KAUSAALINEN PÄÄTTELY JA BAYES-VERKOT ENNAKOINNIN APUVÄLINEENÄ

Marko Ahvenainen & Nina Janasik

Tiivistelmä

Ihmisellä on kyky kuvitella olemattomia. Ilman tätä kykyä mahdollisten ja vaihtoehtoisten tulevaisuuksien tutkiminen olisi hyvin hankalaa. Näkemykset tulevaisuuksista nojaavat päättelyyn. Päättely tapahtuu aina jossain selittävässä viitekehyksessä, joka sallii näkemyksen olla mahdollinen, todennäköinen, uskottava tai toivottava. Kausaalinen päättely ja siihen perustuva todennäköisyyspohjainen malli (Bayes-verkko) mahdollistavat vaihtoehtoisten tulevaisuuksien perustaksi omaksutun selittävän viitekehyksen näkyväksi tekemisen. Bayes-verkko, tulevaisuutta koskevan päättelyn representaationa, ilmentää päättelijöiden näkemystä mallinnettavan systeemin tulevaisuuteen suuntautuneesta kausaalisesta kokonaiskäyttämisestä. Tässä artikkelissa tarkastelemme sitä, miten kausaalista päättelyä ja Bayes-verkkoa voidaan hyödyntää asiantuntijatiedon strukturoinnin ja ennakoinnin apuvälineenä.

Avainsanat: Bayes-verkko, evidenssi, kausaalinen päättely, kontrafaktuaalisuus, mahdolliset maailmat, mallintaminen

1. ”Miksi Miksi -maailma”

Miksi yötaivas on pimeä, vaikka kaiken järjen mukaan näin ei pitäisi olla lukemattomia tähtiä tuikkivassa universumissa? Mikä selittää havaintoaineiston? Tieteessä ilmiö tunnetaan Olbersin paradoksina, jolle on sen esittämisen vuonna 1826 jälkeen annettu useita selityksiä. Erään selityksen keskeinen oletus perustuu kärjistäen siihen, että maailmankaikkeus ei ole äärettömän vanha, joten kaikki valo ei ole ehtinyt saavuttaa maapallolla olevaa havaitsijaa. Katsomalla siis todella kauas näkee mahdollisen tähden, joka ei ole vielä havaitsijalle syttynyt. Toisin sanoen, koska vuorovaikutuksella on äärellinen maksiminopeus (valonnopeus), katsoessaan kauas havainnoija tähyää menneisyyteen, mutta ei vielä koe mahdollisesti tapahtuneen syyn vaikutusta eli sen aiheuttamaa muutosta itsessään, jonka hän tulkitsee näköhavaintona. Voi myös olla,

että mitään tähteä ei ole tai että tähden liikkuessa poispäin havaitsijasta sen lähettämä valo on siirtynyt pois näkyvän valon aallonpituudelta. Miksi-kysymykseen vastaaminen edellyttää, että kuvailemme tai selitämme havainnot eli datan tuottavan prosessin (Pearl 2018). Nobel-palkitun fyysikko Richard Feynmanin ajatusta lainaten, vastaus kysymykseen ”miksi” on aina sidoksissa asiayhteyteen, joka sallii vastauksen olla totta tai riittävä (riittävän järkeenkäypä tai hyödyllinen kelvatakseen vastaukseksi).

”Mahdolliset maailmat” on tulevaisuuden tutkimuksen keskeisiä käsitteellisiä työkaluja. Käsitteellä viitataan tällöin määritelmänsä mukaan mahdolliseen tulevaisuuden asiantilaan, joka voi periaatteessa toteutua jonkin toimijan toiminnan seurauksena tai tämän toiminnastaan riippumatta (Kuusi et al. 2013). Jos esitämme mahdollista maailmaa koskevan ymmärryksemme väitelauseina kuten *”tuloverojen kasvattaminen lisää todennäköisesti työttömyyttä Mikä-Mikä -maan Ihmetys pääkaupungissa”*, ilmaisemme todennäköisyysarviomme lisäksi luottamuksemme kyseiseen hypoteesiin sekä perustelut eli syyt uskoa väite todeksi. Samaan aikaan voidaan kysyä: *”Onko hypoteesi totta tai onko luottamuksemme sen paikkansa pitävyyteen sama kaikissa mahdollisissa maailmoissa?”*. Jos tuloveron nosto voi tapahtua ilman työttömyyden lisääntymistä, päättelemme, että se ei ole riittävä syy työttömyyden lisääntymiselle. Jos työttömyys voi lisääntyä ilman tuloveron kasvattamista, päättelemme, että se ei ole välttämätön syy työttömyydelle. Jos esimerkiksi poliittinen päätöksentekijä haluaa, jostain syystä, nostaa tuloveroa ilman työttömyyden lisääntymistä, eikä se ole perustavanlaatuisesti kiellettyä, niin silloin kysymykseksi jää, miten se tehdään. Järkevän toimijan kohtuullisena tunnusmerkkinä voidaan pitää, että tämä osaa kertoa syyn toiminnalleen ja toimii tavalla, joka edistää tämän tärkeinä pitämiään tavoitteita.

Ei ole olemassa mitään tunnettua yhtälöä, joka sitoisi tulevaisuuden mahdolliset seuraukset yhteen niiden mahdollisten nykyisten syiden kanssa. Se, miten oikeutamme tulevaisuutta koskevat näkemyksemme (kutsutaan sitä vaikka *Miksi Miksi -maailmaksi*) jostain ilmiöstä (esim. *tuloverojen nosto ja sen vaikutus työttömyyteen*) osana jotain empiiristä kontekstia (esim. *Mikä Mikä -maan Ihmetys-pääkaupungissa*), täytyy pilkkoa palasiin. Kausaalimallintamisessa mahdolliset maailmat pilkotaan syihin ja seurauksiin perustuvan päättelyn avulla palasiin. Bayes-verkkoon[1] (todennäköisyysmallin esitys) perustuva kausaalimalli koostuu muuttujista sekä niiden välisistä riippuvuussuhteista, jotka on ilmaistu ehdollisina todennäköisyyksinä. Koska kyseessä on kausaalimalli, verkko on suunnattu (ajan nuoli) ja syklitön (ei takaisinkytkentää ajan nuolta vastaan). Yksittäisistä muuttujien välisistä vuorovaikutussuhteista nousee (matemaattinen) kuvaus systeemin kausaalisesta kokonaiskäyttämisestä, jonka avulla voidaan simuloida (eli laskea) esimerkiksi, minkä muuttujien suhteen käsityksemme tulevaisuudesta on erityisen herkkä tai mitkä muuttujat muuttavat näkemyksen tulevaisuudesta epävarmaksi.

