



Turingin testi, interrogatiivimalli ja tekoäly

ARTO MUTANEN JA ILPO HALONEN

Johdanto

Alan Turing (1912–1954) tunnetaan logiikassa *Turingin teesistä*, joka kytkeytyy artikkeliin "On Computable Numbers, with an Application to the Entscheidungsproblem" vuodelta 1936 ja tekoälyn puolella *Turingin testistä*, jonka hän muotoili artikkelissaan "Computer Machinery and Intelligence" vuodelta 1950. Molemmat näistä ovat edelleen kiinnostavia tutkimuskohteita. Artikkelimme keskittyy nimenomaan tekoälyn problematiikkaan, joten Turingin testi on sikäli keskeisempi. Ei ole kuitenkaan kovin helppo muodostaa selkeää kuvaa siitä, mitä Turing itse tavoitteli testillään, eikä myöskään siitä, mitä testi lopulta testaa. Jotta voisimme paremmin saada selville, mistä Turingin testissä on kyse, on Turingin 1940- ja 1950-lukujen tekstejä tarkasteltava laajemmin. Nämä antavat Turingin ajattelusta rikkaan ja monipuolisen kuvan. Selvästikin Turing etsi ja pyrki kehittämään koneita, jotka ajattelevat ja ovat älykkäitä. Tästä näkökulmasta Turing ei tavoittanut testinsä muotoilullaan sitä, mitä hän tavoitteli.

Saadaksemme paremman kuvan siitä, mitä Turing tavoitteli, tarkastelemme myös hänen 1930-luvun tekstejään, joiden analyysimenetelmä on filosofisesti rikas. On olemassa selviä viitteitä, että Turing etsi vastaavaa analyysimenetelmää myös tekoälyä tarkastelevissa töissään. Kuitenkaan ei ole helppoa löytää sellaista muotoilua, jollaista Turing etsi. Yhtäältä Turing oli mukana kehittämässä pelejä pelaavia koneita, joiden kautta tietyt kysymyksenasettelut tulivat luontevasti esiin, mikä osaltaan on ollut vaikuttamassa Turingin testin muotoon. Toisaalta Turingin aikaan ei ollut loogis-käsitteellisiä keinoja muotoilla testiä sellaiseen muotoon, jollaista hän etsi.

Yksi luonteva muotoilu olisi ollut tuloksena, jos Turing olisi muokannut testinsä sellaiseen muotoon, jossa tietokone osoittaakseen älykkyyttä ei olisi toiminut vastaajana vaan kysyjänä. Näin testistä muotoutuisi luontevasti tiedonetsintäpeli, jossa testattava kone tai muu, jonka älykkyyttä testataan, joutuisi osoittamaan älykkyytensä strategisin kysymyssarjoin, joiden avulla se ratkoisi annettuja ongelmia.

Tuodessamme esiin tällaisten kyselypelien luonnetta tarkastelemme professori Jaakko Hintikan (1929–2015) kehittämää tutkimuksen *interrogatiivimallia*, jonka perusidea on juuri edellä mainittu Turingin testin muunnos kyselypeliksi. Interrogatiivimallin yhteys Turingin ajatteluun on oikeastaan hämmästyttävän läheinen. Itse asiassa Hintikka sovelsi 1970- ja 1980-luvun töissään interrogatiivimallia ja sen taustalla olevia perustavia ideoita tekoälyn analyysiin. Hintikka katsoo, ettei ole hedelmällistä lähteä simuloimaan tietokoneilla inhimillistä ajattelua ikään kuin jo tuntisimme, mitä inhimillinen ajattelu on. Strategian tulisi pikemminkin olla luonteeltaan sellainen, että tarkastelisimme erilaisia älykkäinä pitämiämme toimintoja ja ongelmanratkaisuja, olivatpa ne sitten ihmisten, eläinten tai koneiden suorittamia. Tämä on oikeastaan hyvin lähellä niitä ajatuksia, joita Turing kehitti 1940- ja 1950-luvuilla.

Alan Turing

Alan Turing työskenteli lyhyen elämänsä aikana monen tieteenalan parissa ja oli edelläkävijä muun muassa tietojenkäsittelytieteeseen, tekoälyn ja keinoelämän aloilla. Hänen monipuolisuutensa on huomattavan laajaa ja hänen ideoidensa kantavuus on erittäin suurta. Useat hänen esittämistään ideoista ovat vasta nykyään tulleet ajankohtaisiksi. (Copeland 2004; 2012.) Turingin esitystapa on vaikea. Paikoitellen hänen tekstinsä on raikkaan luontevaa ja miellyttävää, kun taas paikoitellen se on raskasta, teknistä, provokatiivista ja monitulkintaistakin kieltä. Tämä tekee Turingin ideoista kiistanalaisia ja tulkinnanvaraisia.

Tässä artikkelissa keskitymme siihen, mitä Turing ajatteli tekoälystä, jolloin yksi keskeisimmistä käsitteistä on Turing-testi, jonka hän muotoili artikkelissa Turing (1950). Tarkastelemme Turingin testiä seuraavassa alaluvussatarkemmin. Turing tutki tekoälyn problematiikkaa pitkän ajan. Jo 1940-luvun alussa hän kehitti koneita, jotka osasivat pelata pelejä, kuten shakkia. Seuraavassa luodaan lyhyt luettelomainen katsaus Turingin nimenomaan tekoälyyn liittyviin teksteihin, jotta saamme käsityksen Turingin ajattelusta.

Turing (1947) on tekoälyn kehityksen kannalta kiinnostava teksti. Siinä Turing esittelee sekä koneoppimisen että laajemman koneälykkyyden käsitteet. Hän tarkastelee muistin merkitystä oppimisen ja älykkyyden kannalta. Turing (1948) esittelee selvästi edellistä vahvemman näkemyksen koneiden mahdollisesta älykkyydestä. Hän puhuu esimerkiksi ajattelevista koneista ja sähköisistä aivoista. Tässä artikkelissa hän ideoi koneita, jotka ennakoivat konnektionismia. Turing piti vuonna 1951 kaksi radioesitelmää Turing (1951a) ja (1951b), joissa hän täsmentää ajatustaan älykkäistä koneista. Edellä mainitut artikkelit ovat keskeisiä pyrittäessä ymmärtämään Turingin käsitystä tekoälystä. Turing jatkaa asian käsittelyä myös myöhemmissä artikkeleissaan, esimerkiksi Turing (1953). Artikkelimme kannalta on keskeistä huomioda, että Turingin (1936) analyysi ihmislaskennasta on tekoälytutkimuksen kannalta olennaisen tärkeää. Hän muotoilee tämän analyysinsa perusteella niin sa-

notut Turingin koneet, joilla on sama laskentavoima kuin ihmislaskijalla, mikä tunnetaan Turingin teesinä (Copeland 1993, 230–231; Copeland 2004, 40–41; Sieg 1991, 1–2). Myöhemmin hän hyödyntää näitä koneita artikkelissa Turing (1950), jossa hän puhuu näistä erillistilakoneina. Turing (1938) yleistää Turingin koneet siten, että niiden on mahdollista saada laskenta-prosessin aikana lisäinformaatiota. Lisäksi hän muotoilee tässä artikkelissa ns. rajasynteesin käsitteen, joka on osoittautunut hedelmälliseksi niin logiikassa, tekoälytutkimuksessa kuin filosofiassa.

Turingin testi

Turingin testillä tarkoitetaan Turingin artikkelissaan ”Computing Machinery And Intelligence” (1950) muotoilemaa jäljittely-peliä (*imitation game*). Turing muotoilee pelin seuraavalla tavalla:

Ongelman uutta muotoa voidaan kuvailla pelillä, jota sanomme ’jäljittelypeliksi’. Sitä pelaavat kolme ihmistä: mies (A), nainen (B) ja kysymysten esittäjä (C), joka voi edustaa kumpaa tahansa sukupuolta. Kysyjä sijoitetaan eri huoneeseen kuin kaksi muuta. Pelin tavoitteena on, että kysymysten esittäjä saa selville, kumpi vastaajista on mies ja kumpi on nainen. (Turing 1950, 441, lainannut Hodges, 2000, 50–51.)

Turing kuvailee lyhyesti peliä esimerkiksi tarkastelemalla, millaisia strategioita pelaajien olisi hyvä käyttää pelissä. Tämän jälkeen hän antaa pelille uuden muotoilun: ”Esitämme nyt kysymyksen: ’Mitä tapahtuu, kun kone omaksuu tässä pelissä osan A?’ Päätyykö kysymysten esittäjä väärään johtopäätökseen yhtä usein näin pelattaessa, kuin jos peliä pelataan miehen ja naisen välillä?” (ibid.). Tämän hän sanoo korvaavan alkuperäisen kysymyksen ”voivatko koneet ajatella?” (ibid.).

Turingin mukaan tämä kysymys mahdollistaa, tai sen pitäisi mahdollistaa, osapuolten intellektuaalisten piirteiden kriittisen arvioinnin siten, että jos emme kykene erottamaan ihmistä ja konetta tässä suhteessa, niin meidän tulisi myöntää koneella

olevan älykkyyttä (Hodges 2000, 51). Turing muotoili testinsä provokatiiviseen tapaan, mikä tekee testin tulkinnan vaikeaksi. Tämä on myös aiheuttanut selkeitä väärinymmärryksiä. Pelin muotoilu jo sinällään tuo esiin intellektuaalisten kykyjen kannalta epäolennaisen sukupuolen, mikä on aiheuttanut oikeutetuakin kritiikkiä. Emme tässä ryhdy käymään läpi testin herättämää laajaa keskustelua, esimerkiksi kirjasta Moor (2003) saa hyvän kuvan artikkelimme kannalta relevantista keskustelusta.

Testin tarkoituksesta on käyty laajaa keskustelua (kts. esim. Moor 2003), jota Haikonen (2017, 257–258) analysoi. Hän toteaa Turingin testin tietyn provokatiivisen luonteen, jonka myös Hodges (2000, 50) tuo esiin seuraavalla tavalla: ”Jos Turingin tarkoituksena oli ärsyttää humanisti-intellektuelleja tällä seuraileikkijohdatuksellaan luovan mielen ongelmiin, niin hän kyllä onnistui siinä.” Kuitenkaan Turingin idea, mikä se sitten lopulta olikaan, ei kuitenkaan tule selkeästi esiin. Onkin edelleen ”tuhoisia väärinkäsityksiä” testin luonteesta ja tarkoituksesta. Haikonen (2017, 258) antaa tylyn yhteen vetävän arvion sanomalla, että ”Turingin testillä on epäilemättä tiettyä viihdearvoa, mutta muuta hyötyä siitä ei ole”.

Turing oli provokatiivinen kirjoittaja, mutta hän oli myös loogis-käsitteellisesti tarkka. Turingin testi ei ole vain yksi erillinen idea hänen tuotannossaan, vaan se kytkeytyy tiiviisti hänen muuhun tuotantoonsa. Tarkastelemalla testiä yhteydessä hänen muuhun tuotantoonsa, voimme nähdä, mitä hän testillään pyrki tuomaan esiin ja mitä hän ei itse asiassa sanonut, vaikka hänen ehkä olisi pitänyt sanoa.¹ Tarkoituksenamme ei ole antaa ”oikeaa” tulkintaa Turingin testille tai muillekaan Turingin ideoille, vaan katsoa Turingin ajattelun luonnetta yleisemmin ja pyrkiä tuomaan esiin joitakin tekoälyyn liittyviä perustavia ajatuksia, joita on mahdollista edelleen kehittää.

