

Reports from the Department of Philosophy
Vol. 38

DE NATURA RERUM

Scripta in honorem professoris Olli Koistinen sexagesimum
annum complentis

Edited by
Hemmo Laiho & Arto Repo

University of Turku
Finland

Copyright © 2016 Authors

SERIES EDITORS:

Olli Koistinen
Juha Räikkä

Department of Philosophy
University of Turku
FI-20014 Turku
Finland

ISSN 1457-9332
ISBN 978-951-29-6455-0 (PRINT)
ISBN 978-951-29-6456-7 (PDF)

Painosalama Oy, Turku 2016

Sisällys / Contents

Esipuhe	5
Filosofisia kysymyksiä keskiajan teorioissa muutoksesta ja kontinuumista	7
Simo Knuuttila	
Descartes' Punctum Archimedis: The Primality and Unity of Being, the Derivativeness of the Generated Dualities	25
Joseph Almog	
Why Things Matter: Camus' Meursault and Descartes' Causal Principle	59
Jani Sinokki	
Descartes and Spinoza on the Love of God	74
Lilli Alanen	
Spinoza ja moderni kosmologia	98
Juhani Pietarinen	
Spinozan tulkintaoppi	111
Seppo Sajama	
Spinoza on Self-Contentment	126
Daryl de Bruyn	
Remarks on Cognition in Spinoza: Understanding, Sensation, and Belief	134
John Carriero	
Leibniz on Primitive Concepts and Conceiving Reality	148
Peter Myrdal & Arto Repo	
Kantin maailmat	167
Ilkka Niiniluoto	
Kant ja aikajärjestyksen ideaalisuus	176
Valteri Arstila	
Kaappikantilaisuutta?	188
Jussi Haukioja	
Kant ja aistimuksen objektiivisuuden ongelma	195
Matti Saarni	

Maksiimeista ja itsetietoisuudesta Kantin filosofiassa	210
Markus Nikkarla	
Transcendental Anti-Theodicy	220
Sami Pihlström	
Vähäpätöinen kiista	231
Tapio Korte	
Bradley's <i>Reductio</i> of Relations and Formal Ontological Relations	246
Jani Hakkarainen & Markku Keinänen	
Emootioteoriat ja toisen emootioita koskeva tieto	262
Leila Haaparanta	
Habituaalisen toiminnan haaste vapaalle toiminnalle	270
Susanne Uusitalo	
J. V. Snellmanin käsitys oman edun roolista valtiossa	282
Juho Ritola	
J. V. Snellman ja Lorenz Lindelöf äänestysteoreetikoina	301
Eerik Lagerspetz	
Mitä vikaa Granger-kausallisuudessa?	319
Hannu Nurmi	
Metafilosofinen naturalismi ja antinaturalismi uuden ajan filosofiassa	330
Henri Pettersson	
Graciánin viisaudet	349
Juha Räikkä	

Esipuhe

Tämän kirjan kirjoitukset on omistettu professori Olli Koistiselle hänen 60-vuotissyntymäpäivänään 11.5.2016. Kirjoittajat ovat professori Koistisen kollegoita, työtovereita, väitöskirjaohjattavia ja opiskelijoita, entisiä tai nykyisiä.

Kirjoituksista osa on suomeksi, osa englanniksi. Siksi päätimme antaa kirjalle latinankielisen nimen. *De natura rerum* viittaa moneen, mukaan lukien samannimiset vanhat luonnonfilosofiset tekstit, joista varmaankin kuuluisin, Lucretiuksen *De rerum natura*, on käännetty suomeksi *Maailmankaikkeudesta*. Sikäli kuin yhtään professori Koistista tunnemme, tämä, yhtä hyvin kuin Lucretiuksen teoksen mahdollinen kirjaimellisempi käännös *Asioiden luonnosta*, kuvaisi hyvin hänen filosofiakäsitystään. Filosofiasa tutkitaan ihan oikeita, isoja asioita. Filosofit kysyy, millaisia erilaiset oliot ja ilmiöt ovat luonteeltaan – mikä on asioiden luonto.

Kirjan kirjoituksissa tarkastelun kohteena on muun muassa muutos, kausaliteetti, hyvä elämä, tekstien tulkitseminen, inhimillinen toiminta, tunteet, ajattelu, tietoisuus, aika, todellisuus ja maailma.

Kuten sopii odottaa, useissa kirjoituksissa viitataan filosofian klassikoihin, varsinkin Descartesiin, Spinozaan ja Kantiin, jotka ovat olleet Ollin omia rakkaita tutkimuskohteita. Klassikkoteksteihin viitataan filosofian historian tutkimuksessa vakiintuneiden käytäntöjen mukaisesti. Tarkempia selityksiä löytyy kirjoitusten alaviitteistä ja kirjallisuusluetteloista.

Olli, kanssasi on aina ollut ilo työskennellä ja olemme oppineet sinulta paljon. Filosofiksi olet vieläkin melko nuori, joten yhteistyö jatkukoon kauas tulevaisuuteen. Oikein mainiota syntymäpäivää!

Hemmo Laiho ja Arto Repo
San Diegossa ja Turussa huhtikuussa 2016

Filosofisia kysymyksiä keskiajan teorioissa muutoksesta ja kontinuumista

Simo Knuuttila

Antonio Trombetta, 1400-luvulla vaikuttanut italialainen filosofi ja teologi, esitti seuraavan analyysin instantiaalisesta muutoksesta toisensa poissulkevien asiantilojen välillä. Muutos ei voi tapahtua silloin, kun vain toinen tai toinen asiantila vallitsee, koska ne ovat yksittäin aktuaalisia muutosta ennen tai sen jälkeen. Ei voi myöskään ajatella, ettei kumpikaan asiantila ole aktuaalinen muutoshetkellä. Siksi sen, mistä jokin muuttuu (*terminus a quo*), ja sen, miksi se muuttuu (*terminus ad quem*), täytyy olla aktuaalisia samassa fysikaalisessa ajan instanssissa. Ne eivät voi kuitenkaan olla aktuaalisia samassa loogiseen järjestykseen kuuluvassa ”luonnon instanssissa”, vaan muutoksen lähtökohdan täytyy olla aktuaalinen yhdessä luonnon instanssissa ja sen lopputuloksen toisessa. Muutos sisältää reaalisen ristiriidan empiirisessä maailmassa. Se ei kuitenkaan koske ristiriidan lakia logiikassa, joka määrittää ei-temporaalista ja käsitteellistä olemistasoa. Ajan instanssit ovat fysikaalisen teorian mukaan aikakontinuumin ei-jaettavia ja kestottomia jakokohtia, mutta ne eivät jaa sitä niin, että muutosristiriitä häviäisi. Jakamatonta muutosinstanssia vastaavaa tilaa maailmassa ei voi kuvata viittaamatta aktuaaliseen ristiriitaan (p ja $\sim p$). Muutoksen käsitteellisesti erotettavat ristiriitaiset konstituentit ovat samanhetkisiä fysikaalisessa todellisuudessa, mutta ne ovat erillisiä luonnon instansseissa.¹

¹ ”Mutatio instantanea non est sine suis terminis quum sit habitudo inter terminum a quo et terminum ad quem et habitudo seu relatio non est sine extremis; ergo in eodem instanti temporis est terminus a quo ipsius mutationis et terminus ad quem eiusdem. Sed ista non possunt esse in eodem signa nature, quia tum esset contradictio secundum idem indivisibile temporis et nature. Ergo pro diversis signis nature erit terminus a quo and terminus ad quem.” (Antonio Trombetta, *In tractatum formalitatum Scoti sententia*, 4.2.) Seuraava on esimerkki saman ajattelumallin sovellutuksesta 1500-luvun espanjalaisessa vanhurskauttamisoppia koskevassa keskustelussa: ”Respondeo. Illis duobus instantibus [sc. peccati et gratiae] respondet unum instans indivisibile temporis continui, quoniam cum utrumque sit indivisibile et indivisibile additum indivisibili non faciat maius, utrumque indivisibile non signat in tempore extrinseco nisi unum indivisibile, et ideo non solum inter illa duo instantia non mediat tempus intrinsecum, sed neque etiam tempus

Tämä ei ollut uusi teoria. Tosiasiassa Trombetta selittää ja osittain myös lainaa Landulphus Caracciolon 1300-luvun alussa esittämää muutoksen ristiriitoteoriaa, joka herätti omana aikanaan paljon huomiota ja tietysti myös kritiikkiä, koska se poikkesi yliopisto-opettajien omaksumasta Aristoteleen *Fysiikka*-teokseen perustuvasta muutoksen analyysistä, joka oli suunniteltu nimenomaan välttämään ristiriitaisia kuvauksia ja oletuksia asioista. Caracciolon aikalaiset ja samaa teoriaa myöhemmin arvostelleet kriitikot toistivat aristoteelista näkemystä, jonka mukaan muutokset maailmassa voidaan jakaa instantiaalisiin muutoksiin, joihin liittyy muutosta määrittävien asiantilojen kontradiktorinen vastakkaisuus, ja ajallisiin muutosprosesseihin, joissa alkamis- ja loppumishetken määrittäminen sisältää viittauksen kontradiktorisiin kuvauksiin. Muutoshetki on jakamaton ajan jakokohta eli instanssi, jossa itsessään mikään ei voi liikkua tai muuttua. Muutoshetken teoreettinen määrittäminen tapahtuu siksi samaistamalla se edeltävän tilan aktuaalisuuden viimeiseen hetkeen tai myöhemmän, edelliseen nähden vastakkaisen tilan aktuaalisuuden ensimmäiseen hetkeen.

Vastaus ei tyydyttänyt Caraccioloa, koska hän halusi tietää, koska itse muutos tapahtuu, ei sitä edeltävän tai sitä seuraavan tilan aikaa. Näyttää siksi siltä, että jos muutoksia vastakkaisten asiantilojen välillä ollenkaan on, niiden kuvauksessa täytyy viitata kontradiktoristen lauseiden totuuteen samassa ajan hetkessä. Muutoshetken täytyy olla hetki, jossa itse muutos tapahtuu. Ajallinen fysikaalinen maailma on siten ristiriitainen. Logiikassa ei tarvitse silti olettaa ristiriidan lain kumoutumista, koska se koskee ajatonta käsitteellistä järjestystä. Luonnon ristiriidan synnyttämät asiantilat ajan instansseissa voidaan käsittää ajattoman järjestyksen tasolla erilliseksi ja erottaa ne erillisten luonnon instanssien sisällöiksi. (Knuutila & Lehtinen 1979; Knuutila 1986.)²

Luonnon ristiriitoteoria oli osoitus tyytymättömyydestä yleiseksi paradigmatoksi kohoavan aristoteelisen fysiikan liiketeorian yksityiskohtaan, joka koski liikkeen alku- ja loppuhetken määrittämistä aikakontinuumilla. Vaikka ristiriitoteorialla ei ollut monia kannattajia, se säilyi yhtenä linjana muutosta koskevassa keskustelussa. Tosiasiassa kuitenkin suositumpia

extrinsecum” (Petrus de Lorca, *Commentaria et disputationes in Secundam Secundae Divi Thomae*, sect. 3, disp. 22, appendix, n. 31).

² Kretzmann ja hänen seuraajansa viittaavat kontradiktoriseen muutosteoriaan termillä ’Quasi-Aristotelianism’, mikä on harhaanjohtavaa, koska sen kannattajat eivät välttämättä esittäneet, että teoria sisältyisi Aristoteleen fysiikkaan (ks. Kretzmann 1982).

keskustelunaiheita olivat ongelmat, joita aristoteelisen fysiikan keskeinen kontinuumioletus (aika, paikka, jatkuvien kvaliteettien asteet) aiheutti muutoshetken määrittämiselle. Näiden ajateltiin myös olevan Aristoteleen teorian sisäisiä kysymyksiä tai ainakin niihin läheisesti liittyviä, toisin kuin ajatus luonnon ristiriitaisuudesta jakamattomassa ajan hetkessä.

Käsittelen seuraavaksi näitä aiheita tarkastelemalla ensiksi Michael Dummettin esittämää kritiikkiä matemaattisen kontinuumiteorian fysikaalista soveltamista kohtaan artikkelissa ”Is Time a Continuum of Instants” (2000). Kun keskiajan näkemykset muutoshetken määrittämisestä perustuivat kontinuumiteoriaan, tarkoitukseni on selvittää, miten niissä vastataan Dummettin esiin ottamiin kysymyksiin eli mitä vastauksia pidettiin vakuuttavina keskiaikaisessa kontekstissa. Palaan lopuksi muutoksen ristiriitasteoriaan ja vertaan sitä von Wrightin ja Priestin näkemyksiin muutoksen ristiriitaisuudesta.

1. Dummettin käsitys klassisen kontinuumiteorian fysikaalisesta epäasianmukaisuudesta

Dummett aloittaa artikkelinsa kirjoittamalla: ”Oletamme yleensä kaikkien fysikaalisten kvantiteettien määräytyneisyyden eli että kaikilla fysikaalisilla kvantiteeteilla on todellisuudessa absoluuttisesti määrätty koko, jonka ilmaisee jonkin yksikön mukainen reaalityluku.” Hän esittää sitten, että koska emme tosiasiallisesti tiedä tarkasti minkään oletetun kontinuumin suuren kokoa, määräytyneisyysperiaatti edustaa superrealistista metafysiikkaa olettamalla täysin määräytyneitä asiantiloja, jotka vallitsevat, vaikka emme voi edes teoreettisesti tuntea niitä. Dummett ajattelee, että tämä ajatusharha näkyy erityisesti siinä, miten määräytyneisyysperiaatteeseen tukeutuvassa klassisessa teoriassa käsitellään ajallisia rajoja ja mittausta. Klassisella teorialla Dummett tarkoittaa klassisen matematiikan ja fysiikan mukaista ajattelutapaa. Esimerkiksi kestoja mitataan jollakin mittayksiköllä, kuten sekunti, ja jokaisella aikavälillä on täsmällinen kesto, joka ilmaistaan sekunneissa käyttämällä rationaalisia tai irrationaalisia reaalitylukuja. Samaa mallia sovelletaan analogisesti kaikkiin fysikaalisiin suureisiin. Dummett esittää useita ongelmia, joita hänen mukaansa seuraa ajan käsittämisestä niin, että mallina on matemaattinen reaalitylukujen kontinuumi. Ajan ajatellaan silloin koostuvan kestottomista instansseista, jotka ovat järjestyneet tiheäksi lineaariseksi

järjestykseksi, jossa instanssilla ei ole välittömästi seuraavaa instanssia ja instanssien välillä ei ole aukkoja eikä instantiaalista liikettä. Universumin historia jollakin aikavälillä koostuu instantiaalisista fysikaalisista tilakuvauksista, joissa suuruudet on määrätty mittayksiköiden avulla. Humeen viitaten Dummett kutsuu näitä kuvauksia toisistaan loogisesti riippumattomiksi. Niiden välisiä yhteyksiä ilmaistaan kontingenteilla fysikaalisilla laeilla.

Artikkelinsa ensimmäisessä osassa Dummett pyrkii osoittamaan ”klassisen matemaatikon hyväksymillä perusteilla”, että klassinen kontinuumiteoria ei sovellu fysikaaliseen aikaan. Hän tarkastelee kolmentyyppisiä epäjatkuvia matemaattisia funktioita ja niiden yhteyttä erilaisiin fysikaalisiin epäjatkuvuuksiin. Dummettin ensimmäinen esimerkki on intuitiivinen kuva valaistuksen intensiteetin instantiaalisesta siirtymisestä jostakin positiivisesta arvosta nolnaan, jossa se sitten pysyy jonkin aikaa. Tätä voisi kuvata klassisen teorian mukaan funktiolla, joka ilmaisee valaistuksen intensiteetin missä tahansa ajan pisteessä tarkastellulla aikavälillä. Funktiolla on muutoksen kohdalla hyppäysepäjatkuvuus, jonka voi ilmaista kahdella tavalla: pinnalla on positiivinen valaistus jokaisella hetkellä ennen hetkeä t_1 , mutta ei hetkellä t_1 eikä sen jälkeen, tai sillä on positiivinen valaistus myös tällä hetkellä, mutta ei sen jälkeen. Dummett huomauttaa, että tosiasiaassa fysiikassa ei ole kahta tällaista muutokseen liittyvää mahdollisuutta, sillä ei ole mitään fysikaalista tekijää, joka määrittäisi toisen tai toisen näistä, eikä siksi mitään perusteltavaa eroa klassisen teorian erottamien asioiden välillä. Tämä epämääräisyys on seuraus ajan kontinuumiteorian epäasianmukaisuudesta.

Esimerkki toisen tyyppisestä yksinkertaisesta epäjatkuvuudesta on, että jokin taso on määrätyn värinen tai kokoinen jollakin aikavälillä lukuun ottamatta yhtä kestotonta instanssia, jolloin sillä on momentaanisesti toinen väri tai koko. Dummettin mukaan on järjetöntä ajatella tätä fysikaaliseksi tapahtumaksi, vaikka ajatus poistuvasta epäjatkuvuudesta reaalfunktiossa ei sisällä matemaattista ongelmaa. Tällaisissa tapauksissa klassinen teoria kuitenkin tarjoaa kontinuumiteorian johdosta tilaa järjettömille ajatusmalleille, mitä parempi teoria ei tekisi.

Esimerkki kolmannelta epäjatkuvuudesta on kappaleen kiihtyvä oskillaatio kahden position välillä jollakin aikavälillä niin, että sen muutostiheys kasvaa kohti äärettömyyttä. Se vaihtaa silloin suuntaa äärettömästi

ja sen asema on epäjatkuva kuvattavan sarjan raja-arvon kohdalla. Taas on mahdollista määritellä vastaava reaalitylukujen funktio, mutta oskillaatioprosessilla ei ole luonnollista päämäärää ja sen lopputila on edeltäviin tiloihin nähden määräämätön.

Johtopäätöksensä Dummett esittää, että parempi fysikaalinen taustateoria olisi sellainen, joka ei tarjoaisi maaperää kuvitteellisille fysikaalisille epäjatkuvuuksille ja niiden oletetuille vaihtoehdoille kuvauksille. Toisessa osassa hän selittää, miten tilannetta parantaa malli, jossa määritetään kaikille käsiteltäville fysikaalisille suureille tarkkuusalueet, joiden sisään mitattavat suureet sijoittuvat. Tätä voi pitää sumeana realismina tai konstruktivismina, jossa aika ei koostu kestottomista instansseista, vaan erittäin pienistä aikaväleistä. Samoin muut fysikaaliset suureet koostuisivat mittaustarkkuuden rajaamista perusosista, vaikka itse aikaa pidettäisiin kontinuumina kuten klassisessa mallissa. (Dummett 2000 ja 2012.)³

2. Keskiajan teorit alkamis- ja loppumishetkestä

Ensimmäistä Dummettin kriittisistä esimerkeistä voi pitää keskiaikaisen muutoksen raja -keskustelun erityistapauksena. Keskiajan loogisessa semantiikassa kehitettiin 1200-luvun alkupuolelta lähtien analyysisääntöjä termeille *incipit* (alkaa) ja *desinit* (loppuu), ja luonnonfilosofiassa ne liitettiin Aristoteleen fysiikkaan sisältyvään teoriaan muutoksen alkamis- ja loppumishetken määräytymisestä aikakontinuumilla. Tältä pohjalta muodostettiin malleja, joiden tarkoituksena oli selittää aikarajojen mikrostruktuuria muuttuvassa maailmassa. 13. vuosisadalla tuli tavalliseksi esittää erillinen analyysi pysyvien (*res permanens*) ja ajallisesti muuttuvien ja sukessiivisten (*res successiva*) asioiden ajallisista rajoista. Terminologia esiintyy Averroesin latinaksi käännettyissä teoksissa, mutta sen vakiintuminen muutosteorian yhteydessä on silti alkuperältään epäselvä. Pysyvät asiat kuten substanssit olivat jakamattomia kokonaisuuksia, joilla ei ollut eri osia eri aikoina, kun taas sukessiiviset asiat kuten kappaleen liike ja muut ajalliset prosessit olivat sellaisia, etteivät ne voineet olla kokonaisuudessaan aktuaalisia yhdellä hetkellä.

³ Dummettin näkemyksiä arvostelee kriittisesti U. Meyer (2005).

Yleistä ajatusta jatkuvien liikkeiden rajojen eksternaalisuudesta eli ulkoisuudesta tuki Aristoteleen ajatus, ettei ajan instanssissa voi olla liikettä; instanssi oli kestoton menneen ja tulevan jakokohta siten kuin piste geometrisella suoralla. Tämän mukaisesti oli tavallista ajatella, että on olemassa viimeinen hetki ennen liikettä, jolloin on vielä totta sanoa, että kappale on jossakin liikettä edeltävässä tilassa, ja ensimmäinen hetki liikkeen jälkeen, jolloin on totta sanoa, että kappale on liikkeen kautta saavuttamassaan liikkeen jälkeisessä tilassa. Nämä eksternaaliset instanssit muodostavat jatkuvan liikkeen alku- ja loppurajat. Ne ovat liikkeen kestoa ulkoisesti rajaavia instansseja, joissa liikkeen alkutila (*terminus a quo*) ja lopputila (*terminus ad quem*) ovat aktuaaliset. Liikkeen kestolla ei näiden ohella voisi olla sisäisiä (internaalisia) rajoja senkään takia, koska kontinuumilla ei ole vierekkäisiä pisteitä. Alku- ja loppuhetki eivät siten ole hetkiä, joissa jokin muuttuu tai alkaa tai loppuu. Aristoteelisessa teoriassa asiat alkavat tai loppuvat vain jonkin aikavälin sisällä, eivät yhdessä hetkessä. Eksternaaliset alkamis- ja loppumishetket ovat kuitenkin kiinnekohtia, joihin liikkeen alku ja loppu viittaavat.

