingin testiin on tehty ja esimerkiksi muodostettu argumenttikarttoja auttamaan filosofeja ymmärtämään laajaa ja monitasoista keskustelua.⁹ Kuitenkin humanistinen ja historiallinen tutkimus itse keskustelusta on jäänyt vähemmäksi. Tutkimuksen merkitys siis muodostuu siitä, että se osaltaan luo ymmärrystä keskustelusta myös humanistis-historiallisesta näkökulmasta tavanomaisemman filosofisen (argumentteja rakentavan ja itse ongelmaan pureutuvan ja sen ratkaisemiseen pyrkivän) tutkimuksen sijaan. Ajattelevien koneiden teemoja koskevat filosofiset julkaisut tämän vuoksi lähtökohtaisesti keskittyvätkin tarkastelemalleni ajanjaksolle, koska sittemmin julkaistut työt ovat vieneet keskustelua eteenpäin. Nykypäivän filosofian historian tutkimus aihetta käsitellessään tarkastelee yleensä myös argumentteja vuoden 1975 jälkeen, minkä vuoksi jotkin mahdollisesti mielenkiintoiset, mutta pienemmät teemat vuosien 1950–1975 välillä saattavat jäädä myöhempien teemojen vuoksi vähemmälle mielenkiinnolle ja tarkastelulle. Esimerkiksi

⁹ Ks. Horn, 2008.

vuonna 1980 julkaistu John Searlen artikkeli ”Minds, Brains and Programs” (jossa hän esitti kuuluisan ”kiinalaisen huoneen argumenttinsa”, joka esitellään lyhyesti luvussa 3) ikään kuin ”pyyhki pöydän” kaikista aiemmista argumenteista ja määritteli keskustelun uudelleen. Ajattelevia koneita käsittelevässä myöhemmässä keskustelussa ja tutkimuksessa usein tunnutaankin siirryttävän Turingin testin tarkastelusta suoraan Searlen kiinalaisen huoneen argumentin tarkasteluun ohittaen 30 vuotta mielenkiintoista filosofista keskustelua. Tässä työssä kuitenkin palataan keskustelun alkujuurille ja keskitytään pääasiassa siihen, pyrkien siten rekonstruoimaan keskustelun merkittävimmät teemat ja selittämään, miksi ne siellä ilmenevät.

Filosofiassa on pitkä perinne palata ideoiden alkujuurille ja tutkia filosofian historiaa. Usein ajatellaan, että kaikki lähtee Sokrateesta ja monen nykyfilosofinkin unelma varmasti olisi päästä aikamatkalle 2400 vuotta taaksepäin keskustelemaan hänen kanssaan – huolimatta siitä, että Sokratesta usein kuvataan melko ärsyttäväksi ihmiseksi. Sokrateen, Platonin, Aristoteleen ja vaikkapa myöhemmän Immanuel Kantin filosofiasta ajatellaan voitavan yhä löytää jotain uutta, vaikka niitä olisi jo vuosisatoja tulkittu, ja Aristoteleen tapauksessa jopa kanonisoitu katolisen kirkon opiksi. Hieman samassa hengessä tässä työssä palataan ajattelevien koneiden nykykeskustelun filosofisille juurille myös siinä toivossa, että se saattaisi suhteuttaa nykyisiä käsityksiämme uudella tavalla ja tarjota tilaisuuden mahdollisesti löytää uudelleen unohtuneita, mutta lupaavalta vaikuttavia filosofisia teemoja ja argumentteja nykyisen keskustelun edistämiseksi. Suoraa linkittämistä nykykeskusteluun tässä työssä ei kuitenkaan tilanpuutteen vuoksi kyetä tekemään. Tästä työstä saattaisivat siis hyötyä asiasta yleisesti kiinnostuneiden henkilöiden lisäksi osaltaan esimerkiksi historioitsijat, filosofit, mutta myös tekoälyn tutkijat, vaikkakin tietenkin melko rajatussa määrin tutkielman suppeuden vuoksi.

Tutkielman tavoitteena siis on muodostaa käsitys ajattelevia koneita käsittelevästä filosofisesta keskustelusta Englannissa vuosina 1950–1975 erittelemällä aineistossa esiintyviä teemoja ja selittämällä niitä niiden historiallisesta kontekstista käsin ja näin myös selventää muuten varsin hajanaista tieteellistä keskustelua. Tutkielman merkitys muodostuu historiallisen tiedon lisääntymisestä sekä siitä, että se nostaa esiin teemoja, jotka ovat ainakin jossain määrin olleet unohduksissa ja selittää niitä.

1.3 Tieteenhistoriallinen filosofian tutkimus

Tuukka Perhoniemi teoksessaan *Mitan muunnemat – miten määritämme maailmaa, ihmistä ja tietoa*¹⁰ erittelee neljä yleistasoista kysymystä, joihin historiallisia tekstejä ja aineistoja tutkittaessa joutuu pohtimaan vastauksia.

- (1) ”Miten huomioin anakronismeihin liittyvät ongelmat eli myöhemmin esitettyjen ajatusten lisäämisen menneisyyden teksteihin tai tapahtumiin?”
- (2) ”Voiko historiasta löytyä ratkaisuja nykyhetken kysymyksiin?
Millaisia ne ovat ja miten ne löytyvät?”
- (3) ”Miten ymmärrän historiallisten muutosten luonteen ja historian kulun yhtenäisyyden? Ovatko vanhemmat käsitykset alkeellisimpia tai ’väärempiä’ kuin uudemmat?”
- (4) ”Millaisia menneisyyden aineistoja tarkastelen ja mihin niissä kiinnitän huomiota?”¹¹

Varsinkin ensimmäinen kysymys on tämän työn kannalta olennainen, sillä esimerkiksi termi ”kone” on helppo nähdä anakronistisesti nykypäivän näkökulmasta osittain yli puolen vuosisadan takaisissa aineistoissa. Anakronismilla tarkoitetaan asian tai ajatuksen sijoittamista aikaan tai aikakauteen, johon se ei kuulu. Esimerkiksi W. A. Mozartin kutsuminen avantgarden edustajaksi (kuten Eero Tarasti tarkoituksenmukaisesti tekee¹²) on anakronismi, sillä Mozart eli 1700-luvulla ja avantgarde on jossain määrin 1800-luvun ja varsinkin 1900-luvun ilmiö ja käsite. Kuitenkin, kuten Perhoniemi toteaa ja kuten Tarasti havainnollistaa, anakronismi ei aina ole paha ja vältettävä ilmiö, vaan saattaa tuottaa uudenlaista historiallista ymmärrystä. Esimerkiksi Mozartin kuvaileminen avantgardistiksi alleviivaa hänen kokeiluntahtoaan ja pyrkimystään taiteen saavuttamiseen kaupallisuuden ja helppouden sijaan. Tarastin esimerkissä hän toteaa avantgarden suosivan pieniä (ooppera-) saleja ja tapahtuvan virallisen taide-elämän ulkopuolella, ja tämän vuoksi Mozartin ooppera *Taikahuilu* on Tarastin mukaan avantgardistinen – vaikkakin anakronistisesti. Jos Tarasti pyrkisi kokonaan välttämään anakronismin, hän käyttäisi Mozartia kuvatessaan ainoastaan 1700-luvun kontekstin käsitteitä ja termejä. Näin tekee esimerkiksi Umberto Eco

¹⁰ Perhoniemi, 2014.

¹¹ Perhoniemi, 2014, 190.

¹² Tarasti, 2013, 131.

romaanissaan *Ruusun nimi*¹³ 1300-luvun Pohjois-Italian kontekstissa ja rajoittaa romaanin italiankielisen sanaston 1300-luvulla käytettyihin sanoihin.

Anakronismi voi siis auttaa ymmärtämään menneisyyttä nykypäivän kontekstista käsin, mutta toisinaan se saattaa myös haitata sen ymmärtämistä. Oletettavasti tämän vuoksi siitä myös usein varoitetaan. Tämän työn kontekstissa anakronismista kuitenkin saattaisi olla enemmän haittaa kuin hyötyä. Kuten yllä mainitaan, ”kone” on muuttanut merkitystään sen mukaan, kuin koneet ovat kehittyneet ja tämän vuoksi 1950-luvun ”kone” tai vaikkapa ”tietokone” tarkoittaa ainakin jokseenkin eri asiaa kuin 2020-luvun ”kone” ja ”tietokone”. Merkittävin ero lienee se, että 1950-luvun ”koneet” ja ”tietokoneet” olivat yksinkertaisempia ja suurempia sekä ehkä äänekkäämpiäkin kuin 2020-luvulla vastaavilla termeillä kuvattavat asiat. Niiden merkitykset ja käsitteet, saussurelaisittain *signifiét* ovat muuttuneet. 1950-luvun kirjoittajat puhuessaan ”koneista” tarkoittivat suuria, usein mekaanisia, laskukoneita, joiden operoimiseen tarvittiin useita henkilöitä. Tosin, kuten Turingin artikkelista ”Computing Machinery and Intelligence” voi päätellä, hän ainakin odotti niiden kehittyvän monimutkaisemmiksi – jopa niin monimutkaisiksi, että 2000-luvun taitteessa ne jo läpäisisivät matkimispelin – tai anakronistisesti ”Turingin testin”. Perhoniemen kysymykseen (1) vastaisin siis, että pyrin välttämään anakronismia pitämällä mahdollisimman tarkkana mielessä, mitä koneilla, tietokoneilla ja muilla aineistossa esitettävillä asioilla tekstin kirjoittamisajankohtana on kaikkein luultavimmin tarkoitettu. On mahdotonta pyyhkiä kokonaan nykykontekstin vaikutusta käsityksiin, mutta historiallisen tilanteen selvittäminen auttaanee parhaaseen mahdolliseen *totuutta lähestyvään* tulokseen pääsemisen.

Perhoniemen kysymys (2) on työn merkityksen ja merkittävyyden kannalta vain jokseenkin mielenkiintoinen. Seuraavaksi sitä käsitellään tässä lyhyesti, sillä se on enemmänkin teoreettinen, eikä suoranaisesti tämän työn varsinainen merkitystä luova asia, sillä tässä tutkielmassa pyritään ensisijaisesti käsittelemään tutkimuskohdetta ilman pyrkimystä tuoda sitä varsinaisesti nykypäivään. Tämä johtuu siitä, että tällöin pitäisi joko käsitellä myös keskustelua vuosina 1975–2021 tai epämääräisesti vertailla keskustelua vuosina 1950–1975 sekä lähempänä nykypäivää. Tämä ei ole tämän työn tarkoitus.

Yleisesti voidaan sanoa, että historiasta on mahdollista löytää ratkaisumalleja nykypäivän ongelmiin ja tähän myös viitanee populaari sanonta ”se,

¹³ Eco, 1983.

joka ei historiaa tunne, on tuomittu sitä toistamaan”, viitaten oletettavasti erityisesti menneisyyden saatossa tapahtuneisiin virheisiin – tai ainakin päätöksiin, joita nykypäivän kontekstista pidämme virheinä (ja joita ne varmasti ainakin joissain tapauksissa eettisestä näkökulmasta olivatkin). Tässä tapauksessa nykyhetken kysymyksiin voisi löytyä vastauksia tai ainakin mahdollisia uudelleenkäytettäviä argumentteja esimerkiksi koneen toimijuuden suhteen. Jonkinlaisena moraalisenä ongelmana voidaan ajatella, että koneen toimijuus saattaa aiheuttaa moraalisia ja juridisia ongelmia, jos kone aiheuttaa vahinkoa ihmiselle tai omaisuudelle. Voidaan kysyä, oliko autonomiseksi määritelty kone se, joka vahingon aiheutti, vai esimerkiksi ohjelman koodin koodaamista johtanut henkilö. Aineistoni perusteella voisi esimerkiksi argumentoida, että *ajattelu* ja siten refleктоiva toimijuus, jota voidaan pitää vastuussa teoistaan, kuuluu elolliselle olennolle, jolla on myös esimerkiksi tunteita. Toisaalta voisi myös laajentaa *ajattelun* määritelmää koskemaan myös ei-elollisia toimijoita, jotka järkevästi ratkaisevat ongelmia. Voisi mahdollisesti ajatella, että ajattelevien koneiden keskustelu ei kovin paljon ole edistynyt. Tämä taas johtunee suurelta osin kielen rajoitteista ja siis siitä, että kielemme (oli se sitten esimerkiksi suomi, englanti tai vaikka saksa) on ymmärrettävästi pitkään rajoittanut *ajattelun* ihmisen ominaisuudeksi, eikä niinkään eläinten, saati koneiden ominaisuudeksi. Tämä liittyy laajempaan ongelmaan filosofiassa, joka on se, että verrattuna ajatteluamme, kielemme on varsin rajallinen ja siten ei kykene ilmaisemaan ajatuksiamme täysin kattavasti. Ja tämä saattaa päätyä toisinaan siihen, että kieli ohjaa ajattelua, mikä tietenkin on nähtävissä ajattelun rajoittumisena. Näin ainakin Ludwig Wittgenstein ongelman asetti ja *Tractatus logico-philosophicus* -teoksessaan, joka julkaistiin 1921, luuli ratkaisseensa tämän ongelman ja siten filosofian kaikki ongelmat.¹⁴ (Sittenmin hän tosin huomasi, ettei *Tractatus* kaikkia ongelmia sittenkään ratkaissut.) P. M. S. Hacker tekstissään ”Wittgenstein – ihmisluonnosta” toteaa varhaiskauden Wittgensteinin esittäneen filosofian ongelmien johtuvan nimenomaan kielen harhaanjohtavista piirteistä. Tämä johtuu siitä, että kielen erilaiset käsitteet esittäytyvät samankaltaisissa ulkoasuissa, esimerkiksi verbi ”olla” näyttäytyy samankaltaisena tai ”samantasoisena” kuin verbit ”istua” ja ”nukkua”, mutta ”olla”-verbi viittaa paljon syvällisempään, kuin ”istua” tai ”nukkua”.¹⁵ Jälkimmäisiin voi antaa suoraviivaisia vastauksia, joku joko istuu tai ei istu ja nukkuu tai ei nuku, mutta olemisesta ei voi antaa samankaltaisia suoria kyllä–ei-vastauksia.

¹⁴ Wittgenstein, 1963 (alkup. 1921).

¹⁵ Hacker, 2004, 655.

Tämänkaltaisesta kielen kyvyttömyydestä edustaa todellisuutta voidaan siis wittgensteinilaisittain ajatella monien filosofisten kysymysten kumpuavan. Myös tässä työssä termit ja käsitteet ovat aineistossa tärkeässä asemassa ja niiden pohtiminen saattaa auttaa asettamaan nykyistä ajattelevien koneiden keskustelua jonkinlaiseen historialliseen kontekstiin.

Perhoniemen kysymys (3) ei ole kovin relevantti tämän työn puitteissa, sillä se ei pyri juurikaan tuomaan nykyisyyteen käsiteltävän ajanjakson keskustelua. Monet ympäristön tekijät (kuten teknologian kehitys ja yhteiskunnallinen muutos) ovat luultavasti muokanneet keskustelua, mutta se ei vielä tee aiemmista käsityksistä sen virheellisempiä tai oikeampia – ainakaan relativistisesta näkökulmasta käsin. Kirjoitukset heijastavat omaa aikaansa ja aikansa yhteiskuntaa ja teknologiaa eikä niitä voida suoraviivaisesti tuomita vääriksi tai julistaa oikeiksi nykypäivän kontekstista. Kysymys (4), jonka Perhoniemi esittää, on mielenkiintoisempi ja sitä käsitellään tarkemmin luvussa 2.

Richard Rorty esittelee kaksi erilaista tapaa lähestyä historiallista tutkimusta. Hän kutsuu niitä *historialliseksi rekonstruktioksi* ja *rationaaliseksi konstruktioksi*.¹⁶ Tämä työ on pääasiassa historiallista rekonstruktiota, eli tässä käytetään historiakeskeistä lähestymistapaa. Siinä historiallinen konteksti pyritään luomaan mahdollisimman tarkasti ja pyritään tarkastelemaan tapahtumia sen ajan kontekstin ehdoilla. Tällöin esimerkiksi voidaan hieman helpommin välttää jonkin verran harmillisia vahingonomaisia anakronistisia oletuksia ja johtopäätöksiä. Toinen Rortyn kuvaama lähestymistapa on rationaalinen konstruktio eli nykyhetkikeskeinen lähestymistapa, jossa historiaa tarkastellaan nykyhetken ongelmista ja lähtökohdista sekä käsitteistä ja määritteistä käsin. Esimerkiksi sen kysyminen, mitä Aristoteles ajattelisi Turingin matkimispelistä, voisi olla tämänkaltaista nykyhetkikeskeistä filosofista tutkimusta, jossa selvitetään Aristoteleen ajattelua niin laajasti, että siltä pohjalta voisi tehdä johtopäätöksiä hänen ajattelustaan nykypäivän ongelmiin. Tässä anakronismin vaara on suurempi, vaikka kuten yllä on todettu, voi anakronismi toisinaan olla oikeutettu ja hyvä tapa jäsentää menneisyyttä nykyisyydestä käsin. Rationaalisen konstruktion tutkimusotetta ei tässä kuitenkaan lähtökohtaisesti oteta käyttöön.

Kuitenkin siinä missä puhtaan aineistolähtöinenkään tutkimus ei ole mahdollista, kuten yllä todettiin, ei kuitenkaan puhdas historiallinen rekonstruktio ole

¹⁶ Rorty, 1984.

myöskään täysin mahdollinen. On käytännössä mahdotonta käsitellä menneisyyttä puhtaasti menneisyyden itsensä ehdoilla ilman minkäänlaista nykypäivän kontekstin vaikutusta tutkimuskohteen tulkintaan. Tämä saattaa esimerkiksi tapahtua nykypäivän kielen ja käsitysten vaikutuksista tulkintoihin silloin kun vaikutuksia ei täysin pystytä eliminoimaan. Tavoitteena tietenkin aina on mahdollisimman tarkka ulkopuolisten, kuten nykypäivän kontekstin, vaikutusten eliminoiminen, mutta käytännössä se ei ole mahdollista, paitsi ehkä teoreettisella tasolla. Näin siis pieniä rationaalisen rekonstruktion piirteitä esiintyy mahdollisesti tässäkin tutkielmassa.

1.4 Aikaisempi tutkimus

Alan Turingin työtä on tutkittu aiemmin paljon hänen vaikuttavuutensa takia, niin historiallisesti kuin filosofisestikin. Hänestä ja hänen työstään on kirjoitettu useita melko yleisiä kirjoja, kuten esimerkiksi Jack Copelandin ym. teos *The Turing Guide*¹⁷, Robert Epsteinin, Gary Robertsin sekä Grace Beberin toimittama teos *Parsing the Turing Test – Philosophical and Methodological Issues in the Quest for the Thinking Computer*¹⁸, Juliet Floydin sekä Alisa Bokulichin toimittama *Philosophical Explorations of the Legacy of Alan Turing – Turing 100*¹⁹, Simon Lavingtonin toimittama *Alan Turing and His Contemporaries – Building the World's First Computers*²⁰ sekä S. Barry Cooperin ja Jan van Leeuwenin toimittama *Alan Turing – His Work and Impact*²¹.

Tämän työn kaltaista ajattelevien koneiden teemoja 1950–1970-lukujen englantilaisissa akateemisissa filosofian jouluaikakauslehdissä historiallisesti selittävää ja erittelevää tutkimusta ei suoraan ole tehty. Robert Horn esittelee artikkelissaan ”The Turing Test – Mapping and Navigating the Debate”, miten filosofien tarpeisiin on luotu argumenttikarttoja, jotka kattavat koko keskustelun Turingin testistä ja joiden avulla voi esimerkiksi selvittää, mitkä argumentit ovat filosofisessa keskustelussa jo kumottu ja mitkä ovat vielä ratkaisematta ja siten vielä avoinna keskustelulle.²² Myös Turingin

¹⁷ Copeland ym, 2017.

¹⁸ Epstein ym, 2008.

¹⁹ Floyd ja Bokulich, 2017.

²⁰ Lavington, 2012.

²¹ Cooper ja van Leeuwen, 2013.

²² Horn, 2008.

artikkelia ”Computing Machinery...”²³ on käsitelty sen teemoja eritellen ja kontekstualisoiden ja esimerkiksi joitain kommentaareja on kirjoitettu.

Simone Natale ja Andrea Ballatore esittelevät artikkelissaan ”Imagining the thinking machine: Technological myths and the rise of artificial intelligence”²⁴, miten he ovat selvittäneet tekoälyyn liittyviä myyttejä 1950-luvulta 1970-luvulle lehdissä *Scientific American* ja *New Scientist*. Heidän työnsä toimi jossain määrin alkusysäyksenä tälle työlle. Erotuksena heidän työhönsä, käsillä olevassa tutkielmassa tutkitaan kuitenkin ajattelevien koneiden keskustelun *teemoja*, kun he tutkivat, mitä *myyttejä* keskustelussa esiintyi. Teemojen käsite on selkeämpi ja käytännöllisempi, joten sen vuoksi tässä on pitäydytty siinä myytin käsitteen sijaan. Lisäksi heidän tutkimuksensa käsittelee lehtiä, jotka ovat yleistieteellisiä, ja joihin kirjoittaneet olivat monilta eri tieteen aloilta, kun taas tämän työn aineisto koostuu filosofian alan tieteellisissä julkaisuissa julkaistuista artikkeleista.