Turingin vuonna 1936 julkaistu teksti ”laskettavista luvuista ratkaisuongelmaan sovellettuna (”On Computable Numbers, With an Application to the Entscheidungsproblem”) on merkittävä artikkeli logiikan historiassa. Turing tutustui artikkelin

¹ Hintikka (1996) puhuu 1/2- ja 1½-totuuksista tässä merkityksessä.

teemaan Newmanin matematiikan perusteiden kurssilla 1935. Gödel oli vuonna 1931 todistanut, että teorialle, joka sisältää riittävän määrän lukuteoriaa ei ole mahdollista saavuttaa sekä ristiriidattomuutta että täydellisyyttä samanaikaisesti. Ei siis ole mahdollista, että jokaiselle annetulle kielen lauseelle joko se itse tai sen negaatio olisi todistuva. Ristiriidattomuus estää, että molemmat olisivat todistuvia. Siten jokaisessa matemaattisessa järjestelmässä on tosia lauseita, jotka ovat todistumattomia. (Hodges 2000, 13–14.)

Gödel ei ottanut kantaa Hilbertin ratkaisuongelmaan. Ratkaisuongelma edellyttää, että meillä on täsmällinen ratkaisumenetelmän käsite. Jo antiikista lähtien on muotoiltu joihinkin ongelmiin täsmälliset algoritmiset menetelmät, joilla tämä ongelmatyyppi voidaan ratkaista. Kuitenkaan ei ollut täsmällistä ratkaisumenetelmän käsitettä, jolla olisi voitu antaa ratkaisuongelman edellyttämä yleinen ratkaisu. Turingin muotoilu ratkaisumenetelmälle on nerokas: hän muotoilee idealisoidun koneen, Turingin koneen, joka eksplikoi etsityn ratkaisumenetelmän. (Hodges 2000, 14–15.)

Muotoillessaan Turingin koneita Turing analysoi, miten inhimillinen laskija, jota tuolloin kutsuttiin *computeriksi* tai joskus *computeriksi*, suorittaa laskentaa. Tarkoituksena ei ollut näiden ihmisten laskentatoimituksen empiirinen kuvaus tai psykologinen kuvaus laskijan päättelystä, vaan löytää laskennan loogis-käsitteellinen perusta, mikä tekee artikkelista formaalin. Kuitenkin analyysissään hän pyrkii löytämään inhimilliselle laskennalle olennaisimmat tekijät, mikä ennakoi kiinnostavalla tavalla Turingin myöhemmin käyttämää ”introspektiivista analyysia”. Analyysin lopputuloksena Turing muotoilee *mekaanisen ihmisen (computer)*, joka pystyy suorittamaan annetut laskutehtävät (Sieg 2000, 2). Turingin muotoilema mekaaninen laskukone (Turingin kone) kykenee suorittamaan sille ohjelmoidut tehtävät. Turing osoittaa, että Turingin koneet pystyvät suorittamaan samat tehtävät kuin hänen muotoilemansa mekaaniset ihmiset (Turingin teesi).

Turingin käyttämä esimerkki oli päättymätön desimaaliluku, jonka desimaalikehitelmän kone laskee. Siten kone poikkeaa

nykyisin tunnetuista siinä, että oikein toimiessaan koneet (*circle-free*) eivät pysähdy. Tällaisilla koneilla on kiinnostavia loogisia sovellutuksia, esimerkiksi hyperlaskettavuus (Copeland 2002b; Marcer & Fedorec 1994; Bringsjord & Zenzen 2003; Hintikka & Mutanen 1998) ja filosofisia sovellutuksia, esimerkiksi formaali oppimisen teoria (Gold 1967; Oserson, Stob & Weinstein 1986) tai hyperlaskentaan kytkeytyvä mielenfilosofia (Bringsjord & Zenzen 2003). Turingin koneet ovat yksinkertaisia, ja niiden yleistyksenä Turing muotoilee universaalien Turingin koneen, joka on nykyisen tietokoneen teoreettinen edeltäjä. Se on ohjelmitavissa oleva kone, joka periaatteessa voi tehdä minkä tahansa Turingin koneen tehtävän. Universaali kone tuo esiin ohjelmitavuuden, joka mahdollistaa Turingin kehittelevän ajatusta oppivista koneista, joiden oppiminen perustuu koneen kykyyn muuttaa ohjelmaansa.

Siten Turingilla oli jo varhaisissa loogisissa töissään iduillaan ajatus ihmisen ja koneen välisen suhteen tarkastelusta (Copeland 2012, 186). Myöhemmin hän tarkasteli mahdollisuutta opettaa pelejä koneille, jolloin shakki oli tärkeä esimerkki (Copeland 1993; Michie 1961). Näissä Turingin ajatus on samankaltainen kuin hänen artikkelissaan vuodelta 1936: verrataan koneen oppimista ihmisen oppimiseen. Turing (1936, 59) toteaa, että laskevaa ihmistä voidaan verrata koneeseen. Turing (1938, 192–194) pohtii intuition ja kekseliäisyyden roolia matemaattisessa päättelyssä tuoden esiin, että kekseliäisyys voidaan korvata kärsivällisyydellä ja että ei ole mahdollista löytää intuitiota täysin poistavaa formaalia logiikkaa. Myöhemmin hän kuvaa intuition ja kekseliäisyyden suhdetta kirjeessään Newmanille tuomalla esiin, että kekseliäisyys intuition ohjaamana käyttää erilaisia keinoja, kuten geometrisia kuvioita ja piirroksia (Copeland 2004, 135).

Turing ei vielä 1930-luvun artikkeleissaan kehitellyt varsinaisesti tekoälyä, vaikkakin näin jälkikäteen katsottuna näiden artikkeleiden sisällöllä on suuri merkitys myös tekoälyteorian kannalta. Termi tekoäly (*artificial intelligence*) ei ollut Turingilla käytössä, vaan se vakiintui vasta myöhemmin. Eräänä merkittävänä käännekohtana tekoälytermin yleistymiseen on vuonna

1956 järjestetty seminaari "The Dartmouth Summer Research Project on Artificial Intelligence" (Copeland 2004, 353). Turing (1950) käyttää tekoälyä käsitellessään termiä *erillistilakone*, joka viittaa vuoden 1936 artikkelissa määriteltyihin Turingin koneisiin. Varsinaisesti tekoälyyn liittyvään problematiikkaan Turing tarttui vasta 1940-luvulla. Käännekohdan täsmällinen paikantaminen ei ole helppoa, mutta sekä Copeland (2004, 353) että Hodges (2000, 40) paikantavat ajankohdan vuoteen 1941, jolloin Turingin kirjoittama koneälyä käsittelevä käsikirjoitus kiersi Britanniassa hallituksen koodi- ja salakirjoituskoulussa². Onnistuneiden salauskoodien murtamisisissa Turing oli havainnut miten "koneet pyörivät ja ihmiset saavuttivat hämmästyttäviä ja ennen näkemättömiä tuloksia soveltamalla mekaanisia menetelmiä automaattisesti" (Hodges 2000, 40). Samalla Turing alkoi kehittää myös algoritmeja shakin pelaamiseen. (Copeland 2004; Hodges 2000.)

Ensimmäinen tekoälyyn liittyvä Turingin julkaisu on hänen artikkelinsa vuodelta 1947, jossa hän tarkastelee mahdollisuutta opettaa koneita pelaamaan. Tässä artikkelissa tarkasteltu ongelma on selvästi yleisempi kuin vuoden 1936 artikkelissa tarkasteltu inhimillinen laskenta ja sen automatisointi, vaikka vuoden 1936 artikkeli onkin logiikan kannalta erittäin syvällinen ja merkittävä. Koneoppimisessa on keskeistä strategisen ajattelun oppiminen. Pelejä opetettaessa tavoite oli opettaa koneille nimenomaan pelistrategioita (Michie 1961; 1962). Tämä oli Turingilla myös keskeinen ajatus, mikä ilmenee Turingin puhuessa virheiden ja heuristiikkojen roolista koneoppimisessa (Copeland 2012, 191). Lisäksi Turing (1951, 424) vertaa koneen opettamista lasten opettamiseen, mikä muistuttaa vuoden 1936 artikkelissa olevaa ajatusta mekaanisen ihmisen laskennasta. Asiaa voidaan selkeyttää tekemällä käsitteellinen ero algoritmisen mekaanisen ihmisen ajattelun ja älykkään ajattelun välillä,

² Ensimmäisen maailmansodan jälkeen perustettu Government Code and Cypher School (GC&CS) sijaitsi toisen maailmansodan aikana Bletchley Parkissa, jossa siitä tuli kuuluisa saksalaisen Enigma-salakirjoituksen murtamiseksi tehdystä työstä.

jota muiden muassa Niiniluoto (1990) on tarkastellut. Turing ei ollut tarkka tämän erottelun suhteen, mikä tekee hänen tulkitsemisensä vaikeaksi.

Turingin (1936) analyysi ihmisen toiminnan yleisistä piirteistä on tarkastelutavaltaan loogis-käsitteellistä. Tällaista toimintaa on mahdollista simuloida koneiden avulla, jonka osoittavat hänen kehittämänsä Turingin koneet. Olennainen askel oli Turingin ajatus universaalikoneesta, joka kykenisi suorittamaan minkä tahansa tehtävän, joka olisi suoritettavissa jollain Turingin koneella (Hodges 2000, 26). Tämä tuo esiin ohjelmoinnin ajatuksen, joka mahdollistaa myös uudelleenohjelmoinnin. Ohjelmoinnin ja uudelleenohjelmoinnin idea tuo esiin myös oppimisen mahdollisuuden, mikä on Turingille keskeisen tärkeä ajatus. Vuoden 1936 artikkelissa Turingin keskeinen kysymys oli vielä varsin rajoittunut. Vuoden 1947 artikkelissaan "Lecture on the Automatic Computing Engine" hän tarkastelee yleisempää ongelmaa ja näin myös oppivan koneen ajatus tulee selkeämmin esiin. Hänen ajatuksenaan on tarkastella, miten ihminen oppii elämänsä aikana, ja pyrkiä mallintamaan inhimillistä oppimista koneilla. Vuonna 1951 pitämässään radioesitelmässään "Intelligent Machinery, A Heretical Theory" Turing syventää analogiaa ihmisen ja koneen välillä tarkastelemalla virheitä, joita opittaessa tehdään. Metodisesti tämä muistuttaa vuoden 1936 artikkelia, jossa perusajatus on analysoida inhimillistä laskentaa. Tämä on tekoälyn kehittämisen kannalta hedelmällinen lähestymistapa. Turing tarkastelee artikkelissa yksityiskohtaisesti ohjelmoinnin problematiikkaa sekä muistin roolia oppimisessa. Turingin kehitellessä oppivia koneita hän päätyy ideoimaan järjestymättömiä koneita (*unorganized machines*), jotka olennaisesti muistuttavat myöhemmin kehiteltyjä konnektionistisia koneita. (Copeland 2004, 403; Turing 1948, 422-429.)

Järjestymättömät koneet tuovat näkyviin piirteitä, joita "älykkyyttä edellyttäviin" tehtäviin kykenevillä koneilla tulisi olla. Onkin tärkeää kysyä, onko Turingin testi tai sen artikkelissa (Turing, Braithwaite, Jefferson & Newman 1952) kehitetty versio kykenevä arvioimaan näiden organisoimattomien

koneiden älykkyyttä. Ei siis ihme, että Turingin testin tulkinta on edelleen keskustelun kohteena. Ainakin tietty yhteisymmärrys vallitsee, että Turing etsii testillään operationaalista määritelmää älykkyydelle.