Pysyvistä olioista ajateltiin yleisesti, että niillä on olemassaolon ensimmäinen sisäinen alkamishetki. Sen sijaan loppumishetken osalta oli erilaisia mielipiteitä ennen kuin aristoteelinen näkemys eksternaalisesta loppuhetkestä vakiintui. Keski aikaista teoriaa voi pitää pääpiirteiltään aristoteelisena, vaikka esimerkiksi terminologista erottelua pysyvien ja sukseksiivisten asioiden välillä ei sellaisenaan löydy Aristoteleen teoksista.⁴

Dummettin mukaan klassinen malli tarjoaa keskenään poissulkevia mahdollisuuksia asioiden alkamis- ja loppumishetken määrittelylle. Hänen mukaansa tämä on osoitus kontinuumiteorian epäasianmukaisuudesta, koska luonnossa ei ole fysikaalisia perusteita valita näiden vaihtoehtojen välillä. Loogista ja fysikaalista analyysia yhdistävässä keskiaikaisessa mallissa ei ajateltu näin. Sen edustajat katsoivat, että kaikissa instantiaalisen muutoksen tapauksissa oli määrätty instanssi aikaisemman tai myöhemmän asiantilan aktuaalisuudelle ja myös perusteita sen valinnalle. Heidän mukaansa jatkuvan ja siis äärettömästi jaettavan liikkeen rajojen eksternaalisuus perustui siihen, ettei instanssissa voi olla liikettä tai että liikkuva liikkuu, on liikkunut ja tulee

⁴ Keski aikaisista muutoksen rajanmäärityskeskusteluista ks. Kretzmann 1976; Murdoch & Sylla 1978; Murdoch 1979; Murdoch 1982; Knuutila 1986; de Libera 1989; Strobach 1998.

liikkumaan. (Ks. Aristoteles, *Fysiikka* VI.3, 234a24-b9.) Keskiajan ajattelijat eivät siten pitäneet ongelmallisena matemaattisen kontinuumiteorian soveltamista tässä yhteydessä. Heidän mukaansa asioiden luonto osoitti, miten fysikaaliset rajat ja vastaavat alkamisen ja loppumisen teoreettiset käsitteet tuli ymmärtää.

1200-luvun puolivälin käsitys pysyvien asioiden rajoista oli, että ne olivat sisäisiä. Koska pysyvät asiat voivat olla kokonaan aktuaalisia jakamattomassa kestottomassa instanssissa, pidettiin järkevänä ajatella, että niiden olemassaolon aikarajoina olivat aktuaalisuutta mittaavan ajan sisäiset ensimmäiset ja viimeiset instanssit.⁵ Pidettiin luonnonfilosofian ilmeisenä totuutena, että uusi pysyvä olio alkaa olla silloin, kun se ensimmäistä kertaa on. Tämä käsitys oli sopusoinnussa Aristoteleen *Fysiikan* kanssa. Kun tieto Aristoteleen teorian yksityiskohdista vakiintui, pysyvän asian loppumishetken määrittäminen muuttui niin, että sitä ruvettiin pitää ulkoisena.

Pysyvän asian alkamis- ja loppumishetken määrittelyn keskeisenä perusteena alettiin pitää Aristoteleen melko vaikeaselkoista argumenttia *Fysiikka*-teoksen kohdassa VIII.8, 263b9-264a6. Tämän argumentin alkulausetta lainattiin usein perusteena kaikkien muutosten loppuhetken ulkoisuudelle. Aristoteles kirjoittaa: ”On selvää, että ellemme katso sen ajan pisteen, joka erottaa aikaisemman ja myöhemmän, kuuluvan aina vain myöhemmän asian aikaan, jokin sama voisi samanaikaisesti olla ja olla olematta, eikä se voisi olla silloin, kun se on syntynyt. Piste on yhteinen molemmille ajoille, aikaisemmalle ja myöhemmälle, ja se on lukumäärältään yksi ja sama, mutta se ei ole tätä määritelmän mukaan, sillä se on toisen alku ja toisen loppu, mutta sikäli kuin tarkoitetaan asiaa, se liittyy aina myöhempään määreeseen.” Ajateltiin, että aktuaalisuus olemassaolon ensimmäisessä instanssissa koski kaikkia pysyviä asioita ja että niiden häviämistä selitti toisen pysyvän asian syntyminen, jonka olemassaolon ensimmäinen hetki sulki pois edellisen olemassaolon viimeisen instanssin – kontinuumillahan ei ole peräkkäisiä instansseja. Loppumishetket ovat siten kaikkien asioiden suhteen ulkoisia. Pysyvien asioiden yhteydessä voitiin ajatella, että ulkoista hetkeä määrittää myös uuden asian syntymistä edeltävän jatkuvan muutosprosessin viimeinen ulkoinen raja. Aristoteles piti empiirisenä tosiasiana, että asioiden substanssien syntymistä edeltää prosessuaalinen

⁵ Katso tekstejä teoksessa Braakhuis 1979; de Libera 1989; Tabarroni 1989.

muutos toisissa kategorioissa. Esimerkiksi puun muuttumista tuleksi edeltää asteittainen kuumeneminen. Aristoteles ajatteli yleensä (mutta ei aina), että substansseilla ei ole omassa olemisessaan asteita. Keskiajalla ajateltiin vastaavasti, että substanssien syntyminen ja häviäminen samoin kuin muiden pysyvien asioiden muutokset voitiin käsittää niin, että niiden syntymistä mahdollistivat jatkuvat muutokset niissä tekijöissä, jotka kuuluivat pysyvien asioiden olemassaolon välttämättömiin ehtoihin.⁶

Monet kirjoittajat alkaen 1200-luvun puolivälin aristoteelikoista (esimerkiksi Albert Suuri ja Tuomas Akvinolainen) 1300-luvun alun luonnonfilosofeihin (esimerkiksi Thomas Wylton ja Walter Burley) viittasivat juuri tällaiseen perusteluun sille, että pysyvien asioiden muutoshetki määritetään uuden tilan ensimmäiseksi hetkeksi. Se on siten myös olion häviämisen hetki eli sen olemassaolon eksternaalinen loppuraja. Muutoksen yksityiskohtien teoreettinen selvitys tuotti kuitenkin päänvaivaa.⁷ Aegidius Romanus kommentoi Aristoteleen näkemystä *Fysiikan* luvussa VIII.8 toisin, vaikka hän muualla viittaa edellä mainittuun pysyvien ja sukseksiivisten asioiden yhdistämiseen.⁸ Kommentoidessaan *Fysiikan* kohta VIII.8 hän käsittelee muutoshetken määräytymistä yleisemmin muutoksena asian olemisen ja ei-olemisen välillä ja toteaa, että Aristoteleen sääntö suosia aina jälkimmäisen tilan ensimmäistä sisäistä hetkeä on mielivaltainen, koska voitaisiin myös ajatella, että muutoshetkenä pidettäisiin häviävän asian aktuaalisen viimeistä sisäistä instanssia. Hän esittää, ettei ole mitään varsinaista todistusta sille, miten muutoshetki määräytyy, mutta voidaan seurata Averroesin retorista argumenttia, että on asioiden arvon mukaista,

⁶ *Fysiikka* VI.4, 235a35-6; VI.5, 236b2-18; VII.3, 246a6-9; VIII.7, 260b11-15; Knuutila 1986, 247–251.

⁷ Albert Suuri, *Liber de sensu et sensato*, III.2; Tuomas Akvinolainen, *In VIII libros Physicorum Aristotelis Expositio*, VI.5.15; VI.8.9, 14; Thomas Wylton, *Quaestiones libri Physicorum*, VIII.6 (ks. Trifogli 1994, 114–115); Walter Burley, *De primo et ultimo instanti* (ks. Shapiro & Shapiro 1965, 171). Wylton ja Burley ajattelivat, että substanssin tai pysyvän kvalitatiivisen muodon syntymistä edeltävällä jatkuvalla muutoksella ei ole viimeistä sisäistä aikainstanssia, mutta he esittivät myös, että jos edeltävällä muuttuvalla kvaliteetilla on fyysikaalinen maksimi- tai minimiarvo, jonka se saavuttaa, silloin sen muutoksella on sisäinen loppumisinstanssi. Wyltonin ja Burleyn yrityksistä yhdistää muutoksen raja-analyysi ja teoria muuttuvien määreiden asteista ja vaihteluväleistä, ks. Trifogli 1994 ja Sylla 2001.

⁸ Ks. esim. *Theoremata de esse et existentia*, theor. 7 (M. Murrayn käännös teoksessa *Theorems on Existence and Essence*, 41.6).

että preferoidaan pikemmin olemisen kuin ei-olemisen ensimmäistä sisäistä hetkeä.⁹

Aristoteleen esimerkit havaittavista instantiaalisista muutoksista olivat valon aiheuttaman valaistuksen syntyminen (*De anima* II.7, 418b9-26) ja veden äkillinen jäätyminen (*De sensu* 6, 447a1-3; *Fysiikka* VIII, 253b23-8). Hän käsitti molemmat tapaukset luonnollisten potentiaalisuuksien aktualisoitumisiksi, jotka tapahtuvat instantiaalisesti, kun asianmukainen aktivoija on paikalla. Aristoteles ja hänen keskiaikaiset seuraajansa olettivat, että nämä tapaukset voidaan myös käsittää viittaamalla jatkuvien liikkeiden myötävaikutukseen (Sorabji 1983, 410–411). Vaikka keskiajan ajattelijat seurasivat Aristotelesta olettaen ajan, paikan ja kvantiteetin kontinuumiteorian mukaisesti jatkuviksi ja pitivät siten geometrista kontinuumianalyysia keskeisenä teoreettisena mallina instantiaalisen muutoksen selittämisessä, he käsitelivät tässä yhteydessä varsinkin Averroesin fysiikan vaikutuksesta myös ei-jatkuvia kvalitatiivisia muutoksia sekä kvaliteettien intensiteetin vaihteluväliä ja sen maksimi- ja minimiarvoja (ks. Sylla 2015).

Dummettin toinen esimerkki instantiaalisesta poistuvasta epä-jatkuvuudesta muistuttaa jossakin määrin keskusteluja instantiaalisesta muutoksesta ei-kestollisen asian alkamisena ja loppumisena. Esimerkkinä tästä oli perisyntin yliluonnollinen häviäminen Mariasta tämän olemassaolon ensimmäisellä hetkellä. Kysymys Marian perisyntinälaisuudesta oli suosittu keskustelunaihe 1200-luvun lopun teologiassa. Jotkut halusivat tehdä Marian perisyntitilan niin lyhyeksi kuin mahdollista ja esittivät, että se koski hänen olemassaoloaan vain sen ensimmäisessä ei-kestollisessa hetkessä silloin kun hän sai substantiaalisen sielun. Sielulla pysyvänä entiteettinä ajateltiin aristoteelisen teorian mukaisesti olevan olemassaolon ensimmäinen sisäinen instanssi. Tällä hetkellä Maria myös vapautui armon vaikutuksesta perisyntistä. Henri Ghentiläinen, vaikutusvaltainen Pariisin yliopiston

⁹ *Commentaria in VIII libros Physicorum Aristotelis*, VIII, l. 22 (212ra); Trifogli 1993, 100–101. Aegidius esittää, että kun pysyvien asioiden kestoä mitataan eksternaalisella jatkuvalla ajalla, joka perustuu taivaankappaleiden liikkeeseen, fysikaalisten elementtien muutosta voi analysoida siten kuin Aristoteles tekee kohdassa *Fysiikka* VIII.8, mutta pysyviä asioita mittaa sisäisesti myös niiden oma aktuaalisuus niiden yksimuotoisen olemassaolon kullakin hetkellä. Sisäinen mitta ei ole jatkuva vaan diskreetti, ja toisensa poissulkevat muodot, joita mittaa niiden sisäinen aika, voivat olla aktuaalisia samassa ulkoisen ajan instanssissa (*Quodlibeta*, VI.7; Trifogli 1993, 101–106). Samoin vastakohtaiset kvaliteetit, joita mittaa enkeleitä koskeva diskreetti aika, voivat olla aktuaalisia samassa fysikaalisen ajan instanssissa (ks. *De mensura angelorum*, q. 8-10; Porro 2001, 152–157).

teologian professori, oli tämän käsityksen keskeinen teologinen muotoilija. Hän kuitenkin pian korjasi käsitystään, koska sen katsottiin sisältävän ristiriidan. Godefridus de Fontibus kuvaa Henrin ensimmäistä käsitystä siten, että ristiriitaiset lauseet, joiden mukaan Maria oli perisyntyn alainen ja ei ollut sitä, olivat tosia samassa ajan instanssissa, mutta eivät samassa luonnon instanssissa. Luonnon instanssit viittaavat lauseiden eroon ei-temporaalisella käsitteellisellä tasolla. Näyttää siltä, että Henryn varhainen teoria vaikutti myöhempään teoriaan muutoksen risti-riitaisuudesta, joka tosin oli filosofisemmin painottunut kuin Henrin mariologinen näkemys (Brower-Toland 2002; Dumont 1992). Aegidius Romanus kuvaa tätä samaan aikaan tunnettuna teoriana. Vaikka hän ei itse kannattanut sitä, hän oletti, että diskreetissä enkeli ajassa eri hetkiin sijoittuvat ristiriitaiset lauseet voivat olla tosia samassa fysikaalisen ajan hetkessä (*Quodlibeta* VI.20).¹⁰

Muutettuaan ensimmäistä käsitystään Marian perisyntisyydestä Henri puolusti käsitystä, että Maria oli perisyntyn alainen vain olemassaolonsa ensimmäisellä hetkellä ja sen jälkeen armon vaikutuksesta ilman sitä. Hänen sielunsa perisyntisyys oli kestoton instantiaalinen tila. Tämä käsitys johti Henrin miettimään ei-kestollisen tilan alkamisen ja loppumisen määrittelyä ja näiden määritysten suhdetta tätä tilaa seuranneen armon tilan alkuun. Sillähän ei voi olla sisäistä ensimmäistä instanssia, jos instantiaalisella perisyntynillä oli sama sisäinen alkamisinstanssi kuin olemassaololla. Siksi Henri esittää Aristoteleen teoriasta poiketen, että armon tilan alkua määrittää ulkoinen instanssi, jolloin sitä ei vielä ollut ja jonka jälkeen se oli. Instantiaalista tilaa seuraavalla pysyvällä tilalla ei voi olla sisäistä ensimmäistä instanssia, koska instanssit eivät voi olla välittömästi peräkkäisiä kontinuumissa ajassa (*Quodlibet* XV, q. 13). Instantiaalisen tilan sisäiset alkumis- ja loppumisrajat lankeavat yhteen samoin kuin sitä välittömästi korvaavan pysyvän asian ulkoinen alkamisraja.

Nämä olivat kaikki epäaristoteelisia näkemyksiä Henryn teoriassa. Aristoteles ei hyväksynyt pelkästään instantiaalisia kvaliteetteja. Hänen mukaansa kaikkia muuttuvan maailman asioita mittaa aika ja pysyvillä asioilla on olemassaolon ensimmäinen sisäinen instanssi. Koska oliot menettävät niiden vaihtuvat muodot siten, että ne korvautuvat uusilla, edellisten kanssa yhteen sopimattomilla muodoilla, edellisten muodon loppumishetken määrää

¹⁰ Diskreetistä ajasta ks. edellinen nootti.

uuden muodon ensimmäinen sisäinen instanssi. Kappaleen yksittäisen liikkeen vastakkainen liike ei voi olla edelliseen nähden jatkuva, koska edellisen liikkeen loppua määrittää sen tulos, jolla on sisäinen alkamisinstanssi ja joka on siten kestollinen. (*Fysiikka* VIII.8, 262a12-263a3.)

Henri kyseenalaisti tämän tunnetun aristoteelisen käsityksen viittaamalla esimerkkiin, jota monet olivat käsitelleet myös ennen häntä: koska ylöspäin heitetty papu ei pysäytä sen päälle putoavaa myllynkiveä, vaikka se osuisi papuun juuri sen lentoradan käännekohdassa, jatkuva vastakkainen liike on mahdollinen samoin kuin fysikaalisen tilan instantiaalinen aktualisuus.¹¹ Henrin ajatuksena oli, että vaikka pysyvän olion käsite (*res permanens*) yleensä sisältää ajatuksen, että tällainen olio on kestollinen, se tarkoittaa pohjimmiltaan mahdollisuutta olla aktualinen yhdessä hetkessä ilman ajallista peräkkäisyyttä. Siksi pysyvät oliot voi jakaa kestollisiin ja instantiaalisiin (ks. Brower-Toland 2002, 39–41).

Koska Henri ajatteli, että Marian synnittömyys on supranaturaalisesti aiheutunut tila hänen olemassaolonsa ensimmäisen instanssin jälkeen, tavallinen aristoteelinen ajatus jatkuvan tilan ensimmäisen instanssin yhteen lankeavuudesta sen syntyä selittävän edeltävän temporaalisen muutoksen ulkoisen loppumishetken kanssa oli hylättävä ainoana mahdollisuutena. Hänen mariologisessa esimerkissään pysyvällä armontilalla ei ole sisäistä alkamishetkeä. Samoin instantiaalisten pysyvien tilojen alkamis- ja loppumisrajat ovat sisäiset. Thomas Wylton ja Walter Burley käsitelivät alkamis- ja loppumishetken teorioissaan samoja kysymyksiä instantiaalisten tilojen rajoista ja niiden suhteesta niitä edeltävien ja seuraavien tilojen rajoihin (Brower-Toland 2002, 45–46).

Henrin mariologinen esimerkki muistuttaa Dummettin käsittelemää poistuvan epäjatkuvuuden tapausta. Dummettin mielestä fysikaalisten poistuvien epäjatkuvuuksien postulointi osoittaa matemaattisen kontinuumin huonoa vaikutusta fysikaaliseen ajatteluun. Se tarjoaa käsitteellisiä välineitä fysikaalisesti mahdottomien ja järjettömien tapausten tarkasteluun. Henri ei olisi pitänyt tätä vakuuttavana huomautuksena, sillä hän seurasi 1200-luvun teologioiden käsitystä, jonka mukaan fysikaaliset mahdottomuudet eivät rajoittaneet Jumalan kaikkivalttiutta, joka oli määritelty suhteessa käsitteelliseen ristiriidattomuuteen (Knuutila 1993, 100–106). Vaikka

¹¹ Esimerkin aikaisemmasta historiasta keskiajalla ks. Grant 1974, 284–287.

instantiaalinen armontilan syntyminen oli Jumalan kaikkivoimaisuuden synnyttämä yliluonnollinen tapahtuma, Henri yritti lisätä analyysinsä vakuuttavuutta arvioimalla kriittisesti myös oman aikansa fysiikan oletuksia.

Jumalan kaikkivaltius oli usein toistuva käsite 1300-luvun ajatteliijoilla, jotka perustivat liikettä ja muutosta koskevat näkemyksensä olennaisesti matemaattis-loogisille teorioille. Heidän tutkimusotettaan on luonnehdittu metalingvistiseksi luonnonfilosofisten väitelauseiden propositionaaliseksi analyysiksi. Käsiteltävät ongelmat olivat usein keinotekoisia tapauksia, jotka oli muodostettu *secundum imaginationem* ja edustivat siten loogis-metafyysisiä mahdollisuuksia tai Jumalan mahdollisuuksia. Ratkaisut tavoittelivat yleistä sitovuutta matematiikan ja logiikan periaatteiden valossa. (Murdoch 1979 ja 1982.) Parhaiten tunnettuja esimerkkejä uutta matemaattista fysiikkaa kehittävästä liikkeestä ovat Oxfordin kalkulaattoreiksi sanottujen englantilaisten ajattelijoiden tutkimukset, jotka vaikuttivat myös uuden ajan fysiikkaan. Valaisen seuraavassa tätä lähestymistapaa selvittämällä, miten Roger Rosetus käsitteli liikettä noin vuonna 1335 ja tässä yhteydessä esimerkkiä, johon liittyy samantapaisia ajatuksia kuin Dummettin esimerkkiin kolmannesta epäjatkuvuudesta.