Turingin yliaistillinen havaitseminen -teemaa on Turing-tutkimuksessa käsitelty hyvin vähän. David Leavitt käsittelee sitä jonkin verran ja sen mahdollista alkuperää kirjoituksessaan ”Turing and the Paranormal”²⁵, mutta muuten aihe on jäänyt vähäisemmälle huomiolle. Kuten Leavittkin toteaa, se usein ohitetaan muutamalla sanalla tai olankohautuksella.²⁶ Tämä johtunee siitä, että varsinkin filosofit ovat ensisijaisesti kiinnostuneita argumenteista sekä jonkinlaisesta itse totuuden perimmäisestä löytämisestä, ja telepatian pitäminen mahdollisuutena (ja siitä kirjoittaminen filosofian tutkimuksissa) ei ole siksikään kovin uskottavaa. Turingin väitteen alkuperän ja tarkoituksen selvittäminen on historiallisessa mielessä mielenkiintoista, ja hänen henkilöahmoaan mahdollisesti valottavaa, vaikkei se matkimispelin teesiä filosofisessa mielessä edistäisikään filosofian näkökulmasta.

1.5 Käsitteet ja termit

Käsitteellä tässä työssä tarkoitetaan tavalliseen tapaan jonkin asian merkitystä, abstraktia ideaa, kun taas *termillä* käsitteeseen viittaavaa sanaa. Käsitteet *kone* ja *ajattelu* (englanniksi *machine* ja *thinking*) ovat tässä työssä merkittävässä asemassa.

²³ Turing, 1950.

²⁴ Natale & Ballatore, 2020.

²⁵ Leavitt, 2017.

²⁶ Leavitt, 2017, 347.

Ainakin käsitteen *kone* merkityksen voidaan ajatella hieman muuttuneen tämän työn tarkasteluajankohdan alusta (vuodesta 1950) nykypäivään. 1950-luvun koneet – varsinkin tietokoneet – olivat yleensä suurempia ja yksinkertaisempia. Kyse on kuitenkin melko läheisestä menneisyydestä, joten voidaan pitää oletettavana, että lukija on tietoinen käsitteen *kone* lievästä muutoksesta. Vaikka ei olisi, se tuskin ratkaisevalla tavalla häiritseisi tämän työn tutkimuskohteen esittämän *koneen* käsitteen ymmärtämistä – onhan kuitenkin kyse jonkinlaisista tietokoneista. Käsitteiden *kone* ja *ajattelu* tarkempaa analyysia tutkimuksen aineistossa esitellään käsittelyluvuissa.

Menneisyys on se, mitä todellisuudessa ennen nykyhetkeä on tapahtunut ja *historia* on kirjoitusta siitä. Historia ei siis voi täysin tavoittaa menneisyyttä, kuten termikään ei täysin voi tavoittaa käsitettä, mutta totuuden lähestyminen on sille kuitenkin mahdollista tarkentuvien metodien ja historiallisten kontekstien ymmärtämisen lisääntyessä.

Kun tutkimuskohteena on englanninkielisiä tekstejä, on *Oxford English Dictionary* (lyhennettynä ”OED”) nähty tässä tutkielmassa luontevimmaksi sanakirjaksi. Sen nykyinen versio, jota on mahdollista tarkastella internetissä²⁷, ei tietenkään ota tarkalla tavalla huomioon käsitteiden ja termien muutoksia, mutta kuten yllä on huomioitu, merkittävin muutos on tapahtunut käsitteen *kone* kohdalla. Yleensä sanakirjamääritelmiin ei ole tapana viitata, mutta tässä olen selvyuden vuoksi viitannut tarvittaessa.

Tutkielman kielen kannalta on pyritty mahdollisimman tarkkaan selvyyteen, mutta pyritty pitäytymään tieteellisessä tekstityylissä ja tutkimuskohdetta käsittelevissä asianmukaisissa käsitteissä. Tutkielman tekstillinen tyyli saattaa paikoin vaikuttaa haastavalta, mutta tällöin se johtuu ensisijaisesti käsiteltävän asian käsitteellisestä monimutkaisuudesta ja/tai abstraktista luonteesta. Pitkiä virkkeitä on pyritty välttämään, mutta siinä ei aina ole onnistuttu, sillä arvoasteikolla tärkeämpänä on pidetty itse asian mahdollisimman tarkkaa ja havainnollistavaa käsittelyä, vaikka tämä tarkoittaisikin toisinaan haastavia virkkeitä.

²⁷ OED, Linkki: <https://www.lexico.com/>. Linkki tarkistettu 23.4.2021.

2. Aineisto ja menetelmät

2.1 Aineistonkeruu ja aineisto

Tutkielman aineisto kerättiin JSTOR-nimisestä digitaalisesta kirjastosta keväällä 2021. Aineisto rajattiin kattamaan merkittävien englantilaisten akateemisten filosofisten joulun kirjoitukset ajattelevista koneista aikavälillä 1950–1975. Käytännössä rajaus tarkoitti sitä, että tutkielmassa keskityttiin tarkastelemaan Oxford University Pressin sekä Cambridge University Pressin julkaisemia filosofian alan joulunlehtiä, jotka julkaisivat numeroita aktiivisesti ja joiden numerot vuosien 1950 ja 1975 välillä löytyvät JSTOR-kirjastosta. Jos jokin tarkasteltavan julkaisun numero ei löytynyt sieltä, siinä mahdollisesti julkaistu ajattelevia koneita käsittelevä artikkeli ei tällöin päätynyt tämän tutkimuksen aineistoon. Tosin todennäköisyys sille, että JSTOR-kirjastosta puuttuisi numero, jossa on ajattelevia koneita käsittelevä artikkeli, on kohtalaisen pieni, ja voidaan lukea virhemarginaaliin.

Ehdot täyttäviä joulunlehtiä olivat (aakkosjärjestyksessä) *Analysis*, *The British Journal for the Philosophy of Science*, *Mind*, *The Philosophical Quarterly* sekä *Philosophy*. Lisäksi ehdot täytti *Proceedings of the Aristotelian Society*, mutta kyseisessä julkaisussa ei tarkasteluvälin aikana julkaistu yhtään ajattelevia koneita käsittelevää artikkelia, joten sitä ei luettu aineistoksi. Myös esimerkiksi *The Monist* olisi ollut muuten sopiva joulunlehti tarkasteltavaksi, mutta koska tässä tutkielmassa pyritään saamaan kuva keskustelusta koko vuosien 1950–1975 välisenä aikana ja koska *The Monist* joko ei julkaissut 1940– ja 1950-luvuilla (vaan 1890–1936 ja jälleen 1962 alkaen) tai JSTOR ei kata kyseisiä vuosia, ei sitä valittu tässä tutkimuksessa tarkasteltavaksi joulunlehtiksi.

Kun joulunlehdit oli valittu, kävin läpi jokaisen joulunlehdin jokaisen numeron tarkasteluväliltä ja poimin aineistoksi ne artikkelit ja alustavasti myös *Discussion*-osioiden kirjoitukset, joiden otsikossa mainittiin ainakin yksi seuraavista sanoista: ”machine”, ”robot”, ”artificial intelligence”, ”AI”, ”computer”, ”computation”, ”computing”, ”Turing”, ”imitation game” tai ”Turing(’s) test”. Lisäksi sana ”artefact” yhdistettynä esimerkiksi sanaan ”mind” tai ”thinking” tarkoitti, että valitsin artikkelin aineistoon. Kuitenkin artikkelit, joiden otsikoissa oli esimerkiksi ”cybernetics” (tai vastaava) tai ”mechanistic” tai ”technology” jäivät tarkastelun ulkopuolelle, sillä ne eivät riittävän tiiviisti keskittyneet tutkielman aiheeseen. Aineiston ensimmäisen

keräämisvaiheen jälkeen kävin lyhyesti aineiston läpi ja rajasin pois artikkelit, jotka eivät selkeästi pyrkinet käsittelemään ajattelevia koneita, vaikka jokin yllä esitetyistä sanoista olisikin esiintynyt niiden otsikoissa. Esimerkiksi jornaalissa *Proceedings of the Aristotelian Society* on Betty Powellin artikkeli ”Descartes’ Machines”²⁸, joka rajautui pois, sillä se käsitteli, ei ajattelevia koneita 1900-luvun kontekstissa, vaan Descartesin kehittämää koneen määritelmää.

Tällä jornaalien numerojen silmämääräisellä läpikäymisellä painotin tämän tutkielman aineistolähtöisyyttä ja siten tarkasteluun päätyi artikkeleita, jotka tavanomaisessa tekstihaussa olisivat jääneet tarkastelun ulkopuolelle. Esimerkiksi en olisi tekstihaussa tullut ajatelleeksi hakea sanalla ”artefact”, koska se saattaisi viitata hyvin laajaan käsitejoukkoon, mutta koska havaitsin erään artikkelin otsikossa sanan ”artefact” sekä ”mind”, osasin ottaa sen mukaan aineistoon.

Tämän vaiheen jälkeen aineistona oli yhä noin 50 tekstiä, joten rajatakseni tätä realistisemmaksi, keskityin pelkkiin artikkeleihin ja rajasin *Discussion*-osioiden tekstit pois. Tämä rajasi aineistoa huomattavasti järkevämpään kokoon. Jornaalien välillä oli jonkin verran hajontaa artikkelien määrien suhteen: *The British Journal for the Philosophy of Science* -jornaalin artikkelit muodostavat puolet aineistostani. *Analysis*-jornaalista on kolme, *Mind*-jornaalista myös kolme, *The Philosophical Quarterly*stä kaksi ja *Philosophystä* kolme artikkelia.

Tässä tutkielmassa esitellään parhaiten kulloinkin tarkasteltavaa teemaa edustavia tapausesimerkkejä kerätystä aineistosta ja pyritään niiden avulla analysoimaan teemaa ja sen syitä ja vaikutuksia ongelman ratkaisemiseen. Kaikkia kerättyjä artikkeleita ei tässä siis erikseen esitellä, sillä samat teemat esiintyivät useissa artikkeleissa ja olisi siten tarpeetonta käsitellä saman teeman esiintymistä eri artikkeleissa.

Tutkimusongelmaani siitä, mitä filosofit ajattelivat ammatillisesti ja julkaisukelpoisesti ajattelevista koneista vuosina 1950–1975 ja miten se näkyi heidän kirjoittamissaan artikkeleissa, jotka julkaistiin merkittävässä englantilaisissa filosofian alan akateemisissa julkaisuissa, on tämä aineisto melko edustava. Toivottavampaa tietenkin olisi, jos tilan puitteissa pystyisi myös analysoimaan jornaalien *Discussion*-osioiden tekstit, sillä niissä filosofit hieman artikkeleja epämuodollisemmin ilmaisivat ajatuksiaan aiemmin kirjoitetuista artikkeleista sekä niiden mahdollisista ongelmista ja tämä olisi saattanut olla mielenkiintoinen tapa tarkastella hieman epävirallisempaa,

²⁸ Powell, 1970.

mutta kuitenkin akateemiset standardit täyttävää keskustelua ajattelevista koneista. Tilan ja ajan puutteen vuoksi tämä laajempi tarkastelu kuitenkin jää mahdollisia tulevaisuudessa tapahtuvia tarkasteluja varten. Olisin tietenkin voinut rajata journalien määrää ja siten vähentää tarkasteltavien artikkelien määrää ja näin kyetä ottamaan myös *Discussion*-osien tekstit huomioon, mutta katsoin aineiston yhtenäisyyden ja tarkasteltavien journalien suuremman määrän (ja siis paremman yleistettävyyden) vuoksi, että on parempi keskittyä pelkkiin artikkeleihin useammasta journalista. Keskustelu ajattelevista koneista myös tapahtui melko hajautuneesti eri journalien kesken, joten oli luontevampaa käsitellä useampaa journalia ja rajata tutkittava tekstityyppi artikkeleihin. Näin pyrin varmistamaan mahdollisimman sopivan ja edustavan tarkastelun tässä tutkielmassa.

Merkittäviä tutkimuseettisiä ongelmia tämänkaltaisen aineiston kanssa ei juuri tule vastaan. Aineiston muodostavien tekstien kirjoittajat ovat halunneet kirjoittamansa artikkelit mahdollisimman monien nähtäväksi ja ovat aktiivisesti pyrkineet saamaan ne julkisiksi. Tällöin ei ole vaaraa siitä, että tutkittava olisi esimerkiksi ollut tietämätön siitä, että hänen tuottamansa aineisto olisikin julkisempi, kuin mitä hän on ajatellut. Tämä työ ei myöskään tuo aineiston tekstejä julkisemmiksi kuin mitä ne jo ovat, joten tässäkin suhteessa tutkimuseettisiä ja mahdollisia juridisia ongelmia aineiston julkisuudesta tai tekijänoikeuksista ei tämän työn tapauksessa tule eteen. Tässä työssä ei ole tarvetta esimerkiksi anonymisoida tekstejä tai muuten suojella kirjoittajan yksityisyyttä, vaan voi tieteellisen käytänteen mukaisesti yksinkertaisesti viitata aineistoon tekijän nimellä ja tekstin julkaisuvuodella.

2.2 Laadullinen sisällönanalyysi

Laadullisella sisällönanalyysillä tarkoitetaan tässä laadulliseen tutkimukseen kallellaan olevaa sisällönanalyysiä, eli aineiston tarkastelua erityisesti teemoitellen, pyrkien löytämään ja erittelemään aineistosta yhtäläisyyksiä ja eroavaisuuksia ja pohtimalla niiden merkitystä aineistokokonaisuudessa.²⁹ Laadullisen (tai ”kvalitatiivisen”) analyysin erityispiirteenä voidaan pitää pyrkimystä käsitellä aineistoa kokonaisuutena ja

²⁹ Saaranen-Kauppinen ja Puusniekka, 2006.

esimerkiksi tarkastella sen sisäistä logiikkaa pyrkien sen ymmärtämiseen ja selittämiseen.³⁰

Pertti Alasuutari kuvaa laadullista analyysiä kaksivaiheiseksi: ensimmäinen on *havaintojen pelkistäminen* ja toinen *arvoituksen ratkaiseminen*. Havaintojen pelkistäminen on hänen mukaansa ensinnä aineiston tarkastelua valitusta teoreettis-metodologisesta näkökulmasta käsin ja huomioidaan aineistossa lähtökohtaisesti vain sen kannalta olennaiset seikat. Näitä hän kutsuu ”raakahavainnoiksi”. Tämän jälkeen hänen mukaansa raakahavainnoja yhdistetään esimerkiksi etsimällä yhteisiä piirteitä ja nimittäjiä havaintojen kesken. Tämä luonnollisesti vähentää havaintojen määrää.³¹ Tässä työssä havainnoja pelkistetään lukemalla aineistoa ja tekemällä ”raakahavainnoja” teksteistä esimerkiksi niiden sivujen marginaaleihin tai erilliseen muistiinpanovihkoon tai -tiedostoon. Tämän jälkeen vertailemalla raakahavainnoja ja etsimällä niiden kesken yhteisiä piirteitä muodostuu havainnoista kokonaisuuksia, joissa on tiettyjä yhteisiä tekijöitä, esimerkiksi tietynkaltaisia argumentteja tai jonkin tietyn ajattelevia koneita koskevan käsitteen tai teeman käsittelyä.

Arvoituksen ratkaisemisella laadullisessa tutkimuksessa Alasuutari tarkoittaa merkitystulkinnan tekemistä tutkittavasta ilmiöstä niiden havaintokokonaisuuksien perusteella, jotka analyysin *havaintojen pelkistäminen* -osassa ovat muodostuneet.³² Tässä tutkielmassa arvoituksen ratkaiseminen on siis teemojen merkityksien tulkitsemista ja niiden linkittämistä laajempaan filosofiseen keskusteluun varsinkin ajattelevista koneista, mutta myös muihin läheisiin keskusteluihin, kuten mielenfilosofian teemoista käytävään keskusteluun. Toisaalta tiettyjen teemojen jatkuvan esiintymisen selittäminen esimerkiksi vallinneiden filosofisten aatevirtausten avulla voidaan nähdä tämänkaltaisena ”arvoituksen” ratkaisemisena.

Tässä työssä siis erityisesti aineiston muodostavien journaaliartikkelien sisältöjä teemoitteleamalla ja havaintojen yhteyksiä sekä niiden merkityksiä tarkastelemalla pyritään muodostamaan historiallinen konstruktio ajattelevia koneita käsittelevästä keskustelusta englantilaisessa akateemisessa filosofiassa vuosina 1950–1975.

³⁰ Alasuutari, 2011, 30.

³¹ Alasuutari, 2011, 31.

³² Alasuutari, 2011, 34.

3. Ajattelevien koneiden idean taustaa

3.1 Ajattelevien koneiden yleistä historiaa

Ajattelevien koneiden idea on melko vanha ja jossain määrin pyrkimys sellaisen rakentamiseen vaikuttaisi myös melko pitkään olleen ihmisten ajatuksissa. Esimerkiksi keskiaikainen juutalainen Golem-myytti kuvaa ihmisen valmistamaa ihmismäistä konetta, jota voidaan *Sefer Jezira* -kirjan kirjainyhdistelmillä ohjelmoida (anakronistisesti ilmaistuna), eli vaikuttaa sen toimintaan.³³ Toisaalta myös Frankensteinin hirviön tarina ilmentää samankaltaista ihmiskoneen luomisen myyttiä. Myös Leonardo da Vinci hahmotteli vastaavaa ”automaattia” 1400– ja 1500-lukujen vaihteessa.

Kuitenkin vasta 1800-luvulla alkoi tulemaan esiin hieman uskottavampia ja vähemmän myyttisiä tapoja pyrkiä ajattelevan koneen luomiseen. Englantilainen matemaatikko ja insinööri Charles Babbage (1792–1871) kehitti ”analyttisen koneen” (engl. ”Analytical Engine”), joka toimi reikäkorttipohjaisena laskukoneena. Analyttinen kone ja siihen Ada Lovelacen (1815–1852) kehittämä laskuohjelma voidaan katsoa ensimmäisiksi tietokoneeksi ja tietokoneohjelmaksi, mutta ne jäivät lopulta jokseenkin vähemmälle huomiolle.³⁴

Seuraava merkittävä askel ajattelevien koneiden historiassa oli kybernetiikka ja siitä käydyt keskustelut 1940-luvulta lähtien. Kybernetiikan perusajatuksena on, että kybernetiikka ja kyberneettiset koneet pyrkivät taistelemaan kaaosta, entropiaa, vastaan informaation prosessoinnin kautta.³⁵ Jussi Parikka pitää sitä jopa nykykulttuurin ajattelumaailman ja modernin tietokoneellisen ajattelutavan kiteytymänä.³⁶ Hän viittaa tällä siihen, miten kulttuurimme pyrkimyksenä on hallinta ja järjestyksen luominen maailmassa, joka vaikuttaisi muuten vajoavan kaaokseen, kuten entropian termodynamiikasta lainattu perusperiaate vaikuttaisi suljetussa systeemissä osoittavan. Toisaalta erityisesti informaatioajan tavoite on tässä mielessä hallitseminen ja entropiaa vastaan taisteleminen tietojärjestelmien avulla. Parikka käsittelee teemaa siis erityisesti foucaultlaisesta vallan ja vallankäytön kritiikin näkökulmasta.

³³ Mainzer, 2016, 7.

³⁴ Mainzer, 2016, 8–9.

³⁵ Parikka, 2004, 117.

³⁶ Parikka, 2004, 118.

Kybernetiikan uranuurtajana pidetään Norbert Wieneriä, joka pyrki edistämään monitieteistä keskustelua kybernetiikasta ja osaltaan taistelemaan akateemisen maailman eri alojen erityistymistä vastaan. Siis ilmiötä, josta oli viitteitä jo 1940-luvulla, ja joka sittemmin on vain voimistunut, jossa tieteenalat erkanevat toisistaan ja alojen välimaastoihin mahdollisesti syntyy niin syviä kuiluja, että eri alojen tutkijat eivät käytännössä ymmärrä toistensa tieteellistä puhetta. Wiener oli matemaatikko ja kybernetiikka-liikkeessä mukana olevia muita aloja aluksi olivat neurofysiologia sekä ohjaustekniikka. Pian mukaan tuli myös matemaattisen logiikan, automaatioteorian, psykologian sekä sosioekonomian edustajia.³⁷

Kybernetiikan merkittävimmät käsitteet ovat *takaisinkytkentä* tai *palautesilmukka* (engl. *feedback*), *entropia* (engl. *entropy*), sekä *informaatio* (engl. *information*). Takaisinkytkennän, eli palautesilmukan perusajatuksena on, että järjestelmän *lähtösignaali* (engl. *output*) ohjataan silmukkana takaisin järjestelmän *tulosignaali*ksi (engl. *input*). Takaisinkytkentöjä varsinkin kybernetiikan kontekstissa voi olla sekä *positiivisia* että *negatiivisia*. Positiiviset takaisinkytkennät vahvistavat signaalia ja negatiiviset heikentävät sitä.³⁸

Entropialla tarkoitetaan suljetun systeemin sisällä tapahtuvaa energian tasoittumista ja siis pyrkimystä tasapainotilaan tai ”lämpökuolemaan”. Tämä on termodynamiikasta lainattu käsite. Informaatiolla taas viestien lähettämiseen kommunikaatiokanavia käyttäen sen tehostaminen mahdollisimman pitkälle.³⁹ Näitä käsitteitä voisi analysoida pitkällekin, mutta tässä lienee riittävää tuntee ne vain pääperiaatteissaan, jotta ajattelevien koneiden alkutilannetta voisi tämän tutkielman kannalta riittävässä määrin hahmottaa. Varsinaista tutkimusta näistä löytyy tarvittaessa paljonkin.⁴⁰

Kybernetiikan johtavana ajatuksena oli siis näistä lähtökohdista ja käsitteistä lähtien pyrkiä kehittämään järjestelmää, joka ”taistelisi kaaosta vastaan”, kuten Parikka sen kuvailevasti esittää. Tähän ajateltiin päästävän siis takaisinkytkentöjen kautta kehittämällä oppiva järjestelmä, joka kehittyisi paremmaksi ja monimutkaisemmaksi sekä tehokkaammaksi itsestään sen jälkeen, kun se on laitettu alulle tai käyntiin.