Turingin testi syntyi pitkän kehittelyn tuloksena. Hän muotoili jo vuoden 1948 artikkelissaan testin, jolla hän problematisoi älykkyyden käsitettä. Tässä testissä hänellä on tekoälytutkimuksen eräänlainen standardiesimerkki: shakki. Turing pohjustaa testiään tuomalla esiin, että emme pidä jotain älykkäänä, jos sen käytös on ennakoitavissa tai jos toimintaa ohjaa jokin tietty suunnitelma (Turing 1948, 431), mikä tekisi toiminnasta mekanistista. Hän muotoilee testin, jonka avulla voidaan erottaa tällaiset mekaaniset toimijat ja inhimilliset toimijat toisistaan.

Turingin testi on selkeästi kehittyneempi versio tästä aiemmasta testistä. Turingin testiin liittyy selkeitä strategisia piirteitä, joita Turing kuvailee artikkelissaan. Esimerkiksi hän muotoilee seuraavan kysymys-vastaus-keskustelun, jonka on tarkoitus kuvata Turingin testin vaiheita:

K: Kirjoita minulle sonetti *Forth bridgestä*.

V: Minä en ala tähän. En ikinä kirjoittaisi runoutta.

K: Laske yhteen luvut 34957 ja 70764.

V: (Noin 30 sekunnin tauko ja sitten vastaus) 105621.

K: Pelaatko shakkia?

V: Kyllä.

K: Minulla (mustalla) on vain kuningas ruudussa e8 eikä muita nappuloita. Sinulla (valkealla) on vain kuningas ruudussa e6 ja torni ruudussa a1. On sinun siirtosi. Mitä siirrä?

V: (Noin 15 sekunnin tauon jälkeen.) Tornin ruutuun a8 ja matti.

(Turing 1950, 442; lainannut Hodges 2000, 51.)

Tässä on kiinnostavia strategisia vivahteita, esimerkiksi ”inhi-millinen” vastaus ”minä en ala tähän” tai yhteenlaskuun käytetty harkinta-aika ja lopuksi virheellinen vastaus.

1940-luvun lopun ja 1950-luvun alun tekoälytutkimus ”oli intellektuaalista: tietokoneet ajateltiin ihmismielen kaltaisiksi, niitä opetettiin ratkaisemaan monimutkaisia matemaattisia ongelmia ja pelaamaan shakkia” (Niiniluoto 1999, 65). Tämä luonnehdinta pätee osin myös Turingiin, joka kehitti shakkia pelaavia koneita jo 1940-luvun alussa. 1948 hän suunnitteli yhdessä David Campernownen kanssa shakkia pelaavan ”koneen” (Copeland 2004, 563). Turing kuvasi koneiden kehittelymenetelmäänsä ”introspektiiviseksi analyysiksi”, jolla hän tarkoitti, että hän suunnitteli koneet pelaamaan samalla tavoin kuin hän itse pelasi shakkia. (Copeland 2004, 565.) Tämä introspektiivinen analyysi muistuttaa pitkälle sitä, miten hän 1930-luvulla kehitti Turingin koneita.

Tekoälyn tutkijoille pelit – ja erityisesti shakki – ovat olleet jatkuvan kiinnostuksen kohteena. Pelejä pelaavia koneita on tarkasteltu nimenomaan oppivina ja strategisina toimijoina. Hintikka (1990, 308) viittaa von Neumannin kirjaan *The Computer and the Brain*, jossa etsitään koneen ja ihmisen välistä parallelismia. Hintikan mukaan tämä piirre on tekoälytutkimuksen kannalta keskeisen tärkeää: ei tule olettaa, että tietäisimme, miten ihmisäly toimii tai mitä se on. Niiniluoto (1999, 65) viittaa tähän samaan tuodessaan esiin Marvin Minskyn ja Seymour Papertin oivalluksen, ”että AI:n keskeinen pulmakysymys onkin terveen järjen ajattelun ja päättelyn oppiminen”. Turing oli töissään jo huomionnut sekä Hintikan että Niiniluodon esittämät seikat.

Vuoden 1948 artikkelissaan ”Intelligent Machinery” Turing esittelee *ajattelevan koneen*, jolla on ”sähköiset aivot”. Nämä kehittelyt lähestyvät sitä ajatusmallia, jota Hintikka ja Niiniluoto kritisoivat. Kuitenkin samassa artikkelissa Turing luonnehtii *lapsikoneen*, joka oppii ihmislapsen tavoin. Näin ihminen-kone-suhde on problemaattinen ja ihmisen oppimisprosessi on jotain, mitä emme tarkkaan tunne, mutta voimme kehittää konetta, jota voidaan opettaa. Lisäksi tässä artikkelissa hän kehittää

ensimmäisen alustavan testin, jossa tarkoituksena on katsoa miten hyvin kone voi toimia ihmisen tavoin. (Turing 1948, 431). Tässä kyse on siitä voiko inhimillinen tarkastelija peliä seuraamalla havaita, onko shakin pelaaja ihminen vai kone.

Turing (1950) tarkastelee kysymystä, voiko kone ajatella. Kuitenkin tällainen kysymyksenasettelu pitää sisällään oletuksen, että meillä on jo tiedossa, millaista inhimillinen ajattelu on. Ajatteleva kone on riittävän samankaltainen kuin ajatteleva ihminen. Metodisesti tämä poikkeaa sekä Turingin 1936 että 1947 artikkeleista. Voimme kuitenkin ymmärtää myös tämän kysymyksen metodisesti eri tavoin, esimerkiksi tarkastelemalla simulaation käsitettä. Simulaatiosta voidaan ajatella (ainakin) kahdella perustavasti eri tavalla. Ensiksi voidaan ajatella, että simuloitava asia X on tietty ja sitä simuloiva on vain jotain X:n kaltaista, mutta ei todella X:ää. Esimerkkinä voi mainita keino-nahan, joka "simuloi" nahkaa olematta kuitenkaan nahkaa. Copeland (1993, 27) kutsuu tällaista simulaatio₁:ksi. Toisaalta voimme ajatella, että jonkin asian X simulaatio₂ on täsmälleen X:n kaltaista, mutta sen muotoutuminen ei ole normaali. Esimerkkeinä ovat (teollisuudessa tai laboratoriossa valmistetut) synteettiset aineet, jotka todella ovat kyseistä ainetta. Esimerkkinä voisi mainita keinotekoisien hiilen, joka on hiiltä, mutta jota on valmistettu laboratoriossa, tai monet laboratoriossa valmistetut lääkeaineet, jotka siis simuloivat - simulaatio merkityksessä 2 - luonnon kasveissa esiintyviä lääkeaineita. (Copeland 1993, 47.)

Simulaatio on olennaisen tärkeä ajatus. Ei ole varmaa, kumpaa simulaatiota tarkoitetaan puhuttaessa tekoälystä. Ei ole myöskään varmaa, kumpaa Turing itse tarkoitti. Kuitenkin esimerkiksi Haikonen (2017, 258) pitää simulaatiota₂ vääränä. Hän sanoo, että "[p]ikkulapset voivat kuvitella, että nuket tuntevat mielihyvää ja mielihäpä, ja aikuiset ihmiset saattavat kuvitella, että älypuhelimilla on älyä". Haikonen pitää tällaista kuvitelmia "naiivina antropomorfisena harhana". Turing (1953, 569) muotoilee koneen kyvykkyyteen liittyen kuusi kysymystä, joista kuudes liittyy koneen kykyyn tuntea (*have feelings*). Hän vastaa, ettei hän voi olla koskaan varma koneen kyvystä tuntea,

muttei hän myöskään voi olla varma, että toinen ihminen voi tuntea. Näin Turing osoittaa, että edellä oleva Haikosen argumentti voi olla tosi, mutta sen pitäisi tällöin koskea myös toisia ihmisiä.

Kun puhutaan koneiden ajattelusta, niin onko kyse simulaatiosta 1 vai 2? Turingin testissä, jota on muotoiltu uudelleen artikkelissa Turing, Braithwaite, Jefferson & Newman (1952), on kyseessä behavioraalinen tarkastelu, jonka voi luontevasti ymmärtää simulaatio₁:n mukaisesti. Tällöin ajatuksena on, että jos kone pärjää pelissä, niin koneen on katsottava ajattelevan (Copeland 1993, 48). Kuitenkin artikkelissaan Turing (1951) argumentoi tavalla, joka tukee vahvasti simulaatio₂:n mukaista tulkintaa. Turingin testi on eräänlainen keskustelutesti, jossa kuulustelija pyrkii kysymyksiin saamaan selville, onko hänen keskustelukumppaninsa ihminen vai kone.³ Muotoillessaan esimerkkikysymyksiä ja niihin annettuja vastauksia Turing antaa koneen vastata väärin suhteellisen suurten lukujen laskutehtävään, joka ”vaikuttaa” inhimilliseltä. Lisäksi siinä on tyypiltään yksinkertaisia faktisia kysymyksiä, mutta myös strategisesti vaativampia kysymyksiä. Siten Turingin ajatuksena lienee ollut syvälinen keskustelu, joka paljastaa keskustelukumppanin kyvyn ajatella.

Turingin testin avulla pyritään löytämään keino, jolla ulkopuolelta voitaisiin arvioida, onko toimija älykäs. Siten tietokone, joka läpäisee testin, on älykäs simulaatio₂:n merkityksessä. Pyrkinessään rakentamaan älykkäitä koneita hän ei noudata tällaista logiikkaa, vaan hän pykii antamaan koneille kyvyn älykkäiseen toimintaan, johon liittyy keskeisenä piirteenä oppiminen. Turing puhuu tässä merkityksessä elektronisista aivoista tai lapsikoneista, joilla on samankaltainen älyllinen kyky kuin ihmisillä. Lapsikoneet tietyssä merkityksessä tuovat esiin älykkään koneen perustavan ajatuksen. Michie katsoi tämän ajatuksen erityisen kiehtovaksi (Copeland 2012, 186; ks. myös Michie

³ Turingin muotoilema peli ei ole tarkalleen tällainen, vaan sen muotoilu on varsin vaikeasti hahmottuva ja sen luonteesta ja merkityksestä on keskusteltu paljon (kts. esim. Moor 2003).

& Johnston 1984, 79–80). Oikeastaan tähän liittyen voisi argumentoida, että Turingin testin tulisi pikemminkin testata koneen kykyä älykkäänä pitämienne toimintojen (strategiseen) suorittamiseen, kuin meidän kykyymme ulkoapäin arvioida koneen älykkyyttä (Turing 1936; 1953).

Siten Turingin testin huono piirre liittyy siihen, että kysymyksiin vastattaessa ei välttämättä tarvita kovin älykstä strategiaa. Esimerkiksi Copeland (1993, 41) kuvaa Parryn ja Elizan, joiden strategiat (*pattern matching*) ovat yksinkertaisia eivätkä edellytä älyllistä ajattelua. Sen sijaan ne toimivat vastaamalla annettuihin kysymyksiin mekaanisen syntaktisen menetelmän avulla vaikuttaen siten älykkäiltä. Lisäksi Aamoth (2014) kertoo 13-vuotiaasta Ukrainan Odessasta tulevasta Eugene Goostmanista, joka läpäisi Turingin testin. Tosiasiassa Eugene Goostman oli kone, joka läpäisi testin. Se vältti vaikeita kysymyksiä inhimillisellä keinolla, ironialla, mikä osoittaa jälleen keskustelumenetelmään liittyviä strategisia ongelmia. Vardi (2014) tuo esiin päinvastaisen ongelman: eivät kaikki ihmiset ole ”luontevia keskustelijoita”. Vardi sanoo, että ehkä Turing itse ei olisi läpäissyt testiään.