Rosetus tarkastelee kuvitteellisessa tapauksessa enkeliä, joka lentää suoraan pisteestä toiseen siten, että ensimmäinen osa on puolet koko matkasta, toinen on puolet siitä, ja niin edelleen *ad infinitum*. Kun enkeli on lentänyt yhden osan yli, se pysähtyy sen päätepisteeseen yhtä pitkäksi ajaksi kuin kului tämän osan ylittämiseen. Sen jälkeen se jatkaa lentämällä seuraavan osan yli ja pysähtyy taas sen ylittämiseen kuluneen ajan verran jne. Lisäksi oletetaan, että se liikkueessaan liikkuu samalla nopeudella.¹² Rosetus toteaa ensiksi, että enkeli ei tällä tavoin pääse etenemään kovin pitkälle, koska proportionaaliset osat tulevat pian niin pieniksi, että enkelin kognitiiviset kyvyt eivät riitä niiden erottamiseen eikä se enää tiedä, milloin pitäisi lentää ja milloin olla liikkumatta. Hän kysyy sitten, voisiko Jumala auttaa enkeliä pääsemään tällä tavalla matkan loppuun. Rosetus olettaa, että Jumalan

¹² "Sed probo quod in nullo inanimato sunt infinitae partes proportionales. Nam ponatur quod hic sit aliquod continuum inanimatum, et volo quod aliquis angelus moveatur super illum continuum, ita quod in fine cuiuslibet partis proportionalis quiescat totaliter et per tantum tempus per quam <movebatur> super illum continuum, et quod moveatur semper aequae velociter quandocumque <movebatur>. Tunc quaero: aut sic movendo et quiescendo deveniet ad finem istius continui, aut non." (Roger Rosetus, *Lectura super Sententias 3-5*, 4.2 (138-139).)

ääretön tietokyky riittää siihen, että se tuntee aktuaalisesti äärettömiä joukkoja, esimerkiksi äärettömästi jaetun janan proportionaaliset osat. Duns Scotus oli aikaisemmin esittänyt näin (ks. esim. *De primo principio* IV.9, n. 68). Toisaalta äärettömästi jaettavan suureen jakaminen on päättymätön tehtävä. Siksi Rosetus vastaa, että yliluonnollisen avun turvin enkeli pystyisi loputtamasti erottamaan osan toisensa jälkeen, mutta se ei tulisi loppuun tällä tavoin, koska osien sarja on päättymätön. Toisaalta liikkeen tasaisen nopeuden perusteella tarvittava aika olisi kaksi kertaa se, mikä kuluisi matkaan ilman pysähdyksiä. (*Lectura* 4.2 (145–150); ks. myös Hallamaa 1998, 387–388.)

Rosetus kysyy, voisimmeko tietää, onko enkeli liikkeessä vai lepotilassa prosessin viimeisessä osassa ennen päätepisteeseen saapumista. Molemmat vaihtoehdot ovat ongelmallisia. Rosetuksen vastaus on, että koska esimerkissä on kysymys äärettömästä suppenevasta sarjasta, siinä ei ole viimeistä osaa ennen loppua eikä siksi ole Jumalallekaan mahdollista tietää, että ollaan tällaisessa kohdassa. Hänellä on myös muita samantapaisia esimerkkejä, joista seuraava sisältää ehkä humoristiseksi tarkoitetun pelastusopillisen sovellutuksen. Oletetaan, että Jumala hyväksyy ja hylkää liikkujan S joissakin kohdissa sen kulkiessa tasaisella nopeudella matkan A:sta B:hen, joka on jaettu proportionaalisiin osiin kuten edellä. Ensimmäisen osan ylittämisen aikana Jumala hyväksyy S:n, toisen osan aikana ei, kolmannen osan aikana hyväksyy, neljännen ei jne. Hyväksyykö Jumala lopulta S:n, kun se on saavuttanut päätepisteen? Rosetuksen mukaan saattaa hyväksyä tai olla hyväksymättä. Kumpikaan vaihtoehto ei perustu siihen, mitä prosessi aikana tapahtuu – kysymyksen ei-ratkaistavuus perustuu proportionaalisesti jaetun matkan oletettuun äärettömään jaettavuuteen.¹³

Rosetus ei esimerkissään sano, mikä hyväksynnän kohde on, mutta esimerkki on peräisin William Crathornin hiukan aikaisemmin laaditusta *Sentenssikommentaarista*, jossa Crathorn kritisoi teoriaa äärettömästä

¹³ ”Praeterea, volo quod Deus acceptet illud movens pro prima parte proportionali et deacceptet pro secunda parte proportionali et acceptet pro tertia et deacceptet pro quarta, ita quod pro qualibet alterna acceptet et pro qualibet alterna deacceptet. Quo posito quaero tunc aut in illo instanti in quo totum spatium erit pertransitum ab eo, puta in B instanti, erit ipse Deo acceptus vel deacceptus ... de casu illo non sequitur quod erit acceptus nec deacceptus, quia in illo instanti nulla pars proportionalis illius spatii et illius ordinis est primo pertransita” (Roger Rosetus, *Lectura super Sententias*, 4.2 (146–147, 152).

jaettavuudesta ja esittää kuvatun tapauksen osoituksena kontinuumiteorian teologisesta ongelmallisuudesta. Hänen versiossaan joku tekee syntiä ja katu edellä kuvatun mallin mukaan vaihtelevalla tavalla, jolloin ongelmana olisi, ettei Jumalakaan lopussa voisi tietää, ansaitseeko henkilö pelastuksen vai rangaistuksen. Yleensä tuon ajan kirjoittajat ajattelivat, että ajan instanssit eivät ole sen atomistisia osia, mutta erilaisia atomiteorioitakin kehiteltiin. Crathorn oli ensimmäisiä keskiajan luonnonfilosoifeista, jotka esittivät ajan koostuvan kestollisista aika-atomeista. Siksi hän esitti, että edellisessä esimerkissä tulisi olettaa viimeinen osa, jolloin henkilö tulee tai ei tule hyväksytyksi. Tarkoituksena on osoittaa, että atomistinen teoria tarjoaa kontinuumiteoriaa paremman välineistön käsitellä kuvitteellista teologista esimerkkiä. (Ks. William Crathorn, *Quaestionen zum ersten Sentenzenbuch*, I, q. 3; Murdoch 2009, 33.)

Rosetus ei pitänyt mitenkään ongelmallisena sitä, että proportionaalisesti jaettavan äärettömästi suppenevan sarjan viimeinen osa ei ole Jumalan tuntema. Koska on matemaattinen tosiasia, että sellaista ei ole, puuttuva tieto siitä ei rajoita Jumalan kaikkietävyyttä kuten ei myöskään loogisen mahdottomuuden ei-toteutettavuus rajoita Jumalan kaikkivaltiutta. Crathornin idea atomismista ja tiedettävyydestä on lähempänä Dummettin huomautusta, jonka mukaan konstruktivisti, kuten hän, ”ei ajattele todellisuuden sisältävän tai Jumalan luovan mitään, josta rationaaliset luodut eivät periaatteessakaan voi päästä selville” (Dummett 2000, 515). Crathorn on kiinnostunut Jumalan tiedosta, joka koskee atomistisia rakenteita. Ne ovat ihmisille liian pieniä havaittaviksi, mutta hän näyttää olettavan, että käyttämällä järkeään ihmiset voivat saada niistä evidenttiä tietoa.¹⁴

3. Muutos ja ristiriita filosofisessa logiikassa

Keskiajan teoria alkamis- ja loppumishetken määrittelystä on herättänyt jonkin verran kiinnostusta oman aikamme filosofiassa, mutta liikkeen ristiriitateoria ei ole tunnettu, vaikka Priestin parakonsistentin logiikan teoria muutoksesta perustuu samanlaisille ajatuksille. G. H. von Wright esitti kirjassaan *Time, Change and Contradiction* (1969) hiukan toisenlaisen käsityksen reaalisesta ristiriidasta luonnossa. Hänen mukaansa ajanjakso sisältää

¹⁴ Crathornin atomismista ks. Robert 2009.

reaalisen ristiriidan, jos jokin osa mistä tahansa sen jaosta, jolla pyritään eliminoimaan ristiriitaa, sisältää kontradiktorisesti vastakkaisia tiloja. Von Wright käsittelee tällaista ristiriitaa osana modaalilogista aikalogiikkaa, jossa hän kuvaa kokemuksellista aikaa muutokseen perustuvana virtana, jota jakavat ajan osiin kestottomat ja fiktiiviset aikapisteet. Ajattelutapa muistuttaa Aristoteleen aikakäsitystä, mutta hän korostaa, että jos tarkasteltavaan ajan osaan sisältyy jatkuva liike, siihen sisältyy ristiriita, joka on reaalinen, vaikkakaan ei looginen ristiriita. Koska muutos on vastakkaisten tilojen välinen tapahtuma, sitä on mahdotonta eliminoida todellisuudesta jaotteleamalla sen sisältämää ajan osaa pienempiin osiin. Käsitellessään von Wrightin aikalogiikkaa Tero Tulenheimo toteaa, että ”keskiaikaisen muutoksen ristiriita-teorian näkemystä ajan instansseista voi verrata von Wrightin teorian äärellisiin jako-osiiin ja keskiaikaisia luonnon instansseja von Wrightin ajan potentiaalsiin jakopisteisiin.” (Tulenheimo 2011, 111.) Tämä vastaa von Wrightin erottelua loogisen ja reaalisen ristiriidan välillä, mutta eliminoimaton ristiriita sijoittuu von Wrightin teoriassa äärettömästi jaettavaan ajanjaksoon ja keskiajan ristiriitateoriassa taas ajan jakamattomaan instanssiin.

Ajatus instantiaalisesta ristiriidasta sisältyy dialetheistiseen para-konsistentin logiikan versioon, jossa ajatellaan, että eräissä tapauksissa väite ja sen kielto voivat olla molemmat tosia (Priest 2006). Vaikka teorian keskeiset esimerkit koskevat itsereferenssin paradokseja, siinä on käsitelty myös empiirisen maailman ristiriitoja, erityisesti muutosta. Graham Priest kirjoittaa, että kun lähdän huoneesta, olen yhdellä hetkellä huoneessa ja toisella sen ulkopuolella. Hän esittää, että kun oletetaan liike jatkuvaksi, täytyy olla ajan hetki t , jolloin jätän huoneen. ”Olenko sisällä vai ulkona hetkellä t ? Neljä vastausta on tarjolla: (a) olen sisällä, (b) olen ulkona, (c) sekä että, (d) en kumpaakaan. (a) ja (b) ovat symmetrisesti hylättäviä; valita toinen tai toinen olisi täysin mielivaltaista. Tapauksesta (d) voi sanoa, että jos en ole sisällä enkä ulkona, silloin ei ole niin, että olen sisällä ja ei niin, että ei ole niin että olen sisällä. Siksi olen sisällä ja en ole sisällä (tapaus c) tai en ole sisällä ja ei ole niin, että en ole sisällä (mikä seuraa d:stä). Molemmat vaihtoehdot ovat dialetheian mukaisia.” Perusteena ristiriidan hyväksymiselle Priest esittää, että muut vaihtoehdot kieltävät itse ilmiön eli muutoksen aktuaalisuuden (Priest & Berto 2013). Keskiajan muutoksen ristiriitateoria perustui samaan

näkemykseen, että muutoksen täytyy tapahtua instantiaalisesti propositionien p ja $\sim p$ kuvaamien asiantilojen välillä siten, että ne ovat molemmat hetkellisesti tosia. Parakonsistentissa teoriassa ristiriidan laki kielletään muutoksen analyysin yhteydessä viittaamatta keskiaikaiseen tapaan loogiseen ristiriitaan ja heikompaan luonnon ristiriitaan.

Helsingin yliopisto

Kirjallisuus

- Aegidius Romanus (1502). *Commentaria in VIII libros Physicorum Aristotelis*. Venetsia.
- Aegidius Romanus (1503). *De mensura angelorum*. Venetsia.
- Aegidius Romanus (1646). *Quodlibeta*. Louvain.
- Aegidius Romanus (1930). *Theoremata de esse et existentia*. Toim. E. Hocedez. Louvain: Museum Lessianum.
- Aegidius Romanus (1952). *Theorems on Existence and Essence*. Engl. M. Murray. Milwaukee: Marquette University Press.
- Albert Suuri (1890). *Liber De sensu et sensato*. Toim. A. Borgnet, *Opera omnia IX*. Pariisi: Vivès.
- Antonio Trombetta (1502). *In tractatum formalitatum Scoti sententia*. Venetsia.
- Aristoteles (1956). *De anima*. Toim. W. D. Ross. Oxford: Clarendon Press.
- Aristoteles (1955). *Parva naturalia*. Toim. W. D. Ross. Oxford: Clarendon Press.
- Aristoteles (1950). *Physica*. Toim. W. D. Ross. Oxford: Clarendon Press.
- Braakhuis, H. A. G. (1979). *De 13de eeuwse tractaten over syncategorematische termen I: Inleidende studie*. Meppel: Krips.
- Brower-Toland, S. (2002). "Instantaneous Change and the Physics of Sanctification: 'Quasi-Aristotelianism' in Henry of Ghent's Quodlibet XV, q. 13", *Journal of the History of Philosophy* 40, 19–46.
- de Libera, A. (1989). "La problématique de l'instant du changement au XIIIe siècle: Contribution à l'histoire les sophismata physicalia", teoksessa S. Caroti (toim.), *Studies in Medieval Natural Philosophy*. Florence: Olschki, 43–93.
- Dummett, M. (2000). "Is Time a Continuum of Instants?", *Philosophy* 75, 497–515.
- Dummett, M. (2012). "Seeing Things Move", teoksessa M. Frappier, D. H. Brown & R. DiSalle (toim.), *Analysis and Interpretation in the Exact Sciences: Essays in Honour of William Demopoulos*. Dordrecht: Springer, 21–30.
- Dumont, S. (1992). "Time, Contradiction, and Freedom of the Will in the Late Thirteenth Century", *Documenti e Studi sulla tradizione filosofica medievale* 3, 561–591.
- Grant, E. (1974). *A Sourcebook in Medieval Science*. Cambridge, MA: Harvard University Press.
- Hallamaa, O. (1998). "Continuum, Infinity and Analysis in Theology", teoksessa J. A. Aertsen & A. Speer (toim.), *Raum und Raumvorstellungen im Mittelalter*. Miscellanea Mediaevalia 25. Berliini: de Gruyter, 375–388.
- Henri Ghentiläinen (2007). *Quodlibet XV*. Toim. G. J. Etzkorn & G. Wilson. Leuven: Leuven University Press.

- Johannes Duns Scotus (1974). *Abhandlung über das erste Prinzip (De primo principio)*. Toim. W. Kluxen. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Knuutila, S. (1986). "Remarks on the Background of the Fourteenth Century Limit Decision Controversies", teoksessa M. Asztalos (toim.), *The Editing of Theological and Philosophical Texts from the Middle Ages*. Acta Universitatis Stockholmiensis. Studia Latina Stockholmiensia 30. Tukholma: Almqvist & Wiksell, 185–212.
- Knuutila, S. (1993). *Modalities in Medieval Philosophy*. Lontoo: Routledge, 100–106.
- Knuutila, S. & A. I. Lehtinen (1979). "Change and Contradiction: A Fourteenth Century Controversy", *Synthese* 40, 189–207.
- Kretzmann, N. (1976). "Incipit/desinit", teoksessa P. K. Machamer & R. G. Turnbull (toim.), *Motion and Time, Space and Matter*. Columbus: Ohio State University Press, 101–136.
- Kretzmann, N. (1982). "Continuity, Contrariety, Contradiction and Change", teoksessa N. Kretzmann (toim.), *Infinity and Continuity in Ancient and Medieval Thought*. Cornell: Cornell University Press, 270–296.
- Meyer, U. (2005). "Dummett on Time-Continuum", *Philosophy* 80, 135–140.
- Murdoch, J. E. (1979). "Propositional Analysis in Fourteenth-Century Natural Philosophy: A Case Study", *Synthese* 40, 117–146.
- Murdoch, J. E. (1982). "The Analytic Character of Late Medieval Learning: Natural Philosophy without Nature", teoksessa L. D. Roberts (toim.), *Approaches to Nature in the Middle Ages*. Medieval and Renaissance Text and Studies 16. Binghamton: SUNY Press, 183–200.
- Murdoch, J. E. (2009). "Beyond Aristotle: Indivisibles and Infinite Divisibility in the Later Middle Ages", teoksessa C. Grellard & A. Robert (toim.), *Atomism in Late Medieval Philosophy and Theology*. Medieval and Early Modern Science 9. Leiden: Brill, 15–38.
- Murdoch, J. E. & E. D. Sylla (1978). "The Science of Motion", teoksessa D. C. Lindberg (toim.), *Science in the Middle Ages*. Chicago: University of Chicago Press, 241–246.
- Petrus de Lorca (1614). *Commentaria et disputationes in Secundam Secundae Divi Thomae*. Madrid.
- Porro, P. (2001). "Angelic Measures: *Aevum* and Discrete Time", teoksessa P. Porro (toim.), *The Medieval Concept of Time: Studies on the Scholastic Debate and Its Reception in Early Modern Philosophy*. Studien und Texte zur Geistesgeschichte des Mittelalters 75. Leiden: Brill, 131–159.
- Priest, G. (2006). *Doubt Truth to Be a Liar*. Oxford: Oxford University Press.
- Priest, G. & F. Berto (2013). "Dialetheism", teoksessa Edward N. Zalta (toim.), *The Stanford Encyclopedia of Philosophy*. URL = <http://plato.stanford.edu/archives/sum2013/entries/dialetheism/>
- Robert, A. (2009). "William Crathorn's Mereotopological Atomism", teoksessa C. Grellard & A. Robert (toim.), *Atomism in Late Medieval Philosophy and Theology*. Medieval and Early Modern Science 9. Leiden: Brill, 127–162.
- Roger Rosetus (2005). *Lectura super Sententias 3-5*. Toim. O. Hallamaa. Helsinki: Luther-Agricola Society.
- Shapiro, H. & S. Shapiro, "De primo et ultimo instanti des Walter Burley", *Archiv für Geschichte der Philosophie* 47, 157–173.
- Sorabji, R. (1983). *Time Creation and the Continuum*. London: Duckworth.

- Strobach, N. (1998). *The Moment of Change: A Systematic History in the Philosophy of Space and Time*. The New Synthese Historical Library 45. Dordrecht: Kluwer Academic Publishers.
- Sylla, E. D. (2001). "Walter Burley's Physics Commentaries and the Mathematics of Alteration", *Early Science and Medicine* 6, 149–170.
- Sylla, E. D. (2015). "Averroes and Fourteenth-Century Theories of Alteration: *Minima naturalia* and the Distinction between Mathematics and Physics", teoksessa P. Bakker (toim.), *Averroes' Natural Philosophy and Its Reception in the Latin West*. Leuven: Leuven University Press, 141–192.
- Tabarroni, A. (1989). "'Incipit' and 'desinit' in a Thirteenth-Century Sophismata-collection", *Cahiers de l'Institut du Moyen-Âge Grec et Latin, Université de Copenhague* 59, 61–111.
- Thomas Wylton (1994). *Quaestiones libri Physicorum*, artikkelissa C. Trifogli, "Thomas Wylton's Question *An contingit dare ultimum rei permanentis in esse*", *Medieval Philosophy and Theology* 4, 91–141.
- Trifogli, C. (1993). "Giles of Rome on the Instant of Change", *Synthese* 96, 93–114.
- Tulenheimo, T. (2011). "Negation and Temporal Ontology", *Australasian Journal of Philosophy* 89, 101–114.
- Trifogli, C. (1994). "Thomas Wylton's Question *An contingit dare ultimum rei permanentis in esse*", *Medieval Philosophy and Theology* 4, 91–141.
- Tuomas Akvinolainen (1954). *In VIII libros Physicorum Aristotelis Expositio*. Toim. R. Spiazzi. Torino: Marietti.
- von Wright, G.H. (1969). *Time, Change and Contradiction*. Cambridge: Cambridge University Press.
- Walter Burley (1965), *De primo et ultimo instanti*, artikkelissa H. Shapiro & S. Shapiro, "De primo et ultimo instanti des Walter Burley", *Archiv für Geschichte der Philosophie* 47, 157–173.
- William Crathorn (1988). *Quaestiones zum ersten Sentenzenbuch*. Toim. E. Hoffmann. Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters. Münster: Aschendorff.

Descartes' Punctum Archimedis: The Primality and Unity of Being, the Derivativeness of the Generated Dualities¹

Joseph Almog

I would like to understand Descartes. Again. Still.

I would like to go back to Descartes' *Meditations* and understand what this book, and perhaps even his life project(s), was all about. The first and foremost point that is to be made is of “existence” and “uniqueness” (and *unity*)—there exists such a project and it is unique. So, the *Meditations* is not a passing fluke, but the tip of an iceberg, a life-long project.

After 30 years of thinking and teaching him, I can only say Descartes is the most elusive and deepest (at once) thinker I have encountered. Periodically, I feel I understand what he was after but—and this is not inconsistent—I have changed my view about his philosophical DNA every few years. I know that to the end, my end, I will struggle with his thousand faces. Every few years, I discover him afresh; discover not so much that what was said a few years ago was wrong, but that it was just another Russian doll—there is a deeper level to examine that makes better sense of the previous strata.