³⁷ Sayre, 2004, 292.

³⁸ Sayre, 2004, 294.

³⁹ Sayre, 2004, 296–297.

⁴⁰ Sayre, 2004.

Merkittävänä kybernetiikan merkkipaaluna olivat 1940-luvun lopulta 1950-luvun alkupuoliskolle New Yorkissa järjestetyt Macy-konferenssit. Niihin kokoontui monien eri alojen asiantuntijoita useista eri maista. Myös esimerkiksi Alan Turing osallistui niihin 1940-luvun lopulla.

Toisaalla varsinaisen perustavaa laatua olleen tekoälytutkimuksen parissa kesällä 1956 kymmenen tutkijaa työskenteli Dartmouth Collegessa, New Hampshiren osavaltiossa Yhdysvalloissa järjestetyssä workshopissa. Sen oli ideoinut ja järjestänyt pääasiassa matemaatikko John McCarthy, joka oli inspiroitunut Turingin artikkelissaan ”Computing Machinery and Intelligence” esittämästä kysymyksestä ”voivatko koneet ajatella” ja halusi osaltaan olla selvittämässä vastausta kysymykseen.⁴¹ Hän oli saanut Rockefeller Foundation -säätöiltä apurahan ”Dartmouth Summer Project”-nimisen hankkeen järjestämiseen ja tätä pidetään usein varsinaisena *tekoälyn* tutkimuksen alkupisteenä.⁴² McCarthy rahoitushakemuksessaan käytti termiä ”tekoäly” (engl. ”artificial intelligence”) tiettävästi ensimmäistä kertaa. Hän pyrki näin erottamaan oman projektinsa jo aiemmin samankaltaisia ajatuksia pohtineesta kybernetiikasta käyttämällä uutta keksimäänsä tekoälyn termiä.

Dartmouth-projektista katsotaan alkaneen tekoälytutkimuksen ensimmäinen aalto, jonka aikana innostus oli suurta ja sen mukaisesti myös odotukset olivat suuria. Tutkimusta rahoitettiin merkittävästi ja jonkin verran myös teknologia edistyi. Kuitenkin 1970-luvun alussa realiteetit alkoivat tulla vastaan ja innostus aiheeseen jossain määrin lopahti. Äänensävyt muuttuivat kriittisemmiksi ja rahoitusta tekoälytutkimukselle oli vaikea saada. Noin vuonna 1975 katsotaankin alkaneeksi niin kutsuttu ensimmäinen ”tekoälytalvi” (engl. ”AI-winter”), jonka aikana oikein kukaan ei halunnut tulla edes yhdistetyksi tekoälyyn liittyvään tutkimukseen.⁴³ Toisaalta tekoälytutkimuksen ensimmäinen aalto on ajoitettu myös 1950-luvuun puolivälistä 1960-luvun puoliväliin.⁴⁴ Tässä tutkielmassa tarkastellaan vuosia 1950–1975 tekoälytutkimuksen ensimmäisenä jaksona ja vuosi 1975 nähdään luontevana päätepisteenä tämänkokoisen tutkielman aineiston tarkasteluajanjaksolle myös mahdollisen keskustelun hiipumisen havainnoimiseksi. Voi olla paljon odotettua, että edes Turingin artikkelista alkanut keskustelu jatkuisi aktiivisena samoissa lehdissä yli

⁴¹ Mainzer, 2016, 11.

⁴² Bostrom, 2014, 6.

⁴³ Bostrom, 2014, 8.

⁴⁴ Mainzer, 2016, 11.

25 vuotta, sillä tuollaisella aikavälillä ehtii myös monia muita mielenkiintoisia keskusteluteemoja nousemaan esiin.

3.2 Akateeminen filosofia Englannissa 1900-luvun puolivälissä

Englantilaisessa filosofiassa merkittävimmät yliopistot 1900-luvulla olivat Oxfordin ja Cambridgen yliopistot. Niissä analyttisen filosofian perinne eli vahvana ja Englanti on muutenkin kuulunut analyttisen filosofian ydinalueisiin. Merkittävimpiä filosofeja 1900-luvun alkupuolella ja puolivälissä olivat Cambridgessa muiden muassa Ludwig Wittgenstein, G. E. Moore, Bertrand Russell sekä C. D. Broad. Oxfordissa taas ajan merkittäviä filosofeja olivat esimerkiksi A. J. Ayer ja J. L. Austin.

Analyttisessä filosofiassa painottuu erityisesti ongelmien hahmottaminen ja niihin vastausten löytäminen loogisten ja matemaattisten ratkaisutapojen avulla. Analyttisen filosofian erityisiä kiinnostuksenkohteita ovatkin olleet varsinkin tieteenfilosofia ja logiikka, matematiikan filosofia, mielenfilosofia ja jossain määrin myös metafysiikka ja yhteiskuntafilosofiakin sekä etiikka. Analyttinen perinne yleensä erotetaan mannermaisen filosofian perinteestä, joka nimensäkinmukaisesti vaikuttaa erityisesti Ranskan ja Saksan alueella. Varsinkin nykyään erottelut kuitenkin ovat sumentuneet ja varsinkin filosofian laitoksilla analyttinen perinne on nykyään melko voimakas myös esimerkiksi Saksan yliopistoissa. Mannermainen perinne taas on jossain määrin siirtynyt humanistisille aloille taustateorioiden pohjustukseksi. Esimerkiksi Husserlin fenomenologia ja Foucaultin kriittiset teemat ja Frankfurtin koulukunnan ajatukset ovat perinteisesti mannermaiseen filosofiaan luettuja ja nykyään usein humanistisilla ja yhteiskuntatieteellisillä aloilla esiintyviä.

Filosofiaa yleisemmällä tasolla ilmentää keskustelevuus ja argumenttien tärkeys. Filosofia on perimmiltään keskustelua eri kysymyksistä ja monet keskusteluista ovat saaneet alkunsa jo yli 2000 vuotta sitten, eräänä tämän työn kannalta mielenkiintoisena esimerkkinä voisi pitää mielenfilosofista mieli–ruumis-ongelmaa, jossa pyritään selvittämään, onko mieli ruumiista irrallinen ilmiö ja jos on, niin miten se voi vaikuttaa ruumiiseen ja ruumis siihen. (Suomenkielisessä kirjallisuudessa käytetään yhä vanhahtavasti termiä ”ruumis”, vaikka ”keho” ehkä olisi hieman luontevampi termi viitatessaan elolliseen ruumiiseen. Elottomalla ruumiilla tuskin oikein kenenkään mielestä on enää mieltä.)

Filosofian keskukseluveduuden aiheuttama piirre, joka osaltaan erottaa filosofian alaa monista muista akateemisista oppiaineista, on se, että monet filosofian harjoittajat ovat lähes kaikkea kommentoivia filosofeja. Useat siis antavat oman mielipiteensä argumentteineen käynnissä olevaan keskusteluun, vaikka se ei olisi heidän erikoisosaamiseensa liittyvä keskustelu. Tämä oli tosin ilmeisempää tämän työn tarkasteluajankohtana, 1950–1975, kuin nykypäivänä. Nykyään on niin monta keskustelua käynnissä monella taholla, että filosofin on helpompi käydä erikoisosaamiseensa liittyvää keskustelua.

Puhuttaessa vuosien 1950–1975 Englannista, pitää muistaa myös toisen maailmansodan päättymisen 1945 ja sen merkitys jälleenrakentamisen ja uuden optimismin näkökulmista. Toisaalta sota vei kuusi vuotta yhteiskunnan elämästä ja vaikutti moniin traumatisoivasti. Vuoden 1952 alussa kuningas Yrjö VI kuoli ja nykyinen kuningatar Elizabeth II nousi valtaistuimelle. Tämä osaltaan ehkä joidenkin mielissä kuvasi myös uuden aikakauden alkua sotien jälkeen. Vanhaa valtaa edustava Winston Churchill kuitenkin toimi pääministerinä 1950-luvun alussa muutaman vuoden ja yhteiskunnallisessa mielessä edusti jonkinlaista jatkuvuutta aikakausien yli. Tässä työssä ei aineistoa kovin vahvalla tavalla peilata yhteiskunnallisiin tilanteisiin ja kehityskulkuihin, mutta ne on siitä huolimatta hyvä pitää mielessä taustalla olevina tekijöinä aineistona olevien tekstien kirjoittajien elämässä.

3.3 Alan Turing (1912–1954)

Alan Mathison Turing syntyi 23.6.1912 Lontoossa. Hänen vanhempansa olivat ylempää keskiluokkaa ja Turingin koulutus oli tavanomainen ja kiinnostus tieteeseen heräsi jo varhain. Hän opiskeli King's Collegessa, Cambridgessa ja hänet huomattiin nopeasti lahjakkaaksi opiskelijaksi. Vuonna 1935, noin 23-vuotiaana, Turing valittiin King's Collegen Fellow-palkinnon (collegen arvostetuin palkinto) saajaksi. Seuraavana vuonna, 1936, Turing julkaisi artikkelinsa ”On Computable Numbers, with an Application to the Entscheidungsproblem”⁴⁵, jota voidaan pitää hänen kaikkein merkittävimpänä kirjoituksenaan.⁴⁶

⁴⁵ Turing, 1936.

⁴⁶ Hodges, 2019, 2–3.

Artikkelissa ”On Computable Numbers...” Turing määritteli tietokoneen laskentatavan (engl. computation) ja rajan sille, mitä sillä voidaan tehdä. Tämän vuoksi artikkeleita pidetään modernin tietojenkäsittelyopin perustavana tekstinä.⁴⁷

Toisen maailmansodan aikana Turing toimi Blechley Parkissa kryptografian parissa tehtäväänään rikkoa natsi-Saksan käyttämän salauskoneen, Enigman, salaus. Hän tiimeineen onnistui siinä ja sota lopulta päättyi vuonna 1945. Vuonna 1948 hän muutti Manchesterin yliopistoon ja keskittyi tietokoneen ja ihmisen aivojen kykyjen tutkimiseen ja vertailuun. Vuonna 1950 hän julkaisi artikkelin ”Computing Machinery and Intelligence”⁴⁸, jossa hän esitti matkimispelin, jonka avulla hänen mukaansa tarkoitus pystyä määrittelemään, milloin koneen voidaan katsoa ajattelevan.⁴⁹

1952 Turing pidätettiin syytettynä seksuaalisesta suhteesta toisen miehen kanssa ja välttääkseen vankilatuomion hänen piti ottaa estrogeeniannoksia ”seksuaalisten halujensa tukahduttamiseksi”. 7.6.1954 Turing löydettiin kotoaan kuolleen. Kuolinsyyksi määritettiin syanidimyrkytys ja todennäköisenä on pidetty itsemurhaa – ruumiin vierestä löytyi syanidilla myrkytetty omena, josta oli haukattu pala.⁵⁰

Turingin merkittävimpinä saavutuksina voidaan siis pitää tietokoneen laskennallisuuden määrittämistä vuonna 1936, hänen kryptografista työtään toisen maailmansodan aikana, sekä ajattelevia koneita käsittelevän akateemisen keskustelun aloittamista – keskustelun, joka osaltaan on vaikuttanut tekoälyn määrittelemiseen ja myös kehittämiseen.

3.4 Turingin testi ja kiinalaisen huoneen argumentti

Turingin testin, jonka Turing esitti *matkimispelin* (engl. *imitation game*) nimellä artikkelissaan ”Computing Machinery and Intelligence” läpäiseminen on muodostunut tiettyssä mielessä tekoälyn kehittämisen päämääräksi. Kuitenkaan, vaikka monen tietokoneen katsotaan sen läpäisseen, ei vaikuta vielä nykypäivän tilanteessa siltä, että edes kyseiset tietokoneet kuitenkaan vaikuttaisivat inhimillisillä mittareilla sellaisilta, että ne voitaisiin katsoa ajatteleviksi koneiksi. Turingin testi voi kuitenkin toimia jonkinlaisena ensiaskeleena ajattelevien koneiden kehittämisessä.

⁴⁷ Hodges, 2019, 3.

⁴⁸ Turing, 1950.

⁴⁹ Hodges, 2019, 4.

⁵⁰ Hodges, 2019, 4–5.

Turing esittelee matkimispelinsä, eli Turingin testin ensin miehen ja naisen välisenä matkimispelinä: on mies ja nainen (kuulusteltavat A ja B) sekä heidän lisäksi on kuulustelija. Kuulustelija on toisessa huoneessa kuin kuulusteltavat, eikä hänellä ole heihin näköyhteyttä. Kuulustelijan tehtävänä on päätellä, kumpi toisessa huoneessa olevista kuulusteltavista henkilöistä on mies ja kumpi nainen esittämiensä kysymysten ja niihin saamiensa vastausten avulla. Toisen kuulusteltavan (esimerkiksi A:n) tehtävänä on saada kuulustelija luulemaan A:n olevankin itse asiassa B. Tällöin taas B:n tehtävänä on vakuuttaa kuulustelijalle, että A ei puhu totta, vaan yrittää esittäytyä totuudenvastaisesti B:nä. Toisin ilmaistuna tässä asetelmassa mieskuulusteltavan tehtävänä on vakuuttaa kuulusteltava siitä, että mieskuulusteltava onkin naiskuulusteltava. Ja naiskuulusteltavan tehtävänä on totuudessa pitäytyen osoittaa mieskuulusteltavan olevan huijari, sillä kuulustelija tietää, että toinen heistä on. Mieskuulusteltava siis voi läpäistä tämän matkimispelin vain, jos saa uskoteltua kuulustelijalle, että hän itse on naiskuulusteltava, ja naiskuulusteltava onkin pelin huijaava osapuoli ja mieskuulusteltava.

Tätä asetelmaa Turing kehittää edelleen sen varsinaiseen muotoonsa: mies- ja naiskuulusteltavien sijaan kuulusteltavina ovatkin kone ja ihminen. Tässä asetelmassa jälleen kuulustelijan tehtävänä on selvittää, kumpi kuulusteltavista on kone ja kumpi ihminen. Koneen tehtävänä on uskotella kuulustelijalle, että se on ihminen ja kuulusteltavan ihmisen tehtävänä on saada kuulustelija uskomaan, että hän, kuulusteltava ihminen on ihminen ja kone on kone. Koneen katsotaan läpäisseen matkimispelin, tai Turingin testin, jos kuulustelija luulee konetta ihmiseksi, tai ei osaa sanoa, kumpi kuulusteltavista on kone ja kumpi ihminen. Tässä asetelmassa Turing piti ideaalina tilanteena, että kommunikaatio kuulustelijan ja kuulusteltavien välillä tapahtuisi tekstivälitteisesti, esimerkiksi nykyään jonkinlainen chat-toiminto voisi olla juuri tällainen Turingin tarkoittama tekstivälitteinen kommunikaatiotapa.

Turingin testi on siis koneen käyttäytymistä mittaava ja kone läpäisee sen, ja saisi Turingilta arvon ”ajatteleva kone”, jos se tarpeeksi uskottavasti kykenisi tekstivälitteisessä tilanteessa esittämään ihmistä. Tämä on siis hyvin behavioristinen tapa lähestyä ongelmaa. Behaviorismilla tarkoitetaan psykologian teoriaa, jonka mukaan ihmisen tai eläimen – tässä tapauksessa koneen – käytöstä voidaan selittää ehdollistumisen avulla ilman, että vedottaisiin ajatuksiin tai tunteisiin. Turingin testin tapauksessa merkittävä behavioristinen piirre on se, että koneen ajattelun mahdollisuus määritellään ulkoisen käytöksen perusteella, ei niinkään sisäisten ajatusten tai tunteiden

perusteella. Tämä on kuitenkin hyvin käytännönläheinen tapa lähestyä ongelmaa, sillä käytöstä voi empiirisesti havainnoida ja sen uskottavuutta mitata toisen ihmisen reaktioiden perusteella. Turing myös toteaa, että matkimispeli tulisi toteuttaa samalla koneella ja eri kuulustelijoilla useampaan otteeseen sen varmistamiseksi, että koneen ajattelukykyjä puoltava kanta ei olisi yhden kuulustelijan määriteltävissä, sillä ihmisen suhtautuminen toiseen ihmiseen on kuitenkin lopulta melko yksilökohtaista ja toinen saattaa olla helpommin koneen huijattavissa kuin toinen. Tämän vuoksi todellinen Turingin testin läpäiseminen pitäisi määrittää suuremmalla kuulustelijaotoksella kuin yhdellä henkilöllä.

Behavioristinen asenne on muutenkin, kuten myöhemmin tullaan tarkemmin huomaamaan, luonteva Turingille ja hänen kirjoittamisen kontekstilleen 1940-luvun lopun Englannissa, sillä vaikka varsinaisena vallalla olevana tutkimusasenteena väistytvä, oli loogisen positivismin mukainen tieteellinen ajattelu kuitenkin ainakin taustateoreettisena asenteena monen tieteentekijän kohdalla merkittävää pitkälle loogisen positivismin varsinaisen valtakauden, 1920– ja 1930-lukujen jälkeenkin. A. J. Ayer, jonka katsotaan jossain määrin vakiinnuttaneen loogisen positivismin Englantiin 1930-luvulla, totesi myöhemmin 1970-luvulla ajattelevansa, että lähes kaikki siinä oli väärää, mutta siitä huolimatta hän piti sitä henkilökohtaisesti omana teoriataustanaan vielä 1970-luvulla.⁵¹

Turingin testistä ei kuitenkaan voi puhua uskottavasti mainitsematta kiinalaisen huoneen argumenttia. Kiinalaisen huoneen argumentin esitti John Searle vuonna 1980 ja sen perusidea on kritisoida Turingin testin asetelmaa. Searlen artikkeli ”Minds, Brains and Programs”⁵² ei kuulu tämän tutkielman aineistoon sen ajankohdan ja julkaisumaan vuoksi, mutta kiinalaisen huoneen argumentin vaikuttavuuden vuoksi se lienee syytä mainita ja lyhyesti esitellä.

Kiinalaisen huoneen argumentin perusasetelma on hieman samankaltainen kuin Turingin matkimispelin, ei kuitenkaan täysin. On huone, jossa on henkilö, joka puhuu englantia, mutta ei puhu eikä ymmärrä kiinaa (hän vastaa tässä Turingin kuulusteltavana olevaa konetta). Huoneen ulkopuolella on kiinaa äidinkielenään puhuva henkilö (joka vastaa Turingin kuulustelijaa), joka oven ali antaa huoneessa olevalle paperin, jossa on kiinankielistä tekstiä, esimerkiksi kysymyksiä. Huoneessa oleva henkilö näkee paperilla vain kiinalaisia kirjoitusmerkkejä, jotka eivät hänelle merkitse

⁵¹ Hanfling, 2004, 194

⁵² Searle, 1980.

mitään, mutta koska hänellä on ohjekirja (joka vastaa tässä tietokoneen ohjelmaa) siitä, millaisella kirjoitusmerkkijonolla tulee vastata tiettyyn kirjoitusmerkkijonoon, kykenee huoneessa oleva henkilö kirjoittamaan paperille kiinaksi vastaukset, jotka taas ovat merkityksellisiä huoneen ulkopuolella olevalle henkilölle, joka puhuu kiinaa äidinkielenään. Tällä perusteella tämä saattaisi tehdä sen päätelmän, että huoneessa on kiinaa ymmärtävä ja osaava henkilö, vaikka huoneessa olija vain kaavamaisesti seurasi ohjekirjaansa ja kirjoitti sen perusteella vastauksen ymmärtämättä sen paremmin kysymystä kuin antamaansa vastaustakaan.

Tällä argumentilla Searle pyrki kritisoimaan sitä, että Turingin testiä voitaisiin pitää ajattelun tai ymmärryksen takeena koneen suhteen. Vaikka huoneessa oleva henkilö saisikin ohjeen mukaan toimiessaan aikaan ymmärrettävää tekstiä, ei tämä tarkoita, että hän ymmärtäisi tuotoksestaan mitään. Searlen argumentti ei ole behavioristinen, kuten Turingin on, vaan se ottaa huomioon henkilön ajattelun kyvyn ymmärtää merkityssisältöjä, seikka jonka Turing jättää huomiotta.