Keskustelumenetelmä on erittäin luonteva ja hyvä idea älykkyden testaamiseen. Kuitenkin on osoittautunut, että on mahdollista muokata mekaanisia, mutta toimivia keskustelustrategioita. Tällaiset mekaaniset strategiat eivät kuitenkaan osoita keskustelijan älykkyyttä. Voidaankin syystä kritisoida keskustelumenetelmän toimivuutta Turingin esittämässä muodossa. Turingin testi ei tavoita Turingin omia tavoitteita muotoilla älykkyys strategisena ja oivaltavana etsimisenä ja löytämisenä tai oppimisprosessina. Turing (1948, 430–431) puhuu tiedon etsinnästä, joka on älyllisyyden olennainen piirre. Turingin testi mittaa tätä parhaimmillaankin epäsuorasti.

Etsinnälle on mahdollista löytää tehokkaita strategioita, jos konteksti on ”suljettu”⁴, mutta yleisessä tilanteessa ei ole mah-

⁴ Ongelmat tai kysymykset määritellään suljetuiksi, jos niihin on hyvin määritelty vastausten luokka, muuten ne ovat avoimia. Sanomme

dollista löytää mekaanisia etsintästrategioita. Siten tiedon etsijän strateginen toiminta ei ole mekanisoitavissa. Näin ollen olisi luontevaa ajatella, että Turingin testiä kannattaisi ryhtyä kehittämään siihen suuntaan, että testattava ei toimi vastaajana vaan kysyjänä ja tiedon etsijänä. Jos joku agentti kykenee toimimaan aktiivisena tiedon etsijänä erilaisissa tilanteissa, voidaan agenttia hyvin perustein pitää älykkäänä.

Kysymys ”voiko kone ajatella?” on erittäin vaikeasti jäsenytvä. Vastausta siihen voidaan etsiä kahdesta eri suunnasta. Yhtäältä voidaan lähteä tarkastelemaan, miten ihminen tai muut älykkäät olennot suorittavat älykkäitä toimintoja. Itse asiassa Turingin analyysi laskennasta noudattaa tällaista logiikkaa. Toisaalta voidaan ajatella ihmisten älykkyys ikään kuin selviteytyksi ja pyrkiä sitten imitoimaan tällaista älykästä toimintaa. Turingin 1940-luvun lopun ja 1950-luvun töissä painottuu tämä logiikka. Nämä molemmat ovat luontevia ja tärkeitä tapoja lähestyä ongelmaa. Nämä lähestymistavat eivät ole ristiriidassa keskenään ja olisi tärkeää löytää käsitteellinen keino yhdistää nämä lähestymistavat. Oikeastaan Turing itse eri muotoiluissaan vaikuttaa etsivän näitä lähestymistapoja yhdistäviä loogis-käsitteellisiä keinoja.

Turing olisi aikaisempien töidensä nojalla lähtenyt metodisesti etsimään inhimillisen ajattelun loogis-käsitteellisiä piirteitä. Tosiasiassa hän tekikin näin vuoden 1947 artikkelissaan. Kuitenkin Turing päätyi testissään nojaamaan keskustelumene- telmään, jonka tarkoitus oli arvioida koneen (tai muun kohteen) älyllisyyttä. Kuitenkin testi testaa kykeneekö kyselijä erotta- maan ihmisen ja koneen näiden antamien vastausten perus- teella. On vaikea arvioida, miksi Turing ei lähtenyt kehittä- mään ”älyllisen etsinnän logiikkaa” esimerkiksi vuoden 1936 artikkelinsa mukaisesti. Meidän ei ole tarkoituksenmukaista pyrkiä löytämään vastausta, miksi Turing ei tehnyt tällaista, vaan muotoilla loogis-käsitteellinen malli, joka yhdistää nämä kaksi lähestymistapaa.

kontekstin olevan suljettu, jos siihen liittyvä kysymys on suljettu.

Etsiessään vastausta kysymykseen ”voiko kone ajatella” Turing muotoilee Turingin testin, jolla on luonteva tehtävä tietyn tyyppisissä ongelmissa, kuten edellä havaitsimme. Kuitenkaan älyllinen toiminta ei ilmene parhaiten kysymysten vastaajan roolissa, vaan erilaisten ongelmatilanteiden ratkaisijan ja siten strategisten kysymysten esittäjän toiminnassa. Turingin logiikka vuoden 1936 artikkelissa on juuri tämän ajatuksen mukainen. Kuten olemme huomanneet, Turing etsi keinoja muotoilla vastaavaa myös tekoälyn kohdalla. Esimerkiksi radioesitelmässään (Turing 1951b) hän muotoilee yleistetyn Turingin teesin (Copeland 2004, 479), joka yleistää artikkelin Turing (1936) idean tekoälyn analyysiin soveltuvaksi.

Siten Turingin testiä on kiinnostavaa lähteä kehittämään siten, että testattava toimisi tiedon etsijän roolissa. On luontevaa ajatella, että älykkyyttä voidaan arvioida, ei niinkään vastausten antamisen avulla, vaan oivaltavien strategisten kyselyjonojen avulla. Hintikka (1969, 284) kehittää tällaista ideaa tuomalla esiin ”eräiden ajattelijoiden tähdentämää seikkaa, että kokeiden tekeminen merkitsee ’kysymysten osoittamista luonnolle’ siten, että vastausten arvo riippuu olennaisesti siitä, miten teräviä ja miten runsasta käsitteistöä käyttäviä kysymykset ovat”. Tämä huomio on kannaltamme erittäin olennainen ja edellä olevan nojalla on luontevaa ajatella, että Turing itse olisi mielellään kehittänyt testiään tähän suuntaan. Suomalainen sanonta ”kysyvä ei tieltä eksy” antaa hyvän kuvan, miten on luontevaa lähteä kehittämään Turingin etsinnän logiikkaa.

Interrogatiivimalli

Jaakko Hintikka kehitti 1980-luvulta alkaen tutkimuksen interrogatiivimalliksi kutsuttua tieteenfilosofista lähestymistapaa. Kuten jo edellä havaitsimme, tällainen ajatus on Hintikalla ollut mielessä jo 1960-luvulla. Yksi interrogatiivimallin keskeisimmistä ajatuksista on ollut muotoilla yleinen älykkään päättelyn malli, joka luonnehtii inhimillistä älykstä päättelyä niin tie-teessä kuin laajemminkin (Hintikka, Halonen ja Mutanen 2002).

Interrogatiivimalli on luonteeltaan loogis-käsitteellinen tai metodologinen, joten malli ei kuvaa tosiasiallista (empiiristä) inhimillistä päättelyä vaan päättelyn loogis-käsitteellisiä piirteitä. Näin ollen mallin soveltaminen niin yksittäisen tieteilijän päättelyn kuin yleisen tieteellisen päättelyn luonnehtimiseen on mahdollista. Luonnollisesti ei ole mitään estettä tarkastella myöskään (tieto)koneiden päättelyä mallin avulla, mutta samalla tulee kuitenkin huomata, ettei tämä anna suoraan kriteeriä erotella simulaatio_{1:n} ja simulaatio_{2:n} välillä. Kuitenkin interrogatiivimalli soveltuu erittäin hyvin tekoälyn loogis-metodiseen tarkasteluun. Interrogatiivimallin perustana olevalla logiikalla ja filosofisella lähestymistavalla on mahdollista ymmärtää Turingin perusajatuksia tekoälystä.

Interrogatiivimallin teoreettisesti keskeiset elementit ovat: (i) kysymysten logiikka, (ii) mallin loogiset muotoilut, ja (iii) filosofinen logiikka. Interrogatiivimallissa nämä kytkeytyvät läheisesti toisiinsa. Muuntamalla jotain tekijää interrogatiivimallissa koko malli muuntuu. Kukin edellä mainituista elementeistä on itsessään laaja ja kiinnostava filosofinen teema, jonka loogis-käsitteellinen tutkimus on oma tutkimustehtävä. Mainitsemme seuraavassa joitakin piirteitä, jotka ovat interrogatiivimallin ymmärtämisen kannalta olennaisia.

Kysymysten logiikan juuret ulottuvat aina Platoniin ja Aristoteleeseen saakka. Tämä antaa mallille syvällisen filosofisen merkityksen, joka luonnehtii interrogatiivimallin monia piirteitä. Erityisen keskeistä on ymmärtää kyselystrategian keskeinen rooli. Kyselystrategian ehkä keskeisin käytännöllinen esikuva on Sokrates, jonka nerokkaat strategiset kysymysprosessit ovat malliesimerkki strategisista kysymys-vastaus-peleistä. Näiden historiallisesti keskeisin teoretikko on Aristoteles, jonka logiikan teoria on kysymys-vastaus-pelien teoria. Tämän teorian strategioita hän käsitteli *Topiikassa*. Strategian käsite on kysymysten logiikan kohdalla erittäin laaja. Se pitää sisällään niin pääkysymyksen muotoiluun liittyvän problematiikan, informaation lähteiden luonteen analyysin kuin pienten operaatioiden kysymysten strategisen merkityksen pohtimisen

(*sokraattiset kyselypelit*). Yleinen strategian käsite antaa kysymysten logiikalle keskeisen roolin interrogatiivimallissa. Kysymysten logiikan formaalit muotoilut kytkevät sen osaksi laajempaa filosofisen logiikan traditiota. Näin ollen kysymysten logiikka kytkeytyy interrogatiivimallin muihin perustavin puoliin kiinnostavalla tavalla.

Interrogatiivimallilla on erilaisia loogisia muotoiluja. Hintikka on käyttänyt esimerkiksi Abraham Robinsonin malliseurauksen käsitettä, peliteoriaa sekä riippumattomuusystävällistä ensimmäisen kertaluvun logiikkaa (IF-logiikka, *Independence Friendly First Order Logic*). Malliseurauksen käsite tuo selkeästi esiin, että formaalis-loogisesti interrogatiivimallin taustalla oleva logiikka poikkeaa kiinnostavalla tavalla tavanomaisesta ensimmäisen kertaluvun logiikasta. Looginen päättely on luonteeltaan normaalia deduktiota, mutta malliseurauksen käsite tekee siitä ampliatiivista: päättelyyn tulee päättelyprosessin aikana lisäinformaatiota, joka tekee päättelyprosessin tietoa lisääväksi. Lisäksi mallista tuleva lisäinformaatio tulee ”ulkoisesta” lähteestä, joka voidaan tulkita havaintoinformaatioksi tai koikeista tulevaksi informaatioksi. Näin malli soveltuu hyvin empiirisen tutkimuksen päättelymalliksi. Malliseurauus ei kovin hyvin tuo esiin päättelyprosessin dynaamisia piirteitä.

Dynaamisten piirteiden painotuksen kannalta peliteoria on keskeisessä roolissa. Peliteoriassa peli nähdään strategisena prosessina, mikä tulee hyvin esiin peliteoreettisessa semantiikassa (Hintikka & Kulas 1983; Hintikka, Halonen ja Mutanen 2002, 306). Peli sinällään on tavoitteellista toimintaa, jossa pyritään strategisesti saavuttamaan jokin tietty tavoite, joka määrittyy pelin luonteen perusteella. Siten strategia on peliteorian kannalta keskeisessä roolissa. Peliteoria mahdollistaa myös lisäinformaation (*empiirinen informaatio*) ottamisen mukaan tarkasteluihin. Kuitenkin ongelmalliseksi interrogatiivimallin kannalta tulee se, että strategia on pelin ”sisäinen” käsite. Siten strategia ei kata kaikkia niitä merkityksiä, joita Hintikka on pyrkinyt sisällyttämään strategian käsitteeseen. Kuitenkin on mahdollista tarkastella monia erittäin olennaisia ja kiinnostavia yk-

sityiskohtia. Samalla näin tulee selkeäksi yhteys laskettavuuden teoriaa hyödyntävään lähestymistapoihin, joita ovat mm. Kellyn (1996) laskennallinen epistemologia, Hendricksin (2007) monimodaaliteoria ja formaali oppimisen teoria (Osherson, Stob & Weinstein 1986). Hintikan (1986, 65) viitatessa pakottamiseen hän tuo esiin episteemisen logiikan läheisen suhteen topologiaan, joka on eksplisiittisesti esitetty kirjassa Kelly (1996). Strategian käsite on yksi keskeisimmistä interrogatiivimallin käsitteistä. Peliteoriassa strategia tarkoittaa strategiaa pelissä, mikä ei ole sama kuin interrogatiivimallin laajempi strategian käsite.