Tässä artikkelissa esitämme hyvin yleisellä tasolla, miten Bayes-verkkoon pyydystetty asiantuntijoiden yhteinen kausaalinen päättely jostakin valitusta ilmiöstä *(*esim.

[1] Hyviä lähteitä aloittaa tutustuminen Bayes-verkkoihin ovat esimerkiksi Korb & Nicholson (2010) *Bayesian Artificial Intelligence* ja Fenton & Neil (2012) *Risk Assessment and Decision Analysis with Bayesian Networks.*

tuloveron nosto ja sen vaikutukset työttömyyteen) jossain tarkasti rajatussa kontekstissa (esim. *Mikä Mikä -Maan Ihmetys-pääkaupungissa*) voi opastaa meitä mahdollisten maailmojen hahmottamisessa ja tulevaisuuden ennakoinnissa. Ennen kun esittelemme tapaustutkimukseen pohjautuvan BOR-konseptin (sanoista *Bayesian Operation Room*) pääpiirteet, nostamme esiin kolme keskeistä huomiota mallintamisesta ja mallintamisen merkityksestä osana tulevaisuuden ennakointia.

Ensiksi, tulevaisuuden ennakointi edellyttää mallintamisen näkökulmasta jäsentynyttä ja kontekstisidonnaista ns. *kontrafaktuaalista* ajattelua, eli pohdintaa siitä, mikä voisi olla toisin suhteessa siihen maailmaan, jonka nyt ajattelemme tuntevamme. Tätä käsittelemme artikkelin seuraavassa luvussa. Toiseksi, niin ajattelumme maailmasta kuin ajattelumme siitä, mikä voisi olla toisin, perustuu ei niinkään ”evidenssiin” tai ”dataan” itseensä kuin niihin päättelyihin, joita teemme jostain valitusta ilmiökokonaisuudesta eli evidenssistä. Päättelyn muotoja on erilaisia, ja artikkelin kolmannessa luvussa esitämme, että tulevaisuuden ennakoinnin ja erityisesti mallintamisen näkökulmasta ns. abduktiivinen päättely[1], ja sen bayesilainen formalisointi, ovat erityisen tärkeässä asemassa. Kolmanneksi, artikkelin neljännessä luvussa osoitamme, että olemassa olevista mallintamisvaihtoehdoista erityisesti bayesiläinen mallinnus pystyy tuomaan julki eli artikuloimaan näitä päättelyjämme tavalla, joka myös mahdollistaa niiden samanaikaisen kyseenalaistamisen ja muokkaamisen. Näin se samalla tulee artikuloituneeksi niitä neljää eri tapaa, joilla jonkun päättelyn väitetään olevan validi eli pätevä: kyseinen päättely evidenssistä on (ylipäätänsä) *mahdollinen*, se on *todennäköinen* (jonkin subjektiivisen päättelijän mielestä), se on *järkeenkäypä* (eli ei esim. fantastinen tai hullu) ja lopulta se on *uskottava* (mikä liittyy arvioon päättelyn esittäjästä). Käytännössä nämä neljä eri tapaa nivoutuvat miltei erottamattomasti yhteen eri kombinaatioina. Artikkelin viimeisessä luvussa vedämme yhteen edeltävissä luvuissa esiin nostetut langat.

Mahdollisia maailmoja kuvaava kausaalimalli, kuten Bayes-verkko, edustaa julkilausuttua käsitystämme siitä, mikä tuottaa jonkin rajatun systeemin vaiheiden todellista tai mahdollista tilaa koskevan aineiston eli datan.

2. Tapahtumisesta ja sen kuvittelusta

”Logiikka ilman representaatiota on metafysiikkaa.” (Pearl 2018)

Ajattele mahdollista maailmaa, johon sisältyy jokin asiaintila, jonka oletat tai tiedät olevan totta (kutsutaan tätä mahdolliseksi maailmaksi A). Tarkastelun kohteena oleva asiaintila voi olla mitä tahansa, esimerkiksi Titanicin uppoaminen tai hukassa olevat avaimet. Ajattele seuraavaksi mahdollista maailmaa, joka on tuon kyseisen tosiasian vastainen (kutsutaan tätä mahdolliseksi kontrafaktuaaliseksi maailmaksi EI-A1). Poh-

[1] Abduktiivinen päättely on päättelyä havaintojoukkoa koskevaan parhaaseen (todennäköisimpänä tai uskottavimpana pidettyyn) selitykseen (Patokorpi 2009). Bayesin teoreemaa voidaan pitää abduktiivisen päättelyn yleisen muodon esityksenä, joka ilmaisee sen, miten valinta havaintojoukkoa selittävien hypoteesien välillä tapahtuu.

di nyt sitä, miten ja kuinka paljon sinun oli manipuloitava mahdollista maailmaa A täyttääksesi tietämäsi tosiasian kanssa ristiriidassa olevan ehdon. Olisiko esimerkiksi tähystäjän kiikarien löytyminen, tai kapteenin uskomuksen aste laivan uppoamisen mahdollisuuteen suhteessa sen uppoamisen mahdottomuuteen, muuttanut tapahtumien kulun mahdollisesta maailmasta A, jossa Titanic upposi, mahdolliseen kontrafaktuaaliseen maailmaan EI-A1, jossa se ei uponnut? Osaatko arvioida, missä määrin eri muuttujat vaikuttivat lopputulokseen?

Muodostamme jatkuvasti hypoteeseja ja arvioimme eli vertaamme niiden paikkansapitävyyttä suhteessa havaintoihin. Tätä prosessia voidaan kutsua tietyin kriteerein myös tieteen tekemiseksi. Perustavanlaatuisesti kiellettyjen mahdollisuuksien ulkopuolella voit kuvitella lukuisia mahdollisia maailmoja, joissa Titanic upposi tai ei uponnut. Avaimet voivat olla hukassa monessa eri mahdollisessa maailmassa, mutta aivan kuten Titanic upposi, ne löytyvät mahdollisissa maailmoissa vain yhdessä niistä. Fysikaalinen maailma ei tässä mielessä salli ristiriitaa eli vaikka voit kuvitella erilaisia mahdollisia maailmoja, niin tästä huolimatta tapahtuu vain se, mitä tapahtuu (Deutsch 1997; Yanofsky 2019). Voidaan siis sanoa, että mahdollisten maailmojen kuvittelu edellyttää kontrafaktuaalista ajattelua ja kontrafaktuaalinen ajattelu puolestaan edellyttää aktuaalisen maailman, jossa tapahtuu vain se, mitä tapahtuu. Päättelyn näkökulmasta historian ja tulevaisuuden (kognitiivisina manipuloitavina artefakteina) muuttamisella ei ole juurikaan eroa, koska molemmat ovat lähtökohtaisesti pääteltyjä eli mahdollisia.