From the very start—my first reading of the *Meditations* as an undergraduate in the late Rogers Albritton's 300-student class—I was struck by Descartes' *unitary* impulses.²

¹ The paper is dedicated to Olli for his 60th birthday. For acknowledgments, see the last note of the paper.

² I was a first year assistant professor at UCLA and took the class as an undergraduate. I have seen—as mentioned below in the text—excerpts of the *Discours* as a teenager at my school in Paris (Sevres) but not of the *Meditations*. Albritton discussed most of the term meditation I—what he called the problem of *skepticism*—and, towards the end, the last three weeks, he reviewed the mind/body distinction preliminary discussion (What is mind? What is body?) in meditation II. Only in the last week did we leap to meditation VI to see the final installment of the “argument” of the mind/body real distinction and Descartes' conclusion of meditation VI with the throwaway remark that the skeptical drama of the first meditation was just that—made-up (“hyperbolic”) drama. I was hooked.

It took me 18 years to put things together. In my first monograph about Descartes, I tried to temper the dualism my “initiator”, Rogers Albritton, was expounding (as were the clever TAs). I called the monograph “What Am I?”—Descartes’ own question early in meditation II, which he seems to never answer, for he goes on to speak about what (my) mind is and what body is but not explicitly about what I am. I tried to moderate his mind/body reputed *dualism* to the point that most “professional” scholars were saying to me with ironic smiles that though this integration of mind/body I offer may be natural enough a way for the mind and body to *intermingle*, it is surely not dualistic enough to be Descartes’.

Having written that monograph for so long, I had immediately after its appearance regrets that it still makes Descartes too dualistic. My regrets were induced not so much by having claimed for Descartes that the mind and body are *dual*—and thus *two*—things. Rather, it suddenly seemed that the *counting question* of how many they—mind/body—made was the wrong question to *start* with. If you *start* with (my) mind and body, sure enough, they have to make two (and two that are *really* distinct). But by then I was assailed by a new question: Should we—Descartes’ decoders—start with mind and body as the *primal* notions?

By the time I wrote my own self-critique, a sequel to Descartes monograph, I thought mind and body are *derivatives* of the *human being* of which they were the mind and body. (Descartes qualifies his investigation in the title of meditation II as of the nature of the *human* mind; thus I read the mind concerned as *of* a human being.) Stronger yet, the “how many?” question was now (2004–5) not anymore the *fulcrum* of his project. The source question was rather how can there be *thought* in the world, not just my own thinking—my *cogito*—but anybody’s thought? This was *the* lead-question—how can the cosmos/world engender thought? In magnifying the “Cartesian” dualism, we magnify the posited entity over the activity, mind over thinking. I came to think the key is the thinking, mind being a *façon de parler* about this activity—without thinking there is no mind; with it on board and understood well as what I—the human being am *doing* (and, as we see below, *doing-with-the—world*), mind is no longer what dualism *posits*.

Again, I ran into scholarly experts who swore that “Cogito”—understood as *pure* thinking and no world involved—is Descartes’ *Archimedean*

point and the most, automatic, obvious and inalienable fact in the universe (for me). But it was not so on my reading. One minor embarrassment was that, unlike the *Discours*, meditation II didn't have the object level assertion *Cogito*; it had *Sum* ("I exist"). This was odd—an Archimedean point that is not even there?

But even if one forced it into meditation II (in the way the textual (!) "scholars" force it in) there were many other barriers to my thinking being the Archimedean point of anything, even of itself, viz. of plain...thinking. One such barrier—in truth, the ground-zero issue—is that I read meditation I exactly inverted from the "standard model": it is said to show—via the alleged genuine possibility of skeptical scenarios—that *sheer* thinking is prior to and independent of any worldly causal impact. What is thus a *modus ponens* to the standard model—deducing the primality and purity of thinking—is to me an invitation by Descartes to a *modus tollens*—Descartes assails such purely qualitative or purely abstract accounts of thinking, self-standing thinking that segregates apart the world and causation from impacting the thinker. Thinking for Descartes is thinking-of-God-and-world, with the dashes, and *causation* from God-and-World to the thinker as the very maker of the thinking *relation*. There is no "pure" thinking just as there is nothing purely qualitative, not in Descartes. Thinking is a *product* of something logically prior—being; and, as we see in a moment, being with capital B. No Being, no thinking (of it!).

I wrote this sequel monograph *Cogito?* with the added question mark marking the anything but automaticity of any thinking, my own *Cogito* included. The deeper yet—than mind/body—dualism of *thinker versus world*, a thinker/world duality, that promised automatic world-independent free-ride thinking could not be right, I felt, for this unitary thinker, Descartes. It struck me that the thinker/world dualism—itsself so much promoted by introductions that make meditation I an exercise in lone skeptical thinking—was misguided at the seams. It suggested (borrowing from the famous Kantian motif *fiat iustitia e pereat mundus*) that Descartes urges a *fiat ego e pereat mundus* approach—there was a primal moment when (where) I can say "I think" and just stop the world in its tracks, just let the rest of the *mundus* go to hell, and with it consign to oblivion even God. I could just say, "I think", that's my part of the thought/world dualism and now I can leave out the rest of me (my

body) and, in tow, leave out the existence of the world and God. Is not “I think but God does not exist” whether genuinely possible or not, plainly *logically consistent*? All my interlocutors kept saying—nothing in my thinking logically entails anything beyond this much—*Fiat Ego*.

This could not be right, I thought. If I think (at all), God and World exist, whether they-themselves make one or two. My thinking does entail that World/God exist(s). There is not—in Descartes—a dualism of—me/thinker versus World/God.

Now ten years later and after a long journey through Spinoza, I am back with the master, feeling again I did not quite get to the bottom of Descartes even in the corrective *Cogito*? I tempered the mind/body dualism all right and tempered the thought/world dualism but, on both fronts, not enough. Descartes’ was more of a unifier than I allowed and not just about mind and body and respectively, thinker and world. His whole project—and that there is already in his scientific essays of the 1630s such a *whole* project—was about a unified cosmology comprising God, world, mathematics, space, man and all the rest of it. This meant that now 30 years after my Albritton initiation I was back to square one. I now have to restart from scratch in my understanding of Descartes.

So, how do I think of Descartes? I think of him as a sort of Marlon Brando, top-level actor with many “impressive” films. In each the local figure—of that film—is built up well and shows enormous *maitrise de technique*. But I also think the critics who try to use this local “evidence” from this or that film (only) are being misled by not looking for things that Brando does not control, patterns that recur in different films in spite of the masterly local camouflage-technique. A complex man is not quite the sum of his films/papers.

Speaking of the *complexity* of this actor/thinker. I do not pretend to accommodate everything Descartes said, as if pooling it into one large *consistent* set. To my chagrin, professional “analytic” history of philosophy has adopted a certain *positivistic* model of understanding, a model borrowed from court room legal dramas, where we feign that *consistency* and perfect deducibility of all “conclusions” is the point of our thinking—rather than a quest for the fundamental *truth* of the matter. Justification, not truth, is our *raison d’etre* as philosophers of historical figures.

I do not want to take issue with courtroom lawyers—in truth, even in those environments, the insightful practitioners prime truth rather than diluted but consistent at all cost “argumentation”. In any event, understanding a man, let alone a complex man, is not a law school exercise. The very presupposition that what Descartes thought—or even just *said*—is consistent is revelatory about the self-deceptive self-concept of those who make the presupposition. I can only say that if a scholar passes in his own life-long reflections the test of perfect consistency it is a pity because the scholar has not allowed himself/herself to grow, change his/her mind, see deeper. And however matters are with modern thinkers like you and me who operate within modern-legalized philosophy, Descartes was nothing of the sort; Descartes was not a *professional* philosopher, he was a philosopher. In what follows, I try to understand this complex man, not to impress judge Scalia with his kind of well-evidenced verbal pyrotechnics and “argumentation”.³

So, I am not after everything Descartes ever said and surely not in the vein of the lawyer’s game of drowning us with “evidence”. I am rather after *motifs*—in the musical sense of recurring segments and inherent in them, fundamental *themes*—in Descartes, motifs making him rather special, perhaps even distinct. In particular, I am not reflexively into boxing Descartes in this or that familiar “x-ism” box. My Descartes is not a rationalist, not an epistemologist, not even a metaphysician; in general, classifying him as an x-ian or x-ist, is of course gratifying, that is after all what we like best in modern philosophy—classify. But I feel Descartes defies classification. Descartes is not

³ Consider a potent example from the very science—not philosophy!—that gauges consistency and deduction—logic/set theory. The three greatest theorists—at least in my view—of the continuum—Cantor, Gödel, Woodin—have each, and in the space of relatively short time, altered their own earlier view—and with “arguments”—about this most simple (continuum) hypothesis—that the size (“cardinal”) of the real numbers is the second largest infinite size. I believe this is a direct result of the fact they *thought* very hard about the matter and sought the *truth* of the matter. There are of course many such examples of fluctuation at the deepest level of mathematics and natural science and lucky us that there are such thinkers who prime the universe and its truth rather than mere consistency-games. Descartes, like e.g. Cantor and Einstein (in the years 1910–1915 as he was gearing towards the field equations and up to 1917 when he was after the cosmological solutions), was thinking about the most fundamental infrastructure of the universe and with such thoughts came errors, changes-of mind, revisions etc. As we all know, Descartes could enter the courtroom game of consistency checking and deductions—as he does in some letters and in the objections/replies—and beat any adversary hands down. Fortunately, most of the time Descartes did not show us just how legally clever he was but rather actually pursued—with admirable openness—the very infrastructure of Being.

“like” this or that philosopher—though, again, I am aware of our *penchant* for the “x is like y” statements. Descartes is not like his medieval predecessors, nor is he like Leibniz or Locke or Kant or...his successors (who sure often use him as a whipping boy).⁴

Indeed, Descartes is not only to emerge as out of the boxes of rationalism and idealism but he will not even be classed as a...*Cartesian* and will not (even!) be tagged a...*dualist*, two accolades that have become synonymous with his name, making the claim “Descartes is not a dualist” sound like a contradiction. But it is not one. Descartes never says, “I am a mind/body dualist” or “I am a Cartesian epistemologist”. We the apostles say it. By now, we just assume that he is *on ne peut* dualist about mind body in both senses of (1) their separability as substances and (2) the horrid problem of *interaction* between those estranged duals. We also proceed as if his concern right from page 1 of meditation I is *skepticism* and the justification of our beliefs.

I do not so assume, and never did, not even on my first reading in the aforementioned Albritton undergraduate class. Descartes surely points us to *dualities*, indeed many dualities—between mind and body, between thinking and motion (location) in space, between algebra and geometry, between God and the world and between thinker and the world.

But dualities and dualism are not quite the same. Dualism is (1) a *theoretical* frame that sought to take as *primal* the split notions (e.g. *mind* and *body*), (2) erect an iron curtain between the two (pertinent) categories/domains, and (3) only late in the day seek a “synthesis”, making a *composition* out of the two primal separate ingredients. The primal ingredients are the separate notions (substances, ingredients); the “unification” is one that stitches a product or composition.

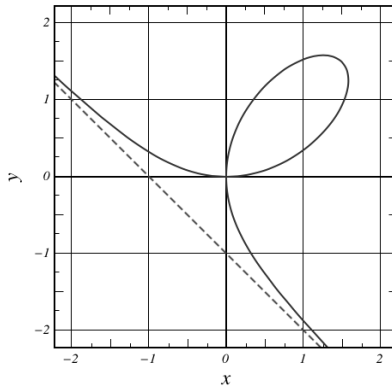
Instead, I will want to argue that on a variety of fronts—for Descartes’ famous mind/body dualism is only one of many dualizations—the very idea of *duality* runs quite the opposite in Descartes—the duality is *generated* by underlying unity; the duals—say, mind and body—are engendered as two

⁴ I have learned an enormous amount from the late Paul Hoffman and my friend John Carriero about Descartes but I do not share their sense that he is infused with medieval notions/ideas. For me, for all his schooling in his youth (and partly *because* of it!), Descartes is a stark original, like David Bowie in this magisterial film, *The Man Who Fell to Earth*. This is often said of Spinoza but I think it is the other way round—Spinoza is infused with medieval ideas, not Descartes.

forms of an underlying key basis. The very idea of the duals is that they are *derivatives* of the underlying unity.

Thus I will argue for Descartes that—there is no duality without a prior generative unity. This may sound abstract but it is an on going policy of his, encompassing his whole cosmology, witnessed in the variety of dualizations below.

So, even this one *ism*, so intimately bound with Descartes, *Dualism* (or “Cartesianism”) is not to be presupposed. No such box of “x-ism”, *assimilating* Descartes to familiar cases should color our reading of this thinker. My algebra teacher in Sevres (Western Paris) got it right when she said as she taught us about the *folium’s* (i) algebraic ($x^3+y^3=6xy$) expression and (ii) geometric expression:



“Don’t trust what your algebra book says about Descartes’ being one of those ‘reducing’ geometry to algebra.” And she added, “Descartes n’est pas du tout comme les autres, il est tout un monde en lui meme.”⁵

1. The Punctum Archimedis and Le Monde (the World)

Now that we mentioned the (le) *monde* we can move on finally to a positive contribution to our title’s quest—the *punctum Archimedis* of Descartes. Descartes’ concern through and through his life project is with the *monde*. If “Archimedean point” connotes to you after years and years of habit, the “axiom” that Descartes’ pursuit is not of the *monde-propre* but rather of

⁵ We return below to Descartes’ unique stance (at the time) about algebraic geometry.

epistemological foundations of beliefs, their justification and certainty, I am saying outright—“my” Descartes’ *punctum Archimedis* is not one in epistemology; it is not a bedrock *belief*. It was not *epistemology* that Archimedes was interested in, it was the whole *mundus*, and it was not epistemology that Descartes was interested in either. Archimedes sought the *infrastructure-mechanics* of the whole world in front of him—the whole planet—and so did Descartes. For Descartes, it is the whole world or nothing.

Descartes—like Archimedes—pursued the *mundus* not *as* the best of all possible worlds or the world-as will and representation or the world *as* the-sum-total of all possibilities or as a transcendental ideal or *as* this or as that. He pursued the world *tout court*, with surely nothing transcendental about it. “It”, the world itself, was for him the most immanent of all, indeed immanence itself. In saying this last, I have not forgotten God, far from it. But Descartes’ simple title of his 1630s *opus* says it *all*—*Le Monde*. No less, no more.

Thus to our modern post-Leibnizian ear accustomed to a plurality of worlds, it is worth emphasizing: For Descartes there is no plurality of “them”—*des mondes*—as in Cusanus or Bruno or the epigone of the many, Leibniz. Instead, in those very years, the 1630s, Descartes’ soulmate, Galileo, speaks of *due sistemi del mondo*—one world, two astronomic systems (theories) of this one—and only—real McCoy. This is Descartes’ focus, as it was a bit later Newton’s, *Le Monde*.

The ubiquitous definite noun phrase is not ubiquitous enough, for in truth the count noun world/monde is not used as a count noun at all by Descartes but rather as a noun ‘growing capitals’; in truth, a proper noun. Make them—worlds—plural and you undo their reality, leaving nothing but representations, models, telltale stories, perhaps mathematized as Leibniz liked to boast with any telltale he came up with, his model-theory of this or that, locally constructed “worlds”, with local *essences* to go with, local Lego pieces to *compose* the local world as a *combination* of such Lego *pieces*, local *essences*.

Just as such a plural *multi-verse* metaphysician is a localist, a singular universe theorist like Descartes is not. Indeed, not only is there one and only global universe for Descartes, the *true and immutable* natures of his are not local essences but are all *global invariants*. I reiterate this key point as many Leibnizers of Descartes miscast Descartes’ meditation II on the wax and meditation

V on triangles and God, as reflecting on “local essences”. Not only are there no local natures/essences for Descartes, in my view, there could not be any.

In this sense, Descartes is not close to the one proximate philosopher I so far did not mention, Spinoza. *De jure* Spinoza flirts with the *possible*—as logically preceding the actual—and flirts with *many* such worlds early in part I of his *Ethics*, in P5, where in what seems, alas, a Leibnizian vein, he argues from the “many” to the “one”, in effect argues by reduction ad *absurdum* to the one-ness of substance and in tow, world.

Descartes never had to *argue/deduce* (for) the existence and one-ness of the universe. Knowing what a theorem is and what an axiom is, he just starts with the absolutely infinitary being.⁶

And so, if there was ever an axiom of metaphysics, in late meditation III, it is writ large: *the absolute infinitary being is*. Spinoza’s “conclusion” (“theorem”) P7, that this ur-being, Being with capital B, or Substance/Nature (with capital “n”!) or God is the one *whose* nature (with little “n”) is to exist (to-be) should have been an *axiom*, not a theorem, for Spinoza, if one indulges in these model theory games of axioms and theorems at all. After all what we say here is a truth of logic: the nature of Being is...being.

Without indulging in axiomatics and theorem-proving, we can say that for Descartes, the *content* of P7 is the starting point of his unitary universal cosmology. He says it out loud and clear late in meditation III. If any “theorem” is involved, it is how all *other*—and as such *local*—being-s are generated by this one ur-Being. This is what I call below Descartes’ *fundamental theorem of Being*, to be read on the model of the fundamental theorem of arithmetic or algebra. But again, we had better not sound too

⁶ As we shall see in a moment, he also does not *deduce* his famous claim of meditation II, “I exist” or *Sum*. But the universe’s own global-*Sum* (“*I-world exist*”) precedes my own local-*Sum*, as we shall see in a moment. They share this much. Neither is deduced from anything else. As for ‘absolutely infinitary being’, let it not be supposed I speak in more modern terms of “quantity”—as in the absolute infinity of the universe of sets as opposed to the “ordinary” infinity (Cantor) of individual infinite (no matter how large) sets (ordinals). It is helpful to not rush to attribute such notions of size to Descartes, who attends rather to what is a *limited* unity and what is an *unlimited* unity. There are connections between the latter topic and the modern “set theoretic” issues of “size” (e.g. all individual sets are essentially limited unities) but this is a separate post-Descartes topic. Note that the so-called “universe” of all sets and ordinals is absolutely infinite in the sense that it does not make a set, from which it has been concluded that it does not make a genuine objectual unity, it is not an “*einheit*” (it is an absolutely infinite multiplicity, a *vielheit*). Descartes universe/God is not only a unity, it is *the mother* unity, the unity that logically-and not -just-causally-precedes all limited local “unities”, each a sub-unity, an image-transform of the ur-unity.

“proof theoretic” when expounding a wisely informal thinker like Descartes. He himself makes the point in ordinary language, late in meditation III and in his conversation with Burman: all local being-s are images—in the sense of *transforms-of*—the “total cause”, the ur-Being.⁷

I am aware that this global to local, *from* Being *to* its-own (in the genitive!) produced images, the local being-s, is not the standard approach to Descartes, by scholars or in the lore. Both scholars and innocent undergraduates—but who knows nowadays, they might just repeat *Wikipedia*—make him the father of skepticism, and/or rationalism, and not least, of a modern justificatory epistemology of beliefs, all local-to-global constructions, *from* my local self-first—often called his *Cogito*—and from local essences/attributes/perfections that I-Cogito-self am meant to be reflecting on as I build up an *aufbau der welt* and lo, an *aufbau der...Gott* (from attributes).

Rather than “duel” my way through this question of global to local vs. local to global, viz. rather than follow Descartes’ dialectical method in the objections/replies, I’d like to follow his method in the main text of the *Meditations* and simply lay—without deductions and legal wranglings back and forth—the picture of unitary Being I see him as painting.⁸

⁷ This brought Olli Koistinen to doubt whether in meditation V Descartes offers an ontological *argument*. Descartes offers an ontological *aperçu*, not a deduction, based on the entry point of late meditation III—the absolutely infinitary being exists. I believe Descartes’ modern French editor, F. Alquié, makes a similar point in an added note to the end of meditation III. I so read Descartes in my *Cogito*? We return to this point below while discussing Koistinen’s God-*cogito*. Regarding Spinoza, I here related to standard interpretations, all reconstructing his *argumentation*. In my own reading, as in my monograph *Everything in its right place*, I suggest that, as in Descartes’ meditation III, in P7, Spinoza starts with Nature/God, the absolutely infinitary being. Those liking ‘legal evidence’ might read P11 as suggesting this deductive argument-free reading.

⁸ I would like to acknowledge a few others who, *contre-courant*, spot the global to local metaphysician and the unifier of all being that Descartes was. I already mentioned in the acknowledgments note Jani Sinokki whose Turku-dissertation of 2016 is rich with such insights about the unity of Nature-science of Descartes in the 1630s and his unified metaphysics—based on the unified science—in the 1640s. Sinokki’s mentor, Olli, is another such Descartes-understander. I have heard it said that Koistinen “Kantianizes” Descartes and Spinoza. I read the influence as running the other way. If you read Olli’s work on true and immutable natures and the God-cogito—*I think therefore God/world exists*—or better in contraposed form—*if God/world does not exist, I don’t think*—you will see the inspiration from global Being theorists Descartes (Spinoza) in Olli’s many insights about Kant—(i) the inevitability of the 1763 global-necessary-being in early Kant and (ii) how it went on to shape Kant’s doctrine of the primality of global space—the primality, the non-conceptuality, the uniqueness, unity and necessity of Space—as showing up early in the *Critique* of 1781 (in the “Transcendental Aesthetic”).