Tässä on nähtävissä kahden erilaisen filosofisen järjestelmän yhteentörmäys. Turing kirjoitti pääasiassa behavioristisesta näkökulmasta artikkelinsa ja nojaa sitä kautta paljon loogisen positivismin perinteeseen, eli pyrkimykseen todentaa tutkimustuloksia empiirisesti, vaikka se jättäisikin toisia – ehkä merkittäviäkin – ominaisuuksia huomiotta. Toisaalta Searle lähestyy samaa kysymystä enemmän semanttisessa mielessä ja siis painottaen sanojen ja merkkien merkityssisältöjen ymmärtämisen tärkeyttä ajattelussa. Turing ei edes yrittänyt muodostaa asetelmaa, jossa nämä olisi otettu huomioon. Tässä taas tulee edustavalla tavalla esiin filosofian tapa taustateorioiden unohtamiseen, jos ne ovat esteenä totuuden ja todellisuuden kannalta merkittävien argumenttien antamiselle. Searle voi kritisoida Turingia aivan erilaisista lähtökohdista käsin kritiikin olematta merkityksetöntä ja erilaisiin taustateorioihin vedoten poiskuitattavissa. Ongelmana tässä vastapuolen argumenttien taustateoreettisten elementtien selvittämisen unohtamisessa saattaa olla se, että argumentoijat keskustelevat oikeastaan aivan eri asioista, eikä kummankaan kritiikki osu maaliin ja vie keskustelua mielekkäällä tavalla eteenpäin.

4. Ajattelevat koneet

4.1 Matkimispeli

Alan Turingin lokakuussa 1950 *Mind*-lehdessä julkaistu artikkeli ”Computing Machinery and Intelligence” oli merkittävä merkkipaalu 1900-luvun jälkipuoliskon keskustelussa ajattelevista koneista. Artikkelissa Turing lähtee kysymyksestä *voivatko koneet ajatella?* (engl. ”*can machines think?*”), mutta toteaa tämän olevan mielenkiinnoton kysymys. Hän esittää, että olisi järkevämpää unohtaa tämä vaikeasti määriteltäviin käsitteisiin tukeutuva kysymys ja keskittyä käytännöllisempään tapaan hahmottaa ongelma. Hänen mukaansa olisi siis parempi pyrkiä selvittämään, voiko kone käyttäytyä uskottavasti ihmisen tavoin. Näin hän pyrki selvittämään edes jotain koneista, kun käsitteet *ajatella* ja *kone* ovat hänen mielestään määritelmällisesti ongelmallisia ja epäselviä, sillä kyseisten käsitteiden tarkkaan – normaalin ja tavanomaisen – merkityksen selvittämiseen tarvittaisiin hänen mukaansa laajaa kyselytutkimusta kansalaisten keskuudessa ja tämä olisi ollut hänen mielestään absurdia. Toisaalta myös tämä näkemys käsitteiden ja termien suhteesta on altis termien käyttötavan muutoksille, jolloin alkuperäinen kysymys *voivatko koneet ajatella* saattaisi tarkoittaa eri kysymystä eri aikoina. Tämän ongelman ratkaisemiseksi Turing pyrki käsittelemään kysymystä käytännöllisen kokeen avulla, kuten myös luvussa 3 on kuvattu.

Turingin artikkelin mielenkiintoinen piirre on se, miten laajasti siinä käsitellään mahdollisia vastaväitteitä tai kumoajia esitetylle argumentille ja pyritään osoittamaan niitä jo etukäteen ongelmallisiksi ennen kuin niitä julkisesti oli edes esitetty. Turing myös tässä suhteessa on ikään kuin aikaansa edellä, sillä tavanomaisempi tapa on käsitellä vasta-argumentteja vasta niiden ilmaannuttua. Turingin tapa ottaa ne huomioon jo alkuperäisen asetelman rinnalla on mielenkiintoinen uusi tapa ohjata keskustelua. Kuitenkin, kuten tutkielmassa tullaan huomaamaan, alkuperäisen artikkelin vastaväitteet saavat melko vähän huomiota matkimispeliä kommentoivissa artikkelissa ainakin tämän työn aineiston puitteissa.

Tässä tiivistettynä Turingin käsittelemät mahdolliset vastaväitteet ja niiden perusajatukset ja Turingin tavat osoittaa ne pitämättömiksi jo ennen esittämistään:

Teologinen vasta-argumentti: ”Ajattelu on ihmisen kuolemattomaan sieluun liittyvä ilmiö ja koneelle ei ihminen voi sellaista luoda, ainoastaan jumala.” Turingin vastaus: väite rajoittaa jumalan kaikkivoivuutta, joten se on sisäisesti ristiriitainen.

”Pää hiekassa” -vasta-argumentti: ”Ajattelevat koneet voisivat aiheuttaa kamalia asioita, toivottavasti koneiden ajattelukyky ei siis toteudu.” Turingin vastaus: tämä ei tarvitse kumoamista, mutta näin ajattelevia ihmisiä voisi lohduttaa esimerkiksi sielunvaelluksen ajatuksella.

Matemaattinen vasta-argumentti: ”Digitaalisissa tietokoneissa on matemaattisen logiikan rajoituksia, kuten Gödelin teoreema (jota Turing ei tästä selvennä – Gödel esitti kaksi epätäydellisyysteoreemaa).” Turingin vastaus: siinä missä koneen ajatteluun saattaa kohdistua loogisia rajoituksia, voi myös ihmisen ajatteluun kohdistua samat rajoitteet.

Tietoisuus: ”Jos kone ei osaa esimerkiksi säveltää sinfoniaa omien alkuperäisten ajatuksiensa ja tunteidensa pohjalta, ei sen voida ajatella olevan tietoinen ja ihmisaivojen kanssa samantasoinen.” Turingin vastaus: toiset mielet -ongelman mukaan emme voi edes tietää ajatteleeko kukaan muu kuin subjekti itse (”minä”).

Useiden kyvyttömyksien vasta-argumentti: ”Kone ei voi olla mukava, ystävällinen, huumorintajuinen – sillä ei siis voi olla monenlaisia erilaisia käyttäytymistapoja, joita ihmisellä on.” Turingin vastaus: riittävän suurella muistikapasiteetilla voi.

Lady Lovelacen vasta-argumentti: ”Kone ei voi tuottaa mitään todellisesti uutta, vaan se voi enintään toistaa jo olemassa olevia asioita, jotka ovat ihmisen aikaansaamia.” Turingin vastaus: ainakin hänen koneensa usein yllättävät hänet itsensäkin tarkkuudella, jota hän ei ole niihin tietoisesti tehnyt.

Hermoston jatkuvuuden vasta-argumentti: ”Kone ei voi matkia ihmisen hermojärjestelmää, joka ei ole digitaalinen, vaan ennemmin eritasoisin, jatkuviin, sähkövarauksiin.”⁵³ Turingin vastaus: koneessa voisi olla jonkinlainen satunnaisuustoiminto, joka arpoo toimintavaihtoehtoja.

Käyttäytymisen epäformaalitus: ”Kaikille tilanteille ei voi olla olemassa valmista toimintatapaa. Tämän vuoksi, koska kaikkia mahdollisia toimintatapoja eri tilanteissa ei voida ennalta koneeseen määritellä, ei kone voi ajatella, kuten ihminen joka eri tilanteissa kykenee arvioimaan mahdollisia erilaisia etenemisvaihtoehtoja.”

⁵³ Myöhemmän tiedon valossa esimerkiksi aivojen synapseissa syötteen tulee yltää tietylle tasolle ennen kuin se lähettää impulssin edelleen. Näin ollen siinä on tietynlainen digitaalisuuden piirre. (McCulloch & Pitts, 1990.)

Turingin vastaus: tässä sekoitetaan käyttäytymisen lait sosiaalisiin normeihin. Jos tässä käsiteltäisiin sosiaalisia normeja, olisi vaikeampi ennustaa koneen toimintaa.

Yliaistillisen havaitsemisen vasta-argumentti: ”Telepatia ja selvänäkeminen saattavat vaikuttaa matkimispeliin, sillä niiden avulla matkimispelin kuulustelija saattaa saada selville ilman kysymyksiäkin, kummassa huoneessa on ihminen ja kummassa kone.” Turingin vastaus: tämä vasta-argumentti on vaikea, koska erityisesti telepatian tilastollinen evidenssi on ”musertavaa”. Näin ollen tulee varmistaa huoneiden telepatiaeristys.

Mielenkiintoinen havainto tämän tutkielman tutkimusaineistosta on se, että Turingin ennalta arvaamia vasta-argumentteja ei niissä juuri nostettu esiin, vaan keskityttiin muiden asioiden tarkasteluun. Turingin oletettu pyrkimys ohjata keskustelua ei siis ainakaan tässä suhteessa näytä onnistuneen ja aineisto koostuukin (varsinkin 1950-luvulle ajoittuva aineisto) pääasiassa yleistä ajattelevien koneiden mahdollisuuden kysymystä erityisesti tavallista kielenkäyttöä painottavan käsiteanalyysin kautta – siis juuri sillä tavalla, minkä Turing artikkelinsa ensimmäisellä sivulla tuomitsi mielenkiinnottomaksi tavaksi puhua koneiden oletetusta mahdollisuudesta ajatella.

Keith Gunderson käsittelee artikkelissaan ”The Imitation Game”, joka julkaistiin vuonna 1964 *Mind*-journaalissa, Turingin esittämää matkimispelin ajatuskoetta ja erityisesti pyrkii esittämään sille vasta-argumentteja, jotka Turing Gundersonin mukaan vaikutti sivuuttavan. Gunderson esittää artikkelinsa toisen osan alussa kysymyksen ”voivatko kivet matkia?” Hän toteaa, että monien mielestä se kuitenkin saattaisi olla liian mieletön kysymys ansaitakseen keskustelua. Tämän jälkeen hän esittää ajatuskokeen ”varpailleastumispelistä”. Siinä on kolme osanottajaa: kuulustelija, sekä kuulusteltavat A ja B. Kuulusteltavista toinen on ihminen ja toinen on kivi. Kuulusteltavat ovat eri huoneessa kuin kuulustelija, mutta huoneiden välisessä seinässä on aukko, josta kuulustelija voi laittaa jalkansa toiseen huoneeseen. Kuulustelijan laittaessa jalkansa toiseen huoneeseen, joko A tai B astuu kuulustelijan varpaille ja kuulustelijan tehtävänä on määritellä, astuiko hänen varpailleen ihminen vai kivi siitä, miten hänen varpailleen astuttiin. Kiven tehtävänä on laskeutua kuulustelijan varpaille sillä tavoin, että kuulustelija luulee ihmisen astuvan varpailleen, ei kiven. (Hän myös selvittää, miten tämä olisi mahdollista toteuttaa teknisesti.) Tämän matkimispeliparodian ajatuksena on esittää, että vaikka kivi voisi huijata ihmistä luulemaan itseään ihmiseksi sillä perusteella, miten se astuu kuulustelijan varpaille, ei tämä vielä tarkoita, että kivi kykenisi ajattelemaan. Gundersonin mukaan siis, vaikka

jokin esittäytyisi ihmismäisenä, ei sitä todellisuudessa voida pitää ihmismäisenä. Näin hän samalla kritisoi behaviorismia, joka ilmenee tässä tapauksessa Turingin artikkelissa, sanomalla, että käytös ei vielä takaa minkään todellista luonnetta käytöksen mukaiseksi.

Samalla Gundersonin parodia kuitenkin hieman ontuu, koska se joko olettaa kivelle yliluonnollisen tavan liikkua itsestään (jolloin kiven kykyä ajatella ei pitäisi nähdä enää niin ongelmallisena) tai vaihtoehtoisesti niin, että kivi on kiinnitetty varpaille astuvaan koneeseen, jolloin kivellä ei tässä asetelmassa ole mitään sijaa, voisihan sen korvata tällöin vaikka omenalla.

4.2 Ajattelevat koneet kielen filosofian ongelmana

Kysymys *voivatko koneet ajatella?* ja sen pohtiminen nousivat aineistossa teeman asemaan huolimatta siitä, että Turing tuomitsi kysymyksen mielettömäksi ja termien ”kone” ja ”ajatella” tavallisen käytön selvittämisen ja analyysin jopa vaaralliseksi.⁵⁴ Turing halusi esittää tavan selvittää empiirisesti, voivatko koneet ajatella ja esittää siis sen vuoksi matkimispelin, jolla hänen mukaansa tämä voitaisiin saavuttaa ilman tarvetta metafyyksiselle mielen käsiteanalyysille tai koneen määrittelylle. Turingin selvästä empiirisestä asenteesta huolimatta monet päätyivät tarkastelemaan asiaa analysoimalla sanoja ”kone” ja ”ajatella” erityisesti niiden tavallisissa käyttötilanteissa ja pyrkien siten ratkaisemaan koneen ajattelun arvoituksen.

Esimerkiksi W. Mays *Philosophy*-journalin artikkelissaan ”Can Machines Think?”⁵⁵ vuonna 1952 käsittelee kysymystä käsiteanalyysin keinoin eritellen kysymyksen termit, erityisesti ”kone” sekä ”ajatella” ja selvittää niiden mahdollisia merkityksiä. Mays lähtee siitä lähtökohdasta, että *sielunelämä* (engl. *mental life*) ja ajattelu ovat ihmisen ominaisuuksia tavanomaisessa kielenkäytössä, eikä niinkään koneen tai eläimen. Hän kuitenkin pitää mahdollisena, että jos tapaisimme olion (kuten Mars-planeetan asukkaan) tai mekanismin, joka vaikuttaisi kaikin puolin käyttäytyvän älykkäästi, voisimme pitää mahdollisena sen kokevan psykologisia kokemuksia.⁵⁶

Mays huomioi Turingin toteamuksen siitä, että vuoteen 2000 mennessä kieli olisi muuttunut sen verran, että saattaisi olla mielekäästä puhua ajattelevista koneista, mutta kuitenkin pitää tätä tarpeettomana spekulointina ja esittää, että on

⁵⁴ Turing, 1950, 433.

⁵⁵ Mays, 1952.

⁵⁶ Mays, 1952, 150.

mielekkäämpää selvittää kielen tilaa ja sanojen merkitystä yleisessä käytössä kontekstissaan 1950-luvun Britanniassa. Hän esittää, että ideaalia olisi tehdä laaja gallup-kysely sanan ”kone” merkityksistä laajalle yleisölle, mutta toteaa sanakirjan lukemisen olevan käytännöllisempää ja lainaa Oxford English Dictionarya. Maysin mukaan se esittää koneen määritelmän yhdistelmänä mekaanisia ja liikkuvia osia, ja jonka vastakohtana on elämän omaaminen (engl. having life), tietoisuus sekä tahto. Tähän Mays toteaa:

”O.E.D:n määritelmä nostaa esiin ainakin yhden asian, sen, että koneen on usein ajateltu olevan jotain, jolla ei ole omaa yksityistä elämää, joka ei unelmoi suorittaessaan tehtäviä, ja jolta puuttuu tietoisuus, älykkyys sekä oma tahto.”⁵⁷

Tämä alleviivaa Maysin pyrkimystä osoittaa, että ainakaan kielellisesti koneella ei voida ajatella olevan kykyä ajatella ainakaan tavallisen kielen käytön mukaisesti ja siten sielunelämä ja ajattelu pitäisi laskea vain ihmisen ominaisuudeksi.⁵⁸

Mays kuitenkin ei jää täysin kiinni 1950-luvun käsitykseen termistä ”kone”. Hän ottaa myös lyhyesti käsittelyyn vanhemman, Karel Čapekin näytelmässään *R.U.R* vuonna 1920 esittelemän termin ”robotti” ja käsittää sen laitteena, joka olisi ihmisen ja koneen välimuoto, samalla hieman kuin Frankensteinin hirviö.⁵⁹ Mays kuitenkin siirtää robotit pian syrjään ja vielä toteaa, että myös termi ”kone” tulisi jättää tässä tarkastelussa sivuun sen tunteellisen sävyn vuoksi ja käyttää neutraalia termiä, kuten englanniksi ”artifice”, jonka Oxford English Dictionary nykypäivänä määrittelee seuraavasti: ”Clever or cunning devices or expedients, especially as used to trick or deceive others”, eli suomeksi ”älykäs tai ovela laite tai apuväline, erityisesti sellainen jota käytetään huijaamaan tai harhauttamaan muita”⁶⁰. Maysille tämä lienee siis neutraali termi, vaikka omasta ajallisesta näkökulmastamme tämä vaikuttaa hieman painottuneelta termiltä. Tämä saattaa myös olla Turing-viittaus, sillä siinä ”artificen” käyttötarkoitus on ihmisten huijaaminen, mikä tietenkin on Turingin matkimispelin perusajatus.

⁵⁷ ”The O.E.D. definition does bring out one thing at least, a machine is usually thought of as something which does not possess a private life of its own, it does not indulge in reverie when at its task, it lacks consciousness, intelligence and will.” Mays, 1952, 149. Käännös minun.

⁵⁸ Mays, 1952, 149–150.

⁵⁹ Mays, 1952, 150.

⁶⁰ Käännös minun.

Mays toteaa myös, että Turingin esittelemä asetelma muuttaisi ”älykkyyden” määritelmäämme eikä vastaisi sen tavanomaista käyttöä. Toisaalta hän myös nostaa esiin, että kone, joka toteuttaisi ihmismäisiä toimintoja, ei olisi enää sellainen, jota 1950-luvun kontekstissa hänen mukaansa kutsuttaisiin koneeksi.⁶¹ Mays kritisoi myös Turingin tapaa inhimillistää koneita käyttämällä kieltä, jossa ”kone tekee valintoja” tai sitä ”rangaistaan tai palkitaan”. Maysin mukaan tämänkaltainen subjektiivinen kielenkäyttö ilman näiden termien tarkkaa uudelleenmäärittelyä koneiden suhteen vain sekoittaa keskustelua. Tämän vuoksi Maysin mukaan ei ole järkevää puhua koneista, joilla olisi psykologisia ominaisuuksia, kuten ajatuksia.⁶²

Myös F. H. George *Philosophy*-journaalissa vuonna 1956 julkaistussa artikkelissa ”Could Machines Be Made to Think?” lähestyy kysymystä ensisijaisesti kielen ja käsiteanalyysin kautta. Hän kritisoi Maysin tapaa pyrkiä määrittelemään termit ”ajatella” ja ”kone” ja päätyy ehdottamaan omaa ratkaisuaan. Georgen mukaan, toisin kuin Turing, Mays käyttää termeistä hyvin rajattua käyttötapaa, jonka Mays esittää yleisenä tai ”luonnollisena” käyttötapana.⁶³ George viittaa tässä *luonnolliseen kieleen* (engl. *natural language*), jolla tarkoitetaan kieltä, joka on luonnollisesti kehittynyt *käytössä (vastakohtana keinotekoiselle kielelle tai tietokonekoodille).⁶⁴ George kuitenkin nostaa esiin kysymyksen siitä, tuleeko luonnollisen kielen ohjata keskustelua – ja missä määrin.⁶⁵ Hän ei siihen kuitenkaan suoraan vastaa.

Hieman myöhemmin, vuonna 1960 *The Philosophical Quarterly*-journaalissa julkaistussa artikkelissaan ”Men and Robots”⁶⁶ A. R. Lacey käsittelee ihmisen ja robotin välisiä eroja erityisesti robotin lähtökohdista käsin. Hän lähtee kysymyksestä ”voivatko koneet ajatella?” (engl. ”can machines think?”) ja erittelee sanojen ”can”, ”machine”, ”think” mahdollisia eri merkityksiä. Hänen mukaansa ”can” voi olla joko empiirinen tai looginen sana. Lacey mukaan empiirisenä sanana ”can” kysyy koneiden kyvyistä, ja koneet, joiden kyvyistä kysytään, voivat olla hänen mukaansa joko koneita, jotka ovat artikkelin kirjoitushetkellä olemassa, tulevaisuuden koneita tai loogisesti mahdollisia koneita, joita ei kuitenkaan voi käytännössä rakentaa. Loogisessa mielessä kysyttäessä taas Lacey mukaan tiedämme koneen kyvyt, mutta ”can” kysyykin, että miten näitä kykyjä pitäisi täsmentää, ”mikä voidaan lukea

⁶¹ Mays, 1952, 150–151.

⁶² Mays, 1952, 1953.

⁶³ George, 1956, 245.

⁶⁴ OED, linkki: https://www.lexico.com/definition/natural_language. Linkki tarkastettu 23.4.2021.

⁶⁵ George, 1956, 245.

⁶⁶ Lacey, 1960.

ajatteluksi?”, kuten Lacey kysyykin. Tämä ”can”-sanana analyysi on mielenkiintoinen, sillä tämä artikkeli on ainoa kielifilosofisesti ajattelevia koneita lähestyvä artikkeli, joka myös analysoi tätä sanaa, muut kirjoittajat ohittavat ”can”-sanana kokonaan ja keskittyvät ilmeisesti mielenkiintoisempina pitämiinsä sanoihin ”machine” ja ”thinking”, joissa kieltämättä on enemmän analysoitavaa ja joiden analyysillä ehkä on mahdollista saada enemmän huomiota. Tässä mielessä Lacey kuitenkin on tunnollisin kielifilosofi tässä kontekstissa.

Termille ”machine” Lacey esittää neljä erilaista merkitystä. ”Machine” on hänen mukaansa ensinnäkin objekti, jonka tarkoitus on toteuttaa tekijänsä sille asettamat tehtävät. Hän kuitenkin ainakin osittain kumoaa tämän väitteen huomauttamalla, että ihminen saattaa olla jumalan omiin tarkoituksiinsa luoma olento, ja toisaalta hänen mukaansa on myös mahdollista, että ihminen hankkii lapsia toteuttamaan heille asetettuja tehtäviä. Ainakin jälkimmäinen on tietysti, kuten Laceykin toteaa, moraalisesti ongelmallinen toimintamalli.