Mahdollisia maailmoja kuvaava kausaalimalli, kuten Bayes-verkko, edustaa julkilausuttua käsitystämme siitä, mikä tuottaa jonkin rajatun systeemin vaiheiden todellista tai mahdollista tilaa koskevan aineiston eli datan. Mahdollisten maailmojen tiedostamisen näkökulmasta merkittävää on, että jos jaamme saman kausaalisen mallin, niin jaamme myös kaikki samat kontrafaktuaaliset arviot (Pearl 2018).

Päätelmiä menneistä ja tulevista – Aineistosta oletuksiin, oletuksista aineistoon

Tulevaisuutta koskevan ajattelun voidaan sanoa perustavanlaatuisesti kontrafaktuaalista, koska vastoin sitä tosiasiaa, että ei ole olemassa tulevaisuuden tapahtumia koskevaa havaittua aineistoa, ”kuvittelemme” sellaista olevan. Jos esimerkiksi havaintojoukkoa selittävän hypoteesin antama ennuste (todennäköisimpänä pidetty tuleva aineisto) on ristiriidassa tulevien havaintojen kanssa, voi olla syytä epäillä hypoteesin tai selittävän viitekehyksen kelpoisuutta. Tämä ei kuitenkaan ole välttämättä osoitus siitä, että hypoteesi on väärä, sillä myös havaintoaineosto voi olla virheellinen. Aineistosta puhuttaessa tulisi faktojen sijaan puhua mieluummin datasta.

3. Evidenssi ja päättelyn muodot

Kausaalinen päättely

Muodostamme jatkuvasti laskelmia tai näkemyksiä mahdollisista vaihtoehtoisista tulevaisuuksista, *"jotka seuraisivat, jos nykyisyys olisi muutamaa yksityiskohtaa lukuunottamatta täsmälleen samanlainen"* (Rovelli 2018). Tästä oletuksesta käsin on luontevaa, että järjestelemme todellisuutta koskevan tietomme syiksi ja seurauksiksi. Kuvaamme ja selitämme maailmaa sanojen kuten "seuraa", "johtuu", "aiheutuu", "estää tai "edistää" avulla (Pearl 2018). Haemme syytä tai syitä sille, "*miksi*" asiat ovat tietyllä eivätkä toisella tavalla. Asioiden ei oleteta vain olevan tai vain tapahtuvan. *"Tulevan tapahtuman syy on sellainen tapahtuma, jota ilman tuleva tapahtuma ei seuraisi maailmassa, joka olisi täsmälleen samanlainen tätä syytä lukuunottamatta"* (Rovelli 2018).

Kausaalisen selittämisen taustaoletus on, että tapahtumattomalla asialla ei voi olla sen seurauksia eli ilman syytä ei saa olla olemassa. Käytämme tätä periaatetta jatkuvasti, koska sillä näyttäisi olevan merkittäviä käytännön hyötyjä. Ilman oletusta siitä, että asiat aiheuttaisivat toisia asioita, tulevaisuuden suunnittelu ja siihen toivottavalla tavalla vaikuttaminen muodostuisi melkoiseksi arpapeliksi. Toimiessamme intentionaalisesti oletamme, että toiminnallamme on vähintäänkin haluttujen seurausten toteutumisen todennäköisyyttä lisäävä vaikutus. Tämä ei tarkoita, että kyseiset toimemme takaisivat tavoitteiden toteutumisen. *Riski* voidaan näin ollen määritellä tavoitteiden saavuttamiseen liittyvänä epävarmuutena tai poikkeamana tavoitteista (Gigerenzer 2015).

Se, että organisoimme maailmaa koskevaa tietoamme syiden ja seurausten avulla, ei tarkoita, että tietäisimme, mitä kausaliteetti (kausaatio) itsessään on.[1] Jos kaksi asiaa liittyy toisiinsa, siitä ei välttämättä seuraa, että toinen on toisen syy. Tunnemme tämän oppilauseena: "*Korrelaatio ei ole kausaatio*". Tilastollinen vastaavuus kahden muuttujan välillä ei siis tarkoita välttämättä syy-seuraussuhdetta niiden välillä. Jotta kausaalisia suhteita ei sekoiteta, on hyvä muistaa, että kausaliteetti tuottaa tilastollisen vastaavuuden, mutta tilastollinen vastaavuus ei tuota kausaliteettia. Tärkeää tämä on erityisesti siitä syystä, että todellisuus itsessään saattaa olla vaikutuksiltaan armelias eikä potki tietämätöntä kintuille kaiken aikaa itsestään muistuttaen. Fyysikko David Deutschin ajatusta mukaillen: *"Teoria litteästä maasta toimii vallan hyvin lentokentän rakentamisessa."* Pallomainen pyörivä todellisuus muuttuu merkitykselliseksi vasta, kun lentää kauas. Totuus ei ole selityksen toimivuuden kannalta aina välttämätöntä.

[1] *Oxford Handbook of Causation* toteaa, että meillä on oikeastaan hyvin vähän yhteisymmärrystä siitä, mitä kausaatio lopulta on. Toisaalta ei ole mitään syytä upota käsiteelliseen suohon vaan voimme käyttää syytä ja seurausta siinä merkityksessä, jonka ymmärrämme arkisessa kielenkäytössämme, kun sanomme (päättelemme) tiettyjen asioiden aiheuttavan joitakin toisia asioita.

Onko kaikella syynsä eli onko riittävän syyn periaate välttämätön?

Perustavanlaatuista oletusta, että "asiat eivät vain tapahdu" kutsutaan riittävän perusteen tai syyn periaatteeksi. Periaate on niin vahva, että saattaa olla vaikeaa edes kuvitella sen olevan muuta kuin tosi. Tieteelliseen tietoon liittyy kuitenkin toinen periaate, jonka mukaan mikään hypoteesi ei voi olla täydellä varmuudella tosi tai epätosi. Jos näin olisi, mikään uusi evidenssi ei voisi tätä vallitsevaa näkemystä enää muuttaa. Voidaankin sanoa, että meillä on syytä uskoa, että hypoteesi "asiat eivät vain ole tai vain tapahdu" on tosi. Toisaalta meillä on myös syytä uskoa, että näin ei välttämättä ole. Fysikaalista todellisuutta kuvaavat luonnonlait eivät nimittäin kerro, miksi asiat ovat niiden kuvaamalla tavalla. Ne ilmaisevat vain, miten suureet muuttuvat suhteessa toisiinsa, mutta eivät miksi (Carroll 2017). Paras ymmärryksemme fysikaalisesta todellisuudesta ei siis vastaa kysymykseen, miksi sitä kuvaavat lait ovat sellaisia kuin ovat.