2. The Unified Frame of Being

In a nutshell, my proposal is that Descartes is interested throughout his life in providing a general-unified frame of Being. How to start? Descartes urges us to look for “the Archimedean point”. What is this point? His statement “I exist” shows up in the early phase of meditation II. I should like it noted that umpteen commentators—surprisingly the most scholastic of scholars, those sporting endless quotes—speak of the *Meditations*’ Archimedean point, the alleged *cogito*, and go on to ponder—already inside the game—whether it is a *deduction or intuition* and other such sudoku-like logical puzzles. But the simple brute fact is that there is no “I think”—yes, no *Cogito*—in meditation II.

There is only “Sum”—I exist—said to be necessarily true as long as I consider it. Very well then, even if you extrapolate from *considering* to “necessarily true as long as I think of it”, there is all the difference in the world between a condition in the background language—as long as I think or dream or hallucinate or sing it—and an object level assertion “I think/dream/hallucinate/sing”. One cannot just dive down from the background language of the theorist—as long as I consider it—to the object language of the thinker and invent an “I consider” or “I think”, as if Descartes said it in his first person, his own mouth. Thus e.g. “*I exist!*” sure *must be* true as long as I dream it or hallucinate it or deny it or sing it or my body/ brain

For another *compagnon de route*, I should mention that ten years ago, when I was writing my “Cogito?” monograph on a little British Columbia island, I was corresponding with a native son of that island and a distinguished universal mathematician of our time, Robert Langlands, a fan of Descartes’ innovations in mathematics/physics (in one breath) and one to trace the unity behind deep correspondences (dualities) so present in Descartes’ initiation of algebraic geometry. Langlands offers, in the context of Descartes’ mathematical-physical work, a sense of an “Archimedean” reading of Descartes. This is expounded beautifully in Langlands’ “Is There Beauty in Mathematical Theories?” (2010) and “Descartes or Fermat?” (2005) (this last is only in Turkish but of which I reconstructed a summary).

Related and affected by the “Langlands program” in contemporary representation theory cum algebraic geometry, there is the first chapter of Ian Hacking’s 2014 book *Why Is There Philosophy of Mathematics At All?*, called “A Cartesian introduction”, also sensing Descartes has been subverted—indeed inverted—by standard philosophical texts. Hacking lingers on the unity behind the duality in Descartes’ algebraic geometry (a theme we expand on below in our slightly different way).

Finally, in his exposition of Spinoza’s famous *Deus sive Natura* unitary metaphysics, Etienne Balibar notes Descartes’ anticipation of the idea in his meditation VI (see Balibar 1998). As mentioned in the acknowledgments, though Carriero and Kambouchner pursue Descartes with a different methodology(-ies), in expounding the unified being outlook, I benefited much from discussion with both.

processes it but that is a long way from me saying in my own language—I dream/hallucinate/deny/sing/have my brain consider my existence.⁹

So we have *Sum* as our *punctum*, and no *Cogito* please. Some—indeed many—find the key in the subject of the sentence, the “I”. This mysterious subjective-inner subject is then taken as the fulcrum of the investigations. Modern philosophy is in narcissistic love with the subjective, the purely qualitative, the inner, all that natural science cannot “understand”, the last refuge of the wizard of Oz-philosopher, and many see Descartes as pointing to that beyond-natural science purely subjective “I”.

Descartes is the most anti-modern-philosophical—in this sense of subjectivity-quest—thinker I can think of. Unlike Locke and his inner forensic self, unlike Kant and his “I think” appended right and left, Descartes says outright in meditation I, which I just read again for weeks word by word in French with Olli, how the book that follows is meant to understand and frame the *sciences of Nature*, his 1630s *Le Monde* all over again, his life project. On my reading Descartes could have said in lieu of “I exist” “René Descartes exists”, that’s where...I start. He speaks “de-re”—of that very object, however referred to in language, and says: that object exists. Indeed, he might not use any term. He might waive his hands about or shriek and jump to get our attention—that object exists. This is the path to the *punctum*.

So no primal ethereal “inner” *punctum soggettivo* for Descartes. But what does my-JA *existence* consist in—what kind of fact is pointed out here? If we spoke linguistically, using *dicta*, which we shouldn’t really, we should ask: What does my “I exist” express, what is the *truth* expressed?

⁹ I am passing in quasi-silence over a further issue. Is the consideration in meditation II of my (objectual) existence—considering it ‘de-re’—or is it of the truth of the *dictum* “I exist”? I believe that Descartes cannot in earnest make the consideration of a linguistic dictum such as “I exist” or—even modulo translation which is anything but trivial here for one who didn’t speak both English and French—“Je pense” or “Cogito” (notice that if various puzzles like Kripke’s Pierre-in-London are brought on board, even a bilingual might run into trouble). In any event, it is not the truth of the dictum (necessary truth as long as...) that is the key but the de-re—de me-JA—fact that I-JA exist that is the key. I will assume this was Descartes’ intent. The truth of the dictum “I exist”—for that matter as long as I consider it—is too weak and vulnerable to parameters kind of a “fact” to get what Descartes wants. What modern logicians like David Kaplan say of the dictum (how in each context of use it is true though the truth locally expressed is not necessary unless used by God) is absolutely correct in my view but is not what Descartes is after. Descartes is not after all contexts and what dicta are invariantly true. He is after this *individual object*—me—and my-own existence.

If in “I exist”, the subject-“I” is not the *punctum Archimedis*, perhaps you expect me to say it is the predicate “exists” that is. But I will not say predicate, there is no “predicate” on Descartes’ page of meditation II and no allusion to this logical reformatory idea. In English, there is no “predicate”, there is just a verb, “exist”. Descartes is not Aristotle or Leibniz or Kant and he does not indulge in revisionary logical form theories, only in using ordinary language as it is. A predicate is something static *attributed* and *true* of me. Existence for Descartes is not something attributed and true of me.

There is only the French or Latin *verb* used. And I submit it is what this linguistic unit, the verb, introduces—Existence and with capital E—that is the *punctum Archimedis* of the Meditations. So I read his “I exist” as—I am part of, related to, a piece of, Existence. I am one of the existent-s. I read the verb not as attributing a feature to me but as relating me to something, note well—some-thing—something that is antecedent to me and larger than me—the human species, the living planet, the world. I am *in* the world, exclaims Descartes, I belong in this larger thing that is Being, with capital B please. In the way I can say that I am part of the human species, I am one of the humans (no predication here but relating me to other objects, Mankind and the plurality of men), I can say I am part of this larger thing yet, the world itself and indeed Being itself.

And this Existence at whose door I now put my foot—I too belong in Existence—this Existence was there before me and will be there after me. I now secured for a moment my being part of it. This Being or Existence with capitals (I use the two interchangeably) is the *punctum Archimedis* of Descartes. More generally, it is the scope of Descartes’ lifelong investigation—*what* there is, *all* that is, and how all those *is-ers* hang together, the unity of the all, thus the unity of all being-s, and thus the unity of Being. For me to be at all is for me to be-long in that logical precursor of me, Being. It is this over-all Being and my place in it that the *Meditations* is out to map.

I think Descartes says it in his own words. His “target” is “Nature in its most general respect”, as in

Et premièrement, il n’y a point de doute que tout ce que la nature m’enseigne contient quelque vérité: car par la nature, considérée en général, je n’entends maintenant autre chose que Dieu même, ou bien

l'ordre et la disposition que Dieu a établie dans les choses créées; et par ma nature en particulier, je n'entends autre chose que la complexion ou l'assemblage de toutes les choses que Dieu m'a données.¹⁰

3. Descartes' Unitarian Motifs

In his account of the unity of Being, Descartes develops the following motifs:

- (M1) The prime absolutely infinitary Being exists.
- (M2) The unity of Nature/God—no God vs. Nature dualism; God is Nature in its most general respect.
- (M3) All local beings are finite images of the prime Being.
- (M4) The unity the universe—no plurality of universes; one and only one Nature/universe.
- (M5) The unity of Nature—*true and immutable* “*natures*” are Nature’s non-local universal invariants.
- (M6) The unity of causation—all determination inside Nature (thinking included) is by causation.
- (M7) The unity of mathematics—algebraic geometry is the language of Nature.
- (M8) The unity of Nature/mathematics—no mathematics vs. Nature dualism—mathematics is the infrastructure of Nature.
- (M9) The unity of Thinking and Being—thinking of object *x* is a second mode of being of *x*.
- (M10) The unity of cognition: no thought vs. perception dualism; thought is more perception.
- (M11) The unity of man—no mind/body dualism; I am one thing, not three, a one-thing that thinks and is embodied.
- (M12) Dualities are generated each by an underlying unity.

Let me amplify on the motifs. I will do so by collecting them in three groups—the primal infinitary Being, Nature and its infrastructure and finally man and his cognition of the universe.

¹⁰ I prefer Descartes in French (as in the 1647 edition) because I have a more direct “feel” to what he conveys. I am no “natural speaker” of Latin but the Latin Descartes is (for me) more stilted.

3.1. The Primal Being

(M1) Descartes does not set out to “prove”—in the sense of a deduction from more primary truths (and perhaps self-evident each) that the absolutely infinitary being exists. That this infinitary being exists is the most primal truth, both in the order of truth-determination and in the order of thinking-determination. If we speak in terms of axioms and theorems, “The infinitary being exists” is an axiom, not a theorem. But speaking of the universal cosmology in such proof theoretic axiomatics is misleading.¹¹

As noted above, the starting point in meditation II, Sum (“I exist”, “J’existe”) involves—*logically*—the existence of the infinitary being. “I exist” is to be read: I am part of Existence (Being), I partake in Existence (Being). The standard “island-me” reading of meditation II is incorrect. It allows the logical consistency—if not possibility—of “I exist but the infinitary being (the world, God) does not”. Not according to our reading: “I exist” entails the world’s/God’s existence. To be is to be an image/product of the world/God.¹²

(M2) The infinitary Being is unique, with no split between it and Nature in its most general respect. One may *mistake* Nature for a mere finite product. This is not to take it in its most general respect—Being itself, all that is—or in Descartes’ language “the full order of created things”. This “order of Being” is not finite, though individual local beings in it are. In like manner, this unity of the all is itself the total cause and is itself self caused.¹³

(M3) All local beings—e.g. this stone, the Loire, you, me (this “I”)—are all “in the image” of the total cause and are made by its generative causation. Thus the fundamental theorem of Being: *every local being is a generated transform (image) of the one prime being.*

¹¹ The fact that Descartes sometimes indulges his critics (to say in their language what really matters to him) is no “evidence” this is the fundamental mode of presentation of his universal cosmology. Descartes own deep algebraic geometry is surely not axiomatic but in philosophical examples he speaks the language of Euclidean axioms etc.

¹² Descartes makes the point most clearly in meditation III: the finite existence assertion—my existence reflects (witnesses) the infinitary Existence. His French editor Alquié acknowledges it in an adjoined footnote.

¹³ For more on causation, see M6 below.

3.2. The Structure Theory of Nature

(M4)–(M6) Descartes’ true and immutable natures (with little “n”) are not separate local “essences” or “real definitions” in the vein of Aristotle or Aquinas, or later, Locke’s nominal or real essence. The true and immutable natures are *of* the fundamental-prime Being, invariants of Being’s processes of production—whenever a local being is made in the image of the ur-Being, the transformation preserves the invariants. The “always (necessarily) true” of triangles having 180 interior angles sum or that this much is itself fixed by the exterior angles (and in turn the curvature) are little “n” true/immutable natures of the capital “n” Nature. A little “n” nature is a transform of Nature. So much for (M5).¹⁴

As for causation (M6), causation is not the *cement* of the universe’s full order of created things. “Cement” suggests—as indeed Hume conceived of the separate local pieces—*gluing* together initially unglued things, each given separately (at least logically and in one’s imagination). Descartes’ causation *finds* all beings as bound at the outset, simply as part of what it is to be a being—to be a local being is to be an image/transform of the primary Being. This transforming force “applies”—first and foremost—to Being itself, the primal case of causation is self-causation (e.g. thinking by (in) God is such a case of second mode of being, of God (as thinker) having himself in mind by self causation).¹⁵

When a finite being like me has God in mind or the sun or myself (I think: *the sun is bright* or *I am a man* or *I think*), I have in my mind or better yet, in my thinking, the objects of thought—God, sun, JA. This is a second mode of being of those objects—God, the sun, JA—in Almog’s mind, on top of their first mode of being, in the heavens or here on earth.

To Elizabeth’s presupposition that causation seems to apply only to matter, Descartes points out that there is one and only one global causation enveloping all Being. When one speaks of the motion of the earth caused by the sun and when one speaks of the motion of my arm *raised* by myself we

¹⁴ Descartes may have been the first to know that it is the exterior angles sum via the curvature that determines the interior angles sum (see Gottlieb 1996). That true and immutable natures are not local, object-by-object “essences” has been emphasized by Olli in his work on Descartes (see Koistinen 2014).

¹⁵ On thinking understood as a just more (caused) being, a second mode of being generated by causation from the first mode, see (M9) below.

have two cases of this one global causation, even if only the latter—my raising my arm—involves thinking. Causation in Descartes—as is clear in the case of Being itself—is not, as is often made to be, a *reductive* relation, restricted to “just motion” in extension. The reductive relation is according to Descartes what Elizabeth presupposes, leaving no room to (forms of) *thinking*, which, for Descartes, is more causation (but now enriched at the outset as a global universal primary Being force).¹⁶

Causation for Descartes is not an abstract *relation*, as if of the logical form aRb , a relation “relating” two separate unities, the cause and the effect, given prior altogether to the world’s causation. Causation is rather for Descartes a unified process *engendering* objects like such local “relata” $a, b, c\dots$ In our local speech, for obvious epistemic reasons of available access and evidence, we may restrict attention (“zoom in”) and abstract local units (objects, events). We thereby focus on one end of the *process (the cause, the effect)* but for Descartes it is the generative *process* in the global field—the universe—that is the very maker of the local objects (events).

From this perspective, the influential modern Humean picture starting with scattered local objects/events subsequently cemented by an “add-on” “logicized” relation of causation, is an abstract “model” re-constructing the real-universe structure. In the model, we can practice localism about the “objects” or “events” and mere “relations” subsequently added “on” them. In reality, the local things are all transforms and products of the unfolding of Nature by universal process causation.¹⁷

(M4), (M7)–(M8): Mathematics involves for Descartes two separate “levels” of unification—*intra*-mathematical and *trans*-mathematical. The first is taken up in the present annotation, the second in the next.

Descartes is not a bottom-up combinatorialist in the foundations of mathematics, as was Leibniz, and as are many in the now very influential set theoretic foundations, emerging in the late nineteenth century quest for a language of elementary ingredients from which we make combinations-ensembles. In such combinatorialist foundations, we make up a “universe” of such local combinations/ensembles, but it-itself, the universe, is not an

¹⁶ See below M9 on thinking as a second mode of caused-being.

¹⁷ For a development of Descartes “process in a global field view” of causation rather than Humean localist view of causation as a relation between events, see Sinokki 2016.

objectual unity (an “*einheit*”) but a mere *manifold* of such limited combinations. This is the modern set theoretic route.

Descartes is not seeking such bottom-up localist combinatorial foundations. His language is not that of (second order) logic or set theory but of algebraic geometry. Philosophical texts often speak in a high school algebra class mode viewing Descartes as *reducing* geometry to algebra, in effect making Descartes a prequel to Leibniz, viz. looking at the algebraic language of operations of combinations as the foundational language, the *ars combinatoria*.

This inverts Descartes top-down globalist and anti-reductionist methodology. His conception of algebraic geometry is of a totally new and opposite—*non-reductive* to algebra—insight of *correspondence*. A correspondence between geometry and algebra is not a reduction; it is a revelation of a surface duality behind which lies a deep *generative unity*—the unity that generated the duals as two manifestations thereof. This is true for the emergence of deep modern algebra/geometry dualities (see immediately below) but it also holds a more general point about dualities beyond mathematics—a duality is not a reduction of one of the duals to the other; a duality is a manifestation of a double-headed symbiosis of an underlying *generative unity*.¹⁸

If anything, the correspondence reveals how deeply geometric are the symbolic algebraic formulas (whether we think of the *solution-set* or later, by the twentieth century’s Grothendieck’s carving notion of *scheme*, of the formulas-set themselves as geometric deep down). Descartes is a *geometre* in the tradition of Galois, Poincare, Weil and Grothendieck, all thinking top down, eventually of the moduli space of curves (varieties).¹⁹

¹⁸ That all Descartes’ dualities are such symbiotic correspondences reflecting a generative unity is a point we make in our last annotation below, M12. On the importance of correspondences—and no reductions—in Descartes’ algebraic geometry and its modern development, see also Hacking’s chapter *op. cit.*

¹⁹ Let me mention the following theme without really developing it here except for indicating an “intention” (the theme is developed in the monograph project (with Sinokki) about Descartes the unitarian metaphysician-scientist). (1) It is interesting to note that a host of landmark conjectures in modern Diophantine geometry and algebraic geometry have a very strong “Parfum de Descartes” as one of the conjecture issuers pointed out. In the book projected I linger on Mordell’s conjecture (now a theorem by Faltings) regarding the class of Diophantine equations of genus 2 or higher having only finitely many solutions (hence the classification uses a topological invariant; this is the “parfum de Descartes” viz. to unfold the nature of the class of equations by such an invariant) (2) The Weil conjectures that opt again—on yet a more general scale—to understand the “nature” of curves (and counting the numbers of solutions of algebraic equations) by way of topological invariants, specifically the correlation between cohomological structure of

The misreading of Descartes as if he was seeking a symbolic logic language of operations as a *reduct*-language misses Descartes idea of a symbiotic correspondence—e.g. the geometric folium (the figure) is revealed to have a formula correspondent (in time, a group theoretic base controlling the equations) and the equations are revealed to be geometric in nature, with deep topological invariants *determining* that “nature”, when the solution-sets are projected—taken as a manifold—in a global space (e.g. the complex plane).²⁰

This is not all. There is more unification yet at work in (M4), (M7)–(M8).

The foregoing split between the bottom-up local to global combinatorialist and the top-down globalist impacts the next level of unification vs. splitting, this time between mathematics and the study of “physical” nature.

What I view as the Leibnizian localist and combinatorialist tradition seeks what is in effect not a *universe* theory but a *model* theory, a spectrum of models, viz. combinations from the local ingredients and admitting a *multi-*

algebraic varieties over the complex numbers, and the arithmetic structure (number of solutions) of algebraic varieties over finite fields. (3) The Fontaine-Mazur conjecture according to which (p-adically based) Galois representations (on the absolute Galois group over an extension field K) come from “geometry”, which for our purposes here can be said to mean—arise from just the kind of cohomological invariants—from étale cohomology—key to the Weil conjectures. Related and at the basis of the last mentioned two ideas is the quest for a punctum archimidis of every algebraic variety, with a very strong “Parfum of Descartes”, an ur-level unity called the *motive* (Grothendieck: motif) behind the algebraic variety, as if the geometric crux of the variety (in a simple case, the “Jacobian” of the curve). See for an introduction Mazur 2004. In his *Récoltes et Semailles*, Grothendieck says—in his beautiful and informal prose, echoing words and style that sound like authentic Descartes in his appendix to the *Discours*: “Parmi toutes les choses mathématiques que j’avais eu le privilège de découvrir et d’amener au jour, cette réalité des motifs m’apparaît encore comme la plus fascinante, la plus chargée de mystère - au coeur même de l’identité profonde entre “la géométrie” et “l’arithmétique”. Et le “yoga des motifs” [...] est peut-être le plus puissant instrument de découverte que j’aie dégagé dans cette première période de ma vie de mathématicien.” More information and “hunches” on where Descartes would have found himself in the modern rich intersection of algebraic geometry, representation theory and (algebraic!) automorphic forms theory is given by Hacking (2014). Like Hacking, I am not native to these complex landscapes and I rely, like him, on native sources, as well as a sort of foreign language understanding and hunches. Dieudonné, who knew Descartes’ *La Geometrie* in and out, is a most formidable pedagogue, even for half hobbling, half blind “tourists” like myself (see his historically rich Dieudonné 1975).

²⁰ As mentioned, two others developing this picture of Descartes as a universal “geometre” are the aforementioned work of Robert Langlands (2005; 2010) and, affected by it, Ian Hacking’s first chapter of op. cit.

verse foundations, wherein a spectrum of possible universes precedes the real universe, which is *posterior* and is just one (or isomorphic to one) of the items in the multi-verse spectrum. The outlook encourages the idea that mathematics is given prior to and independently of the physical world and consists of a spectrum of such combinations/models. In studying Nature or the physical world, we *apply* this earlier-already-given mathematics. The outlook invites a mathematics/Nature dualism and with it hard questions like Wigner’s famous sense of magic viz. the unreasonable effectiveness of mathematics in the study of Nature—mathematics is hatched in advance and prior to Nature; how come, by what magic, do we find it so useful in studying this posterior locus, the physical world?