Toinen sanan ”machine” merkitys Laceyille on se, että se on tehty epäorgaanisista materiaaleista⁶⁷, ja kolmas, että kone on objekti, joka toimii ”tieteen lakien” mukaan, ei niinkään periaatteita seuraten. Lacey kumoaa nämä määritelmät toteamalla toisesta merkityksestä, että asia ei ole aivan niin yksinkertainen, ja kolmannesta, että esimerkiksi ”periaatteiden seuraaminen” ei ole kovin yksinkertaisesti määriteltävissä. Neljäs sanan ”machine” merkitys Laceyille on se, että se on toiminnan objekti, jolla ei ole omia päämääriä. Lacey toteaa tämän viimeisen olevan lupaavin sanan ”machine” määritelmä, mutta jättää senkin hieman epämääräiseksi toteamalla, ettei ole selvää, mitä *päämäärien omaamisella* tarkoitetaan (engl. ”*having ends on its own*”).

Sanan ”machine” lyhyen määrittely-yrityksen jälkeen Lacey keskittyy sanan ”think” analysointiin. Se voi olla hänen mukaansa ainakin kaksimerkityksinen: se saattaa tarkoittaa älyllisten tai laskennallisten tehtävien suorittamista, tai sitten sillä saatetaan hänen mukaansa viitata monenlaisiin mielen toimintoihin.

Laceyn artikkelin päämääränä on esittää, että kone, joka ajattelisi ja varsinkin tuntisi ja kokisi asioita, olisikin oikeastaan ihminen. Tällä Lacey pyrkii kritisoimaan imitaation tai matkimisen merkitystä koneen älyllisen kehittyneisyyden

⁶⁷ Filosofisesti mielenkiintoinen viimeaikainen kehitys on ollut suurellekin yleisölle esitetyt ”xenobotit”, jotka ovat orgaanisesta materiaalista valmistettuja pieniä robotteja. Tämä viimeaikainen kehityskulku viimeistään vaikuttaisi kumoavan Laceyn määritelmän, että kone olisi epäorgaanisista materiaaleista valmistettu.

mittarina ja omaksuu antibehavioristisen näkökulman ongelmaan nähden. Hän huomauttaa, että puhuessaan koneen kohdalla jonkinlaisesta toiminnasta, joka vaatisi älyllisiä kykyjä, kuten että kone valitsee, käyttää hän verbin kohdalla lainausmerkkejä, siis ”kone ’valitsee’”, koska hänen mukaansa kone ei aidosti valitse mitään, vaan toteuttaa ainoastaan ihmisen koodaamia käskyjä ja vaikuttaa valitsevana.

Artikkeli on jaettu kolmeen osaan, joista kahdessa ensimmäisessä käsitellään pääargumenttia ja kolmannessa pohditaan hieman koneiden mahdollisen mielen suhdetta filosofiseen *toiset mielet -ongelmaan* (engl. *problem of other minds*), jonka teemaa käsitellään tarkemmin myöhemmin.

Tavallista kieltä painottava analyysitapa saattaa ainakin osittain selittyä vallalla olleesta filosofisesta kiinnostuksesta tavallisen kielen filosofiaan. Ilmari Kortelaisen ja Joose Järvenkylän mukaan filosofista keskustelua Isossa-Britanniassa alkoi 1940-luvun lopulta alkaen hallita Oxfordin yliopistosta alkunsa saanut tavallisen kielen filosofian suuntaus.⁶⁸ Tavallisen kielen filosofia syntyi vastaliikkeenä G. E. Mooren ja Bertrand Russelin 1900-luvun alussa Cambridgessa kehittelemälle loogiselle analyysille, joka painotti tutkittavan ongelman erittelyä sekä analysointia ja jonka mukaan filosofian oli tuotettava ”aidosti uusia tuloksia”.⁶⁹ 1940-luvulla siihen olivat antaneet osansa myös looginen positivismi ja Wittgenstein varsinkin varhaiskautensa filosofialla (*Tractatuksella*). Oli siis muodostumassa pientä koulukuntaeroa Cambridgen filosofian ja Oxfordin filosofian välille siten, että Cambridgessä painotettiin Russelin, Mooren ja Wienin piirin oppeja, kun taas Oxfordissa alettiin nuoremman filosofisukupolven toimesta suuntautua sen selvittämiseen, miten käytämme kieltä tavallisissa elämäntilanteissa.⁷⁰ Heidän mielestään loogista analyysiä edustavat filosofiset lauseet ovat luonnollisen kielen vastaisia (Russelin mukaan kielen harhaanjohtamisen osoittamisella voidaan ohittaa monet filosofiset ongelmat – samaa esitti Wittgenstein *Tractatuksessa*), ja filosofiaa tulisi siis kehittää ottamalla luonnollisen kielen, eli tavallisen kielenkäytön tavat huomioon.⁷¹

P. F. Strawson julkaisi vuonna 1950 *Mind*-journaalissa (sattumalta Turingin artikkelin ”Computing Machinery...” sisältämää numeroa edeltäneessä numerossa) artikkelin ”On Referring”⁷², jota voidaan pitää merkittävänä tavallisen kielen filosofian ideoiden esiintuojana. Strawsonin mukaan emme voi pitää lauseita

⁶⁸ Kortelainen ja Järvenkylä, 2013, 19.

⁶⁹ Kortelainen ja Järvenkylä, 2013, 12–13.

⁷⁰ Kortelainen ja Järvenkylä, 2013, 19.

⁷¹ Kortelainen ja Järvenkylä, 2013, 15.

⁷² Strawson, 1950.

tosina tai epätosina, tai että jokin sana tai fraasi varsinaisesti viittaisi johonkin objektiin. Hänen mukaansa ainoastaan lauseiden käyttötavat voivat olla tosia tai epätosia ja ainoastaan sanojen ja fraasien käyttötavat viittaavat. Michael Morris selittää tätä Strawsonin ideaa esittämällä, että esimerkiksi lause ”minä kirjoitan tällä hetkellä tätä lausetta tietokoneen näppäimistöllä” on tosi silloin, kun lauseen ”minä” viittaa *minuun*, kirjoittajaan. Jos taas *lukija* sanoisi saman lauseen, olisi se epätosi. Hän ei siis voi käyttää samaa lausetta eri tilanteessa ollessaan (lukiessaan), mutta se ei kuitenkaan tee lauseesta itsestään totta tai epätotta, lauseen käyttötapa ja -tilanne ratkaisevat.⁷³

Strawsonin kirjoitus herätti mielenkiintoa ja jo aiemmin samankaltaisia tavallisen kielenkäytön filosofiaa painottavia ajatuksia oli levinnyt erityisesti Oxfordin yliopistolla. Tämä vaikutti filosofien mielenkiinnon kohteisiin ja erityisesti heidän tapoihinsa tarkastella filosofisia kysymyksiä, kuten kysymystä ajattelevien koneiden mahdollisuudesta.

Turing (joka opiskeli ja toimi pitkään Cambridgessa) edustaa kielifilosofisessa mielessä enemmän Russellin ja Wittgensteinin esittämää kantaa siten, että koska luonnollinen kieli saattaa olla harhaanjohtavaa kysyttäessä, että voisivatko koneet ajatella, tulisi asiaa tavallisen käyttötavan käsiteanalyysin sijaan tarkastella empiirisesti todennettavissa olevalla keinolla. Tämä hänen mielestään oli toteutettavissa matkimispelin avulla ja siten ohittaa luonnollisen kielen epämääräisyydet.

Tässä voidaan siis nähdä, että Turingin tapa käsitellä ongelmaa oli hyvin erilainen, kuin mikä oli vallalla oleva tapa. Turing tiedostaa tämän ja toteaaakin jo artikkelinsa alussa, että hänen ehkä odotettaisiin tekevän gallup-kysely sanojen käytöstä ja siten pyrkiä ongelmaa ratkaisemaan. Hän kuitenkin pitää tätä absurdina ja päättyy mieluummin behavioristisen matkimispelinsä esittämiseen. Turing oli ensisijaisesti matemaatikko, ja Cambridgessä pitkään toiminut, joten ehkä tämän vuoksi hänen artikkelissaan näkyy tämä analyttinen vaikutus. Kuitenkin voidaan katsoa, että Oxfordlaisen tavallisen kielen filosofian laaja vaikutus, joka Kortelaisen ja Järvenkylän mukaan alkoi hallita brittiläisessä filosofiassa jo 1947 alkaen on vaikuttanut ainakin Turingin artikkelia seuranneeseen keskusteluun käsiteanalyysin tasolla.⁷⁴

⁷³ Morris, 2007, 61.

⁷⁴ Kortelainen ja Järvenkylä, 2013, 19.

4.3 Elävät koneet

Eläviä organismeja (varsinkin ihmisiä) ja koneita vertailevia artikkeleita esiintyi erityisesti *The British Journal for the Philosophy of Science*ssä. Erityisesti ihmisaivojen toimintatapaa ja koneen toimintatapaa vertailtiin monessa eri artikkelissa.

Artikkelissaan ”Living and Lifeless Machines”⁷⁵ (*The British Journal for the Philosophy of Science*-journaalissa vuonna 1954) Reginald O. Kapp tarkastelee elävän organismin ja koneen välisiä eroavaisuuksia ja yhtäläisyyksiä. Hän toteaa, että hänen mielestään keskustelussa ei ole tullut esiin, mikä on *koneelle luonteenomaista* (engl. *what characterises a machine*). Hänen mukaansa, jotta voidaan määritellä, onko elävä organismi kone, on kyettävä ensin antamaan kriteerit sille, mikä on kone ja mikä ei ole kone. (Tämä siis kuuluu samalla myös jossain määrin koneen määritelmää pohtivien artikkelien joukkoon.) Hän päätyy esittämään koneen olevan sellainen, jota insinööri kutsuisi koneeksi (ja kannattaa tältä osin tavallisen kielen määritelmiä, tosin hieman erikoistuneemmassa mielessä – hän kun ei kysy määritelmää tavalliselta vastaantulijalta kadulla tai gallup-kyselyllä, kuten esimerkiksi Mays ehdottaa, vaan koneisiin erikoistuneelta henkilöltä ⁷⁶). Toisaalta hän myös asettaa koneelle sellaiset ehdot, että sen on suoritettava haluttuja toimintoja.⁷⁷

Tämän jälkeen hän vertailee koneen toimintaperiaatteita, kuten polttoaineen tarvetta, liikettä ja niin edelleen, vastaavan kaltaisiin inhimillisiin tarpeisiin ja toimintaperiaatteisiin. Hänen mukaansa esimerkiksi ihmisen kehon liikkuvat osat vaikuttavat hieman samankaltaisilta kuin koneen liikkuvat osat. Tämä ei kuitenkaan hänen mukaansa vielä tarkoita, että koneet ja ihmiset olisivat samastettavissa. Hän päättää analyysinsä lopulta toteamalla, että vaikka sekä koneessa että ihmisessä on tarve *syönteelle* (engl. *input*), on erotuksena se, että koneen syöte on ajallisesti satunnaisempaa (rajoittunut hetkiin, jolloin konetta käytetään), ja se kohdistuu rajallisempiin toimintoihin. Ihmisessä syöte taas esimerkiksi on *jatkovaa* (ihminen havainnoi ympäristöään jatkuvasti) ja *läpäisevää* siinä mielessä, että se mahdollisesti koskettaa monta kehon osa-aluetta (esimerkiksi yllättävä ääni saattaa hätkähdyttää sekä fyysisesti että psyykkisesti ja esimerkiksi tuoda mieleen muistoja aiemmasta hätkähdyttävästä tilanteesta.) Näin hän erottelee elolliset ja elottomat koneet ja toteaa, että niissä ei kovin merkittäviä eroja ole, ja että asiaa tulisi tarkastella laajemmin.

⁷⁵ Kapp, 1954.

⁷⁶ Mays, 1952.

⁷⁷ Kapp, 1954, 93.

Kaksiosaisessa artikkelissaan ”Do Computers Think? (I) ja (II)”⁷⁸, jotka julkaistiin kahdessa peräkkäisessä *The British Journal for the Philosophy of Science* numerossa vuonna 1956, Mario Bunge käsittelee aivojen ja koneiden eroja. Hänen erityinen tapansa lähestyä ongelmaa on vertailla ihmisen kykyä abstraktiin matematiikkaan ja koneen fyysistä laskentakykyä todeten, että niissä ei ole kyse samasta asiasta, vaikka ne jossain määrin hänen mukaansa korreloisivatkin keskenään.⁷⁹ Bunge myös painottaa, että koneet representoivat mentaalisia funktioita toteuttamatta niitä kuitenkaan. Hänen mukaansa ihminen on se, joka kykenee muodostamaan yhteyksiä koneen antamien tulosten ja abstraktien asioiden välille.

Bungen artikkeli edustaa ihmisten ja koneiden välisten matemaattisten ”kykyjen” teemaa siten, että siinä hän osoittaa eroja tietokoneen ja ihmisen toimintatapojen välillä pyrkien kyseenalaistamaan koneen kykyä ihmismäiseen ajatteluun. Hän osoittaa, että koneen toimittamat laskutoimitukset ovat enemmänkin mekaanista laskentaa käytännöllisellä tasolla, kun taas ihmisen matematiikka käsittelee ideoita abstraktimmin. Näin hän pyrki osoittamaan, että koneet eivät teoriassakaan ole kykeneviä ajatteluun, sillä tietokoneiden toimintaperiaate laskennallisuuden mielessä hänen ajallisesta näkökulmastakaan käsin ei ollut juurikaan muuttumassa.

T. R. Miles artikkelissaan ”On the Difference between Men and Machine”⁸⁰ esittelee uuden käsitteen: *homo mechanisma*. Artikkelin julkaistiin *The British Journal for the Philosophy of Science*ssä vuonna 1957. Miles aloittaa artikkelin toteamalla kysymyksen ”mikä on koneen ja ihmisen välinen ero?” (engl. ”what is the difference between a man and a machine?”) tavallisen merkityksen olevan ilmeinen: kone näyttää erilaiselta kuin ihminen ja *homo mechanisma* toimii tässä siis ajatuskokeena siitä, miten konseptit ”mieli”, ”ihminen”, ”kone” ja ”tietoisuus” toimisivat, jos kone, joka kaikilta ulkoisilta puoliltaan olisi kuin ihminen, kyettäisiin luomaan. Hän käsittelee kysymystä ensin mahdollisten argumenttien kautta, mutta päättyy lopulta itse esittämään, että kyseiset konseptit toimisivat kyllä siinäkin tilanteessa, että *homo mechanisma* olisi mahdollista toteuttaa. Hän toteaa lopulta, että hänen mielestään ihmisen ja koneen merkittävin ero lopulta on, että ihminen kykenee suhteuttamaan maailmasta tekemänsä havainnot *kehonsa mallin* tai *kehoskeemansa* (engl. *body-schema*) avulla, mutta kone – edes *homo mechanisma* – ei siihen samalla tavoin kykene, koska sillä ei ole tämänkaltaista kehoskeemaa. Kehoskeemalla hän viittaa aiempaan keskusteluun

⁷⁸ Bunge, 1956.

⁷⁹ Bunge, 1956, 215.

⁸⁰ Miles, 1957.

ihmisen ryhdistä ja lihasmuistista ja siten pyrkii kehittämään samantapaisen intuitiiviseen kehon tuntemukseen perustuvan teoretisoinnin siitä, millainen suhde ihmisellä on kehoonsa ja esittää siis, että kone ei kykene samaan, sillä kone on ihmisen rakentama ja siksi sillä ei ole samankaltaista suhdetta kehonsa historiaan kuin ihmisellä on.

Tällä artikkelilla Miles pyrki aloittamaan keskustelua kehoskeeman sopivuudesta tähän kontekstiin, mutta ainakaan tämän tutkielman aineistossa ei käsitettä käsitelty tämän laajemmin. Ainakin siis sen puitteissa voidaan katsoa, ettei se saanut juurikaan vastakaikua, osittain ehkä monimutkaisuutensa vuoksi. Artikkelin julkaistiin 1957, jolloin oli yleisellä tasolla paljon odotettua ajatella, että ihmisen kaltainen *homo mechanisma* olisi mahdollista toteuttaa ja, että mikä sen looginen todennäköisyys olisi. Ajatuskoetta ja käsitettä voidaan siis katsoa pidetyn ainakin aineiston puitteissa mielenkiinnottomina.

Tässä kuitenkin nousee esiin se, miten filosofit olivat loogisin ajatuskuluin melko paljon aikaansa edellä. Samalla, kun ensimmäisiä kaupallisempaan tarkoitukseen suunniteltuja tietokoneita rakennettiin, oli filosofien mielet kiinnittyneet jo biologisiin koneisiin. Hieman samaan tapaan kuin Turing puhui ajattelevista koneista vuonna 1950 ja ennusti teknologian kehitystä 50 vuotta eteenpäin yllättävän tarkasti. Tästä voisi jonkinlaisena johtopäätöksenä ajatella, että voisi olla myös nykypäivänä mielenkiintoista lukea filosofien tekstejä tulevaisuuden koneista, koska filosofinen ajattelu – vaikkakin ehkä nykypäivän realiteeteista irrallaan – saattaa löytää loogisia mahdollisuuksia, joita voi olla mahdollista toteuttaa teknologian kehittyessä lähitulevaisuudessa.

4.4 Ajattelevat koneet mielen filosofian ongelmana: toiset mielet -ongelma

Akateemisissa filosofisissa aikakauslehdissä käytävästä keskustelusta näyttää nousevan esiin eräs erityinen laajahko teema: *toiset mielet –ongelma*. Anita Avramides toteaa *Stanford Encyclopedia of Philosophy* -kirjoituksessaan toiset mielet -ongelman olleen paljon puheenaiheena varsinkin mielenfilosofiassa 1900-luvun puolivälin paikkeilla.⁸¹ Tämä näkyy selkeästi myös tämän tutkielman aineistossa. Esimerkiksi Lacey *The Philosophical Quarterly*ssä vuonna 1960 julkaistussa artikkelissaan “Men and

⁸¹ Avramides 2020, 1.

Robots”⁸² omistaa koko kolmannen osan toiset mielet -ongelman pohtimiseen erityisesti robottien näkökulmasta. Hänen mielestään ongelmana on sen osoittaminen, että toisilla ihmisillä on mieli – tai että he ovat tietoisia olioita, puhumattakaan koneista. Hän päätyy toteamaan, että lopulta lienee syytä uskoa, että myös muut ihmiset ovat tietoisia toimijoita, ei vain ”minä”, ja että koneiden tietoisuutta voisi määritellä niiden toiminnan kautta behavioristisessa mielessä – siis esimerkiksi samaan tapaan, kuin miten Turing asian esitti artikkelissaan ”Computing Machinery and Intelligence”⁸³.

Kuten Avramides toteaa, ja mikä lienee muutenkin ilmeistä, on että toisten mielten ongelma pohjautuu epistemologiseen ongelmaan, joka äärimmäisyyksissään saattaisi johtaa *solipsismiin*, eli teoriaan siitä, että vain ”minä” voidaan perustellusti ajatella olevan olemassa. Tämä ajatus juontuu Avramidesin mukaan Descartesin ajatteluun ja hänen radikaaliin skeptisistiseen lähtökohtaan, jonka pohjalle hän *Mietiskelyt*-teoksessaan⁸⁴ rakentaa episteemisen järjestelmänsä.⁸⁵ Tähän liittyy, kuten Avramideskin tuo esille,⁸⁶ Descartesin kysymys siitä, voiko edes subjekti itse olla varma, että hän on tietoinen, eikä esimerkiksi pahan demonin vallassa siten, että tämä demoni saa subjektin ajattelemaan ja kokemaan tiettyjä asioita, esimerkiksi omassa tapaukseni sen, että istun lauantai-iltana työpöytäni ääressä kirjoittamassa tätä lausetta tietokoneella ja työpöydälläni on kasa papereita ja puolikas olut. Descartesin vastaus kuitenkin on, että koska hän voi olla *aivan varma* siitä, että paha demoni ei ole häntä saanut valtaansa, voi hän lopulta todeta, että ainakin hänen oma mielensä on olemassa.⁸⁷ Näin ollen subjekti, ”minä”, descarteslaisen ajatuksen mukaan voi todeta kyllä itse olevansa olemassa, mutta tämä ei vielä tuo varmuutta siitä, että myös muilla ihmisillä olisi mieli tai tietoisuus, eli että he eivät ole *filosofisia zombeja* (ihmisiä, jotka behavioristisessa mielessä näyttävät toimivan aivan tavanomaisten ihmisten tavoin, mutta joilla kuitenkin ei argumentin vuoksi ole mieltä tai tietoisuutta).

Myös Turing käsitteli toisten mielten ongelmaa, erityisesti solipsismin kautta artikkelissaan ”Computing Minds...”. Hän tiivistää keskustelun ytimen muutamaa riviin:

”[...] ainoa tapa olla varma siitä, että kone ajattelee, on *olla* kyseinen kone ja tuntea itsensä ajattelevan. Tällöin olisi mahdollista myös kuvailla näitä

⁸² Lacey, 1960, 71.

⁸³ Turing, 1950.

⁸⁴ Descartes, 2015.

⁸⁵ Avramides, 2020, 3.

⁸⁶ Avramides, 2020, 3.