Abduktiivinen ja bayesilainen päättely

Käsitteiden *bayesilainen mallintaminen*, *mahdolliset maailmat* ja *kausaliteetti* lisäksi tässä artikkelissa esitettyyn lähestymistapaan liittyy keskeisesti lähtöoletus, jonka mukaan käsitystämme tulevaisuudesta ei oikeuta evidenssi sinänsä, vaan päättely evidenssistä. Samasta havainnoista voidaan johtaa erilaisia mahdollisia maailmoja vailla ristiriitaa. Tämä siitä syystä, että hahmottamamme mahdolliset maailmat nojaavat aina johonkin havaintoa selittävään viitekehykseen.

Mikä on syy uskoa, että kaikki joutsenet olisivat valkoisia?

Mustia joutsenia on, koska käytämme valkoista joutsenen määrittelyyn ilman loogista syytä sille, että joutsenet ovat valkoisia.

Induktion ongelma (tai sen ongelmattomuus) on hiertänyt joidenkin tieteentekijöiden mieltä siitä lähtien kun David Hume esitti, että induktion pätevyydelle ei ole mitään loogista syytä. Toisin sanoen, jos ei ole mitään loogista syytä sille, että joutsenet ovat valkoisia, ei ole mitään loogista syytä uskoa väitteeseen, että kaikki joutsenet ovat valkoisia, vaikka kaikki tähän mennessä nähdyt joutsenet ovat valkoisia (Yanofsky 2019). Aivan kuten induktio myös abduktio lähtee liikenteeseen joukosta havaintoa, mutta toisin kun induktio, abduktio ei pyri yleistämään johtopäätöstä. Abduktioon liittyvä kysymys on pikemminkin: mitkä hypoteesit voivat selittää havaintojoukon.[1]

[1] Kahdesta teoriasta, jotka ennustavat samat havainnot, mutta eroavat olettamuksiltaan toisistaan, on mahdoton, tieteen keinoin, sanoa kumpi on oikea (Feynman 1965).

Toisin sanoen, siihen liittyvä kysymys on, ”mitkä kaikki prosessit voivat tuottaa havaintoaineiston?”. Bayesilainen päättely formalisoi abduktiivisen päättelyn ja auttaa valitsemaan vallitsevista hypoteeseistä evidenssin valossa uskottavimman (Carroll 2017). Seuraavaksi tarkastelemme, miten tämä tarkemmin ottaen tapahtuu.

Miten bayesilainen päättely toimii

Oletetaan arkinen skenaario, jossa löydän (minä olen päättelijä) jääkaapista valmisaterian, jonka parasta ennen päiväys on mennyt vanhaksi neljä päivää sitten. Onko ateria vielä syötävä? Henkilökohtaisen kokemukseni pohjalta minulla on syy uskoa, että neljä päivää on vielä turvallinen ylitys ja ateria on syötäväksi kelpaava. Ennusteeni on, että syömäkelpoisuuden todennäköisyys on 80 %. Todennäköisyyden voidaan ajatella olevan osuus, joka vastaa tietyn vaihtoehdon mahdollisuutta kaikista mahdollisista vaihtoehdoista: siis tietty mahdollinen pala kaikkien mahdollisuuksien kakkua. Tavallaan siis ajatellen jääkaapissani olevan joukon identtisiä aterioita, joista 4/5 osaa olisi annetuilla kriteereillä syömäkelpoisia.[1] Minulla ei kuitenkaan ole identtisiä annoksia vaan ainoastaan tämä yksittäispakattu yksittäistapaus, jolla on ainutlaatuinen historia ja tulevaisuus ravintona tai roskana. Vaikka minulla ei ole kyseisestä valmisateriasta mitään aiempaa kokemusta, uskon siis muiden kokemuksieni pohjalta ruuan olevan syötävää melko varmasti. Merkitään tätä 80 %:n priori-todennäköisyyttä (priori viittaa todennäköisyyteen ennen lisätietoa) termillä P(H). Tämä on lähtökohtainen uskomukseni aste sille, että hypoteesini *”valmisateria on vielä syötävää”* on totta.

Uutta informaatiota ilmenee, kun avaan paketin ja nuuhkaisen ruokaa. Haju on hieman outo. Kysyn itseltäni, että miltä tämän pitäisi tuoksua oletuksilla a) syötävää tai b) ei syötävää. Minun on siis muodostettava näkemys, joka ilmaisee, millä todennäköisyydellä asiantila “ruoka on syötävää” antaa kyseisen aistihavainnon. Merkitään tätä P(E|H), joka on todennäköisyys sille, että evidenssi (outo haju) ilmaantuu ehdolla hypoteesi (on syötävää) on totta. Arvion tämän todennäköisyyden olevan noin 50 %. Ajattelen siis, että evidenssi sopii noin puoleen syötäväksi kelpaavista aterioista. Toisin sanoen, vertauskuvainnollisesti puhuen, 100:sta identtisestä kuvitteellisesta ateriasta, joista 80 oli uskoakseni syötäviä, ajattelen siis evidenssin sopivan 40 ateriaan. Tarkastelen myös silmämääräisesti ruoka-annosta, mutta en havaitse mitään huolestuttavaa. En tosin ole varma, miltä ruuan pitäisi näyttää oletuksilla a) syötävää b) ei syötävää. Teoriani sen osalta, mikä on visuaalisesti huolestuttavaa, on melko puutteellinen, joten jätän sen huomioimatta. Miten uusi informaatio vaikuttaa käsitykseeni siitä, onko ruoka vielä syötävää?