Descartes’ outlook is quite the other way—it would be unreasonable if mathematics did not describe Nature, just as it would be unreasonable if my DNA did not control the morphogenesis of my body—the “structure” theory is *generated* by the very body (universe) it is the structure-of. Structure is not something hatched independently and then—as Wigner likes to say and all of us follow him—subsequently *applied* to or *realized* in the “body” it describes. Structure is not “applied” or “realized” in Descartes; structure of a being *x* is a *generated* by-product of the very being *x* whose structure it is. This covers Descartes’ idea of true and immutable natures (of Nature)—they are not independently given local essences subsequently applied to Nature, they are generated as universal invariants of Nature.²¹

On Descartes’ outlook, Nature is prior to mathematics or better generative of and symbiotic with it (in the way my DNA is generative of and symbiotic with my body). Mathematics is engendered by Nature/the universe, which is a one-and-only uni-verse encompassing all (including mathematics). There is no “other realm” (other universe, spectrum of models/worlds) mathematics independently lives in. Mathematics is not independent—it is the *generated* infrastructure of the one and only absolute universe. In this sense, Descartes is not a *multiverse* model theorist. He of course can “construct” models for a certain group of principles (where some base principles hold but not others). Such “models” are not parallel “uni-verses” but rather sub-structures of this and only universe. The so-called common conception (much emphasized by Leibniz) of “physics” as a “mere” description of this local

²¹ Recall here Koistinen’s (2014) view of true and immutable natures.

world (though possibly not of different others) is not at all Descartes' view—physics is the *structure theory* of the one-and-only universe and as such, physics is not a restriction of mathematics but rather physics and mathematics are one and universal. We can of course consider abstractions of this base universe but they are restrictions/fragments of the primal universe.

3.3. The Unity of Man and His Cognition of Nature

(M9) For Descartes my *thinking* of the sun is a second mode of *being* of the sun—being that is not in the heavens (this much is its first mode of being) but an “image”—by causation—of this first mode of being, in me, a thinker. Thinking of object *x* is a second mode of being of *x* generated by the being *x* and its causal effect on me, the thinker.

This key idea unifies thinking and being by making my thinking of an object, just like my perceiving, an image-transform (“image” in the sense of product, not the sense of qualitative photo) of the object in me. And thus, my thinking of *x* is a generated product of the being *x* action on me. Descartes radically reverses the “flow diagram” often assumed in separatist thinking accounts—and all dubbed...“Cartesian” accounts.

In the separatist/Cartesian picture, thinking “takes place” in my mind (and is not dependent on “reception” of signals by my world-embedded body). It consists of grasping “contents” (sometimes “ideas”) prior to and independently of the objects thought-about and prior to any causal impact from the world on me-the thinker. With the contents “in our heads”, we look for matching between contents and candidate satisfiers. Classical accounts made the “contents” impressionistic and qualitative and looked for qualitative resemblance—e.g. something that looks like a large orange ball on the horizon (or the zenith)—the sun. Recent amendments of this Cartesian account have kept the flow diagram but substituted qualitative ideas and qualitative resemblance by matching of a “structural essence”-information that is not “qualitatively given”. Still the flow diagram remains the same—we proceed in two stages, from mind to contentful ideas (the first stage) and then, the second stage, by resemblance, “finding” an object to match.

Descartes' account reverses the flow diagram and does away with the two-stage matching. The being thought of—the sun, me, God—precedes the thinking of...*it*. The thinking-of-it is induced by the object's first mode of

being—the object generates in the receiving thinker a second mode of being of that originating-cause—the sun, me, God. The receiving thinker *refers back*—*re-ferrer* says Descartes to indicate as in the Latin/French, the initial *ferrying* by causation of the object unto the thinker and now the *re-ferrying* back to the object by the thinker.

On this reading of Descartes, I am thinking of the sun independently of any matching—by the sun—of internal-in the head content. Indeed, my “qualitative” or “structural” contents may be wrong of the sun (Descartes’ own “physical” theory of the sun/earth system were in error). It is still me thinking—as in me perceiving—of the *sun even if I get its features wrongly*. I am thinking-of-it because it is *it*—the sun—that *came* to be in my mind by causally modifying it. This structural analogy between receiving the sun in perception or in thinking leads us to our next annotation.

(M10) There is no perception vs. thinking duality in Descartes but a unique *process* by which the thinker (his mind) *comes to have* in it the object of reference, the object of the aforementioned *re-ferrer* relation. An object, say the sun, has been ferried to me, and I-thinker now re-fer back to it. This much is the fundamental and unitary cognitive relation—coming to have—note the progressive tensing—an object in mind. Descartes speaks of that causative object coming to have a second mode of being. This much is the *nature* of the relation of thinking, *what* thinking is—objects coming to be related to thinkers that do a re-ferrying back.

There is the *what* the thinking relation is and there is next, the adverbial *how*, indicating the way the thinking takes place. The verb—think—and the ad-verb—how the thinking occurs—should not be fused.

How the thought-of-the-object impacts me, whether the impact involves the inducing of a qualitative image—and in a particular sense modality (by hearing? by vision? etc.)—is not of the *nature* of the relation but of the adverbial manner of that relation (process). Thus all those having in mind the sun or the now long dead and gone Aristotle or all those thinking of the un-visualizable God, think of the *same object* and are in the same cognitive state—thinking of that object x, viz. having that object x in a second mode of being in the plurality of thinker-s’ minds. It is the object x that determines the cognitive state of the many thinkers that are thus x-impacted.

For one such impacted thinker, the impact may be accompanied by a vivid image that gets the surface features of the item x correctly; another thinker may have an image but rather qualitatively incorrect of the object x (though x is still the same object thought of). A yet third cognizer may be thinking of the very same object and be in that very state—thinking of x —even if no image is generated. Those who shook Aristotle’s hand in ancient Athens and me-JA millennia later think of the same Aristotle. We are both in this one state—thinking of Aristotle—however we “qualitatively” embellish this thinking-state with “qualitative” presentations. If Moses saw directly God on the mountain, he is still co-thinking of God with me, I who had no occasion to form a qualitative impression. We both think *of* God because God has impacted our two minds. *How* we think of Him may show “qualitative differences”.

It is quite clear that subsequent dissections of the relation “ x thinks of (about) y ”—by rationalists like Leibniz, empiricists like Locke, synthesizers like Kant, let alone modern analyses such as Frege and Russell (two different modern analyses by the way)—built into the very nature of the cognitive relation—into the *thinks-of*—the *adverbial* information—the *how*—I mentioned. This is often read back into Descartes—as if he was the *pere fondateur* of this “Cartesian” picture that perception (acquaintance) must involve an image, whereas abstract *thinking* of x must not. I do not believe this is correct of Descartes.

Quite the opposite, Descartes allows that I can *perceive*—and uses the relevant French and Latin cognates—myself, God, abstract mathematical entities (and *kinds* of entities) without any qualitative imagery. Descartes also allows that I am in a position to *re-ferrer* to Aristotle whether I can bring up in my mind an image of that man or not. The word “Aristotle”, involving no qualitative image but carrying/ferrying Aristotle to me, is enough for the referrer to Aristotle to take place.

An important case of such qualitative-image-free perceiving is of—Le Monde. It is often assumed, by those insisting on qualitative images, that the world is “too big” for me to have a right qualitative image of. I thus cannot perceive it (and if Kant is right, I cannot form of a *concept* of it either, so in short, since it’d better be one of these two *ways*, I cannot think of the *mundus*, period).

Descartes points out that I can perceive the ocean without being in the space-station and taking a photo of the full expanse; the latter day astronauts think of the very same ocean of which Descartes in 1641 could not have a full image of. Descartes points out that I see the little wave hitting the beach and *thereby* I have perceived the Atlantic Ocean. Idem for perceiving this or that local *manifestation* of the universe, and thereby perceiving it, the universe, without a qualitative image of its full expanse. Indeed, not only can I and do I perceive the universe, I *cannot but* perceive it. In any perception I take in, of my fingernail or your lemon tree, I perceive the universe. It constantly impacts my perception.²²

This freedom from qualitative imagery allows two thinkers who are differently equipped “qualitatively” to think the same—the sun is hot. The thinker here in Europe and the aboriginal in Australia surely had different qualitative presentations of the world—and of space with capital S and in it of the sun—but they thought of the same if they thought—the world is large or ancient or scary and the sun is hot.

And there is more. As indicated, he who saw Aristotle directly and I who may have only heard of him second hand or received his name *perceive* that very man. What is the key is *which object* here has a second mode of being by way of causal impact. Whether *the how* involves qualitatively clear imagery, confused imagery or no imagery at all is relevant to many presentational-details of my impressions but not to whether I think/perceive of Aristotle.

Expanding further this observation, we may dissect the common assumption—automatically repeated in teaching of meditation I—that when I dream, hallucinate and see wide-awake a certain image of a lemon tree, I have in all three cases the same qualitative image. This is fine but shows in no way what is also said in the same breath in those introductions to the first meditation viz. that the three states are one and the same underlying cognitive state, described as the same...thinking.

²² Descartes uses the verb “perceive”, in the French 1647, of the wax: “Mais ce qui est à remarquer sa perception, ou bien l’action par laquelle on l’aperçoit, n’est point une vision, ni un attouchement, ni une imagination, et ne l’a jamais été, quoiqu’il le semblât ainsi auparavant, mais seulement une inspection de l’esprit, laquelle peut être imparfaite et confuse, comme elle était auparavant, ou bien claire et distincte, comme elle est à présent, selon que mon attention se porte plus ou moins aux choses qui sont en elle, et dont elle est composée.”

Not for Descartes. First, in the cases of the dream and hallucination, I may not at all be having a *lemon*—that kind of fruit **K**—let alone a *particular* lemon *x*, in mind, simply because the origin of the cognitive episodes is not a real lemon but the pixel-ing/photo-montage inside my head of the qualitative image of a lemon-lookalike (a “twin”). But even if all three states are in effect causally impacted by this one worldly lemon in the garden, the cognitive states are still deeply different: to think is not to dream is not to hallucinate. The invariant qualitative *how*—the lemon-look—does not make the *what*—we have on hand three different *kinds* of cognitive states.

And so it goes in abstract matters—the what and the how of cognition must be separated. I mention three examples that are Descartes’ favorites—God, me, mathematical objects—and a fourth, given as his own example—the wax. In all four cases, I cannot form a picture—with a qualitative image—of the object, or if I do, it—the how—is *distracting*—the visual presentation of the *folium*, my own narcissistic Dorian gray imaging of myself in the mirror, Michelangelo’s depicting of God on the ceiling of the chapel and the smell and taste of the piece of wax.

Descartes does not say—we do not *perceive* the wax because the qualitative presentation alters; indeed, he asserts exactly the opposite: I *perceive* the same wax, *even though* the qualitative presentation has altered. And so it goes—the incantation in the biblical first commandment to not go by God’s image (but go by our being “in his image” in this other sense of causal-transform) is very much in Descartes’ sense: we thinkers are beings in the image of God and in thinking of Him, my thinking—the second mode of His being—is an image—not in the qualitative sense but in the causal product sense—in me, of Him-God.

The same lesson applies in mathematics (geometry)—one need not go as far as the *chiliagon*, so much stressed in the textbooks to “see” that Qualitative-images—henceforth Q-images—are not the key to the Causal-imaging—henceforth C-images—effected by algebraic-coding. We can see the point by holding on to the “simple” figures of triangles where (1) reputedly we *can* form Q-images. Well this is not *what* we are thinking/perceiving. It is the C-image,

the triangle as a product of the surrounding space and its curvature that we think/perceive, without “qualitative visualization”.²³

Thinking/perceiving does not work by Q-images. It works by C-imaging, whereby an object—be it the sun or God—is impacting causally the thinker and makes this last *come to have* in his mind the object in “second mode of being”, come to have as an “image” (effect/product) the original object.

In meditation I, Descartes attacks the inside-out model of thinking by Q-images. Once one operates on the level of the Q-image, one has gone astray on a host of dimensions—hallucinating, dreaming and veridically-seeing are classified as processing the same “thought”; secondly, thinking different thoughts about Q-twins poses a “puzzle” of individuation; and not least, there is a level of “representation”—the Q-image medium—that can engender skeptical fears—what if I had this Q-image “thought” and there was no world in the back of it?

The modern solution followed by many “Cartesian” scholars is to block this gap by exchanging the Q-image by a “structural” essence that is supposed to guarantee better resemblance and no skeptical “holes” instead of real causes.

But we only reproduce the same “problem” one level down. Not only will there be now structural twins—instead of Q-twins—but also the very casting of the issue has preserved the fundamental mistake: we still start with the *media* given prior to and independent of the world (God) and cast around for some “condition of resemblance” that would fix/determine the referent of the medium.

Instead, the ingenuity of Descartes’ reversal of the “mechanism” of thinking lies in its now operating by causation outside-in—there is no “fixing” or “determination” of under-determined “internal” media to be done. It is all fixed/determined at the outset: if it is God (the sun, this twin sister, not that one) that generated the second mode of its own being (*viz.* my thinking of *that* object), then it is *that* object that is being thought of. What else demands here fixing?

In the world’s metaphysics, everything about the relation “JA thinks of the sun” is determined/fixed by the unfolding of causality in the world. We of

²³ I remind us again that Descartes knew that it is the exterior angle sum and the ambient curvature that makes the triangle’s interior angle sum.

course may need *epistemic* identification (not to be confused with ontic determination)—we need to figure out *who* it is I am thinking-of, *whom* do I have in mind—and whether I have a real worldly object in mind at all. I have no “a priori”—in the sense of *purely internal* in the head scanning—guarantees. But I do not have such internal a priori guarantees precisely *because* it is the real world that *originates* and causes my cognitive state as a second mode of being of that originating-cause, e.g. the sun.

4. Two Addenda about Thinking/Perception

Two final addenda to this outside-in mechanism of thinking as modified-being are called for. They are in effect two errata on “concessions” I have made en route that I should like to amend. The first regards the *apriority* vs. *certainty* of successfully thinking, the second about the nature of *sensation*.

So, first, regarding the certainty of successfully thinking, I said that there may not be a purely internal Q-level certification *that* I am thinking and of *whom* I am thinking. This should not entail or even suggest that it is not *certain* (for me) that I am thinking and of this particular item, say this piece of wax in front of me. Certainty is not formal derivability (or logically consequence) from Q-images. As Moore noted, I am as certain as could be—know for certain—that I have two hands, though this is not *deducible* from Q-images alone. In meditation VI, Descartes knows for *certain* that there are material bodies, that God exists and that he has pain in his leg (when he does). If he were a soldier with an amputated leg, he’d have the same Q-image (let us suppose). But he is not such a soldier and his mechanism of receiving by perception the pain from the leg has worked by reliable causation all right. Yes, Descartes does not know a priori—by internal scans alone—that he has not lost a leg. But he is *certain* he has the two of them on him and the nagging pain is in the right one.

Secondly, regarding *sensation*. We have a tendency—even among the most quotation-bound scholars—to read back into Descartes later doctrines about “consciousness” and *qualia*, as if Descartes was Nagel and Kripke writing on pain or Kant speaking of self-consciousness as the key (key to what?). I am not familiar with more than one and a half place in the *Meditations* where Descartes focuses on self-*consciousness* and even there, it is not consciousness of a self but just awareness of full human me-JA, (or him, René

Descartes), the *full* me. Descartes is not Locke and he has no *forensic* self, invented *on top* of the nature-made human being, René Descartes.

In like vein, for Descartes, our sensations of pain or cold are not the “what it’s like” or *qualia* so prized in modern philosophy of mind. For Descartes, the sensation of cold is my *perceiving* the cold e.g. on my skin and my *sensing* the pain in the leg (and it is *in the leg* that I sense it to be). I perceive those *bodily* phenomena. I may perceive them clearly, sometimes clearly and distinctly, often confusedly. But I *perceive* them *in* my body, not scan them in the qualitative internalities of my mind. And so, Descartes is not a modern *qualia*-friend. He never says that we could have pain without the corresponding brain state. How could we given the *nature* of pain and the nature of our sensing of it?

My sensing the pain (in the leg) is the leg’s second mode of being in my perception. Of course, as noted above and counter to Kripke’s internalizing and apriorizing my knowledge of the purely qualitative “essence” of the pain, for Descartes, I do not know a priori (in the internal sense) that my leg is in pain. But now that you kick my leg, I am certain—*know* for certain—that my leg is in pain. I am not the amputated soldier and I know this much for certain too.

(M11) I come now to our last two—the most famous and imposing—Cartesian dualities—of thinker vs. world and then, of mind vs. body.

First, the thinker vs. the world he thinks of. In *Cogito?* I urged that in Descartes’ account of thinking we find a two-tiered *God-cogito*. First, “I think” entails “I think of God/world”; secondly, “I think (of God/world)” entails “God/world exists”. Olli Koistinen (2014) has simplified this in his writings to a *God-Cogito*: If I think then God exists.

It is important not to view Descartes as “merely” asserting: If I think of *x*, then *x* exists. He does assert this general conditional—counter to his many “intentional object” interpreters. It is clear why it is a fundamental truth for him: if to think of *x* is to have *x* in one’s mind in a second mode of being for *x*, clearly *x* has to have the first mode of being to impact my mind, later, in the second mode. So, if I think of *x* at all—which is not ascertainable by my own mind by sheer introspection (as assumed by modern writers)—*x* must have *made* me think of it by a process extending from *x* to me (my mind). Note that this does not mean that *x* goes on existing as I *come to have* it in mind. I do

think regularly about Descartes but he is long gone. However, he did have to be part of the universe to initiate the process through which he came to be in a second mode of being in my mind.

All of this is true for any object x I think of. What is distinct of my thinking of God—and I believe this is the point of Koistinen’s God-Cogito too—is that in *any* thinking of mine, God’s existence comes up in my mind because even when I am thinking *locally*—e.g. of Descartes, the Mt. Blanc or my fingernail—the objects in question—indeed any objects—are finite manifestations in the image of Him. And thus, He, like the vast ocean, is present in a second mode of being in my mind as I focus locally on this or that tiny wave hitting the shore.

The oft-cited “external world skepticism”—understood here about *thought*, not about *knowledge* (though this derivative one is also covered by the remarks that follow)—is not just “false”, it is—again—logically—incoherent. To think, on my reading of Descartes, is to think-the-world. The idea that I would be thinking without a world at all (as in the *Malin Genie* paranoia) or even without me taking-in the world—without the world having a second mode of being in me—is not just false or necessarily false (not possible); it is not *consistent*. If I had not taken the world in, I would not be thinking at all.²⁴

Moving on to our second duality, we reach the most resistant duality of all, the mind/body dualism. Let me reiterate that I am aware of the fact that Descartes says the mind and body are substances (“complete things”) and are really distinct (however we dissect this relation). But I do not think this makes him a *dualist*, not even the very mild “integrative dualist” I read him to be two decades ago in my *What Am I?* I misunderstood what Descartes says and implanted it too rapidly in a form of dualism. Descartes response to Elizabeth (and in a well-known earlier variation to Regius) says it all:

When I wrote my letter I thought you still had at the forefront of your mind the reasons which prove that (a) the soul is distinct from the body; and I didn’t want to ask you to push them aside so as to bring to the fore the notion of (b) their union that everyone always experiences within

²⁴ The remarks above regarding Descartes on sensation as perception suggest that there could not, for Descartes, be the modern philosopher’s favorite—*purely* qualitative sensations. Sensation is of—is a taking in by my body—causal sources, including sources in that very body.

himself without philosophizing—simply by knowing that he is a single person who has both body and thought whose natures are such that this thought can move the body and can sense what happens to the body. (Descartes 2015, 6.)

What does Descartes say here? It is not that I—this single person I know without philosophizing—am made mind-less or body-less. But we must understand the ontological *provenance* of my mind and body (and we better speak of someone’s—e.g. my own—mind and body rather than the abstract “concepts” of mind and of body, an indulgence Descartes allows himself in meditation II). My mind is to be read in the possessive and genitive: *my* (JA’s) mind, the mind of me, this human being, JA. Such is also my body—the body of this human being. The *body* and *thought* mentioned in the letter are *of* the “single person”. There are no doubt “contexts” where we investigate some aspects of me—why am I depressed vs. why would I fall from the roof at a certain velocity, just like a rock, towards the earth’s gravity center?²⁵

Yes, physics focuses on the latter, psychology on the former. University departments are based on this idea of ‘divide and conquer’ but we should not confuse a focus for the purpose of a certain investigation with what is possible or actual, in the structure of the things proper. The compartmentalized univers-ity is not the universe.

Investigating a subject matter *insofar* as it is this or that is no pronouncement about the very *nature* of this subject. I can surely investigate numbers insofar as they are given to me by the provable theorems of formal arithmetic. But this is no suggestion the consistency of the proof system (itself a number theoretic truth) is not of the very *nature* of the numbers. My restricted investigation is exactly what it is: investigating the numbers *insofar* as the provable theorems of formal arithmetic go. No less, but also no more. Idem for me-JA—investigated restrictedly insofar as I am a psychological system vs. insofar as I am a free falling body.