⁸⁷ Descartes, 2015, 44.

tuntemuksia maailmalle, mutta kenenkään ei oikeutetusti tulisi huomioida tätä millään tavalla. Samaan tapaan tämän näkemyksen mukaan on mahdotonta tietää, että kukaan *ihminen* ajattelee, ilman että on kyseinen ihminen. Kyse on solipsistisesta näkemyksestä. Se saattaa olla kaikkein loogisin näkemys, mutta se tekee ajatusten kommunikoimisen vaikeaksi. A on taipuvainen ajattelemaan ”A ajattelee, mutta B ei” samalla, kun B uskoo ”B ajattelee, mutta A ei” Sen sijaan, että tästä väiteltäisiin jatkuvasti, on tavallista ylläpitää kohteliasta konventiota, että kaikki ajattelevat.”⁸⁸

Tämä on melko kattava ja tiivis esitys toiset mielet -ongelmasta ja samalla osoittaa hyvin, miten Turing pyrkii välttämään filosofisiin ongelmiin jumittumista ja käytännönläheisesti tuo ajatuksiaan esiin. Hänen päällimmäinen tavoitteensa on muodosta empiirinen määritelmä ajattelevalle koneelle ja hän vaikuttaisi näkevän filosofisten ongelmien olevan enemmänkin esteenä tämän toteuttamiselle. Sen vuoksi hän tässä, kuten myös kommentissaan ajattelevista koneista artikkelinsa alussa, pyrkii sivuuttamaan koko kysymyksen ilman kysymyksen tarkempaa käsittelyä. Vaikka asenne sinänsä on käytännönläheinen ja sinä mielessä kunnioitettava, juuri tämänkaltainen filosofisten ongelmien sivuuttaminen potentiaalisesti antaa monille filosofian alan kirjoittajille hyviä kritiikin kohteita. Tosin aineistossa tätä Turingin aiheen sivuutusta ei kuitenkaan merkittäväällä tasolla käsitelty.

⁸⁸ “[...] the only way by which one could be sure that a machine thinks is to be the machine and to feel oneself thinking. One could then describe these feelings to the world, but of course no one would be justified in taking any notice. Likewise according to this view the only way to know that a man thinks is to be that particular man. It is in fact the solipsist point of view. It may be the most logical view to hold but it makes communication of ideas difficult. A is liable to believe ‘A thinks but B does not’ whilst B believes ‘B thinks but A does not’. Instead of arguing continually over this point it is usual to have the polite convention that everyone thinks.” Turing, 1950, 446. Käännös minun.

5. Turing ja yliaistillinen havaitseminen

5.1 Yliaistillisuuden argumentti

Turingin ”Computing Machinery and Intelligence” –artikkelin yhdeksäs vastaväite, *yliaistillisuuden argumentti* (engl. *the argument from extra-sensory perception*) on mielenkiintoinen, koska se ainakin intuitiivisella tasolla vaikuttaa hyvin paljon eroavan Turingin akateemisesta kontekstista ja yleisestä työnsä luonteesta. Aihe ei myöskään ole saanut juurikaan huomiota Turing-tutkijoiden suunnalta, minkä vuoksi sitä on mielekästä käsitellä tässä hieman laajemmin.⁸⁹ Joissain tapauksissa kyseinen vasta-argumentti on jopa kokonaan jätetty pois ”Computing Machinery and Intelligence” –artikkelin uusintapainoksista.⁹⁰ Tässä luvussa teemaa tarkastellaan hieman muita ajattelevien koneiden teemoja laajemmin, sillä se vaikuttaisi erikoisuutensa lisäksi myös luovan kontekstia koko Turingin artikkelille ja jopa matkimispelin idealle. Mielenkiintoinen piirre on myös se, että tämän tutkimuksen aineistossa ei tähän Turingin vastaväitteeseen viitattu missään artikkelissa. Tämä on siis enemmänkin ”anti-teema” teemojen joukossa ja siten samalla mielenkiintoisesti luo poissaolollaan ymmärrystä tutkimuskohteesta.

Turing artikkelinsa ”Computing Machinery...” vasta-argumenttiosiossa käsitteli argumenttina (9) sen mahdollisuutta, että yliaistillisuus, esimerkiksi telepatia, aiheuttaisi ongelmia matkimispelin toteuttamiselle käytännössä. Hän lähtee siitä oletuksesta, että lukija tuntee yliaistillisuuden idean sekä käsitteet telepatia, selvänäkö, ennaltatietäminen sekä psykokineesi. *Yliaistillisuudella* (engl. *extra-sensory perception*) tarkoitetaan oletettua tapaa aistia asioita muilla tavoin, kuin tunnettujen aistien välityksellä.⁹¹ *Telepatia* (engl. *telepathy*) taas on oletettua ajatusten ja ideoiden välittämistä muiden, kuin tunnettujen aistien välityksellä⁹², *selvänäkö* (engl. *clairvoyance*) viittaa oletettuun tapaan aistia asioita tai tapahtumia tulevaisuudesta tai tavallisten aistillisten kontaktien tuolle puolen⁹³. *Ennaltatietäminen* (engl. *precognition*) hieman samaan tapaan kuin selvänäkö, sananmukaisesti tarkoittaa

⁸⁹ Leavitt, 2017, 348.

⁹⁰ Hofstadter, 1990, 599.

⁹¹ OED, linkki: https://www.lexico.com/definition/extrasensory_perception. Linkki tarkastettu 23.4.2021.

⁹² OED, linkki: <https://www.lexico.com/definition/telepathy>. Linkki tarkastettu 23.4.2021.

⁹³ OED, linkki: <https://www.lexico.com/definition/clairvoyance>. Linkki tarkastettu 23.4.2021.

asioiden tietämistä ennalta, erityisesti yliaistillisessa mielessä.⁹⁴ *Psykokineesillä* (engl. *psycho-kinesis*) tarkoitetaan oletettua kykyä siirtää asioita pelkällä ajatuksen voimalla.⁹⁵

Turing ei suoraan viittaa mihinkään tekstiin yliaistillisuutta käsitellessään, vaan jättää lukijan vastuulle olla tietoinen kyseisistä yliaistillisuutta käsittelevistä tutkimuksista, mikä saattaa ensisilmäyksellä tuntua kohtuuttomalta odotukselta. Turing myös toteaa, että nämä yliaistilliset ilmiöt vaikuttaisivat todistavan vääräksi kaikki tieteelliset käsityksemme ja toteaa tilastollisen näytön, erityisesti telepatian osalta, olevan ylivoimainen.⁹⁶ Hän harmittelee, että niiden sijoittaminen yksilön omaan maailmankatsomukseen on kuitenkin vaikeaa ja sanoo, että niiden hyväksymisestä ei ole suuri askel otettavaksi kummitusten ja mörköjen olemassaolon hyväksymiseen. Turing kuitenkin myös antaa ymmärtää, että asiaa ei ollut täysin aukottomasti vielä todistettu, mutta tutkimustulokset olivat niin selvää suuntaa antavia, että Turing odotti tämän olevan vain ajan kysymys.⁹⁷

Turingin mukaan erityinen ongelma varsinkin telepatiasta aiheutuu matkimispelille, jos kuulustelija ja kuulusteltava ihminen kykenevät kommunikoidaan telepaattisesti keskenään. Telepatia 1900-luvun alkupuolen Englannissa usein yhdistettiin tiettyihin henkilöihin ja toisinaan sen ajateltiin olevan myös tiedostamatonta⁹⁸, joten Turing ei artikkelissaan luultavasti oleta kaikkien ihmisten olevan telepatiaan kykeneviä ehkä tietämättäänkin ja siten asettavan koneen epäedulliseen asemaan, mutta nostaa tämän mahdollisuuden kuitenkin tarkasteltavaksi. Turing ei kovin laajasti pohdi, miten telepatian ongelmasta päästäisiin, vaan ehdottaa lopuksi vain, että testi toteutettaisiin ”telepatian kestävässä huoneessa” (engl. ”telepathy-proof room”). Tämä osaltaan alleviivaa sitä, miten tämä on Turingille uusi ongelma, jonka selättämistä hän ei kovin kattavasti ole ehtinyt miettiä. Toisaalta voi myös olla, että Turing pitää ongelmaa niin vakavana matkimispelille, että hän ei kunnollista ratkaisua ongelmaan kyennyt löytämään. Seuraavassa alaluvussa selvitetään yliaistillisen havainnoimisen ymmärtämiseen pyrkineitä tutkimuksia ja tämän jälkeen käsitellään sitä, miksi yliaistillisuuden huomioiminen on Turingin tapauksessa huomiota herättävää, mutta toisaalta mahdollisesti ymmärrettävää.

⁹⁴ OED, linkki: <https://www.lexico.com/definition/precognition>. Linkki tarkastettu 23.4.2021.

⁹⁵ OED, linkki: <https://www.lexico.com/definition/psychokinesis>. Linkki tarkastettu 23.4.2021.

⁹⁶ Turing, 1950, 453.

⁹⁷ Turing, 1950, 454.

⁹⁸ Leavitt, 2017, 351.

5.2 Yliaistillisuuden tutkimukset 1900-luvun alkupuolella

Yliaistillisuus populaarina ilmiönä on vanha, mutta ensimmäistä kertaa sitä pyrittiin tuomaan akateemiseen kontekstiin 1800-luvun lopulla. Lontoossa perustettiin 1882 yhdistys nimeltä ”Society for Psychical Research” (SPR). Sen tarkoituksena oli tutkia yliaistillisia ja spirituaalisia ilmiöitä samaan ”eksaktiin ja kiihkottomaan” tapaan, jolla tieteellisessä tutkimuksessa ratkaistaan ongelmia.⁹⁹ Se siis pyrki joko osoittamaan nämä ilmiöt vääriksi tai (mikä ehkä todennäköisempää) antamaan niille uskottavuutta tieteelliselläkin kentällä. Tässä alaluvussa esitellään 1900-luvun alun anlogo-amerikkalaisen kontekstin kahden merkittävimmän ylikuonnollista aistimista tutkineen tieteentekijän koeasetelmia sekä C. D. Broadin merkittävää näitä tutkimuksia ylistävää kirjoitusta *Philosophy*-journaalissa vuonna 1944.

5.2.1 J. B. Rhine (1895–1980)

Tieteen kentälle yliaistillista havaitsemista käsittelevät tutkimukset pääsivät ensimmäisenä Yhdysvalloissa, Duke Universityssä, Pohjois-Carolinan osavaltiossa 1930-luvulla. Tämän toteuttaja oli Joseph Banks Rhine, joka oli yhdysvaltalainen botanisti ja parapsykologi, eli yliaistillista havaitsemista tutkiva henkilö. Hän teki kokeiluja ns. *Zener-korteilla* ja nämä kokeilut muodostivat hänen työnsä kulmakiven.¹⁰⁰ Zener-kortit ovat Karl Zenerin kehittämä korttijärjestelmä, jossa on kokoelma viittä eri symbolia – suorakulmiota, ympyrää, ristiä, tähteä sekä aaltoviivoja – esittäviä kortteja, kutakin viisi kappaletta. Yhdessä pakassa on siis 25 korttia. Rhinen menetelmä oli se, että tutkittavan tuli arvata seuraava kortti ennen, kuin se nostettiin tai vaihtoehtoisesti arvata koko pakan korttien järjestys. Tällä menetelmällä hän pyrki selvittämään tutkittavan kykyjä ennaltatietämykseen sekä varmasti jossain määrin selvänäköön, sillä kyseiset käsitteet vaikuttavat olevan ainakin osittain päällekkäiset. Toisaalta Rhine myös ajatteli telepatian olevan selvänäön kanssa päällekkäinen käsite, eli että telepatiaa ja selvänäköä ei voi täysin erottaa toisistaan.¹⁰¹ Tämän vuoksi tutkittavan tuli arvata seuraava kortti ennen sen nostamista, jotta yliaistillinen aistiminen voitaisiin rajata ennaltatietämykseen ilman mahdollista telepaattista havainnointia. Jos kortteja nostava näkisi kortin ennen tutkittavan arvaamista, voisihan ajatella, että tutkittava arvaisikin kortin oikein telepatian vaikuttamana, eikä ennaltatietämyksen. Yliaistillinen

⁹⁹ Leavitt, 2017, 347.

¹⁰⁰ Mackenzie, 1981, 650.

¹⁰¹ Mackenzie, 1981, 650.

havaitseminen usein nähtiin tahdosta riippumattomana ja alitajuisena, joten oli tärkeää varmistaa, että mahdollinen telepaattinen yhteys ei häiritsisi tutkimusta ja vääristäisi sen tuloksia.

Rhinen korttitutkimuksia kuitenkin kritisoitiin psykologien suunnalta voimakkaasti esimerkiksi huomauttamalla, että tutkittavan on mahdollista nähdä kortin symboli kortin takapuolelta raskaan painojäljen vuoksi. Ensimmäiset Zener-kortit painettiin melko raskaasti ja sen vuoksi toisinaan saattoi olla mahdollista nähdä painojälki kortin takapuolelta, mutta koska Rhine pyrki yliaistillisen havainnoimisen osoittamiseen tieteellisin metodein, niin hän käytti kevyemmin painettuja kortteja, joiden kuviota ei ollut mahdollista nähdä kortin takapuolelta.¹⁰² Myös esimerkiksi tulosten kirjaamista ja tilastojen käyttöä myöhemmin kritisoitiin.¹⁰³ Kuitenkin Rhinen tutkimusten merkittävin ongelma oli, että kukaan muu amerikkalainen tutkija ei pystynyt toistamaan tutkimustuloksia.¹⁰⁴

5.2.2 S. G. Soal (1899–1975)

Vaikka Rhinen tutkimukset eivät saaneet varsin hyvää vastaanottoa yhdysvaltalaisessa keskustelussa, eikä niiden tuloksia kyetty siellä toistamaan, toisin oli Englannissa, jossa S. G. Soal omaksui Rhinen asetelman ja kehitti sitä tieteellisesti vielä hieman kunnianhimoisemmaksi.

Samuel Soal oli englantilainen matemaatikko ja parapsykologi. Hän luennoi matematiikasta Queen Mary Collegessa, Lontoon yliopistossa ja oli jäsenenä yhdistyksessä Society for Psychical Research, joka pyrki tutkimaan yliaistillisiä ilmiöitä.¹⁰⁵ Soal pyrki myös toistamaan Rhinen tutkimuksia vuonna 1934 omaksuen samalla tiettyjä episteemisestä näkökulmasta metodia tarkentavia elementtejä, kuten ottamalla useita Zener-korttipakkoja käyttöön ja sekoittamalla ne keskenään, näin pyrkien kasvattamaan korttien satunnaisuutta. Tulokset eivät olleet lupaavia, paitsi kunnes W. Whateley Carington – Soalin kollega ja SPR:n jäsen – ehdotti, että Soal vertaisi tutkittavan vastauksia, ei päällimmäisenä olevaan korttiin (johon vastaus oli alun perin osoitettu), vaan sitä edeltäneeseen sekä seuranneeseen ja sitten laskisi yhteen osumat.¹⁰⁶

¹⁰² Mackenzie, 1981, 651.

¹⁰³ Mackenzie, 1981, 650–651.

¹⁰⁴ Levitt, 2017, 349.

¹⁰⁵ Leavitt, 2017, 347.

¹⁰⁶ Leavitt, 2017, 349.

Soal myös uudisti kortteja vaihtamalla symbolit eläinkuvioihin – elefanttiin, kirahviin, leijonaan, pelikaaniin ja seepraan – vaihtelun vuoksi (hän oli kyllästynyt Zenerin ”kuviin kuvioihin”). Soal myös uudisti tutkimusasetelmaa ottamalla käyttöön Rhinen yksinkertaisempaa tutkittava–tutkija-asetelmaa monimutkaisemman asetelman, jossa on kaksi tutkittavaa, joista toinen oletettavasti omaa telepaattisia kykyjä ja niillä pyrkii selvittämään toisessa huoneessa olevan toisen tutkittavan näkemiä korttien kuvia. Asetelma oli jossain määrin samankaltainen kuin Turingilla myöhemmin matkimispelissä, joten voidaan ehkä pitää mahdollisena, että Turing olisi saanut inspiraatiota matkimispelinsä asetelmaan näistä Soalin kokeista – voimmehan olettaa hänen olleen niistä tietoinen.¹⁰⁷ Tämä kuitenkin saattaa olla liian pitkälle menevä johtopäätös tässä tehtäväksi – kuitenkin voitaneen todeta, että tämä saattaisi olla mielenkiintoinen vertailevan tutkimuksen kohde, jos haluaisi tarkemmin selvittää Turingin matkimispelin asetelman taustoja.

5.2.3 C. D. Broad *Philosophy*ssä 1944

Filosofi C. D. Broad, joka oli tuolloin Trinity Collegen ”Fellow” Cambridgessä, julkaisi *Philosophy*ssä vuonna 1944 kirjoituksen ”The Experimental Establishment of Telepathic Precognition”¹⁰⁸. Kirjoitus ylisti Soalin tutkimuksia ja totesi niiden olevan erityisen merkittäviä psykologeille ja filosofeille. Kirjoitus on 15-sivuisena melko pitkä Discussion-osion teksti *Philosophy*ssä (yleensä ne ovat keskimäärin noin 2 sivua pitkiä), joten myös tämä korkeatasoinen akateeminen filosofian alan jurnaali antoi yliaistillisen havaitsemisen ajatukselle harvinaisen paljon tilaa. Tätä selittää ehkä se, että Broad oli aikansa merkittävä filosofi, jolloin lienee ymmärrettävää, että jurnaali kuin jurnaali oli valmis julkaisemaan mitä tahansa hän halusikin julkaista. Broad myös totesi Soalin tutkimuksista, että niiden todistusvoima on ”tilastollisesti ylivoimainen” (engl. ”statistically overwhelming”), ilmaisu, jota Turingkin käyttää, kuten myös Leavitt mainitsee.¹⁰⁹

Broadin mukaan Soalin tulokset olivat merkittäviä ainakin kolmella tavalla: (1) tarkoituksellisen huijaamisen mahdollisuudet ovat hänen mukaansa eliminoitu, (2) tuloksia käsitellään yksinkertaisella ja suoraviivaisella tavalla ja todennäköisyydet sattumia vastaan ovat kasvaneet ”astronomisen suuriksi”, (3)

¹⁰⁷ Leavitt, 2017, 350–351.

¹⁰⁸ Broad, 1944.

¹⁰⁹ Leavitt, 2017, 351.

olosuhteita on muuteltu ja korrelaatioita eri olosuhteissa saaduille tuloksille on löydetty.¹¹⁰

5.3 Turing ja yliaistillinen havaitseminen matkimispelissä

Miten nämä yliaistillisen havaitsemisen tutkimukset sitten päätyivät mahdollisena vastaargumenttina Turingin artikkeliin? Turing Cambridgessä toimineena matemaatikkona ei 1940-luvun Britannian kohtalaisen loogis-positivistisessa akateemisessa kontekstissa ole todennäköisin yliaistillisten ilmiöiden kannattaja. Hän kuitenkin ottaa tämän ilmiön tosissaan, mikä on mielenkiintoinen asia, sillä hänen työnsä loogis-positivistisen asenteen lisäksi hänen artikkelinsa ”Computing Machinery...” muut osat ovat hyvin pitkälti materialistiseen maailmankuvaan nojaavia.

Looginen positivismi on *positivismin* kehittyneempi muoto, jota kehiteltiin erityisesti ”Wienin piirissä” (saks. ”Der Wiener Kreis”) 1900-luvun alkupuolella. Wienin piirin perustavoitteena oli päästä ”metafyysikasta” eroon tieteessä ja filosofiassa ja yhdistää tieteenalat metodologisesti yhden tieteellisen metodin alle. Auguste Comte (1798–1857), jota pidetään sosiologian alullepanijana ja joka merkittäväällä tavalla edisti positivismin ohjelmaa sen alkuvaiheessa, piti positivismissa tärkeänä, että myös ihmisen toimintaa ja yhteiskuntaa käsittelevissä tieteissä teologiset ja metafysiset ainekset puhdistetaan pois.¹¹¹ Käytännössä tämä siis tarkoitti erityisesti empiirisen tutkimuksen sekä logiikan painottamista ja esimerkiksi mielenfilosofiassa sitä, että ihmisen mieli nähdään materiaalisesti ihmisen aivojen ilmiönä, eikä spirituaalisena henkenä, joka on yhteydessä toisiin ulottuvuuksiin, kuten esimerkiksi Descartes asian monien mielestä esitti.

A. J. Ayer oli englantilainen Wienin piirin osallistuja. Hänen kirjansa *Language, Truth, and Logic*¹¹² sai paljon huomiota ja nousi toisen maailmansodan jälkeen ”bestselleriksi” Britanniassa.¹¹³ Tämä vaikutti yleisiin käsityksiin loogisesta positivismista ja tieteenfilosofiasta Englannissa.

Loogisen positivismin merkittävänä käytännöllisenä osana voidaan pitää *verifikaatioperiaatetta*. Sen mukaan esitetyt väitteet on oltava yhdistettävissä tehtyihin

¹¹⁰ Broad, 1944, 261.

¹¹¹ Sintonen, 1986, 4.

¹¹² Ayer, 1936.