Varsinainen kysymykseni kohdistuu todennäköisyyteen P(H|E), joka on todennäköisyys sille, että hypoteesini (on syötävää) on totta tietyllä evidenssillä (outo haju). Lienee ilmeistä, että uusi informaatio voi olla vaikuttamatta, lisätä tai vähentää uskoani hypoteesin (syötävää) paikkansapitävyyteen. On siis löydettävä kerroin, joka kertoo alkutilanteen uskomukseni asteelle P(H), tuleeko sen uuden tiedon valossa pysyä en-

[1] Jos objektiiviset todennäköisyydet ovat tiedossa, rationaalinen toimija sovittaa uskomuksen asteensa näihin (David Lewis, ”Principal Principle”).

nallaan, heiketä vai voimistua. Minulla on tähän mennessä todennäköisyydet sille, että hypoteesini *"on syötävää"* (ollessaan totta) antaa evidenssin (outo haju) eli P(E|H) ja todennäköisyys sille, että hypoteesi on totta ennen evidenssin ilmaantumista eli P(H). Kertomalla P(E|H) ja P(H) keskenään saadaan päivitetty pala kakkua, mutta koko kakku on vielä päivittämättä eli en tiedä, kuinka suuri osa kakusta päivitetty pala on.

Koko kakku on evidenssin ilmenemisen todennäköisyys P(E), kun kaikki "tiedetyt" mahdollisuudet otetaan huomioon. Toisin sanoen koko kakku on todennäköisyys sille, että evidenssi on totta. Kärjistäen voisi ajatella, että kyseessä on todennäköisyys sille, että havaittu tosiasia on tosiasia. Kaikki tiedetyt mahdollisuudet ovat edellä irrotettu pala ja loput kakusta yhteensä. Loput kakusta saadaan laskemalla, tai arvioimalla, evidenssin ilmaantumisen todennäköisyys oletuksella hypoteesi ei ole totta. Oletan tämän todennäköisyyden olevan 80 % eli evidenssi (outo haju) sopii valtaosaan tapauksista "ei syötävä". On toki mahdollista, että ateria on syömäkelvoton ilman outoa hajuakin. Tällöin mahdollisesti syömäkelvottomien aterioiden koko joukko on todennäköisesti suurempi kuin se, mihin evidenssi sopii. Olen päätynyt päättelyssäni siihen, että uuden informaation päivittämässä kakussani on siis kaikkiaan 40+0,8×20 eli 56 palaa. Näin ollen päivitetty mahdollisuusosuus, että ateria on uuden evidenssin (outo haju) valossa syömäkelpoinen kaikkien mahdollisuuksien joukosta on 40/56 eli noin 71 % (ks. taulukko 1). Evidenssi siis heikensi uskoani aterian syömäkelpoisuuteen. Koska priori-todennäköisyys P(H) oli niin vahva, heikentymistä oli vähemmän kuin pelkän hajun perusteella olisi ollut ehkä syytä olettaa.

Tietynlaisesta hämäryydestään huolimatta Bayesin teoreema tarjoaa rationaaliselle toimijalle analyyttisen apuvälineen tarkastella sitä, miten valita eri hypoteesien tai johtopäätösten väliltä.

Edellä päättelyn näkökulmasta läpikäyty Bayesin teoreeman voidaan esittää yleisessä muodossaan seuraavasti (mukaillen von Baeyer 2003):

> *Todennäköisyys sille, että hypoteesi on totta annetulla evidenssillä, P(H|E), on jokin kerroin, (k), kertaa priori-todennäköisyys sille, että hypoteesi on totta, P(H). Kerroin, (k), on todennäköisyys sille, että hypoteesi antaa evidenssin, (E|H), jaettuna evidenssin ilmaantumisen todennäköisyydellä, T(E).*

Taulukko 1. Edellä esitetty esimerkki Bayesin teoreemaan mukaisesta uskomuksen asteen (todennäköisyys) määräytymisestä lähtöhypoteesin paikkansapitävyyteen riippuen siitä, mitä informaatiota päättelijällä on käytössään.

Priori-todennäköisyys		
Mikä on todennäköisyys, että jääkaapista otettu neljä päivää viimeisen käyttöpäivän ylittänyt eines on vielä syötävä?	P(H)	80 %
Uutta informaatiota ilmenee (outo haju)		
Mikä on todennäköisyys, että avattaessa eines haisee oudolta ja on silti syötävä?	P(E|H)	50 %
Mikä on todennäköisyys sille, että avattaessa eines haisee oudolta ja ei ole syötävä?	P(E|ei-H)	80 %
Priori-todennäköisyys päivittyy uuden informaation myötä posteriori-todennäköisyydeksi		
Miten todennäköistä tai uskottavaa on, että eines on syötävä, kun se haisee oudolta ja sen viimeisen käyttöpäivän päiväys on mennyt neljä päivää sitten vanhaksi?	$\frac{P(H)*P(E\|H)}{(P(H)*P(E\|H) + P(E\|ei–H)*[1–P(H)])}$	71 %

Bayesin teoreema, $P(H|E) = \frac{P(H)*P(E|H)}{P(E)}$ on formaali esitys sille, miten uskomuksen määrä hypoteesin paikkansapitävyyteen muuttuu uuden informaation ilmaantumisen myötä. Vaikka sen matematiikkaa on periaatteessa yksinkertaista ja selkeää, se mistä luvut lopulta tulevat ei ole välttämättä aina yhtä selkeää. Tähän on hyvin yksinkertainen syy: usein on mahdoton tietää, mikä on mahdollista, jonka seurauksena on mahdotonta tietää, kuinka suuri osa kaikista mahdollisuuksista on se osa, johon evidenssi sopii. Tietynlaisesta hämäryydestään huolimatta Bayesin teoreema tarjoaa rationaaliselle toimijalle analyyttisen apuvälineen tarkastella sitä, miten valita eri hypoteesien tai johtopäätösten väliltä. Se tekee näkyväksi, miten havaintoaineisto, sekä tästä havaintoaineistosta esitetyt ja mahdollisiksi, todennäköisiksi, järkeenkäyviksi ja/tai uskottaviksi määritellyt tai arvioidut päättelyt, *yhdessä* ehdollistavat ajatteluamme.

Bayesilainen päättely on siis järkeilyä siitä, kuinka todennäköisenä pidetään annetulla evidenssillä tiettyä mahdollisuutta kaikkien mahdollisuuksien joukosta. Se mitä puolestaan kutsumme päättelyssämme kausaliteetiksi, noudattaa usein samaa päättelyn rakennetta: Esimerkiksi P(a|b) on todennäköisyys sille, että a tapahtuu ehdolla b on tapahtunut. Edellä esitetyssä päättelyssä esimerkiksi kausaalisesti ruuan pilaantuminen voi aiheuttaa mahdollisesti oudon hajun, ei päinvastoin. Kausaalimallinnuksella päästään kiinni kokonaisvaltaiseen miksi-näkemykseen, joka edustaa käsitystämme asioiden tapahtumisesta ajan nuolen suuntaan etenevässä syklittömässä syiden ja seurausten verkossa. Malli siis edustaa ajattelemaamme oikeutustamme tulevaisuutta koskevalle näkemykselle.