This should be enough to dissipate the ethos of Descartes allegedly separating two primal self-standing objects only accidentally stitched late in the day, as if cohabiting in this housing project, the human being. But there is

²⁵ ‘Human being’, ‘man’ and ‘person’ are used here interchangeably.

more to be said here against the alleged dualism. I will keep here to two simple facts.²⁶

First, the appearance of separability by “university departments”—psychology vs. physics—is in the end misleading. Yes, we have a division of labor between local investigations and not everyone can be a theorist of everything. But as the psychologist ponders my depression or pain he realizes that understanding the human brain (and the environment impacting it) is of the “essence” of the investigation. In like manner, the physiologist who investigates the fibers firing or the lack of certain key “happy-making” molecules in the brain soon discovers that a mere geometric or biochemical description of the brain states is not a full description. That in human brains *these* states are felt as pain or experienced as depression is again of the very nature of the states.

The philosopher’s insistence that he can imagine the brain firing without the pain felt is just that—imagining—and in my view not of the firing of *these* brain fibers with their genuine nature (which is to be felt as pain!) included. Sure enough I can say I imagine myself without imagining my human mom and dad, let alone the tiny sperm and egg I came from. But I can’t really so imagine. I cannot imagine real-me as an object just brought by a stork or by your parents or by another pair of sperm and egg of my actual parents. I can write stories and shoot movies presenting such verbal tracts but that is all—I have not taken that object, me, with its nature and no mere front, and imagine of it all these scenarios. I can Q-image a host of corresponding things and misdescribe them *saying* I have imagined historical-me as so and so. Idem for my pain and depression and the necessarily (stronger yet, by nature) correspondent c-fibers and d-fibers firing. Modern philosophy (in the vein of Leibniz and Hume and Kripke) is over-impressed with its power to imagine (I say: to Q-image) a host of *recherche* combinations. Such Q-imaginings are not revelatory of Nature, but of its little ‘n’ natures and what is possible for them.

(M12) My final observation about unitary-Descartes is more methodological. It concerns the very idea of a duality. We are concerned here not just with the mind/body duality but dualities (in Descartes) in general.²⁷

²⁶ I have written earlier on these matters in connection to modern materialism vs. dualism issues (the work of Kripke vs. Descartes) in Almog 2002 and 2010.

I'll put his duality-within-unity point made to Elizabeth in my own language—the *my mind/my body* duality makes sense only as a duality *within-the-host-unity* and as *generated-by-the-host*, the underlying human being. We do not have an *assemblage*, a mind from here and a body from there, each given as itself and separately, only to compose later from them a unity of composition, me-JA. The very categories—my (human) mind and my human body—are to be read: that human being's mind and body.

In this sense—noted by Descartes in his fourth replies—the duality is like the duality of my two hands (left and right) or the arm (hand) and body or, for that matter, my brain and my body. All are not cases of ingredients given in advance and separately like the independent and prior ingredients of a chocolate cake at the bakery, given first separately, only to “move on” next to the construction-making of the cake. The left and right hands duality is generated by the underlying body, it is a dual-system serving that body whose hands they are. Dual-systems pervade Nature: a human body, an animal (“wolf”) body, the ecology and atmosphere of the Amazon basin, the electromagnetic field, the absolutely-infinitely-many ordinals and the corresponding absolutely-infinitely many sets, let alone Descartes' anticipating the correspondence between the algebraic structure of a variety (over finite fields) and the geometric/topological structure of the variety in the complex plane. All the foregoing are symbiotic systems of dualities, each generated by the underlying unity, ultimately the *ur* mother unity, the uni-verse, the unity that lives *by way* of the interactions of the duals in the symbiosis. A *correspondence*—of this structural kind—is the signature of a generative unity.

In sum, I would say that dualities there are many in Descartes, a whole manifold of complementary dualities. But their very presence witnesses the *unity* of Being that Descartes saw so starkly. A duality is the symbiotic manifestation of the unity that generated it and sustains itself by way of the interaction of the duals. This is true of algebra and geometry when taken in the most general respect—as in the aforementioned “Parfum de Descartes” conjectures; this is true of the human mind and body in their most general

²⁷ The idea of dualities as *generated* by a deeper underlying unity shows up in Almog 2007.

respect viz. as occurring in me, the human being; and it is true at the generator of it all, Nature, when taken in its most general respect viz. God.²⁸

University of Turku

Bibliography

- Almog, Joseph (2002). “Pains and Brains”, *Philosophical Topics* 30 (1), 1–29.
- Almog, Joseph (2007). “The Cosmic Ensemble: Reflections on the Nature–Mathematics Symbiosis”, *Midwest Studies In Philosophy* 31 (1), 344–371.
- Almog, Joseph (2010). “Dualistic Materialism”, in Robert C. Koons & George Bealer (eds.), *The Waning of Materialism*. New York: Oxford University Press.
- Balibar, Etienne (1998). *Spinoza and Politics*. Transl. Peter Snowdon. London: Verso.
- Descartes, René (1967–88). *Œuvres philosophiques*. Ed. Ferdinand Alquié. Paris: Garnier.
- Descartes, René (2015). *Correspondence between Descartes and Princess Elisabeth*. Ed. & transl. Jonathan Bennett. URL = http://www.earlymoderntexts.com/assets/pdfs/descartes1643_1.pdf
- Dieudonné, Jean (1975). “On the History of the Weil Conjectures”, *The Mathematical Intelligencer* 10, 7–21.
- Gottlieb, Daniel Henry (1996). “All the Way with Gauss-Bonnet and the Sociology of Mathematics”, *The American Mathematical Monthly* 103 (6), 457–469.
- Grothendieck, Alexandre. *Récoltes et Semailles. Réflexions et témoignage sur un passé de mathématicien*. URL = <http://lipn.univ-paris13.fr/~duchamp/Books&more/Grothendieck/RS/pdf/RetS.pdf>
- Hacking, Ian (2014). *Why Is There Philosophy of Mathematics At All?* New York: Cambridge University Press.
- Koistinen, Olli (204). “The Fifth Meditation: Externality and True and Immutable Natures”, in David Cunning (ed.), *The Cambridge Companion to Descartes’ Meditations*. New York: Cambridge University Press, 223–239.
- Langlands, Robert P. (2005) “Descartes ile Fermat”, *Matematik Dünyası* 2, 54–61.

²⁸ I owe thanks to Jani Sinokki’s dissertation (Sinokki 2016) and related conversations about a joint Descartes monograph project we are engaged in about the unitarian (non-dualist) Descartes. I also benefited from conversations with Lars Vinx, Hemmo Laiho, Tapio Korte, Jessica Pepp, Sarah Coolidge, Diego Comba, Mandel Cabrera, and Juha Räikkä, as well as long term off and on discussions with Etienne Balibar, Harry Frankfurt, Barbara Herman, Howard Wettstein, Paolo Leonardi and Emmanuela Scribano. The ongoing correspondence with two organic Descartes scholars—John Carriero and Denis Kambouchner—has enriched my thinking in umpteen ways, even if we have sometimes differences about what this master-thinker was after. Reading Descartes’ modern French editor, F. Alquié, has been very helpful. His annotations of meditation II and especially III are insightful. Finally, day-by-day joint teaching and joint reflection with Olli about Descartes, as well as Spinoza and Kant, have been most formative for me. Olli’s unitarian Descartes (and the reflection of Descartes’ themes in Spinoza’s unification-s) keeps stimulating my own dialog (“dialogue a distance”) with Descartes. And of course the robust human friendship only makes the pursuit richer.

- Langlands, Robert P. (2010) “Is There Beauty in Mathematical Theories?”, University of Notre Dame, January 2010. URL = http://publications.ias.edu/sites/default/files/ND_0.pdf
- Mazur, Barry (2004). “What is a Motive?”, *Notices of the AMS* 51 (10), 1214–1216.
- Sinokki, Jani (2016). *Descartes’ Metaphysics of Thinking*. Turku: University of Turku.
- Spinoza, Benedictus de (1985). *The Collected Works of Spinoza*. Ed. & transl. Edwin Curley. Princeton, NJ: Princeton University Press.

Why Things Matter: Camus' Meursault and Descartes' Causal Principle

Jani Sinokki

There is hardly another story that so vividly shows something about *significance*¹ than Camus' *The Stranger* (*L'Étranger*, 1942). However, what it shows about significance is not that easy to tell. One option is that via its peculiar narrator, Meursault, the story aims at establishing that the universe in itself is inherently *insignificant*. In the novel, Meursault is described as completely normal and functional human as anybody: he has a job, a girlfriend, and a libido, among other things. Yet he can attach no significance whatsoever to things like the death of his mother, his girlfriend's wish for him to marry her, and even his killing a man. Nor do those happenings arouse any ordinary emotional responses in him. His case can be taken as an argument for the unlikely possibility of completely significance-free human existence. Taken this way, the novel would seem to conclude that significance is merely a human projection on an otherwise indifferent reality, and therefore it is something distinct from things and happenings we for some reason happen to think of as having significance.

In this essay I examine another possibility. The story can also be read as an argument to a conclusion that is the opposite of the above-mentioned one. Meursault's character can be treated as a thought experiment about a thinking being that is disconnected from the reality in an important way. In other respects – biologically – he is just like any of us, but the difference lies in his inability to connect onto just those features of reality that make things have significance for us. Taken this way, the novel forms a *reductio ad absurdum* of the possibility of the existence of a being that is in other respects a complete human being, but for which there is no significance.

¹ *Significance* refers to whatever it is that makes things meaningful and worth something to us. Meaning and value are obviously closely connected concepts, but as in this essay I have specifically tried to avoid heavy theoretical commitments, and as those terms are subjects of many existing philosophical theories, I will limit my examination to the less-burdened, though admittedly bit awkward, term *significance*.

The story also includes the key to seeing what it is that makes up the significance. Meursault's deprived state is in fact constituted by his inability to map ordinary causal relations. He does not merely fail to see the significance of things, but his failure to see it is because he cannot see how things past, present, or future can or might affect him. That is, he fails to understand everyday causes and effects. To see this, I will take up a largely ignored feature of Descartes' metaphysics. Descartes thinks that reality is the totality of things bound together by causality. To be real – to exist – is to be part of that causally bound nexus. In the light of Descartes' view, Meursault's state is paradoxical: bodily he is part of our reality and acts along it, but his thinking remains an isolated island outside reality and cannot quite reach what is real.

With Meursault's case and Descartes' conception of causality and reality we have a rather simple way to understand the world and its significance, that is, why things in the world matter to us. In the end of the paper I will take up another insight from Descartes to defend the general line of the essay. Essentially I will shift the burden of proof onto those views that would make significance something more than a part of, or even something outside, the causal tissue of reality. But before we get there, I will discuss the case of Meursault and then move on to analyze his actual condition with the help of Descartes.

1. Meursault

The Stranger's main character, Meursault, is in most respects an ordinary person. But a striking feature is his first-person description of his inability to perceive, experience, or attach any significance to anything. That is, according to his testimony, not only does he lack the usual emotional responses that seem “normal” in certain situations, but he is also unable to understand how any of events, or even his own choices, might make any difference *for him*. This way or that way, it is all the same for him. I will recapitulate below the story's main points.

Meursault, an Algerian pied-noir, has received a letter informing him that his mother has died. He attends to the wake and the funeral. A bit oddly, the closest he comes to showing emotions – or observing them in himself – is by being mildly annoyed by his becoming conscious of his apologetic attitude about his relation with his mother. Otherwise, everything is all the same to

him; completely neutral and indifferent. The day after the funeral he goes swimming and meets an old acquaintance, Marie, with whom he soon develops a relationship. Despite his spending much time with her, he reports the affair as having little consequence to him. When Marie asks him to marry her, he says the matter is not here nor there for him, nor does he love her, but still he might as well please her by accepting her wish. In his job in an office he shows disinterest towards his current position as well as his future prospects that involve big changes in his life, including his possibly moving to Paris. Regardless of his indifference, at no point Meursault understands himself unhappy, but rather to the contrary.²

Meursault develops an amicable, though again for him insignificant, relationship with his neighbor, Raymond Sintes. Raymond is a hot-tempered dandy, and, a rumor has it, a pimp. Meursault agrees to write a letter to help Raymond to get back at his girlfriend – an Arab girl Raymond has been keeping – in a rather nasty way. The plan goes awry and Raymond beats up the girl. Meursault ends up testifying, still indifferently, for Raymond at the police station. As a result of this episode, girl's brother seeks revenge on Raymond. After a confrontation on a beach, Meursault accidentally bumps into the man again and ends up killing him. Meursault describes closely the chain of events leading to the killing: Though he realized that he should walk away when he encountered the man resting on the beach, he instead ended up taking a step further, maybe because of the discomfort caused by the scorching sun. The step in turn caused the Arab to draw his blade, and at the same time sweat from Meursault's brows filled his eyes, making him fire his gun. But after clearing his eyes, and seeing that the man lies motionless, he shot the man four more times.

The rest of the book focuses on Meursault's trial. At first, Meursault does not understand the consequences of his killing the man, but thinks that his interrogators will soon set him free. Rather peculiarly the subsequent hearing and the trial focus more on Meursault's character than on his crime. It seems that Meursault's unexplained decision to shoot several rounds to his victim would have been enough to get him convicted of murder, even despite

² Meursault tells, in the end of the first part, that he had been happy on the beach (Camus 1942, 39). Again at the end of the second, final part, he comes to realize about his life in general that "I'd been happy, and that I was happy still" (ibid., 76).

the second class status of moors in Algiers at the time (from which the attitudes towards Muslims in Europe do not seem to have progressed that much, it seems). The court understandably has troubles believing that Meursault's killing a man who also has a grudge against Meursault's companion – a notorious pimp, it is testified – happens by chance. Meursault's friendship with Raymond and their being on the beach earlier together, his using Raymond's gun in the killing, the letter to the victim's sister that is written by Meursault, and his testifying for Raymond at the police station make it quite unbelievable that the shooting was not premeditated.

Even in the face of this overwhelming evidence, the court ruling is more based on the testimonies about his callous behavior at his mother's funeral, his perceived atheism, and on the fact that he is not willing or able to neither explain his indifference nor make an effort to hide that indifference. In effect, he seems to be condemned less for the murder and more for being the way a human being should not, or cannot, be. Essentially, he is *sentenced for being non-human*.

Meursault's hidden crime – not being human – is reflected by the severity of his punishment. For his continental heritage and the fact that his victim is an Arab, Meursault would be likely to suffer milder consequences than a moor killing a moor, or a French man killing another, would suffer. Yet Meursault receives the capital punishment.

It seems that in the end Meursault himself comes to accept this underlying declaration of his non-humanity:

For all to be accomplished, for me to feel less lonely, all that remained to hope was that on the day of my execution there should be a huge crowd of spectators and that they should greet me with howls of execration. (Camus 1942, 76.)

This finale of the novel recognizes Meursault's being hopelessly unlike those who will greet him with howls. He finds comfort in the very fact that there is an unbridgeable gulf between him and *the others*, which no amount of diplomacy could ever remove, for his existence itself is threatening to what is

essential to the existence of those *others*. Because of this, the *others* must hate Meursault for being the way he is, and he finally comes to terms with this.

We might think Meursault as a version of philosophical zombies: he is physically just like a human, but his psychology is missing something crucial. Where philosophical zombies are imagined to lack all conscious experiences altogether, Meursault's case seems like a much more nuanced thought experiment that proceeds along same lines. Being a conscious being in a human body seems obviously a necessary condition for being a functional human, but Meursault's case suggests that this is not sufficient. If we accept the verdict of his not being a human, as I think we should (and which does not entail that we accept his punishment), then it seems that besides mere consciousness a certain kind of connection to significance must be essential to being human.

2. Causality

I wish now to direct our attention for a moment to another French writer, Descartes. He is, among other things, famous for dividing the reality into two realms, that of thinking beings and that of extended bodies. Human body was for Descartes mechanistically operated automaton, to which thinking – enjoying another, higher kind of being, than body – was related in some way, though it is not that clear how he thought these two things in the end are related.

I do not want to enter here into disputes about interpretation and scholarship. I would like to point only that there is a feature in Descartes' conception of how thinking and the human body are related that is often ignored. The point in question is Descartes' causal principle that is considered most extensively in the *Meditations* as a part of the 3rd Meditation argument for the existence of God.

Descartes first makes a claim about causality in general, for he holds it as self-evident that

there must be at least as much <reality> in the efficient and total cause as in the effect of that cause. For where, I ask, could the effect get its reality from, if not from the cause? And how could the cause give it to the effect unless it possessed it? (CSM 2: 28; AT 7: 28.)

One way to understand what this means is to think it analogically to one billiard ball hitting another: the kinetic energy transferred to the other ball must come from the first; otherwise it would have come from nothing, which is impossible. It is like this for anything, for any being or property, there is. It is clear that for Descartes causation requires that there is a continuity or transference of something – motion, energy, etc. – from the cause to the effect. Crucially, describing causation as a mere relation between the cause and effect does not capture the view presented here. The cause and effect must form a continuum of existence where the cause imparts some of its reality to the effect, if there is causation at all. The much later Humean question how we do or can know about causal relations between things is completely another matter, and should not be confused with Descartes' view.

Descartes' consideration is aimed at pointing out that this principle “there must be at least as much <reality> in the efficient and total cause as in the effect of that cause” is a universal, logical principle. The denial of it amounts to admitting uncaused things and events, which denial would be detrimental to all human cognition. I explain briefly.

If we accept that some things lack causes (which is not the same as accepting that there are things that have causes unknown to us), we accept that world is inherently arbitrary. Anything might happen at any time with nothing causing it; anything could come from nothing. Accepting this view would be absurd. In fact, accepting that some things lack causes would preempt any attempt to infer or deduce or judge anything. Any argument towards the conclusion that some things lack causes has to assume the opposite of this conclusion, Descartes' causal principle. The reason is that any piece of logical argumentation has to eventually rely on coherence, and without assuming Descartes' principle there is no such thing available, so argumentation without the principle is impossible. Since the negation “some things lack causes” of Descartes' principle cannot even be coherently argued for, his principle is on solid ground.

An interesting feature of Descartes' principle – especially in the light of how Descartes' “dualism” has been understood in later philosophy – is that Descartes argues that this principle also governs the existence of thinking:

But it is also true that the idea of heat, or of a stone, cannot exist in me unless it is put there by some cause which contains at least as much reality as I conceive to be in the heat or in the stone. For although this cause does not transfer any of its actual or formal reality to my idea, it should not on that account be supposed that it must be less real. (CSM 2: 28; AT 7: 41.)

Again, I do not wish to enter interpretational questions. What matters here is that just like outside our minds everything has a cause – assuming the opposite is absurd, as we noted – the same applies to thinking. Also everything we think of has a cause that imparts its reality to the thought. Of course, and just as was the case above, the point is not to say that we always *know* those causes. What we know is that even in the case of thoughts everything has a cause that is responsible for the existence of that thought, for its being in the way it is. There are no thoughts that lack causes: the thinker herself and her dispositions, and her causal relations with other things, at least, are part of the total cause of a thought she entertains. Of course, even if we would not be able to know all the relevant causes adequately, we would still know that for the way our thinking *now* is, there is a cause.

These aspects of Descartes causal principle result in the following picture. The causes imparting reality to their effects form a network of causally interrelated things. We, beings who think in, and of, the world, are part and parcel with that network. This means that the *reality* – the totality of things there exists – is bound together by causality. To be real is simply to be part of this causal structure, for anything that is not part of some causal line somewhere in it, is simply nothing. And nothing can come out of nothing, that is, have effects on things that are *something*.

I wish to suggest that we view ourselves with this framework. It does not matter that we do not know exactly what kind of causal relations are at play on the levels of nature we cannot observe directly. We can simply take the basic principle – that for everything there is a cause – and to view our own existence as a piece of the general fabric of reality bound by causality.³

³ I would like to diffuse a possible worry. Traditionally many have thought that causality and freedom of will fit poorly together. Descartes was not one of them, but instead thought that freedom of will is self-evident. So, for him any theory that denies free will can be rejected on the

3. Causal Detachment

I would now like to analyze the significance and its lack in Meursault with conception of reality borrowed from Descartes. Just to be sure, everything Meursault himself describes obeys causality in the novel in accordance to the principle. There are no uncaused events that Meursault faces. All his actions have definite causes though some of them like his pulling the trigger repeatedly might lack his reasoning or conscious decision. Still, his actions are not uncaused in any sense. He merely fails to see how at least some of them come about. Everything up to his punishment has a definite causal structure leading up to it. In the end, Meursault's inability to explain the four more shots and the pause in between the first and the subsequent shots affects the investigating Magistrate's opinion and the jury among other things. The same is true of Meursault's unwillingness during the investigation to explain his actions or to tell white lies about his non-existent feelings, which unwillingness again affects the verdict he faces.