¹¹³ Creath, 2021, 12.

havaintoihin ja näin siis verifioidaan paikkansa pitäväksi tai pitämättömiksi. Havaintojen paikkansa pitämättömyys tarkoitti tässä sitä, että ne nähtiin mielettöminä tai merkityksettöminä. Loogisten positivistien mukaan oli siis kahdenlaisia merkityksellisiä väitteitä, ne, jotka ovat verifioidut havaintojen perusteella sekä ne, jotka loogisesti tai matemaattisesti voitiin osoittaa tosiksi *a priori* (ilman havaintoa).¹¹⁴

Vaikka saattaa olla liian pitkälle menevää arvailua lähteä Turingia luokittelemaan loogiseksi positivistiksi pelkän asuinpaikan ja ajankohdan perusteella, on hänen artikkelissaan ”Computing Machinery and Intelligence” tiettyjä loogisen positivismin piirteitä havaittavissa. Turingilla näyttäisi olevan pyrkimys esittää menetelmänsä siten, että se jättää *ajattelun* käsitteestä metafyyssisen puolen pois ja keskittyy siihen, mitä voidaan selkeästi havainnoida, siis koneen käyttäytymiseen. Tämä on siis behavioristinen lähestymistapa, mutta myös looginen positivismi ja varsinkin sen verifikationistinen puoli näyttäisi vaikuttavan koko matkimispelin asetelman taustalla.

Turingin behavioristiset ja verifikationistiset elementit viittaisivat siihen, että hän pyrki nimenomaan loogisen positivismiin viitekehyksestä käsin tuomaan esiin argumenttinsa siitä, milloin kone voitaisiin määritellä ajattelevaksi. Tähän luo kuitenkin vahvaa kontrastia se, että Turing ottaa huomioon yliaistillisen havaitsemisen mahdollisena vasta-argumenttina loogis-positivistiselle matkimispelilleen. Toki akateemisessa maailmassa alkoi 1940-luvulla nousemaan jonkin verran antipositivistisia ajatuksia, esimerkiksi Cambridgessa vaikuttaneen Ludwig Wittgensteinin myöhäisfilosofiasta (joka osui 1940-luvulle) on löydettävissä jonkin verran antipositivistista ajattelua, vaikka hänen 1910-luvulle sijoittuva varhaisfilosofiansa vaikuttikin paljon loogisen positivismiin kehitykseen ja Wienin piirin ajatteluun. Kuitenkin se, että Turing nostaa esimerkiksi telepatian mahdolliseksi merkittäväksi kumoajaksi teesilleen, herättää kysymyksen, mistä Turing sai ajatuksen yliaistillisesta havaitsemisesta ja miten hän sisällyttää sen siihen kokonaisuuteen, joka on artikkeli ”Computing Machinery...”.

Ottaen huomioon, miten paljon yliaistillisen havaitsemisen tutkimukset saivat palstatilaa 1940-luvulla esimerkiksi *Philosophy*ssä, tulee hieman ymmärrettävämmäksi, miksi Turing otti yliaistillisen havaitsemisen mahdollisuuden niin vakavasti huomioon artikkelissaan ”Computing Machinery...”. Yliaistillinen havaitseminen oli 1940-luvun lopulla vahvasti esillä myös populaarissa mediassa, kuten BBC:n radiokanavalla ja *The Times*issa – jossa syyskuun 1949 aikana termi ”telepatia”

¹¹⁴ Hanfling, 2004, 196.

esiintyi 16 kertaa.¹¹⁵ On siis tämän kannalta jokseenkin ymmärrettävää, että Turing oli huomionnut asian lehdistön kirjoituksista ja muusta mediasta sekä *Philosophyn* julkaisemasta Broadin kirjoituksesta. David Leavittin mukaan Broad toimi Cambridgessa samaan aikaan, kuin Turing ja tästä päätellen, sekä mainiten, että molemmat olivat homoseksuaaleja, päättelee Leavitt Turingin oppineen yliaistillista havaitsemista käsittelevistä tutkimuksista juuri Broadilta.¹¹⁶ Tämä kuitenkin saattaa olla hieman pitkälle menevä johtopäätös ja todennäköisempää onkin, että hän on jossain akateemisen elämän epäformaalimmissa tilanteissa kuullut tästä, esimerkiksi Ratio Clubin tilaisuudessa, jonka jäsen Turing oli.¹¹⁷ Ratio Club oli brittiläinen kybernetiikkaan keskittynyt keskusteluryhmä, jossa myös muita aiheita käsiteltiin, myös telepatiaa.¹¹⁸ Telepatia oli jopa erään kokoontumiskerran aiheena, mutta vasta vuonna 1951.¹¹⁹ Tämän vuoksi ainakaan kyseinen kokoontumiskerta ei ole voinut toimia Turingin aloitteena yliaistillisen havaitsemisen idealle. On kuitenkin luontevaa ajatella, että teema olisi ollut jo aiemmin epävirallisempänä keskustelunaiheena myös Ratio Clubilla.

On siis mahdollista ajatella, että Turing olisi saanut idean yliaistillisen havaitsemisen teeman sisällyttämisestä artikkeliinsa mahdollisen vastaväitteen muodossa joko mediasta, C. D. Broadilta henkilökohtaisesti tai tämän *Philosophyssa* julkaistun tekstin kautta tai esimerkiksi Ratio Clubin keskustelujen kautta. Mielenkiintoista on, miten paljon Turing kuitenkin antaa painoarvoa tälle mahdolliselle vasta-argumentille ja ei edes kunnolla pyri sitä kumoamaan. Mitä jos telepatia olisikin osoittautunut todeksi? Tätä Turing pitää hyvin mahdollisena. Silloinhan hänen koko matkimispelinsä olisi osoittautunut mahdottomaksi toteuttaa tiukan empirismin periaatteiden mukaan ja sulkien pois kaikki mahdolliset telepaattiset vaikutuskeinot. Turing toteaa kyllä, että kokeen voisi järjestää telepatian kestävässä huoneessa, mutta millainen se tarkalleen ottaen olisi? Telepatialla kun tarkoitetaan oletettua kykyä välittää ajatuksia ja ideoita muilla keinoin, kuin vain tunnettujen aistien avulla, niin se tarkoittaa käytännössä dualismin hyväksymistä. Tämä taas tarkoittaisi siis telepaattisten kykyjen olevan ”ylemmäntasoisia” kuin, mitä fyysinen maailma on, jolloin telepatian kestävä huoneenkin tulisi olla jollain tavalla yliluonnollinen. Turing ei esitä, millainen tämän telepatian kestävä huoneen tulisi olla.

¹¹⁵ Leavitt, 2017, 351,

¹¹⁶ Leavitt, 2017, 350–351.

¹¹⁷ Dewhurst, 2019, 39.

¹¹⁸ Dewhurst, 2019, 39.

¹¹⁹ Husbands ja Holland, 2008, 116.

Turing pitää telepatian ongelmana erityisesti sitä, että tällöin kuulustelija voisi käyttää tilannetta hyväkseen ja selvittää sen avulla, kumpi kuulusteltavista on ihminen ja kumpi kone. Tämä on mielenkiintoisesti päinvastainen asetelma kuin matkimispelissä muuten, jossa kone pyrkii huijaamaan kuulustelijaa. Turing ei kuitenkaan selvitä, mikä olisi kuulustelijan motiivi huijata telepaattisten kykyjensä avulla. Koneella on selvä ”motiivi” huijata (jos koneeseen tässä kontekstissa voidaan yhdistää motiiveja), se saisi ”ajattelevan koneen” tunnustuksen. Kuitenkaan kuulustelija ei juurikaan hyötyisi tilanteesta, paitsi siinä tapauksessa, että hän ei halua antaa ”ajattelevan koneen” tunnustusta koneelle. Kuitenkin Turing saattaa tässä ajatella, kuten usein ajateltiin, että telepatia olisi ollut enemmänkin intuitiivinen ominaisuus, jolle kuulustelija ei voisi mitään.

Turingin suhtautumisessa telepatiaan on kuitenkin artikkelin kannalta merkittävämpikin ongelma. Jos Turing hyväksyy telepatian ja muun yliaistillisen havaitsemisen mahdollisuutena, niin eikö tämä tarkoittaisi silloin, kuten yllä on mainittu, että ihminen olisi dualistinen olento, jolla on keho, sekä fyysisen maailman ”ylittävä” mieli.¹²⁰ Turingin matkimispeli lähtee siitä oletuksesta, että kone kykenee uskottavasti matkimaan ihmistä ja siten sen voidaan katsoa ajattelevan, mutta jos ihmisen ajatteleva mieli ei ole fyysinen, siis esimerkiksi jonkinlainen aivojen sivutuote, koko matkimispelin asetelma vaikuttaa hieman ristiriitaiselta ja olisi mahdotonta rakentaa ajatteleva kone. Tämä paljastaa Turingin epäjohdonmukaisuuden artikkelissaan ”Computing Machinery...”. Vaikka artikkeli on merkittävä merkkipaalu tekoälytutkimuksessa ja keskustelussa ajattelevista koneista, sisältää se perustavanlaatuisen sisäisen epäjohdonmukaisuuden lähtiessään liikkeelle siitä, että ajatteleva kone on mahdollista toteuttaa, mutta samalla hyväksyen mahdollisuuden materiaalin ylittävään mieleen ja ajatteluun, jollaista ihminen itse ei oletettavasti kykenisi kuitenkaan toteuttamaan.

Lisäksi artikkelin sisäistä ristiriitaa alleviivaa se, että Turing artikkelin muissa osissa lähestyy kysymystä ajattelevista koneista ja matkimispelistä hyvin behavioristisesta näkökulmasta, jonka taustalla voi helposti nähdä loogisen positivismin vaikutuksen, kuten yllä on osoitettu. Johdonmukaista selittävää tekijää tälle on vaikea keksiä ja lieneekin uskottavinta todeta, että Turing luultavasti halusi olla avoin uusille vakuuttavalta vaikuttaville tieteellisille tuloksille – olivathan ”tutkimustulokset”

¹²⁰ Tämän myös Stevan Harnad huomaa kommentaarissaan Turingin artikkeliin ”Computing Machinery and Intelligence”. (Turing, 2008, 57, Harnadin kommentti.)

yliaistillisesta havaitsemisesta empiirisiä ja siten jossain määrin Turingin tapaiseen havaintojen kautta maailmaa jäsentävään tiedekäsitykseen yhteensopivia.

Artikkelin sisäinen epäjohdonmukaisuus materialismin ja dualismin osalta saattaa kuitenkin selittää, miksi toisinaan kyseinen yliaistillisuuden vasta-argumentti on kokonaan jätetty pois artikkelin uusintapainoksista.¹²¹ Turing ei kuitenkaan ollut koulutukseltaan filosofi, vaan matemaatikko, joten periaatteessa voi olla mahdollista ajatella, että hän ei ollut perehtynyt metafysiikkaan ja dualismin kysymyksiin ja siksi tällainen ristiriita on päässyt artikkeliin mukaan.

Turing ei kuitenkaan väitä yliaistillista havaitsemista todeksi, vaan enemmänkin toteaa, että jos se todistettaisiin, saattaisi tulla eteen esitetyn kaltainen ongelma erityisesti telepatian suhteen, johon tulisi silloin siinä tilanteessa selvittää ratkaisua matkimispeli-koetta järjestettäessä. Turing siis enemmänkin näki tässä mahdollisen ongelman ja pyrki nostamaan keskustelua asiasta, jotta koeasetelmaa mahdollisesti tulevaisuudessa suunnittelevat tutkijat ottaisivat tämänkin asian huomioon.

Mielenkiintoinen havainto tämän tutkimuksen aineistosta on se, että tähän yliaistillisuuden vasta-argumenttiin ei puututa Turingin artikkelia muuten paljon käsittelevässä filosofisessa keskustelussa. Keskustelua käytiin vilkkaasti jo 1950-luvun alussa, jolloin yliaistillinen havaitseminen oli vielä ainakin jossain määrin mahdollisena pidetty ilmiö – ainakin Ratio Club järjesti siitä vielä esitelmä- ja keskustelutilaisuuden vuonna 1951.¹²² Todennäköisin syy sille, että asiaa ei otettu käsiteltäväksi ainakaan aineistona olevissa jouluaikaleissa saattaa olla, että monien mielestä oli mielenkiintoisempaa keskustella esimerkiksi tavallisen kielen filosofian näkökulmasta siitä, mikä on ”koneen” ja ”ajattelun” määritelmä, tai muista teemoista, joita tässä työssä on esitelty. Yliluonnollinen havaitseminen ajattelevia koneita käsittelevässä keskustelussa on kuitenkin Turingin aloitteesta huolimatta melko rajallinen teema, josta on luultavasti ollut vaikea keksiä filosofista keskusteltavaa, varsinkin kun tutkimustulokset olivat vielä alustavia, kuten Turingin antaa ymmärtää, ja koska aihe oli potentiaalisesti poleeminen, haluttiin ehkä odottaa tarkempia tutkimustuloksia, joita ei koskaan tullut.

Yliaistillisen havaitsemisen teeman käsittely kuitenkin auttaa ymmärtämään tämän tutkimuksen aiheita, ajattelevien koneiden teemaa osoittamalla,

¹²¹ Leavitt, 2017, 347.

¹²² Husbands ja Holland, 2008, 116.

että filosofian julkaisuissa, jotka tässä tutkimuksessa muodostavat tutkimusaineiston, on pyritty ainakin artikkelien tasolla keskittymään kohtalaisen perinteiseen filosofiseen keskusteluun, joka käsittelee perinteisiä filosofian teemoja, kuten mielenfilosofiaa ja kielifilosofiaa. Poikkeuksena tästä on tietenkin Turingin artikkeli ”Computing Machinery...”, joka esitteli keskusteluun ajattelevat koneet ja niiden behavioristisen tunnistusmenetelmän uutena teemana.

6. Lopuksi

Tässä tutkimuksessa kartoitettiin ajattelevien koneiden teemoja englantilaisissa akateemisissa filosofian julkaisuissa vuosina 1950–1975. Tarkoituksena oli selvittää, mitä filosofit ajattelivat ajattelevista koneista ja mitä eri teemoja keskustelussa nousi esiin ja miten ne liittyivät muuhun ajan filosofiseen keskusteluun. Keskustelujen teemat keskittyivät varsinkin käsitteiden ”kone” ja ”ajatella” analyysiin ja monen filosofin pyrkimyksenä vaikutti olevan osoittaa, että kone ei kykene ajattelemaan – ei ainakaan samalla tavoin kuin ihminen. Tämä liittyi osaltaan ajan henkeen, jossa kielifilosofia oli monien mielenkiinnon aiheena ja monet pyrkivät tätä ongelmaa ratkaisemaan tavallisen kielen käsiteanalyysin avulla.

Alan Turingin vuonna 1950 *Mind*-journaalissa artikkelissaan ”Computing Machinery and Intelligence”¹²³ esittelemä *matkimispelin* ajatus esiintyi aineistossa teemaana jonkin verran, varsinkin matkimispeliä suoraan kommentoivissa teksteissä, kuten aiemmin esimerkkinä käytetyssä Keith Gundersonin artikkelissa ”The Imitation Game”¹²⁴, joka julkaistiin *Mind*-journaalissa 14 vuotta Turingin artikkelin jälkeen, vuonna 1964. Siinä Gunderson pyrkii kritisoimaan Turingin matkimispelin asetelmaa ”varpailleastumispelin” parodian avulla siten, että vaikka jokin vaikuttaisi uskottavasti ihmismäiseltä, ei se vielä tarkoita, että se olisi todellisuudessa ihmismäistä. Samalla tämä oli myös behaviorismin kritiikkiä siten, että vaikka jokin asia käyttäytyisi tietyllä tavalla, esimerkiksi ihmismäisesti, ei takaa sille vielä mitään todellisia ominaisuuksia, kuten ajattelukykyä, ainoastaan kyvyn matkia tiettyä ominaisuutta ja matkiminen ei tässä nähdä riittäväksi ehdoksi ominaisuuden omaamiselle.

Toinen aiheesta noussut teema oli ihmisten ja koneiden ominaisuuksia vertaileva keskustelu. Se on luonteva jatke keskustelulle ja teemalle siitä, voivatko koneet ajatella. Ajattelu on luontevinta nähdä ensisijaisesti ihmisen ominaisuutena ja siten artikkeleissa käsiteltiin sitä, millainen koneen tulisi olla, jotta se voisi ajatella. Keskustelussa siis tarkasteltiin sitä, minkälaisia ominaisuuksia ihminen omaa ja vertailtiin niitä ominaisuuksiin, joita koneilla 1900-luvun jälkipuoliskon alussa oli, tai mitä niillä loogisesti ajateltuna on mahdollista kuvitella olevan tulevaisuudessa. Tällä perusteella usein pyrittiin määrittämään jonkinlaista vastausta kysymykseen ”voivatko

¹²³ Turing, 1950.

¹²⁴ Gunderson, 1964.

koneet ajatella?”. Eräs esimerkki oli T. R. Milesin artikkelissaan ”On the Difference between Men and Machine”¹²⁵ vuodelta 1957 esittelemä ajatuskoe *homo mechanismasta*. Se on ihmisen luoma biologinen ja ihmisenkaltainen olento, joka kuitenkin on kone. Tällä ajatuskokeella Miles pyrki tarkastelemaan koneen mahdollisuuksia ihmismäiseen olemiseen, ja esimerkiksi ajatteluun tilanteessa, jossa se fyysisiltä ominaisuuksiltaan ei eroa ihmisestä. Miles kuitenkin totesi tässäkin erona olevan sen, että koneella ei ole samankaltaista yhteistä historiaa kehonsa kanssa, kuin ihmisellä. Ihmisen keho on kulkenut ihmisen elämässä mukana sen alusta asti ja ihminen on luonut tiiviin suhteen siihen tämän vuoksi. Juuri luodulla koneella taas ei tämänkaltaista suhdetta siihen ole. Näin hän päätyi esittämään, että koska kone on kerralla rakennettu valmiiksi, ei sillä ole samanlaista suhdetta kehoonsa, kuin ihmisellä yleisesti ottaen on ja näin sen ei voida ajatella hänen mukaansa olevan täysin inhimillinen, vaikka esimerkiksi ajattelun suhteen se muuten muistuttaisikin ihmistä.

Toisaalta ihmisen ja koneen ominaisuuksia vertailevat artikkelit osoittivat, miten filosofit tai ainakin filosofisissa jurnaaleissa julkaistut tekstit saattavat loogisten päättelyketjujen vuoksi olla paljonkin aikaansa edellä esittämiensä mahdollisten ratkaisujen kanssa. Ratkaisut ongelmiin olivat luonteeltaan ensisijaisesti loogisia, mikä saattaisi alleviivata logiikan merkitystä filosofiassa, mutta samalla myös teknologian kehityksessä myös muuten kuin tietojenkäsittelytieteen harjoittaman logiikan suhteen. Sitä voidaan pitää ainakin jossain merkityksessä järkevän innovatiivisena tapana tarkastella erityisesti teknisiä mahdollisuuksia, vaikka ne eivät tarkasteluhetken historiallisessa kontekstissa vaikuttaisikaan toteutettavilta käytännössä. Tämä kuitenkin antaisi mahdollista suuntaa tulevaisuuden innovaatioille. Tämän vuoksi yhtenä johtopäätöksenä tässä työssä voidaan myös pitää logiikan tutkimuksen merkittävyyden alleviivaamista myös arkipäiväisempien ongelmien ratkaisemisessa, mutta myös uusien innovaatioiden kehittämisessä.

Ajattelevien koneiden ja kielifilosofian välinen teema ”koneen” ja ”ajattelun” määritelmistä ja käyttötavoista tavallisessa kielenkäytössä nousi aineistosta vahvasti esiin. Monissa artikkeleissa, myös niissä, jotka tässä teemoiteltiin esimerkiksi pääasiassa elävien koneiden teeman alle, käsiteltiin usein myös tavallisen kielenkäytön näkökulmasta ”koneen” ja ”ajattelun” määritelmiä ja käyttötapoja. Ajattelevien koneiden kielifilosofista käsittelyä esiintyi esimerkiksi W. Maysin artikkelissa ”Can

¹²⁵ Miles, 1957.

Machines Think?”¹²⁶ vuodelta 1952. Siinä hän erityisesti tavallisen kielen käyttötapoja painottaen tarkastelee termien ”machine” ja ”think” käyttötapoja todeten lopulta, että ainakaan siinä mielessä koneen ei voi ajatella ajattelevan, sillä kone määritelmällisesti ei ole ihmisen tavoin ajatteleva olio, vaan suoritteita toteuttava laite. Tavallisen kielen käyttötapojen painottaminen oli kuitenkin luontevaa erityisesti 1950-luvun (mutta myös jossain määrin 1960– ja 1970-lukujen) kontekstissa, sillä 1940– ja 1950-lukujen vaihteessa oli tavallista kieltä painottava filosofinen suuntaus noussut Englannissa ja erityisesti Oxfordissa merkittäväksi suuntaukseksi. Tämän vuoksi sen avulla pyrittiin hahmottamaan ongelmaa kuin ongelmaa tarkastelemalla ensin sitä, miten ongelman ydinsanoja käytettiin tavallisessa kielenkäytössä.

Siinä missä tavallisen kielen analysointia käytettiin ajattelevien koneiden ongelman ratkaisemisessa, käytettiin ajattelevia koneita myös ihmisen ja varsinkin ihmisen mielen ymmärtämiseen pyrkimisen välineenä. Näin ollen tällä uudella esimerkillä pyrittiin ratkaisemaan vanhoja mielenfilosofisia ongelmia. Varsinkin toiset mielet -ongelma nousi aineistossa teemaksi. Ongelmana siinä on sen perustelevinen, miksi yksilön pitäisi ajatella, että toisella yksilöllä olisi mieli ja ajatuksia, sillä se ei kovin yksinkertaiselta vaikuta. Aineistossa esimerkiksi Turing artikkelissaan ”Computing Machinery...”¹²⁷ kuitenkin pyrki käytännönläheiseen tapaan ratkaisemaan tätäkin filosofista ongelmaa edes jossain määrin toteamalla, että sitä ei voi ratkaista, mutta ihmisten välisen toiminnan mahdollistamiseksi on kohteliasta ajatella, että myös toisilla ihmisillä olisi mielet ja ajatuksia.¹²⁸ Toiset mielet -ongelma ei tämän uuden lähestymistavan avulla kuitenkaan lähtenyt ratkeamaan, sillä se on lopulta episteemisessä mielessä hyvin haastava ratkaista.