Hintikalle oli olennaisen tärkeätä löytää logiikan esitys, joka kykenisi esittämään käsitteellisesti rikkaan dynaamisen struktuurin, joka esittäisi interrogatiivimallin olennaisia loogis-käsitteellisiä puolia. Hintikan kehittämä riippumattomuusystävällinen logiikka (IF-logiikka) on osoittautunut tässä suhteessa erittäin hedelmälliseksi. (Hintikka 1996; Väänänen 2007.) Ensimmäisen kertaluvun logiikassa kvanttorit ovat lineaarisesti järjestyneet, mikä aiheuttaa sen, että sisemmät kvanttorit kuuluvat aina ulompien kvanttorien vaikutusalaan. Tämä ei vastaa luonnollisen kielen rakennetta. IF-logiikassa oleva riippumattomuusoperaattori mahdollista tällaisen riippumattomuuden eksplikoinnin. Tämä operaattori on osoittautunut sekä loogisesti että filosofisesti erittäin tärkeäksi. (Ks. esim. Hintikka ja Sandu 1996.) Logiikan kannalta on kiinnostavaa esimerkiksi se, että siinä on mahdollista esittää niin sanotut *Henkin-kvanttorit* (haarautuvat kvanttorit), mikä osoittaa, että IF-logiikan ilmaisuvoima on suurempi kuin ensimmäisen kertaluvun logiikan ilmaisuvoima. Filosofisesti on kiinnostavaa, että riippumattomuusoperaattorin avulla on mahdollista esittää episteemisten käsitteiden hienovarainen analyysi.

IF-logiikka tuo interrogatiivimallin osaksi edellä kuvattuja Hintikan epistemologian ja kielifilosofian keskeisiä piirteitä. Näin interrogatiivimallin loogis-teoreettinen rakenne jäsentyy selkeäksi ja yhtenäiseksi teoreettiseksi kokonaisuudeksi. Erityisesti tämä on osoittautunut episteemisen logiikan kohdalla erittäin tärkeäksi. (Hintikka, Halonen & Mutanen 2002.)

Episteeminen logiikka

Interrogatiivimallin tavoite on kuvata yleistä inhimillistä rationaalisuutta, jonka yksi keskeinen piirre on kyky hankkia tietoa (Hintikka, Halonen & Mutanen 2002). Siten on selvää, että episteeminen logiikka on interrogatiivimallin kannalta keskeinen ala. Hintikan kirja *Knowledge and Belief* (1962) on historiallisesti merkittävä julkaisu episteemisestä logiikasta, joka on kiinnostava nykyisen filosofisen tekoälytutkimuksen kannalta. Kirjan julkaisun aikoihin episteeminen logiikka, ja yleisemmin modaalilogiikka, kehittyi nopeasti. Hintikka (1962) oli alallaan merkittävä työ, joka nojaa Hintikan laajaan aiempaan tutkimukseen (Copeland 2002a, 123–126) ja jossa esitetyt episteemisen logiikan tutkimukset ovat vaikuttaneet tekoälyn logiikkaan (Thomason 2018).

Perusidea episteemisten käsitteiden analyysissä on, että nämä ovat modaalikäsitteitä eli niiden analyysi edellyttää useamman mahdollisen maailman loogisesti samanaikaisen tarkastelun. Siten analysoitaessa lausetta "A tietää, että p" jouddumme tarkastelemaan useita mahdollisia maailmoja. "A tietää, että p" tarkoittaa sitä, että p on tosi kaikissa niissä maailmoissa, jotka ovat yhteensopivia kaiken sen kanssa, mitä A tietää. Näitä voidaan kutsua A:n tietomaailmoiksi. Tietomaailmojen välille on määritetty saavutettavuusrelaatio, joka yhdistää kaikki tietomaailmat. Tiedon käsitteen keskeinen ominaisuus on, että tieto on totta. Siten edellä mainitun relaation tulee olla refleksiivinen. (Copeland 2002a.)

Hintikka oli yksi keskeisistä modaalilogiikan kehittäjistä 1950-luvun lopulla ja 1960-luvun alussa. Hän kehitti mahdollisten maailmojen semantiikkana tunnettua modaalikäsitteiden semanttista teoriaa. Tässä keskeisiä ajatuksia ovat modaliteettien ymmärtäminen kvantifiointina yli eri vaihtoehtojen (mahdollisten maailmojen) ja eri mahdollisten maailmojen välillä vallitseva saavutettavuusrelaatio. Näiden avulla oli mahdollista antaa loogisesti täsmällinen luonnehdinta eri modaalilogiikan systeemeille, mikä mahdollistaa yksityiskohtaisemman modaalilogiikkojen loogisen tarkastelun. (ibid., 123–126)

Hintikka (1962) tekee yksityiskohtaisen episteemisten käsitteiden loogis-käsitteellisen analyysin, joka antaa näille käsitteille intuitiivisesti selkeän ja loogisesti täsmällisen luonnehdinnan. Työssä semanttisesti keskeiset käsitteet ovat *mallijoukon* ja *mallisysteemin* käsitteet. Mallijoukko on mahdollisen maailman osittainen kuvaus, joka tehdään kielellisesti. Mallijoukon ja mallisysteemin käsitteet ovat osoittautuneet loogisesti ja filosofisesti hedelmällisiksi. Esimerkiksi interrogatiivimallin kannalta keskeinen (malliteoreettinen) pakottaminen tulee luontevasti liitettyksi tähän kontekstiin. Myöhemmin Hintikka käyttää pääsääntöisesti tavanomaista malliteoreettista käsitteistöä.

Episteeminen logiikka ei ole vain episteemisten käsitteiden keskeisten ominaisuuksien analyysia, vaan samalla myös tiedon hankinnan analyysia. Tiedon hankinta on etsintää ja perustelua, dynaamista toimintaa, jossa tiedon hankkija on aktiivinen toimija (Hendricks 2007, 102–103). Jo kirjassa *Knowledge and Belief* on havaittavissa tiettyjä dynaamisuuden piirteitä, jotka tulevat esiin esimerkiksi mallijoukkojen konstruktioissa. Hintikan filosofiaa on luonnehtinut dynaamisuuden painottaminen oikeastaan kaiken aikaa. Jälkikäteen tulkittuna jo hänen väitöskirjassaan (Hintikka 1953) tulee dynaamisuus mukaan. Erityisesti peliteoreettinen semantiikka toi dynaamisuuden keskiöön (kts. esim. Hintikka ja Sandu 1997).

Kuitenkin episteemisessä logiikassa dynaamisuus otetaan systemaattisesti huomioon vasta myöhemmin niin sanotun toisen sukupolven episteemisessä logiikassa (Hintikka 2003; Hendricks, Jørgensen ja Pedersen 2003; Hendricks 2007). Dynaaminen episteeminen logiikka on interrogatiivimallin keskeinen logiikka. Kuitenkaan ei ole selvää, miten dynaamisuus rakennetaan episteemiseen logiikkaan. Peliteoreettinen semantiikka tuo dynaamisuuden pelien muodossa. Tämä ei kuitenkaan tuo ideaa esille riittävän yleisessä muodossa. Filosofian historia tarjoaa luontevan ratkaisun sokraattisten kyselypelien muodossa: Hintikalle (1986, 66) kysymysten logiikka on keskeisin episteemisen logiikan sovellutus. Kysymysten logiikka tarkastelee, miten kysyjä hankkii tavoittelemaansa tietoa. Siten kysymysten logiikka on todellista dynaamista episteemistä logiikkaa.

Älykkyyden kriteeristöä

Tarkastellaan lyhyesti joitakin kannaltamme keskeisiä piirteitä kysymysten logiikassa. Henkilön kysyessä jonkin yksinkertaisen kysymyksen, esimerkiksi "kuka asuu tässä talossa?", hän olettaa, että talossa asuu joku. Tätä kutsutaan kysymyksen pre-suppositioksi. Ollakseen järkevää kysyä, on tiedettävä, että pre-suppositio on tosi eli että kysymykselle on olemassa jokin, todenmukainen vastaus. Kysymys määrittää, minkä tyyppiset vastaukset voivat olla vastauksia. Talossa asukkaina voi olla joku henkilö, joten vastauksena esimerkikikysymykseen voidaan antaa henkilöiden nimiä. Jos kysyjä saa vastaukseksi nimen "d", niin miten voimme ratkaista, onko vastaus riittävä? Vastaus on riittävä (konklusiivinen), jos se antaa kysyjälle hänen tarvitsemansa tiedon eli jos hän vastauksen perusteella tietää, kuka talossa asuu, mikä on kannaltamme erittäin olennainen episteeminen ehto. Olennaista ei ole vain vastaus ja sen tuoma lisäinformaatio, vaan laajemmin kysymys-vastaus-suhde tai oikeastaan kysymys-vastaus-prosessi. Näin kysymysten logiikka kytkeytyy episteemiseen logiikkaan tuoden siihen dynaamisen painotuksen.

Tarkastellaan seuraavassa yksityiskohtaisemmin, miten kysymys-vastaus-prosessi tuottaa kysyjälle hänen toivomansa tiedon. Vastauksen hänen kysymykseensä "kuka asuu tässä talossa?" tulisi antaa kysyjälle tietoa siitä kuka talossa asuu. Kysyjän tulisi voida todenmukaisesti sanoa, että "minä (M) tiedän, kuka talossa asuu", jonka looginen muoto on

$$(1) \quad (\exists x)K_M A(x).$$

Tämä ilmaisee juuri sen, mitä intuitiivisesti ajattelemme: kysymyksen edellytyksenä on, että tiedän jonkun todellisen henkilön asuvan talossa. Yleensä ajattelemme, että vastaukseksi riittää henkilön nimi. Olkoon vastauksena henkilön nimi "Ilpo". Oletetaan lisäksi, että vastaus on tosi. Nyt vastaus antaa kysyjälle tiedon, että Ilpo asuu talossa. Intuitiivisesti tämä vaikuttaa luonteelta. Formaalisti tämä merkitään seuraavasti:

$$(2) \quad K_M A(\text{Ilpo}).$$

Tämä ei kuitenkaan välttämättä ole se, miten intuitiivisesti ajatteleme. Jos kysyjä ei tiedä, kuka Ilpo on, hän ei ole saavuttanut tavoittelevansa tietoa eli (2):sta ei seuraa (1). Vaikka olemme havainneet tämän loogisten kaavojen avulla, niin tämä asia on todellinen myös arkiajattelussamme. On selvää, että jos kysyjä ei tiedä, kuka Ilpo on (hän ei tunne Ilpoa), niin vastaus "Ilpo" ei juurikaan auta kysyjää.

On monia käytännön tilanteita, joissa emme etsi vastaukseksi sitä henkilöä, johon nimi viittaa. Esimerkiksi kysymykseen "kuka vastaa Suomen ulkopoliitikasta?" vastaukseksi voi aivan hyvin riittää "tasavallan presidentti", vaikka kysyjä ei tietäisikään, kuka presidentti on. (Hintikka 2007, 78.)