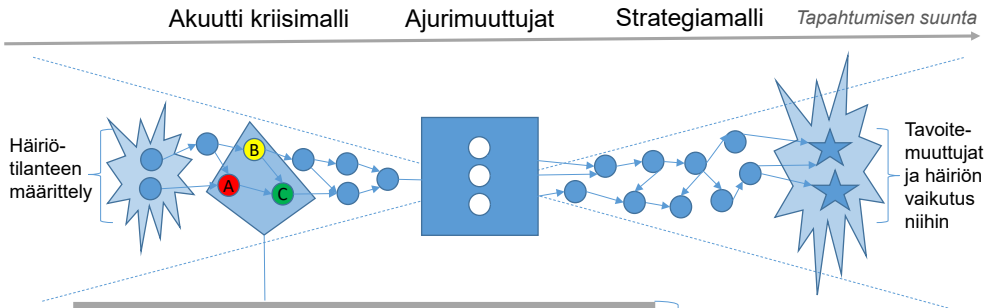
Jos tietäisimme asioiden todellisen laidan, emme tarvitsi teorioita ja hypoteeseja, tai päättelyä, jolla valitaan hypoteeseista parhaiten evidenssiä selittävä vaihtoehto. Koska emme tiedä ja koska aineisto ei puhu puolestaan, on tosiasiat (data) selitettävä. Aineisto on siis tässä mielessä sokea eikä sen avulla voida vastata kysymykseen, miksi aineisto on sellainen kuin se on. Jos oletamme, kuten induktiossa oletamme maailmankaikkeuden olevan ajan suhteen yhtenäinen, saatamme ajatella tulevaisuuden muistuttavan tulkitun säännönmukaisuuden puitteissa menneisyyttä (Yanofsky 2019). Tähän uskoon ei ole syytä tuudittautua, vaan selitystä etsiessämme tarkkaavaisuus tulisi kääntää edellä toistettuun kysymykseen siitä: mitkä prosessit tuottavat aineiston. Tämä ajatus on BOR-konseptin ytimessä.

4. Bayes-verkkojen soveltaminen käytäntöön: tapaustutkimus BOR (Bayesian Operation Room)

Esitämme seuraavaksi vaiheittain pääkohdat siitä, miten kausaalista mallinnusta ja Bayes-verkkoja voidaan hyödyntää osallistavasti asiantuntijatiedon strukturoinnissa. Esimerkki perustuu tapaustutkimukseen, jossa on haluttu ennakoida yllättävän häiriötä tuottavan muutostekijän (eli analogian mukaisesti ”*tuloverotuksen nosto*”) vaikutusmekanismeja organisaation (eli analogian mukaisesti *Mikä Mikä -maan* pääkaupungin) kykyyn toteuttaa kaupunkistrategiaansa (”*tuloverotuksen noston [todennäköiset] vaikutukset työttömyyteen*”). BOR-konsepti on kehitetty osana Suomen Akatemian yhteydessä toimivan Strategisen tutkimuksen neuvoston rahoittamaa WISE-tutkimushanketta (wiseproject.fi). Tekemisen tapa eli se, miten asiantuntijoiden kanssa toimitaan, on erottamaton osa BOR-konseptia. Jätämme sen tässä yhteydessä kuitenkin vähemmälle huomiolle ja keskitymme itse malliin liittyviin näkökohtiin. Tässä luvussa esitetty BOR-konseptin kuvaus on tiivistetty versio artikkelissa ”*Tulevaisuusresilienssi ja strateginen ennakointi: kriisinkestävyyden harjoittelua bayeslaisella kausaalimallinnuksella*” (Ahvenainen et al. 2021) esitetystä kuvauksesta.

BOR-konseptin keskiössä oleva Bayes-verkko on todennäköisyyspohjainen malli, joka yhdistää systeemianalyysiä ja kognitiivista mallinnusta. Malli ilmaisee tekijöidensä näkemyksen siitä, miten tarkasteltava maailma toimii rajatuissa, käsiteltävän kysymyksen kannalta relevanteissa puitteissa.

BOR-konseptin keskiössä oleva Bayes-verkko on todennäköisyyspohjainen malli, joka yhdistää systeemianalyysiä ja kognitiivista mallinnusta. Toisin sanoen, malli ilmaisee tekijöidensä näkemyksen siitä, miten tarkasteltava maailma toimii rajatuissa, käsiteltävän kysymyksen kannalta relevanteissa puitteissa. Malli koostuu tarkastelun kohteena olevan systeemin graafisesta kuvauksesta, jossa määritellään laadullisesti muuttujat sekä niiden väliset riippuvuudet. Tämän lisäksi malli sisältää muuttujille määriteltyjä todennäköisyystaulukoita (ks. kuva 1), joissa määritellään ehdollisin todennäköisyysjakaumin muuttujan määrällinen riippuvuus siihen vaikuttavista muista muuttujista.



Ehdollinen todennäköisyystaulukko muuttujalle C				
Jos A =	$A_{[true]}$		$A_{[false]}$	
Jos B =	$B_{[true]}$	$B_{[false]}$	$B_{[true]}$	$B_{[false]}$
$P(C_{[1]})$	0.35	0.80	0.05	0.15
$P(C_{[2]})$	0.50	0.15	0.15	0.50
$P(C_{[3]})$	0.15	0.05	0.80	0.35
	P(C|A[t],B[t])	P(C|A[t],B[f])	P(C|A[f],B[t])	P(C|A[f],B[f])

Muuttujien tilojen välisten todennäköisyyspohjaisien riippuvuuksien määrittelyn periaate

Kuva 1. BOR-konseptin Bayes-verkon yleiskuvaus (Ahvenainen et al. 2021).