Tutkielman viidennessä luvussa käsiteltiin Turingin yhdeksättä vasta-argumenttia (yliaistillisen havaitsemisen vasta-argumenttia) ajatuskokeelleen ja pyrittiin selvittämään sen taustoja ja suhdetta Turingin yleisempään tapaan katsoa maailmaa. Tässä huomattiin ristiriitaisuuksia, sillä hänen muu julkaistu työ, esimerkiksi kyseessä olevan artikkelin ”Computing Machinery...”¹²⁹ toiset osat painottavat loogisen positivismin ajatusta metafysiikan eliminoimisesta ja behavioristista näkökulmaa empiiristen tulosten saamiseksi. Tämänlainen materialistinen maailmankuva nähtiin loogisesti sallivan koneiden kyvyn ajatella, sillä tällöin myös ihmisen mieli olisi

¹²⁶ Mays, 1952.

¹²⁷ Turing, 1950.

¹²⁸ Turing, 1950, 446.

¹²⁹ Turing, 1950.

fyysinen ilmiö, joka ainakin teoriassa olisi mahdollista toteuttaa koneissa. Yhdeksäs vastaväite taas hyväksyy yliluonnollisen mielen, mikä taas on hieman epäjohdonmukaista. Luvussa käsiteltiin myös aiempia yliaistillista havaitsemista käsitelleitä ”tieteellisiä tutkimuksia” ja niiden esiintymistä filosofian keskusteluissa. Jonkinasteista samankaltaisuutta Turingin testin asetelman ja telepatiakokeiden asetelman kanssa nähtiin, mutta yhteneväisyys lopulta on niin pientä, että olisi hieman pitkälle menevä johtopäätös todeta, että Turing olisi saanut idean matkimiskokeeseensa juuri näistä telepatian mahdollisuutta tutkineista koeasetelmista.

Tutkielman käsittelemiä teemoja siis olivat Turingin testi (eli matkimispeli), ”koneen” ja ”ajattelun” tarkoitusta selvittävät pyrkimykset, erityisesti tavallisen kielenkäytön näkökulmasta, koneiden ja ihmisten vertaileminen varsinkin mielen ominaisuuksien osalta, sekä koneiden asettuminen toiset mielet -ongelmaan. Nämä ovat jossain määrin erillisiä teemoja ainakin pinnallisella tasolla tarkasteltuna, mutta niissä on monissa samankaltainen pohjavire, nimittäin pyrkimys ihmisen mielen ja ajattelun ymmärtämiseen ja määrittämiseen koneiden ajattelun käsittelemisen kautta. Keskustelussa tarkennettiin ymmärrystä ajattelusta niin koneen, kuin samalla myös ihmisen suhteen ja pohdittiin sitä, mikä tekee ihmisestä ja ihmisen ajattelun sekä mielen ominaisuuksista erityisiä – tässä suhteessa koneisiin. Keskustelussa myös pohdittiin vanhaa toisten mielten ongelmaa ja sitä, tuoko ajattelevien koneiden idea siihen mitään uutta näkökulmaa. Myös tietenkin käsiteltiin sitä, voiko ajattelua mitata behavioristisesti. Monessa tapauksessa kirjoittajat pyrkivät erityisesti osoittamaan, että koneet eivät voi ajatella, ja että niillä ei voi olla mieltä, mutta näkökulmana tämä oli ainakin siinä mielessä oikeutettu, että se auttoi täsmentämään sitä, mistä oikeastaan puhutaan, kun puhutaan ajattelevista koneista.

Yliaistillista havaitsemista ja ajattelevia koneita käsittelevän keskustelun merkittävin suhde on ensisijaisesti se tosiasia, että Turing otti yliaistillisen havainnoimisen käsittelyyn artikkelissaan ”Computing Machinery...”. Mutta toisaalta se käsittelee ihmisen mielen oletettuja kykyjä ja siten yleensä ihmisen ajattelua. Vaikka yliaistillista havainnoimista ei nykypäivän näkökulmasta ole lähtökohtaisesti luontevaa pitää mahdollisena, edustaa se osaltaan 1900-luvun alun ja puolenvälin pyrkimyksiä ihmisen mielen ymmärtämiseen myös tältä osin. Ilmiötä myös pyrittiin selvittämään ensisijaisesti empiirisiin menetelmin, koeasetelmilla, joissa sitä voidaan tilastollisesti pyrkiä havainnoimaan ja siten osoittamaan joko puoltavia tai hylkääviä johtopäätöksiä. 1900-luvun puolessa välissä yliaistillista havaitsemista käsitelleet tutkimukset olivat

antaneet lupaavia tuloksia esimerkiksi telepatian puolesta, minkä vuoksi asiaa jonkin verran myös oli luontevaa pitää mielessä myös ajattelevista koneista puhuttaessa. Alan Turing näin tekikin, mutta tässä tutkielmassa kerätyn aineiston perusteella se kuitenkin ei saanut kovin suurta kannatusta. Kuitenkin jo se, että Turing asian nosti esiin, on ajattelevien koneiden teemojen kannalta merkittävää, koska potentiaalinen kysymys aiheeseen liittyen olisi voinut olla esimerkiksi ”jos ihminen on materiaallinen olento ja kykenee telepatiaan, eikö ole myös mahdollista, että koneen kehittyessä ajatuksien ja mielen suhteen ihmisen tasolle, myös sillä olisi mahdollisesti telepaattisia kykyjä?” Kysymys on sellainen, minkä olisi voinut ehkä kuvitella aineiston perusteella Turingin kysyvän, mutta hän pitääkin sitä ongelmana matkimispelinsä asetelmalle ja siis koneelle mahdottomana ilmiönä. Teemallisesti tämä siis menee lähelle ihmisen ja koneen eroja käsittelevää teemaa, koska siinä pohditaan – tai ainakin Turing pohtii – asioita, jotka ovat oletuksen mukaan ihmisille mahdollisia, mutta koneille mahdottomia. Tässä Turing tulee siis lähelle sen sanomista, että kone ei oikeastaan olekaan kovin samanlainen, kuin ihminen sen mielen ja ajattelun kyvyiltä.

Tässä tutkielmassa tarkasteltiin tutkimuskohdetta tieteenfilosofisten suuntausten osalta hermeneutiikan ja konstruktivismin asettamista perustuksista käsin. Ne ovat tieteenfilosofisina suuntauksina pääasiassa ohjaamassa tutkielman muotoa ja mielenkiinnon kohdetta sekä sen perustavaa tulkinnan tapaa. Ne eivät siis ole tässä katsottu menetelmällisiksi tai muuten käytännönläheisiksi tavoiksi lähestyä tutkimuskohdetta. Hermeneutiikan vaikutusta tutkielmassa on pyrkimys tutkimusaiheen mahdollisimman kokonaisvaltaiseen ymmärtämiseen huolimatta siitä, että tutkielman tarkastelemat teemat vaikuttaisivatkin alustavasti jokseenkin erillisiltä. Sen vaikutusta on esimerkiksi tutkielman tapa pyrkiä hahmottamaan teemojen keskinäisiä yhteyksiä sekä konteksteja ja tulkita sekä selventää niitä ja niiden taustoja laadullisen sisällönanalyysin avulla.

Toisaalta konstruktivismin vaikutus tässä tutkielmassa on hieman pienempi, mutta silti nähtävissä siinä, miten erityisesti tieteellinen – tai tässä tapauksessa filosofinen – keskustelu muodostaa ideoita sekä argumentaatiota ja niiden kautta tieteellistä todellisuutta, joka tässä tutkielmassa on tulkittu teemoina ja niiden tietynkaltaisina esiintymisen tapoina. Konstruktivismin painotus tässä tutkielmassa ei siis ole järin merkittävä, mutta sen olemassaolo taustalla vaikuttavana teoreettisena asenteena on kuitenkin mainittava, sillä sen katsotaan tässä luovan merkitystä itse tieteelliselle – ja tässä tapauksessa filosofiselle – keskustelulle.

Myös sosiaalisella konstruktionismilla on taustateoreettinen vaikutuksensa tässä tutkielmassa, mutta lähinnä sen *realistisella* muodolla, siis sillä joka myöntää sekä *raakojen tosiasioiden*, että *institutionaalisten tosiasioiden*, kuten John Searle ne erottelee ¹³⁰, olemassaolon. Raakoja tosiasioita ovat siis vuoret, joet sekä esimerkiksi luonnonlait (siis asiat, jotka ovat ihmisestä riippumattomia) ja institutionaalisia tosiasioita esimerkiksi raha, yliopistot sekä journalit (jotka ovat ihmisen toiminnasta riippuvaisia). Tässä tutkielmassa filosofinen keskustelu ajattelevista koneista on nähty institutionaalisenä tosiasiana, joka osaltaan muokkaa todellisuutta. Esimerkiksi Turingin testi on jossain määrin muodostunut tekoälyn määrittelyn erääksi työvälineeksi, jolloin Turingin esittämä matkimispelin asetelma tosiasiasa voidaan nähdä määrittävän sitä, mitä koneen ajatteluksi tai tekoälyksi kutsumme. Näin se siis voidaan nähdä ikään kuin rakentavan todellisuutta, tai ainakin ohjaavan sen tulkintaa tietyllä tavalla, joka ei ilman sitä olisi mahdollinen. Samalla myös keskustelu ajattelevista koneista on vaikuttanut osaltaan esimerkiksi tekoälytutkimuksen alkuun ja sen teemoihin, joten osaltaan tässä tarkasteltu keskustelu on ainakin sosiaalisen konstruktionismin kautta nähtynä muokannut tekoälytutkimusta ja siis myös tekoälyn kehittämistä ja itse tekoälyä tiettyyn suuntaan.

Tutkielma on kuitenkin perustavalta ajatukseltaan historiatutkimusta, jossa edellä mainitut tieteenfilosofiset taustalla vaikuttavat teoriat on otettu eksplisiittisesti huomioon. Historiatutkimuksen tutkimusstrategian ajatuksena tässä työssä on ollut selvittää tutkimuskohteen ilmenemistä erityisesti sen teemojen kautta ja pyrkiä selittämään esiin nousseita teemoja ja niiden yhteyksiä toisiinsa sekä niitä ympäröivään kontekstiin. Näin on pyritty luomaan historiallista rekonstruktiota siitä, millainen oli filosofinen keskustelu ajattelevista koneista englantilaisissa filosofian jurnaaleissa vuosina 1950–1975.

¹³⁰ Searle, 1995.

Lähteet

Internet-sivut tarkastettu 23.4.2021.

Tutkimusaineisto

Bunge, Mario: Do Computers Think? (I). *The British Journal for the Philosophy of Science*, volume 7, 1956.

Bunge, Mario: Do Computers Think? (II). *The British Journal for the Philosophy of Science*, volume 7, 1956.

George, Frank Honywill: Could Machines Be Made to Think? *Philosophy*, volume 31, 1956.

Gunderson, Keith: Robots, Consciousness, and Programmed Behaviour. *The British Journal for the Philosophy of Science*, volume 19, 1968.

Gunderson, Keith: The Imitation Game. *Mind*, volume 73, 1964.

Kapp, Reginald: Living and Lifeless Machines. *The British Journal for the Philosophy of Science*, volume 5, 1954.

Lacey, Alan Robert: Men and Robots. *The Philosophical Quarterly*, volume 10, 1960.

Mays, Wolfe: Can Machines Think? *Philosophy*, volume 27, 1952.

Miles, Thomas Richard: On the Difference between Men and Machines. *The British Journal for the Philosophy of Science*, volume 7, 1957.

Turing, Alan: Computing Machinery and Intelligence. *Mind*, volume 59, 1950.

Tutkimuskirjallisuus

Alasuutari, Pertti: *Laadullinen tutkimus 2.0*. Vastapaino, Tampere, 2011.

Avramides, Anita: Other Minds. *Stanford Encyclopedia of Philosophy*, winter 2020 edition, Stanford, 2020.

Bostrom, Nick: *Superintelligence – Paths, Dangers, Strategies*. Oxford University Press, Oxford, 2014.

Broad, Charlie Dunbar: The Experimental Establishment of Telepathic Precognition. *Philosophy*, volume 19, 1944.

Cooper, Barry S. & Leeuwen, Jan van (toim.): *Alan Turing – His Work and Impact*. Elsevier, Waltham, Oxford, Amsterdam, 2013.

Copeland, Jack; Bowen, Jonathan; Sprevak, Mark & Wilson Robin: *The Turing Guide*. Oxford University Press, Oxford, 2017.

Creath, Richard: Logical Empirism. *Stanford Encyclopedia of Philosophy*, spring 2021 edition, Stanford, 2021.

Descartes, Rene: *Teokset II – Mietiskelyjä ensimmäisestä filosofiasta, kirjeitä 1640–1641*. Gaudeamus, Helsinki, 2015 (2002).

Dewhurst, Joe: British Cybernetics. Teoksessa Sprevak, Mark & Colombo, Matteo: *The Routledge Handbook of the Computational Mind*. Routledge, London, New York, 2019.

Eco, Umberto: *Ruusin nimi*. WSOY, Helsinki, 1983.

Epstein, Robert; Roberts, Gary & Beber, Grace (toim.): *Parsing the Turing Test – Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Springer, Berlin, 2008.

Floyd, Juliet & Bokulich, Alisa (toim.): *Philosophical Explorations of the Legacy of Alan Turing – Turing 100*. Springer, Cham, 2017.

Hacker, Peter: Wittgenstein – ihmislunnosta. Teoksessa Monk, Ray & Raphael, Frederic (toim.): *Suuret filosofit*. Otava, Helsinki, 2004.

Hanfling, Oswald: Logical positivism. Teoksessa Shanker, Stuart: *Philosophy of Science, Logic and Mathematics in the Twentieth Century*. Routledge History of Philosophy, Volume IX. Routledge, London, New York, 2004.

Hietala, Marjatta: Mitä tutkia ja miten? Teoksessa Autio, Sari; Katajala-Peltomaa, Sari & Vuolanto, Ville (toim.): *Historioitsijan arki & tutkimuksen prosessi*. Vastapaino, Tampere, 2001.

Hodges, Andrew: Alan Turing. *Stanford Encyclopedia of Philosophy*, winter 2019 edition, Stanford, 2019.

Hofstadter, Douglas: *Gödel, Escher, Bach – An Eternal Golden Braid*. Vintage, New York, 1990.

Horn, Robert: The Turing Test: Mapping and Navigating the Debate. Teoksessa Epstein, Robert; Roberts, Gary; Beber, Grace (toim.): *Parsing the Turing Test – Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Springer, Berlin, 2008.

Husbands, Phil & Holland, Owen: The Ratio Club: A Hub of British Cybernetics. Teoksessa Husbands, Phil; Holland, Owen ja Wheeler Michael (toim.): *The Mechanical Mind in History*. MIT Press, Cambridge (MA), 2008.

Kortelainen, Ilmari & Järvenkylä Joose: Näkökulmia tavallisen kielen filosofiaan. Teoksessa Järvenkylä, Joose & Kortelainen, Ilmari (toim.): *Tavallisen kielen filosofia*. Gaudeamus, Helsinki, 2013.

Lavington, Simon (toim.): *Alan Turing and His Contemporaries – Building the World's First Computers*. British Informatics Society Limited, Swindon, 2012.

Leavitt, David: Turing and the paranormal. Teoksessa Copeland, Jack; Bowen, Jonathan B.; Sprevak, Mark; Wilson, Robin; ym: *The Turing Guide*. Oxford University Press, Oxford, 2017.

Lähdestmäki ym 2009a: Lähdesmäki, Tuuli; Hurme, Pertti; Koskimaa, Raine; Mikkola, Leena; Himberg, Tommi: Hermeneutiikka. *Menetelmäpolkuja humanisteille*. Humanistinen tiedekunta, Jyväskylän yliopisto, Jyväskylä, 2009. Linkki: <https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/menetelmapolku/tieteenfilosofiset-suuntaukset/hermeneutiikka>.

Lähdestmäki ym 2009b: Lähdesmäki, Tuuli; Hurme, Pertti; Koskimaa, Raine; Mikkola, Leena; Himberg, Tommi: Historiatutkimus. *Menetelmäpolkuja humanisteille*. Humanistinen tiedekunta, Jyväskylän yliopisto, Jyväskylä, 2009. Linkki: <https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/menetelmapolku/tutkimusstrategiat/historiatutkimus>.

Lähdestmäki ym 2009c: Lähdesmäki, Tuuli; Hurme, Pertti; Koskimaa, Raine; Mikkola, Leena; Himberg, Tommi: Konstruktivismi. *Menetelmäpolkuja humanisteille*. Humanistinen tiedekunta, Jyväskylän yliopisto, Jyväskylä, 2009. Linkki: <https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/menetelmapolku/tieteenfilosofiset-suuntaukset/konstruktivismi>.

Lähdestmäki ym 2009d: Lähdesmäki, Tuuli; Hurme, Pertti; Koskimaa, Raine; Mikkola, Leena; Himberg, Tommi: Sosiaalinen konstruktionismi. *Menetelmäpolkuja humanisteille*. Humanistinen tiedekunta, Jyväskylän yliopisto, Jyväskylä, 2009. Linkki: <https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/menetelmapolku/tieteenfilosofiset-suuntaukset/sosiaalinen-konstruktionismi>.

Mackenzie, Brian: Joseph Banks Rhine: 1895–1980. *The American Journal of Psychology*, volume 94, 1981.

Mainzer, Klaus: *Künstliche Intelligenz – Wann übernehmen die Maschinen?* Springer, Berlin, Heidelberg, 2016.

McCulloch, Warren & Pitts, Walter: A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematica Biology*, volume 52, 1990.

Morris, Michael: *An Introduction to the Philosophy of Language*. Cambridge University Press, Cambridge, 2007.

Natale, Simone & Ballatore, Andrea: Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *The International Journal of Research into New Media Technologies*, volume 26, 2020.

Oxford English Dictionary: Clairvoyance.
Linkki: <https://www.lexico.com/definition/clairvoyance>.

Oxford English Dictionary: Extrasensory perception.
Linkki: https://www.lexico.com/definition/extrasensory_perception.

Oxford English Dictionary: Natural language.
Linkki: https://www.lexico.com/definition/natural_language.

Oxford English Dictionary: Precognition.
Linkki: <https://www.lexico.com/definition/precognition>.

Oxford English Dictionary: Psychokinesis.
Linkki: <https://www.lexico.com/definition/psychokinesis>.

Oxford English Dictionary: Telepathy.
Linkki: <https://www.lexico.com/definition/telepathy>.

Parikka, Jussi: *Koneoppi – ihmisen, teknologian ja median kytkennät. Kulttuurituotannon ja maisemantutkimuksen laitoksen julkaisuja I*, Turun yliopisto, Pori, 2004.

Perhoniemi, Tuukka: *Mitan muunnelmat – Miten määritämme maailmaa, ihmistä ja tietoa*. Vastapaino, Tampere, 2014.

Powell, Betty: Descartes' Machines. *Proceedings of the Aristotelian Society*. volume 71, 1970.

Rorty, Richard: The Historiography of Philosophy: Four Genres. *Philosophy in History*, volume 67, 1984.

Saaranen-Kauppinen, Anita & Puusniekka, Anna: Sisällönanalyysi. *KvaliMOTV - Menetelmäopetuksen tietovaranto*. Yhteiskuntatieteellinen tietoaarkisto, Tampere, 2006.
Linkki: https://www.fsd.tuni.fi/menetelmaopetus/kvali/L7_3_2.html.

Sayre, Kenneth: Cybernetics. Teoksessa Shanker, Stuart: *Philosophy of Science, Logic and Mathematics in the Twentieth Century*. Routledge History of Philosophy, Volume IX. Routledge, London, New York, 2004.

Searle, John: *The Construction of Social Reality*. Penguin Books, London, 1995.

Searle, John: Minds, Brains, and Programs. *Behavioral and Brain Sciences*, volume 3, 1980.

Sintonen, Matti: Positivismi. Teoksessa Niiniluoto, Ilkka & Saarinen, Esa (toim.): *Vuosisatamme filosofia*. WSOY, Helsinki, 1986.

Strawson, Peter Frederick: On Referring. *Mind*, volume 59, 1950.

Tarasti, Eero: *Musiikki humanismina – Suomen saloilta Pariisin salonkeihin. Esseitä vuosilta 2003–2013*. University Press of Eastern Finland, Joensuu, Kuopio, 2013.

Turing, Alan: On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, volume 2, 1936.

Turing, Alan: Computing Machinery and Intelligence (kommentoitu painos). Teoksessa Epstein, Robert; Roberts, Gary; Beber, Grace (toim.): *Parsing the Turing Test – Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Springer, Berlin, 2008.

Uebel, Thomas: Vienna Circle. *Stanford Encyclopedia of Philosophy*, summer 2020 edition, Stanford, 2020.

Wittgenstein, Ludwig: *Tractatus logico-philosophicus – Logisch-philosophische Abhandlung*. Suhrkamp, Frankfurt a. M., 1963 (alkup. 1921).