Edellä sanoimme, että jollei kysyjä tunne Ilpoa, niin vastaus "Ilpo" ei anna riittävää informaatiota kysyjälle. Kysyjän tulee joko etukäteen tietää, kuka Ilpo on tai sitten kyselyprosessin tulee antaa riittävä määrä informaatiota, jotta hän kyselyprosessin lopussa tuntee Ilpon. Tätä tietoa sanomme kysymyksen *konklusiivisuusehdoksi*, joka voidaan merkitä seuraavasti:

$$(3) \quad (\exists x)K_M(\text{Ilpo}=x).$$

Nyt (3):sta ja (2):sta yhdessä seuraa (1). Siten vastaus Ilpo on konklusiivinen, jos sen lisäksi, että hän tietää Ilpon asuvan talossa, hän tietää myös kuka Ilpo on. Kuitenkaan kysyjän ei tarvitse kysymystä esittäessään vielä tietää, kuka Ilpo on. Kysymys-vastaus-prosessin tulee tuottaa kysyjälle tarvittava tieto. Näin kysymys-vastaus-prosessi tuottaa tarvittavan tiedon.

Ehdossa (3) ei ole kyse vain pienestä asiasta, vaan erittäin olennaisesta seikasta. Kysyjä haluaa paikantaa Ilpon hänen episteemisessä kentässään, ei vain formaalein seikoin, vaan paikantaa hänet sosiaalisesti ja moraalisesti, kuten Hintikka (1975, 148) asian muotoilee. Tämä mahdollistaa analyysin sille, miten hyvä annettu vastaus on eli kuinka lähelle konklusiivista vastausta annettu vastaus "Ilpo" tulee ja toisaalta, mitä kysyjä voi tehdä täydentääkseen annetun vastauksen konklusiiviseksi (ibid., 149), johon edellä viittasimme puhuessamme IF-logiikasta.

Tarkastellaan lähemmin, miten voimme arvioida vastauksen riittävyyttä. Voimme olettaa, että vastaus "Ilpo" on oikea eli $A(\text{Ilpo})$ on tosi. Kuten ehto (3) osoittaa, niin todenmukainen vastaus ei kuitenkaan välttämättä ole riittävä, ellei kysyjä pysty episteemisesti paikantamaan annettua vastausta. Hintikka (1975, 149–153) analysoi tilannetta kytkemällä annetun vastauksen kysyjän taustatietoon. Olkoon T kysyjän tiedot kysymyksen esittämishetkellä. Tällöin vastaus "Ilpo" lisää henkilön tietoon lauseen $A(\text{Ilpo})$ eli hänen tietonsa on vastauksen jälkeen $T \& A(\text{Ilpo})$. Intuitiivisesti kysyjä tietää nyt kaiken aiemmin lisäksi, että Ilpo asuu talossa. Episteemisessä logiikassa tietoa tarkastellaan suhteessa tiedon agentin tietomaailmoihin (Cope-land 2002a; Hintikka 1962). Siten tämä muotoillaan episteemisessä logiikassa siten, että kaikissa kysyjän tietomaailmassa pätee " $A(\text{Ilpo})$ ", mikä vastaus supistaa henkilön episteemisiä vaihtoehtoja. Tämä tarkoittaa, että tarkastelemalla tietomaailmoja, voimme arvioida ja analysoida hänen tietojensa tarkkuutta ja täsmällisyyttä. Tavoitteena kysyjällä on saavuttaa todenmukaista tietoa. On selvää, että jotkut vastaukset ovat tässä suhteessa parempia kuin toiset, joten tarvitsemme luonnehdinnan, mitä todenmukaisuuden aste tarkoittaa, jolle kuitenkin on erittäin vaikeaa antaa intuitiivisesti tyydyttävää luonnehdintaa. Tässä asiassa episteeminen logiikka osoittaa intuitiivisen voimansa. Vastauksen todenmukaisuudenastetta luonnehtiaksemme tarvitsemme joitakin loogisia määritelmiä. Olemme havainneet, että kysyjä ei välttämättä tiedä, kuka Ilpo on. Termillä "Ilpo" on jokin referentti kussakin mahdollisessa maailmassa: olkoon $|\text{Ilpo}|_w$ termin "Ilpo" referentti mahdollisessa maailmassa w . Tämä voidaan määrittää missä tahansa mahdollisessa maailmassa, joten myös aktuaalisessa maailmassa w_0 . Etsitty mitta, joka määrittää vastauksen objektiivisen menestyksellisuuden, voidaan määrittää todennäköisyysmitan P avulla seuraavasti:

$$(4) \quad P(\text{Ilpo} = |\text{Ilpo}|_{w_0} \mid (A(\text{Ilpo}) \& T)).$$

Mitan objektiivisuus tulee siitä, että mitta on määritetty suhteessa aktuaaliseen maailmaan. Mitta siis kertoo, kuinka todennäköisesti kysyjän tieto on todennukaista. Mitän luontevuus on selvää. Etsiessämme tietoa tavoittelemme nimenomaan todennukaista tietoa, jonka saavuttamista mitta esittää.

Tietomme on aina puutteellista, emmekä yleisesti kykene tunnistamaan mikä tiedoistamme ei ole (tai mikä on) todennukaista. Siten edellä oleva mitta objektiivisuudessaan ei huomioi, että emme kykene erottelemaan maailmoja, jotka kuuluvat tietomaailmoihimme. Sillä ne kaikki ovat yhteensopivia kaiken sen kanssa, minkä tiedämme. Siten on luontevaa etsiä määritelmä, joka on "käytännöllisempi", jolloin määrittelyyn tulee väistämättä tiettyä subjektiivisuutta. Tällöin määritetään edellä oleva mitta suhteessa johonkin luokkaan mahdollisia maailmoja. Siten edellä viitattu subjektiivisuus tarkoittaa esimerkiksi, että mahdollisten maailmojen luokka on kysyjän tietomaailmat. Siten saamme seuraavan määritelmän:

$$(5) \quad \max_{i \in I} P(I|p_0 = |I|p_0 |_{w_i} | (A(I|p_0) \& T)),$$

missä mahdolliset maailmat w_i ($i \in I$) ovat kaikki henkilön episteemisesti mahdolliset maailmat. Tämä antaa mitan sille, miten hyvin vastauksen antama tieto sopii yhteen henkilön muun tiedon kanssa yhteen, jolloin tämä mittaa vastauksen subjektiivista konklusiivisuutta. (Hintikka 1975, 149-151.)

Analyysi ei kerro tarkemmin, miten kysyjä voi täydentää vastauksen konklusiiviseksi, sillä ei ole olemassa jotain tiettyä keinoa saavuttaa vastauksen konklusiivisuutta eli saada ehto $I|p_0 = |I|p_0 |_{w_0}$ toteutetuksi. Oikeastaan ainoa tapa on kyselyprosessin strateginen kehittäminen, joka on keskeinen asia Hintikan interrogatiivimallissa.

Strategian käsite on olennaisen tärkeä ja edellä oleva kaava (5) tuo hyvin esiin, miksi strategia jää helposti hyvin abstraktiksi tai konkreettiseksi "pelin sisäiseksi" strategiaksi. Ei ole mitään yleistä menetelmää, jota soveltamalla olisi mahdollista löytää paras tai edes välttämättä toimiva strategia. Yksilön tietomaailmat ovat keskenään "samanarvoisia", joten kaavassa (5) esiintyviä vaihtoehtoja ei ole mahdollista erotella. Meillä on

kuitenkin monia erilaisissa tilanteissa toimivia strategioita, kuten esimerkiksi eksperimentaalinen menetelmä monien tietuongelmien ratkaisuun (Hintikka 1987b; 1988a). Monia esimerkkejä toimivista etsintä strategioista löytyy kirjallisuudesta (kts. esim. Osherson, Stob ja Weinstein 1986; Hendricks 2001; Martin ja Osherson 1998).

Tiedonetsintästrategian tärkeys on ilmeinen. Lisäksi se, ettei meillä ole yleistä keinoa löytää menestyksekkästä strategiaa, ei vähennä strategian etsinnän merkitystä. On olennaisen tärkeää pyrkiä jäsentämään strategiaa niin yleiseltä kannalta kuin täsmällisestä, loogis-käsitteelliseltä kannalta. Hintikan filosofia on tästä erinomainen esimerkki. Hän tarkastelee strategiaa monista erilaisista näkökulmista (ks. esim. Hintikka 1986; 1987a; 1989), mikä tekee interrogatiivimallista filosofisesti niin kiinnostavan ja tärkeän. Meillä ei ole yleistä keinoa löytää parasta strategiaa tiedonetsintä peleihin. Hintikka (1969, 282) tuo tämän esiin puhuessaan geometrysten konstruktioiden löytämisen strategian mahdottomuudesta: "Koska edellisten löytäminen on yleisessä muodossaan aivan kirjaimellisesti laskematon toimitus, on lähellä ajatus, että yleisen tieteellisen metodin tasolla asioista keskusteltaessa myös ilmiöiden eksperimentaalinen erittelemine on asia, jolle ei voida antaa täsmällisiä yleisiä sääntöjä." Se, ettei ole keinoa löytää menestysstrategiaa, on kannaltamme tärkeää siinä mielessä, että tämä tekee tiedon etsinnän malliesimerkilliseksi järkevyyden ilmentymäksi (Hintikka, Halonen ja Mutanen 2002). Nämä huomiot tulevat esiin, kun analysoimme Turingin ajattelua.

Siten on selvää, miksi Hintikka (1986, 77) sanoo interrogatiivimallia toiseksi keskeiseksi episteemisen logiikan sovellutukseksi. Interrogatiivimallissa kyse on yksittäisen kysymyksen analyysin sijaan tarkastella strategisia kysymys-vastaus-prosesseja. Edellä huomasimme, että kysymyksen konklusiivisuusehto tekee jo yksittäisestä kysymyksestä prosessin, jossa kysyjä etsii riittävää määrää tietoa (informaatiota), jotta hän saavuttaa episteemisen tavoitteensa. Interrogatiivimallissa

tämä näkemys yleistetään tieteelliseen tiedonhankintaan ja lopulta kaikkeen rationaaliseen järkeilyyn (Hintikka, Halonen & Mutanen 2002).

Kuten aikaisemmin todettiin, interrogatiivimallin keskeisenä filosofisena ja loogisena historiallisena esikuvana on sokraattinen kyselymenetelmä, jota Sokrates harjoitti käytännössä Ateenassa. Platon kirjasi menetelmän omista dialogeistaan strategisina tiedonetsintä prosesseina, jotka eivät aina päättyneet toivottuun lopputulokseen, vaan kysymys jää vaille vastausta. Kuitenkin, esimerkiksi *Menonissa*, on kuvattuna menestykselias kyselyprosessi. Aristoteles sen sijaan muotoilee systemaattista teoriaa tällaisille kysymys-vastaus-prosesseille. Tästä näkökulmasta hänen syllogistiikkansa on osa laajempaa kysymys-vastaus-prosessien teoriaa, jossa keskitytään kysymyksiin, joihin vastaamiseen on jo riittävästi tietoa. Interrogatiivimallin yksi keskeisimmistä käsitteistä on strategian käsite: strategia kuvaa sekä koko kyselyprosessia että vastauksen etsimisprosessia. (Hintikka 2007; Hintikka, Halonen ja Mutanen 2002.)