Mallinnuksen ensimmäisessä vaiheessa, prognoosissa, muodostettiin kausaalinen näkemys siitä, miten mahdollinen häiriötilanne vaikuttaa ajurimuuttujiin ja näistä edelleen kaupungin strategisiin tavoitteisiin. Ajurimuuttujilla tarkoitetaan tässä yhteydessä niitä mallin muuttujia, joihin mallissa halutaan muutostekijöinä, suhteessa lopputilaa ilmaiseviin strategisiin tulosmuuttujiin, kiinnittää erityistä huomiota. Tässä mallissa ajurimuuttujat koskivat ympäristön tilaa ja ihmisten terveyttä. Osallistujien tehtävänä oli yhdessä keskustellen ennakoida (piirtäen) häiriön pitkän aikavälin vaikutuksia edustavia kausaaliketjuja (vaikutusmekanismeja), jotka mahdollisesti vaarantaisivat kaupungin strategisten tavoitteiden toteutumisen. Kun analysoitavan systeemin graafinen kuvaus oli valmis, muuttujien välisiä riippuvuuksia määriteltiin tarkemmin ja malli muutettiin numeeriseksi Bayes-verkoksi. Riippuvuussuhteita määriteltäessä arvioitiin suhteen suuntaa (negatiivinen/positiivinen riippuvuus), suhteellista voimakkuutta (heikko/kohtalainen/voimakas vaikutus) ja tähän liittyvää epävarmuuden määrää (vähäinen/kohtalainen/erityisen suuri epävarmuus).

Numeerista Bayes-verkkoa käytettiin mallinnetun systeemin diagnosoinnissa. Tämä tarkoitti käytännössä sitä, että mallin kuvaamassa häiriytyneessä tulevaisuudessa kuljettiin (taannustettiin) takaisinpäin häiriön syihin. Näin voitiin muodostaa näkemys siitä, miten mallin edustamia mahdollisia maailmoja voi manipuloida niin, että strategian vaarantumisen todennäköisyys pienenee. Diagnostisessa päättelyssä analysoidaan tiettyyn tapahtumaan tai havaintoon todennäköisesti johtaneita olosuhteita kausaaliketjujen suunnan vastaisesti. Koska kausaaliset suhteet on mallissa kvantifioitu, on mahdollista simuloida tulevaisuuden herkkyyttä suhteessa eri muuttujiin. Näin saatiin esiin avainmuuttujat, joiden tilojen muutokset olivat merkittävimmin kytköksissä

Mallinnuksen käyttötarkoitus oli luoda analyyttinen ja systemaattinen kuva siitä, mitkä valinnat ovat organisaatiolle olemassa ja mitkä mahdolliset toimenpiteet niitä edustavat, jos halutaan parantaa kriisinsietokykyä.

muutoksiin strategisia tavoitteita kuvaavien muuttujien tiloissa. Asettamalla strategisia tavoitteita kuvanneet muuttujat huonoimpaan mahdolliseen tilaansa, pystyttiin mallin todennäköisyysjakaumien muutoksista arvioimaan, millaiset tekijät tulevaisuudessa todennäköisimmin vaarantaisivat strategiset tavoitteet häiriön seurauksena.

Diagnoosin tulosten pohjalta luotiin skenaariot siitä, miltä tulevaisuus näyttäisi silloin, jos päädytään kunkin tavoitteen osalta huonoimpaan mahdolliseen tilanteeseen. Näin saatiin esiin osallistujien ajattelua edustava näkemys siitä, miten strategian toteuttaminen voisi vaarantua. Skenaarioiden pohjalta osallistujat keskustelivat yhdessä siitä, mihin toimiin voidaan ja tulisi ennalta ryhtyä, jotta haitat mallin kuvaamassa häiriötilanteessa jäisivät mahdollisimman vähäisiksi. Keskustelun tuloksena syntyi toimenpide-ehdotuksia, joilla kaupunki voisi ennakoiden lisätä omaa kriisinsietokykyään.

BOR-konseptissa malli on osa menetelmä- ja yhteistyöalustan arkkitehtuuria, jonka avulla pyritään saavuttamaan asiantuntijatiedosta jaettu käsitys siitä, miten mahdollinen häiriö toimii osana organisaation strategisia pyrkimyksiä. Tosin sanoen, miksi ja miten yllättävä ja äkillinen muutosvoima kauaskantautuu? Mallinnuksen käyttötarkoitus oli luoda analyyttinen ja systemaattinen kuva siitä, mitkä valinnat ovat organisaatiolle olemassa ja mitkä mahdolliset toimenpiteet niitä edustavat, jos halutaan parantaa kriisinsietokykyä.

5. Lopuksi

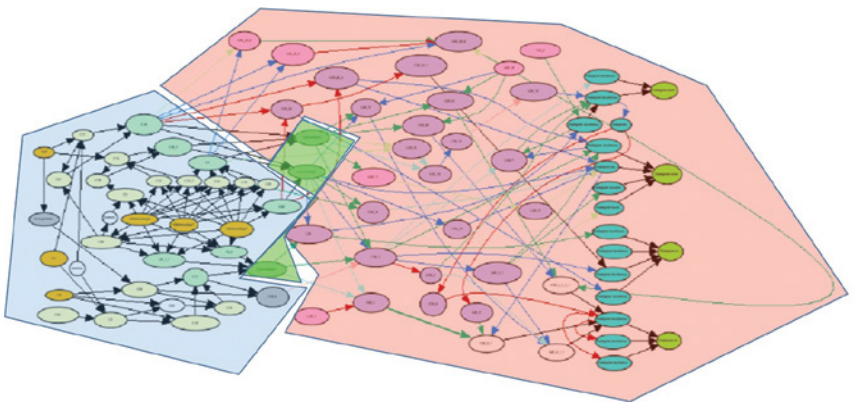
Miksi sitten ylipäätänsä lähteä kysymään nimenomaan bayesilaiseen päättelyyn ja siihen perustuvaan kausaalisen mallinnuksen keinoin i) millaisilta jonkin ilmiön (esim. tuloverojen nosto) ii) mahdolliset vaikutukset (eli *Mikä Mikä -maa* kokonaisuudessaan) iii) johonkin toiseen ilmiöön (esim. *työttömyyden tasoon*) iv) näyttäytyvät jossakin tietyssä kontekstissa (esim. *Mikä Mikä -maan Ihmetys-pääkaupungissa*) v) jonkun tietyn asiantuntijajoukon (eli ne, jotka on kerätty BOR-harjoitukseen) kollektiivisissa silmissä, tai paremmin, kollektiivisessa mielessä?