Interrogatiivimallilla on lukuisia sovellutusalueita. Lähtökohtaisesti malli kuvaa tiedonhankintaprosessien logiikkaa. Luonteva sovellutus on tieteellisen tutkimuksen logiikka (metodologia). Tällä sektorilla malli on osoittautunut toimivaksi. Sen avulla on voinut antaa yleisen metodologisen luonnehdinnan tieteelliselle tutkimukselle (Hintikka 2007). Toisaalta mallin avulla on ollut mahdollista analysoida monia metodologian ja tieteenfilosofian erityiskysymyksiä, kuten tieteellisen päättelyn teoriaa (Hintikka 1988a), oppimisen ja opettamisen teoriaa (Hintikka 1982), induktion logiikkaa (Hintikka 1992), todennäköisyyspäättelyn teoriaa (Hintikka 1987b), eksperimentaalisen tutkimuksen logiikkaa (Hintikka 1988a), yleistä rationaalista päättelyä (Hintikka, Halonen ja Mutanen 2002), mutta myös metodologian erityiskysymyksiä, kuten identifioituvuuden käsitteen rooli metodologiassa (Hintikka 1991) tai tieteellinen selittäminen (Halonen ja Hintikka 2005).

Tekoälytutkimuksen perusteiden kannalta edellä olevat kaavat (4) ja (5) ovat olennaisia. Ne osoittavat, että on mahdollista

luonnehtia täsmällisesti etsintäprosessin onnistuneisuutta. Kaavan (4) nojalla voimme luonnehtia prosessin objektiivista onnistuneisuutta, jolloin arvioinnissa katsotaan tuottaako etsintäprosessi oikean tuloksen. Kaavan (5) avulla voimme luonnehtia subjektiivista onnistuneisuutta, millä tarkoitetaan mitan relativisoimista agentin tietomaailmoihin. Intuitiivisesti voidaan sanoa, että kaavan (4) avulla voidaan luonnehtia etsintäprosessin objektiivista luovuutta ja kaavan (5) avulla luovuutta käyttää agentin omaa tietoa. (Hintikka 1975, 153–156; Hintikka, Halonen ja Mutanen 2002.)

Turing ja tiedonetsintäpelit

On selvää, että interrogatiivimallin pohjalta on mahdollista kehittää ”operationaalinen älykkyystesti”, jossa älykkyys tulee esiin kyvyssä etsiä tietoa. Kuten olemme havainneet, tiedon etsintä edellyttää aitoa metodista (strategista) kykyä, joten tiedonetsinnässä ei ole mahdollista menestyä Copelandin (1993, 41) kuvaamalla Parryn ja Elizan käyttämällä triviaaleilla strategioilla (*pattern matching*). Myöskään tiedonetsintää ei saada menestykselliseksi trivialisoimalla se ironian avulla, kuten Eugene Goostman teki. Lisäksi myöskään etsintätesti ei edellyttäisi testattavan olevan jouheva keskustelija, joten Vardin (2014) kuvaama ongelma ei esiinny. Turing (1936) analysoi inhimillistä laskentaa tavalla, joka merkittävässä määrin ennakoii nimenomaan tiedonetsintätestin kaltaista mallia. Voikin ihmetellä, miksi Turing ei vienyt testiään tähän suuntaan. Turing oli jo 1940-luvun alussa kehittämässä tekoälyä. Erityisesti tällöin olivat suuren kiinnostuksen kohteena koneet, jotka osasivat pelata pelejä, erityisesti shakkia. Pelien ohjelmien tuli olla ainakin jossain määrin strategisia. Turingin universaalikoneen idea mahdollisti niin ohjelmoinnin kuin ainakin jonkinlaisen strategian asettamisen koneille. Turingin keskustelu oppivista koneista on kaiken aikaa strategista. Pelaavien koneiden kohdalla oli kaiken aikaa selvää, että koneen tuli oppia strateginen tapa pelata, mikä tulee hyvin esiin esimerkiksi artikkelista Michie (1961). Jo 1940- ja 1950-luvuilla pohdittiin laajalti,

voiko kone kehittyä shakissa ihmisen kaltaiseksi pelaajaksi tai jopa mestaritason pelaajaksi (Copeland 2004, 352, 562–563). Tällöin ongelmana oli koneen strateginen luovuus: oliko sitä, vai olivatko koneet vain ”mekaanisia” pelaajia (Turing 1950, 455). Kuten edellä havaitsimme Turing (1948) muotoilee testin, jolla tarkkailija pyrkii tunnistamaan, onko pelaaja ihminen vai kone, joka antaa luontevalla tavalla vastauksen tällaiseen kysymykseen. Selvästikin Turingin testi, yleisenä keskustelutestinä, on tämän kiinnostava yleistys (Turing 1950). Ei ole kuitenkaan helppoa tulkita testiä. Oikeastaan testi pyrkii vastaamaan kahden eri kysymyksen: yhtäältä toiminnan älykkyyden arviointi ja tunnistaminen ja toisaalta älykkään toiminnan analyysi ja älykkään toiminnan konstruointi (synteesi). Kuitenkin testi on osoittautunut erittäin kiinnostavaksi ja hedelmälliseksi.

Vaikka oppivat koneet olivat strategisia, niin ei oikeastaan ollut sopivaa strategian käsitettä, jonka varaan olisi ollut mahdollista kehittää tietoa etsiviä koneita. Loogisen (matemaattisen) ja empiirisen tutkimuksen paralleelisuus ei ollut selkeästi ymmärretty asia. Tämän paralleelisuuden syvä merkitys tulee esiin *strategiateoreemasta* (Hintikka, Halonen ja Mutanen 2002, 306, 330). Strategiateoreema tuo esiin, että vaikka ei ole olemassa ratkaisumenetelmää logiikassa, niin voimme hyvin arvioida matemaattisen päättelyn luovuutta. Strategiateoreema sanoo, että tämä sama voidaan tehdä myös empiirisen tiedon etsinnän kohdalla. Turing (1936) on tässä suhteessa erittäin kiehtova. Siinä Turing ihmislaskentaa analysoimalla eksplikoi koneen laskentastrategian, joka on ”ekvivalentti” ihmislaskijan strategian kanssa. Hän yleistää tämän koskemaan oraakkelikoneita vuoden 1938 artikkelissaan. Oraakkelikoneet voivat saada laskentaprosessin aikana lisäinformaatiota oraakkelista, jota kone käyttää laskentaprosessissa hyväkseen. Turing (1938) muotoilema looginen rajajärjestelmä (*limit system*) on kiinnostava yleistys loogisesta päättelystä rajapäättelynä, jonka laskenta-voima on Π^0_1 . Tätä rajapäättelyn (*limiting lemma*) ideaa on sovellettu formaalissa oppimisen teoriassa (Gold 1965; 1967; Putnam 1965) ja siihen nojautuvassa tieteenfilosofiassa (Kelly 1996;

Hendricks 2001). Clote (1986) yleistää rajateoreeman kompleksisuusteoriaa koskevaksi loogiseksi tulokseksi. Formaalin oppimisen teorialla ja siten myös Turingin (1938) ideoilla on kiinnostavia yhteyksiä interrogatiivimalliin. Erityisesti formaalin oppimisen teoriassa on tarkasteltu erilaisia mahdollisuuksia muotoilla tiedon etsintästrategioita.

Siten Turingilla oli jo 1930-luvun töissä tarvittavat elementit, joiden varaan hän olisi voinut muotoilla testinsä tiedon etsintätestinä. Kuitenkin ne ongelmat, joiden parissa hän työskenteli peliä kehittäessään, estivät häntä näkemästä tätä mahdollisuutta. Tämä ilmenee esimerkiksi katsottaessa niitä vastaväitteitä, joita Turing (1950) tarkastelee. Hän viittaa Gödelin epätäydellisyystulokseen ja omaan vuoden 1936 artikkeliinsa, mutta ei tuo esiin oman työnsä strategista puolta. Turingin 1940- ja 1950-luvun töissä on runsaasti viitteitä, että hän etsi tekoälyn muotoa, jota voitaisiin luonnehtia simulaatioksi₂.

Loppusanat

Olemme edellä painottaneet strategisen ajattelun keskeisyyttä. Strateginen ajattelu ei kuitenkaan voi olla siinä määrin yksinkertaista, että sitä voitaisiin simuloida mekaanisin keinoin. Tällaisenkkin strategian oppimiseen liittyy luovuutta ja oppimista, kuten esimerkiksi Michien (1961) esittämät koneiden pelistrategioiden oppimistulokset osoittavat. Strategian, osoittaakseen älykkyyttä, tulee olla riittävän kompleksista. Tosiasiassa sen ei voi sallia olevan mekanisoitavissa. Tässä merkityksessä interrogatiivimalliin liittyvä strateginen ajattelu osoittaa aitoa älykkyyttä

Turing etsi älykkyyttä ja luovuutta. Hänen menetelmänsä etsinnässä olivat kiinnostavat ja filosofisesti rikkaat. Kuitenkin testaaminen edellyttää, että testi on selkeästi mitattavissa, mikä rajoittaa testin suunnittelua. Turingin ajatuksen kehittäminen suuntautui lopulta testin sellaiseen muotoiluun, joka lopulta oli mekanisoitavissa ja siten testi ei kerro aivan niin paljon, kuin Turing siltä ilmeisestikin odotti. Kuitenkin hänen kehittelynsä on täynnä kiinnostavia luonnehdintoja, joita on tärkeää tutkia ja

kehittelä edelleen. Yksi suunta, jonka olemme tässä artikkelissa havainneet, on se, että Turingin testiä olisi kiinnostavaa lähteä kehittämään nimenomaan interrogatiivimallin pohjalta tiedon etsinnän osoittaessa älykkyyttä. Tällä olisi suora yhteys Turingin kehittämiin ideoihin ja hänen pitkään kehittämäänsä metodiseen orientaatioon. Testi olisi Turingin hengessä nimenomaan metodinen testi, jonka periaatteessa voisi hyväksytysti suorittaa niin ihmiset kuin koneetkin. Kuitenkin testin läpäisy edellyttäisi älykkääksi ymmärtämäämme toimintaa. Testi on periaatteessa sellainen, ettei tiedon etsinnän strategiaa ole mahdollista mekanisoida, mikä on testin kannalta hyvä ominaisuus.⁵

*Merisotakoulu & Maanpuolustuskorkeakoulu
Helsingin yliopisto*

Kirjallisuus

- Aamoth, Doug (2014) "The Turing Test", *Time* 183(24), 18.
- Bringsjord, Selmer ja Zenzen, Michael (2003) *Superminds: People Harness Hypercomputation, and More*. Dordrecht: Kluwer Academic Publishers.
- Copeland, Jack (1993/2006) *Artificial Intelligence: A Philosophical Introduction*. Oxford: Blackwell.
- Copeland, Jack B. (2002a) "The Genesis of Possible Worlds Semantics", *Journal of Philosophical Logic* 31, 99–137.
- Copeland, Jack B. (2002b), "Hypercomputation", *Minds and Machines* 12, 461–502.
- Copeland, Jack B. (toim.) (2004) *The Essential Turing: The Ideas That Gave Birth to the Computer Age*. Oxford: Clarendon Press.
- Copeland, Jack B. (2012) *Turing: Pioneer of the Information Age*. Oxford: Oxford University Press.
- Epstein, Robert, Roberts, Gary ja Beber, Grace (toim.) (2008) *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Dordrecht: Springer.

⁵ Kiitämme *Ajatuksen* nimettömiä arvioijia hyödyllisistä kommentista, jotka auttoivat meitä parantamaan artikkelimme aikaisempia versioita.

- French, Robert M. (2012) "Dusting Off the Turing Test", *Science* 336, 164–165.
- Godwin, Dwayne ja Cwam, Jorge (2014) "Turing's Test", *Scientific American MIND* 25 (6), 76.
- Gold, E. Mark (1965) "Limiting Recursion", *Journal of Symbolic Logic* 30 (1), 28–48.
- Gold, E. Mark (1967) "Language Identification in the Limit", *Information and Control* 10, 447–474.
- Haikonen, Pentti O. A. (2017) *Tietoisuus, tekoäly ja robotit*. Helsinki: Art House.
- Halonen, Ilpo ja Hintikka, Jaakko (2005) "Toward a Theory of the Process of Explanation", *Synthese* 143 (1/2), 5–61.
- Hendricks, Vincent Fella (2001) *The Convergence of Scientific Knowledge - a View from the Limit*. Dordrecht: Kluwer Academic Publishers.
- Hendricks, Vincent Fella (2007) *Mainstream and Formal Epistemology*. New York: Cambridge University Press.
- Hendricks, Vincent Fella, Jørgensen, K. F. ja Pedersen, Stig Andur (toim.) (2003), *Knowledge Contributors*. Dordrecht: Kluwer Academic Publishers.
- Hintikka, Jaakko (1962) *Knowledge and Belief. An Introduction to the Logic of the Two Notions*. Ithaca: Cornell UP.
- Hintikka, Jaakko (1968) "Language-Games for Quantifiers" teoksessa Nicholas Rescher (toim.), *Studies in Logical Theory*. Oxford: Blackwell Publishers, 46–73.
- Hintikka, Jaakko (1969) *Tieto on valtaa ja muita aatehistoriallisia esseitä*. Porvoo: WSOY.
- Hintikka, Jaakko (1973) *Logic, language-games and information: Kantian themes in the philosophy of logic*, Oxford: Oxford UP.
- Hintikka, Jaakko (1975) "Answers to Questions" teoksessa *The intentions of intentionality and other new models for modalities*. Dordrecht: D. Reidel, 137–158.
- Hintikka, Jaakko (1976) *The Semantics of Questions and the Questions of Semantics*. Acta Philosophica Fennica 28 (4). Helsinki: Philosophical Society of Finland.
- Hintikka, Jaakko (1979a) "Language-Games" teoksessa E. Saarinen (toim.), *Game-Theoretical Semantics*. Dordrecht: D. Reidel, 1–26.
- Hintikka, Jaakko (1979b) "Quantifiers in Logic and Quantifiers in Natural Languages" teoksessa E. Saarinen (toim.), *Game-Theoretical Semantics*. Dordrecht: D. Reidel, 27–47.

- Hintikka, Jaakko (1984) "The Logic of Science as Model-Oriented Logic" teoksessa P. Asquith ja P. Kitcher (toim.), *PSA 1984, 1*. East Lansing: Philosophy of Science Association, 177-185.
- Hintikka, Jaakko (1985) "A Spectrum of Logics of Questioning", *Philosophica* 35, 135-150.
- Hintikka, Jaakko (1986) "Reasoning About Knowledge in Philosophy: The Paradigm of Epistemic Logic" teoksessa J. Y. Halpern (toim.), *Theoretical Aspects of Reasoning About Knowledge*. Los Altos: Morgan Kaufmann Publishes, 63-80. (Myös teoksessa teoksessa J. Hintikka ja M. B. Hintikka 1989, 17-35.)
- Hintikka, Jaakko (1987a) "Language Understanding and Strategic Meaning", *Synthese* 73(3), 497-529.
- Hintikka, Jaakko (1987b) "The interrogative approach to inquiry and probabilistic inference", *Erkenntnis* 26 (3), 429-442.
- Hintikka, Jaakko (1988a) "What Is the Logic of Experimental Inquiry?", *Synthese* 74(2), 173-188.
- Hintikka, Jaakko (1988b) "On the Development of the Model-Theoretic Viewpoint in Logical Theory", *Synthese* 77(1), 1-36.
- Hintikka, Jaakko (1989) "Questioning as a Philosophical Method" teoksessa J. Hintikka ja M. B. Hintikka 1989, 215-233.
- Hintikka, Jaakko (1990) "The Languages of Human Thought and the Languages of Artificial Intelligence" teoksessa L. Haaparanta, M. Kusch ja I. Niiniluoto (toim.), *Language, Knowledge, and Intentionality: Perspectives on the Philosophy of Jaakko Hintikka*. Acta Philosophica Fennica 49, 307-330.
- Hintikka, Jaakko (1991) "Towards a General Theory of Identifiability" teoksessa J. H. Fetzer et al. (toim.) *Definitions and Definability: Philosophical Perspectives*. Dordrecht: Kluwer Academic Publishers, 161-183.
- Hintikka, Jaakko (1992) "The Concept of Induction in the Light of the Interrogative Approach to Inquiry", teoksessa John Earman (toim.), *Inference, Explanation and Other Frustrations: Essays in the Philosophy of Science*. Berkeley: University of California Press, 23-43.
- Hintikka, Jaakko (1996) *The Principles of Mathematics Revisited*. Cambridge: Cambridge University Press.
- Hintikka, Jaakko (2001) *Filosofian köyhyys ja rikkaus. Nykyfilosofian karitustusta*. Helsinki: Art House.

- Hintikka, Jaakko (2003) "A Second Generation Epistemic Logic and its General Significance", teoksessa V. F. Hendricks et al. (toim.), 2003, 33–56.
- Hintikka, Jaakko (2007) *Socratic Epistemology*. New York: Cambridge University Press.
- Hintikka, Jaakko ja Bachman, James (1991) *What If ...?: Toward Excellence in Reasoning*. Mountain View, CA: Mayfield Publishing Company.
- Hintikka, Jaakko ja Hintikka, Merrill B. (1989) *The Logic of Epistemology and the Epistemology of Logic*. Dordrecht: Kluwer Academic Publishers.
- Hintikka, Jaakko ja Kulas, Jack (1983) *The Game of Language: Studies in Game-Theoretical Semantics and Its Applications*. Dordrecht: D. Reidel.
- Hintikka, Jaakko ja Mutanen, Arto (1998) "An Alternative Concept of Computability" teoksessa J. Hintikka, *Language, Truth and Logic in Mathematics*. Dordrecht: Kluwer Academic, 174–188.
- Hintikka, Jaakko ja Sandu, Gabriel (1996) "A Revolution in Logic?", *Nordic Journal of Philosophical Logic* 1 (2), 169–183. (Myös suomeksi teoksessa J. Hintikka (2001).)
- Hintikka, Jaakko ja Sandu, Gabriel (1997) "Game-Theoretical Semantics" teoksessa J. Van Benthem ja A. Ter Meulen (toim.), *Handbook of Logic and Language*. Amsterdam: Elsevier Science, 361–410.
- Hintikka, Jaakko, Halonen, Ilpo ja Mutanen, Arto (2002) "Interrogative Logic as a General Theory of Reasoning" teoksessa R. H. Johnson ja J. Woods (toim.), *Handbook of Practical Reasoning*. Dordrecht: Kluwer Academic Publishers, 295–337.
- Hodges, Andrew (2000/1997) *Turing: luonnonfilosofi*. Helsinki: Otava. (Alkuperäinen laitos: *Turing: a Natural Philosopher*, Lontoo: Phoenix 1997.)
- Hodges, Andrew (2008) "Alan Turing and the Turing Test" teoksessa R. Epstein et al. (toim.), *teoksen nimi*. (2008), 13–22. Myös internetissä: <<https://www.turing.org.uk/publications/testbook.html>>.
- Kelly, Kevin (1996) *The Logic of Reliable Inquiry*. Oxford: Oxford University Press.
- Marcer, P. J. ja Fedorec, A. M. (1994) *Alternative Forms of Computation and New Routes to Parallelism*. Lontoo: University of Greenwich.
- Martin, Eric ja Osherson, Daniel (1998) *Elements of Scientific Inquiry*. Cambridge, MA: M.I.T. Press.
- Michie, Donald (1961) "Trial and error", *Science Survey* part 2, Harmondsworth: Penguin, 129–145.

- Michie, Donald (1962) "Puzzle-Learning versus Game-Learning in Studies of Behaviour", teoksessa Good, I. J. (toim.), *The Scientist Speculates*. Lontoo: Heinemann, 90–100.
- Michie, Donald ja Johnston, Rory (1984) *The Creative Computer: Machine Intelligence and Human Knowledge*. Harmondsworth: Viking.
- Moor, James H. (toim.) (2003) *The Turing Test: The Elusive Standard of Artificial Intelligence*. Dordrecht: Kluwer Academic Publishers.
- Niiniluoto, Ilkka (1990) "Äly, tahto, tunne ja tekoäly" teoksessa I. Niiniluoto, *Maailma, minä ja kulttuuri*. Keuruu: Otava, 122–139.
- Niiniluoto, Ilkka (1999) *Critical Scientific Realism*. Oxford: Oxford University Press.
- Osherson, Daniel, Stob, Michael ja Weinstein, Scott (1986) *Systems that Learn*. Cambridge, MA: M.I.T. Press.
- Putnam, Hilary (1965) "Trial and Error Predicates and a Solution to a Problem of Mostowski", *Journal of Symbolic Logic* 30, 49–57.
- Rapaport, William J. (2003) "How to Pass a Turing Test", teoksessa J. H. Moor (toim.) 2003, 467–490.
- Russell, Stuart ja Norvig, Peter (2003) *Artificial Intelligence: A Modern Approach* (Second Edition). Upper Saddle River, New Jersey: Prentice Hall.
- Sandu, Gabriel (2015) *Logic, Language and Games*. Acta Philosophica Fennica 91. Helsinki: Philosophical Society of Finland.
- Sieg, Wilfried (1991) *Mechanical Procedures and Mathematical Experience*, Report No. CMU-PHIL-20, Pittsburgh, Pennsylvania 15213-3890 .
- Sieg, Wilfried (2002) "Calculations by Man and Machine: Conceptual Analysis" teoksessa W. Sieg, R. Sommer ja C. Talcott (toim.), *Reflections on the Foundations of Mathematics*. New York: A K Peters/CRC Press, 390–409.
- Thomason, Richmond (2018) "Logic and Artificial Intelligence", *The Stanford Encyclopedia of Philosophy* (talven 2018 laitos), toim. E. N. Zalta. <<https://plato.stanford.edu/archives/win2018/entries/logic-ai/>>.
- Turing, Alan (1936) "On Computable Numbers, with an Application to the Entscheidungsproblem" teoksessa Copeland 2004, 58–90.
- Turing, Alan (1938) "Systems of Logic Based on Ordinals" teoksessa Copeland 2004, 146–204.
- Turing, Alan (1945), "Proposed electronic calculator" teoksessa Copeland, B. Jack, *Alan Turing's Electronic Brain: The Struggle to Build the ACE, the World's Fastest Computer*, Oxford: Oxford University Press, 2012, 369-454.

- Turing, Alan (1947) "Lecture on the Automatic Computing Engine" teoksessa Copeland 2004, 378–394.
- Turing, Alan (1948) "Intelligent Machinery" teoksessa Copeland 2004, 410–432.
- Turing, Alan (1950) "Computer Machinery and Intelligence" teoksessa Copeland 2004, 441–464.
- Turing, Alan (1951a) "Intelligent machinery, A Heretical Theory" teoksessa Copeland 2004, 472–475.
- Turing, Alan (1951b) "Can Digital Computers Think?" teoksessa Copeland 2004, 482–486.
- Turing, Alan (1953) "Chess" teoksessa Copeland 2004, 569–575.
- Turing, Alan, Braithwaite, Richard, Jefferson, Geoffrey ja Newman, Max (1952) "Can Automatic Calculating Machines Be Said to Think?" teoksessa Copeland 2004, 494–506.
- Vardi, Moshe Y. (2014) "Would Turing Have Passed the Turing Test?", *Communications of the ACM* 57 (9), 5.
- Väänänen, Jouko (2007) *Dependence Logic: A New Approach to Independence Friendly Logic*. Cambridge: Cambridge University Press.