Tulevaisuudentutkimuksen näkökulmasta ensimmäinen vastaus tähän liittyy yleiseen kysymykseen siitä, miksi mallinnuksia, bayesilaisia tai muita, ylipäätänsä tehdään: koska meillä ei ole muita keinoja vastata sellaisiin kysymyksiin, joihin ei voida vastata suorilla havainnoilla. Tulevaisuutta koskevat kysymykset ovat yleensä juuri tällaisia. Mallit, ja yleisemmin ottaen skenaariot, on ajateltavissa eräänlaisina kollektiivisina tulevaisuuteen suuntautuneina (tilanteesta riippuen, mahdollisina, todennäköisinä pidettyinä, järkeenkäypinä ja/tai uskottavina) kertomuksina tai narratiiveina, jotka mahdollistavat inhimillisen kollektiivisen intentionaalisen toiminnan normatiivista arviointia eli ”harkintaa”. Esim. hallitusten välisen ilmastopaneelin (IPCC) mallit mahdollistavat vastauksen kysymykseen ”mitä tapahtuisi, jos jatkaisimme nykyistä elämäntapaamme, ja mitä tapahtuisi, jos emme tekisi niin?”. Tämä kysymys osaltaan mahdollistaa kollektiivisen vastauksen siihen, haluammeko tehdä näin vaiko ei.

Suhteessa tähän yleiseen mallien luonnehdintaan, nimenomaan bayesilaisella mallintamisella on kuitenkin tähän harkintaprosessien mahdollistamiseen liittyviä erityis-

piirteitä. Siinä missä kaikki mallit antavat jonkinlaisen vastauksen kysymykseen ”mikä tapahtuisi, jos tekisimme niin eikä näin?”, yllä kuvaavamme bayesilainen mallintaminen kysyy samanaikaisesti, ”miksi ajattelette, että olisi juuri näin, eikä jotenkin toisin?” Toisin sanoen, itse mallintamisprosessi on suunniteltu siten, että sekä asiantuntijoiden kontrafaktuaalinen ajattelu ja päättely jostain ilmiöstä ja sen mahdollisista vaikutuksista jossain kontekstissa että sen perustelut tulevat samanaikaisesti esiin.

Mallintamisprosessin lopputulemana on formaalisesti esitetty kokonaisnäkemys siitä, miten jokin asiantuntijajoukko päättelee, että jokin ilmiö kausaalisesti vaikuttaa johonkin toiseen ilmiöön jossain kontekstissa, ja millä tavoin tämä asiantuntiajoukko perustelee päättelyjensä validiutta tai pätevyyttä. Näitä päättelyitä voidaan luonnehtia – ja harjoituksessa myös de facto luonnehdittiin – *mahdollisiksi/mahdottomiksi* (”ei ole mahdollista, että tulovero nostaa työttömyyttä”), *todennäköisiksi/epätodennäköisiksi* (”on todennäköisesti näin”), *järkeenkäyviksi/ei-järkeenkäyviksi* (”ei todellakaan käy järkeen, että näin tapahtuisi, vaikka mahdollista onkin”) tai *(epä)uskottaviksi* (”ainahan se on mahdollista, mutta miksi ihmeessä uskoisimme näin tapahtuvan, kun on paljon muitakin selittäviä tekijöitä maailmassa”), ks. kuva 2. Käytännössä nämä neljä eri tapaa oikeuttaa ”päättelyitä evidenssistä” nivoutuvat – ja nivoutuivat myös de facto BOR-harjoituksessa – miltei erottamattomasti yhteen eri kombinaatioina. Kuvassa 2 on esitetty havainnollistus siitä, miltä nimenomaan tämä hypoteesien ja niiden perusteluista kudottu tilkkutäkki tarkemmin ottaen näytti.



Kuva 2. BOR-mallin graafinen kuvaus (akuutti kriisimalli pohjustettu sinisellä, ajurimuuttujat vihreällä ja strategiamalli punaisella).

Tämä tulevaisuuden ennakoinnin uusi konsepti mahdollistaa lopulta myös kysymyksen siitä, olemmeko asiantuntijoina kollektiivisesti todellakin huomioineet kaikki ilmiöön liittyvät ja vaikuttavat asiat päättelyissämme evidenssistä, vai onko mahdollisesti jotain jäänyt huomioimatta. Toisin sanoen: bayesilainen mallinnus mahdollistaa ei vain *kollektiivisen harkinnan* jostain valitusta ilmiöstä ja sen mahdollisista vaikutuksista johonkin toiseen ilmiöön (tämän tekevät kaikki mallit), vaan myös *kriittisen itsereflektion tästä omasta pohdinnasta* ja siihen liittyvistä usein ”itsestään selvistä” olet-

tamuksista. Se on näin ollen myös vahva vastalääke ns. ryhmäajattelulle ja näin ollen myös sille, että saatamme olla kollektiivisesti pahanlaatuisesti väärässä jonkin meille tärkeän ilmiön osalta. Ludwig Wittgensteinia muokkaillen, bayesilainen mallintaminen näyttää lasipulloon joutuneelle kärpäselle, miltä sen vanginnut kapistus tarkemmin ottaen näyttää.

Lähdeluettelo

Ahvenainen, Marko – Janasik, Nina – Lehikoinen, Annukka & Reinekoski, Tapio (2021) Tulevaisuusresilienssi ja strateginen ennakointi: kriisinkestävyyden harjoittelua bayeslaisella kausaalimallinnuksella. *Futura* 4/2021.

Carroll, Sean (2016) *The Big Picture. On the Origins of Life, Meaning, and the Universe Itself*. Dutton, New York.

Deutsch, David (1997) *Todellisuuden rakenne*. Terra Cognita, Helsinki.

Feynman, Richard (2017) *The Character of Physical Law*. With New Foreword by Frank Wilczek. The MIT Press.

Gigerenzer, Gerd (2015) *Riskitietoisuus*. Terra Cognita Oy, Helsinki.

Kuusi, Osmo – Bergman, Timo & Salminen, Hazel (2013) Tulevaisuudentutkimuksen käsitteitä. Teoksessa Kuusi, Osmo – Bergman, Timo & Salminen, Hazel (toim.) *Miten tutkimme tulevaisuuksia?* Tulevaisuuden tutkimuksen seura ry, Helsinki.

Lewis, David (1980) A Subjectivist's Guide to Objective Chance. Reprinted in *Philosophical Papers*, Vol. 2, 1986, Oxford University Press, Oxford.

Pearl, Judea & Dana Mackenzie (2018) *The Book of Why. The New Science of Cause and Effect.* Basic Books, New York.

Patokorpi, Erkki & Ahvenainen, Marko (2009) Developing an abduction-based method for futures research. *Futures*, 41(3), 126–139.

Rovelli, Carlo (2018) *Ajan luonne*. 2. painos. Ursan julkaisuja 160, Tallinna.

von Bayer, Hans Christian (2005) *Information. The New Language of Science*. Harvard University Press, Cambridge.

Yanofsky, Noson S. (2019) *Perustellun tiedon ulkorajat. Mitä tiede, matematiikka ja logiikka eivät voi kertoa.* Terra Cognita Oy, Helsinki.