According to Descartes' view, the following is true even of Meursault. Meursault is a part of the reality, which in itself is a complex network of causal relations reaching over time. His life forms a part of that network, though in the perspective of the totality a quite small part indeed. Still, he is part of the existence, so his life and actions affect other things, and other things affect his life, necessarily.

This fact is reflected by even the most ordinary affairs. For instance, Meursault's decision to marry or not to marry Marie affects both of their lives. Once he has to make the decision there is no escaping the consequences of the choice forced on him. Whatever he chooses, it will affect his, and also Marie's, life. If he opts for marriage, some future actions are no longer options for him. If he later chooses conduct not compatible with Marie's and others' expectations from a married man, there will be further consequences that would be wanting had he not married her. The point is that everything in his life from conducting daily routines and organizing spare time to possibilities of

grounds that it denies something that is self-evidently true. While this view itself might be thought suspect, it seems that Descartes had no problems in fitting it together with his view of causation. The reason lies likely in the fact that his view of causation as reality transmission is more organic than deterministic, and that uncaused action in any case hardly would constitute freely chosen action. We should thus *prima facie* hold on to both freedom of will and his logical principle and suppose no contradiction between them.

his self-realization and his flourishing as a being will be affected more or less by the choice he makes.

What we have here is an example of how mundane, yet complex, matter causality is. It is just the complexity of consequences of this kind that make even easy living occasionally difficult. It is not easy, or many times not even possible, to see in advance relevant consequences of one's actions and choices. Not to mention of fitting one's choices and desires to those of others. But regardless of this, it remains certain that there are always consequences. This is just what it is to exist, to be part of the causal tissue that is the reality.

Meursault's actual impairment comes across in a short piece of conversation he describes as follows (Camus 1942, 28–29):

Marie came that evening and asked me if I'd marry her. I said I didn't mind; if she was keen on it, we'd get married.

Then she asked me again if I loved her. I replied, much as before, that her question meant nothing or next to nothing—but I supposed I didn't.

“If that's how you feel,” she said, “why marry me?”

I explained that it had no importance really, but, if it would give her pleasure, we could get married right away. I pointed out that, anyhow, the suggestion came from her; as for me, I'd merely said, “Yes.”

Then she remarked that marriage was a serious matter.

To which I answered: “No.”

She kept silent after that, staring at me in a curious way. Then she asked:

“Suppose another girl had asked you to marry her—I mean, a girl you liked in the same way as you like me—would you have said ‘Yes’ to her, too?”

“Naturally.”

At first sight it might seem that Meursault is merely not giving any value to social institutions like marriage, or that he is lacking emotions towards Marie. It might seem that Meursault simply has no preference over one life course over the other that will inevitably follow on the basis of his choice. However, the roots of his problem seem much deeper. What he does not seem to consider are just the ordinary consequences of his choices. If he would consider those consequences, it would remain very difficult for him to rationally think that the choices he makes have absolutely *no* difference at all. There are obvious differences in the consequences, so he should be able to prefer one to the other, even on whimsical grounds. It should be noted that this is a matter that is completely independent of any possible emotional responses considering those consequences might arouse in him. The differences are something that Meursault should be able to detect even if he would lack emotions altogether. So, it seems that his indifference is in some way grounded in his not even considering his options adequately for one or the other reason.

This limitation in his considerations actually comes across all over his narrative. While incarcerated, Meursault is questioned on the shooting. He first believes that he will be released once the investigators know how things went down. But matters get complicated when he has to explain the incident, and is asked by the Magistrate for a reason for his firing the gun (Camus 1942, 42–43):

“Why did you fire five consecutive shots?”

I thought for a bit; then explained that they weren’t quite consecutive. I fired one at first, and the other four after a short interval.

“Why did you pause between the first and second shot?”

I seemed to see it hovering again before my eyes, the red glow of the beach, and to feel that fiery breath on my cheeks—and, this time, I made no answer.

During the silence that followed, the magistrate kept fidgeting, running his fingers through his hair, half rising, then sitting down again. Finally,

planting his elbows on the desk, he bent toward me with a queer expression.

“But why, why did you go on firing at a prostrate man?”

Again I found nothing to reply.

The magistrate drew his hand across his forehead and repeated in a slightly different tone:

“I ask you ‘Why?’ I insist on your telling me.” I still kept silent.

The only explanation Meursault is later able to come up with for his pulling the trigger again is that he did it because of the scorching sun (Camus 1942, 64). He does not even consider that it might make a difference whether he just shot five times or instead four more times *after* seeing that the man is defenseless from the first shot. Meursault’s condition on the basis of his own testimony is almost like an animal that has no capacity to recognize how his actions lead to consequences. He certainly has a preference of not being incarcerated or sentenced to death, but on all occasions he fails to act towards those goals. It is not that he merely does not consider any consequences his actions in the light of his preferences; it is that he does not understand those consequences. All along it is this inability to recognize ordinary causal relations that creates his indifference. If he would be able to see those consequences, he would be able to prefer one option over the other. At the minimum, he could at least prefer options that more likely are productive to his continued existence, and then act accordingly.

It is interesting to view Meursault as a being that is disconnected from the reality at a very fundamental level. That he is a thinking being, almost like any human being, but still not quite part of the same reality as we are. His understanding is not properly connected to the same causal network that constitutes the reality, but remains an outsider to it. He is tied to the world through his body, but his thinking is not quite able to penetrate into it. The more naturalistic explanations of his condition, say, a brain damage that impairs his cognitive abilities, are not that interesting. Such naturalistic explanations are uninteresting just because they appeal themselves to an

ordinary causal mechanism. And there are so many people around us that suffer from some such impairment, but still are not like Meursault. Meursault's only problem lies with understanding significance, but people suffering from brain damage often struggle with a host of other kinds of problems.

The whole novel can be read as a thought experiment. It invites us to imagine how the world would seem to us if we were mere bystanders in it, outside observers lacking a true access to it or not taking part in it. In effect, it describes truly an absurd entity, an as-if person that observes the world though is completely detached from the ordinary causal relations of the world. It is much like Meursault's consciousness would form another, distinct universe parallel to ours, though not quite connected with it the way we are. Whereas our thinking is part and parcel with the causal relations of the world, for Meursault, those relations appear like merely contiguous events he can observe, yet has no means to glue those events together properly.⁴ In this light, the thought experiment explains why understanding significance requires nothing beyond understanding causality: things become meaningful when you see how they connect.

So, the declaration underlying Meursault's sentence, according to which he is non-human, can point out something of the essence of what it is to be human. It is that we, thinking human beings, are not merely detached observers of the world. If we were, we would suffer the same condition that has befallen Meursault, the strange character a kin to philosophical zombie. We could not get to worldly causality, because our thinking would be in an important way distinct from it. We could maybe observe the world, but we could not understand it. In this way we can think of Camus' Meursault as a piece of philosophical fantasy that portrays a thinking being disconnected from the world. In effect, it forms a *reductio ad absurdum* of a view according to

⁴ The closest Meursault comes to understanding that there is such a thing as causality to his actions is after pulling the trigger: "And each successive shot was another loud, fateful rap on the door of my undoing" (Camus 1942, 39). He dimly realizes that he has done something that has severe consequences, but as the following trial and his hearing portray, he cannot understand those consequences.

which our thinking is distinct from the ordinary causal network that makes up reality (as well as its significance).⁵

The view discussed here might be objected by saying that maybe even in the condition of complete detachment we could project something like causality (or significance) on the worldly happenings. But then again, that seems to entail that we could at least perceive those possible deviant cases in which our projection fails, where our projection does not match the world. But this, in turn, would entail that we can after all get to the causality, though maybe only indirectly. Therefore the objection fails, for it applies after all only to cases that assume at least some kind of connection to the real causality, and not complete detachment of one's thinking from the other system of causal relations, the world. So, even after the objection we, in one way or the other, would be part of the world.

4. A Further Consideration

This far I have presented a case for thinking of the world and us in it, and significance things in the world have for us, in terms of causality. I pointed out that Descartes' causal principle cannot be easily contested, for rejecting it would require rejecting what is necessary for any argument. As that causal principle – understood properly as describing causation as a kind of continuous reality transmission as opposed to a mere relation between distinct things – entails that the reality is an interrelated causal system, we have a good reason to view even ourselves and our lives as parts of a larger whole, as opposed to our being singular, isolated thinking things. Meursault, in turn, can be taken as a literary presentation of a *reductio* of such a position that treats us as somehow detached from the rest of the causally-bound reality. With these conceptions we can then see that understanding significance, and probably related concepts of meaning and value, too, can be understood by focusing on the causal nature of the reality that produces us.

⁵ Though I would like to avoid underscoring, it seems that any theory of thinking that relies solely on features like *representationality* or *intentionality* of thought as an explanation of the mechanism of our access to the reality “outside” us is a suitable target to the *reductio* here described. However, discussing this more general question about the relation between thought and world would lead us too far away from the topic of this paper.

However, it seems to me that some views we have about ourselves, about the nature of our thinking especially, are so deep-ingrained to our philosophical imagination that I anticipate some opposition, from many directions. Therefore I would like put one final thought forward to defend my view. This thought is basically just an attempt to shift the burden of proof to the side it rightfully belongs, at least when our philosophical imagination is not misleading us to thinking otherwise. The thought in question is also from Descartes. He applies it to the objection about his philosophy making thinking and body distinct, and thus incapable of causal relations. More specifically, the questions posed to him are how thinking could have effects on the body, and how physical objects could affect thinking, given that thinking is immaterial while bodies material according to him. Here's his short answer, of which we should take heed:

The whole problem contained in such questions arises simply from a supposition that is false and cannot in any way be proved, namely that, if the soul and the body are two substances whose nature is different, this prevents them from being able to act on each other. (CSM2: 275; AT 7: 213.)

The underlying thought, as I see it, is based on Descartes' conception of causality and reality as a causally interrelated whole. If we begin by making a division inside a group of things that are interrelated – part of the same reality – how could we end up thinking that those things are not after all interrelated? But this is just how the worries Descartes is answering proceed. If someone claims that the possibility of some division or grouping of things in fact shows that certain relations between the things divided cannot exist after the division, while they were there before making it, the one making the claim certainly has to bear the burden of proof. Descartes points out just this. Further, it seem like an impossible task to show that thinking is *not* causally connected the corporeal body against all appearances to the contrary.

I would like to point out that Descartes' answer applies to the case made above for understanding causality, and also for significance. Descartes' causal principle – understood the way suggested in this paper – seems like a natural starting point for thinking about what is real. By far the simplest explanation

of how we (and our thoughts) connect to things is that they all are part of the same causal network that forms the reality. If someone were to argue otherwise, I think he or she should have to try to convince us of the need to draw something apart from that interconnected network, be that something thinking or significance, or something like moral values, or culture. Until good reasons are provided, we should stick with Descartes' reasonable causal principle and the picture of reality it engenders. That is, we should think that everything that is real – that exists – is connected in different ways by causality, and we should device our views and theories about those pieces of reality accordingly.⁶

University of Turku

Bibliography

- Camus, Albert (1942). *L'Étranger*. Transl. Stuart Gilbert. New York: Random House.
- Descartes, René (1988). *The Philosophical Writings of Descartes*. 3 vols. Transl. John Cottingham, Robert Stoothoff, and Dugald Murdoch (Volume 3 includes Anthony Kenny as a translator). Cambridge: Cambridge University Press. 1988. [CSM]
- Descartes, René (1983). *Oeuvres De Descartes*. 11 vols. Ed. Charles Adam & Paul Tannery. Paris: Librairie Philosophique J. Vrin. [AT]

⁶ This paper is indebted to Olli Koistinen for numerous discussions about Meursault, Camus, and Descartes, among other things. I would also like to express my gratitude to Joseph Almog for discussions on the same topics.

Descartes and Spinoza on the Love of God

Lilli Alanen

Spinoza was one of the fiercest critics of Descartes's dualist philosophy yet took over many of its ideas and conceptual apparatus in developing his secular ethical doctrine based on a new metaphysical framework consistent with the new science. He endorses not only Descartes's mechanistic physiology, but also some kind of explanatory dualism (attribute dualism) and some version of a Cartesian account of the passions as complex psychophysical phenomena. He sees the explanation of the passions as central to his moral project but seeks a coherent alternative to Descartes's account where the mastery of passions depends not on the postulation of a free power of decision of the will that Spinoza rejects but on the exercise of reason alone. One intriguing example of Spinoza's debt to Descartes is his definition of love that will be examined here. Spinoza's account of affects in the *Short Treatise*, according to many scholars a first version of the *Ethics*, is directly inspired by Descartes's theory of passions that Spinoza read in Latin translation. Spinoza later distances himself from Descartes's view. This paper considers some interesting similarities and differences between the view of our two philosophers on love and the highest good. After some introductory remarks on philosophical views of love I first summarize Descartes's account of love and of God as an object of love. I then look at how Spinoza uses and transforms these notions in the framework of his radically new understanding of human being, psychology and the nature of agency. I end with a brief discussion of Spinoza's doctrine of intellectual love of God and its implications.¹

¹ This paper is offered to Olli in testimony of a shared passion for early modern philosophy with an ensuing endeavor to increased understanding of the doctrines of two of its seminal thinkers. I have learnt from and enjoyed his insightful and thought-provoking work on Spinoza as well as from inspiring discussions over the years. The paper is extracted from a longer essay "Perfecting Nature: Descartes and Spinoza on the Highest Good" (in progress), which gives fuller arguments and evidence for some of its claims that could not be developed here.

1. Reflecting on the Topic of Philosophical Treatments of Love

Annette Baier comments that those “who have told us the ... most insightful things about love are poets and novelists”, whereas philosophers who “are supposed to be lovers of a sort, tend to be all thumbs when it comes to handling love.” Baier contrasts what she calls the theological tradition to a biological one represented by her hero Hume. The former is marked by a certain ambivalence that she spots in Plato, Augustine, Descartes and Kant, who are said to combine “a very strong definition of what the real thing would be, and a claim that no love between human persons will satisfy this definition.” According to the gloomy view she ascribes to them “human loves are doomed to failure. At their worst, they degenerate into mutual loathing, betrayal, and enmity; at their best they are interrupted by death and end in separation and bereavement.” The moral these “misamorists” draw from their experience, is to turn their love to God, a steadier, incorruptible object of love, protected from the risks and contingencies of earthly love. Descartes, she recognizes, is heretic in his adherence to this tradition (Baier 1994, 37). Spinoza is not mentioned at all in this context, but the opening paragraph in *Treatise on the Emendation of the Intellect* (II/5) offers this striking illustration:

After experience had taught me that all things which regularly occur in ordinary life are empty and futile, and I saw that all the things which were the cause or object of my fear had nothing of good and bad in themselves, except insofar as [my] mind was moved by them, I resolved at last to try to find out whether there was anything which would be the true good, ... which alone would affect the mind, ... which, once found and acquired, would continuously give me the greatest joy, to eternity.²

Spinoza finds a secure, eternal, object of love in what he calls God or nature. The intellectual love of God that we are taught to strive for in the *Ethics* is said to arise from the highest kind of knowledge, which comes with the highest

² I have used *The Collected Works of Spinoza*, edited and translated by Edwin Curley. Citations of passages from the *Ethics* are in the following form: the initial number indicates the Part of the *Ethics*; ‘p’ indicates a proposition, ‘a’ indicates an axiom, ‘s’ indicates a scholium, ‘c’ indicates a corollary, and ‘d’ indicates a definition (when immediately following a part number) or a demonstration (when immediately following a proposition number).

kind of joy or blessedness (5p32–p37). As Spinoza describes it, the intellectual love of God, which is also a kind of self-love – both god’s love of god itself and the self-love of the human mind – answer pretty well to the requirements set on the true good in the passage quoted: it is, as we will see, sufficient to affect the mind all by itself, and it is such that once found and acquired, it will continuously give you the greatest joy, to eternity. The way Spinoza develops it may seem closer to a pre-Christian, Aristotelian tradition than to that of Christian theology, but it also has, as I will show, surprising antecedents in thoughts Descartes develops privately in a letter to his friend Chanut.

2. God as an Object of Human Love

Having completed his first proofs of God’s existence in the Third Meditation, before proceeding to what he is about to deduce from it, Descartes reflects:

I should like to pause here and spend some time in the contemplation of God: to reflect on his attributes, and to gaze with wonder and adoration on the beauty of this immense light, so far as the eye of my darkened intellect can bear it. For just as we believe through faith that the supreme happiness of next life consists solely in the contemplation of the divine majesty, *so experience tells us that this same contemplation, albeit much less perfect, enables us to know the greatest joy of which we are capable in this life.* (AT VII 52; CSM II 36; italics added.)³

Puzzled, perhaps, by this unorthodox statement Pierre Chanut, French Ambassador in Stockholm, asked his friend Descartes to explain how it is possible to love God using natural reason alone without the aid of religion and divine grace. How can an object that is infinite in every sense inspire anything else than “astonishment (*étonnement*, here: wonder or amazement) and a very respectful confusion”? (Letter to Descartes, 1st of December 1646, AT X 611.)

³ Cf. the comments on this passage by John Carriero (2014, 20–23). In citing Descartes I use AT to refer to the standard edition of Descartes’ works by C. Adam and P. Tannery, *Oeuvres de Descartes*, and CSM for the English translation by J. Cottingham, R. Stoothoff and D. Murdoch of *The Philosophical Writings of Descartes*, Volumes I and II; and CSMK for *The Philosophical Writings of Descartes*, Volume III: The Correspondence, by the same translators plus A. Kenny.

The question is of great importance for early modern philosophy, and the way Chanut formulates it is interesting. How can an object that is infinite and surpasses us in every way, one of which we are totally dependent and have no control over, inspire any love, or work as our highest good in general? It can be addressed to Descartes as much as to Spinoza, even when their respective views of God are very different. Descartes's God is a version of a traditional, Christian God, a transcendent, omnipotent and infinite being, which albeit object of indubitable knowledge, surpasses our limited and imperfect human understanding. Spinoza's God again is immanent in nature of which we are finite parts. Although God-Nature is infinite and omnipotent – an infinite and eternal cause of itself – Spinoza thinks that the finite human mind not only can in principle know and understand God but that it also actually knows and understands God at least to some extent at all times. There is no radical incommensurability between the divine intellect and our finite minds for Spinoza – a view that Descartes would not accept. These differences in their conception of God are reflected in how Descartes and Spinoza respectively conceive of our highest good, what it is and, to the extent it involves love of God, how we come to love God. To understand their answers, we must consider their respective views of love.

3. Descartes on Love

Descartes distinguishes between clear and distinct love “which is purely rational (*raisonnable*)”, and love which is a confused thought and “a passion (*une passion*)” properly (To Chanut, February 1 1647, AT IV 601; CSMK 306).

The first in my view consists simply in that when our soul notices (*s'aperçoit de*) some good, present or absent, which it judges to be fitting for itself, it joins itself willingly to it, that is to say, it considers itself and the good in question as a whole of which it is one part and the good another. (AT IV 601; CSMK 306.)

This is Descartes's own peculiar definition of love, though reminiscent, perhaps, of Aristophanes' tale in Plato's *Symposium*. Love as Descartes here defines it is purely intellectual, and is based on (a) the perception of a good

and (b) a judgment that it is fitting for oneself. A wholly cognitive affair, it would seem, were it not for the fact that the cognitive act here includes (c) an act of will uniting the soul with the good so perceived: in Descartes gloss on it, the will right away ‘considers’ itself as forming a whole with the good in question. The act of will, as explained in the *Passions de l’âme*, is not the same as desire, which always relates to the future. It is instead an assent (*consentement*) whereby one “considers oneself from this very moment as if joined with what one loves in such a way that one imagines a whole, thinking of oneself merely as a part of it, and the thing loved another” (Art. 80, AT XI 387; CSM I 356). Love, so defined, is really a kind of judgment.

Descartes derives important consequences from this definition for ethics – for the “morale” developed in correspondence (with Elisabeth, Chanut, Queen Christina) and in the *Passions*. Thus, to Elisabeth he wrote, September 15, 1645, that one should always put the interest of the whole of which one is part before one’s own particular interests, whether the good one is thus joined to is another person, one’s family, society, king or nation (AT IV 293). To Chanut he writes:

It is the nature of love to make one consider oneself together with the object loved as a whole of which one is but a part, and to transfer the care one usually has for oneself to the preservation of this whole in such a way as to keep for oneself in particular only a part of it [of the care], which is as great or little as one deems one’s own part of the whole to which one has given one’s affection. From all this it is obvious that our love for God should be, beyond comparison, the greatest and most perfect of all our loves. (AT IV 611–612; CSMK 311.)

Although Descartes does not (as Spinoza does in *Ethics*, Part Four) dwell on this point, self-love is here the starting point, so the care one has for oneself is simply transferred to the whole one has thus formed through one’s will and of which one henceforth considers oneself as part. The whole so formed constitutes, as it were, a larger self. Depending on the value of the object as compared to yourself, be it a material object, another person, a community or God, the part that it constitutes of the whole you now identify with, your love

and care for it will be greater or lesser than the one you have for yourself, and the greater it is, the less you will hesitate to sacrifice yourself for it.

As to the Scholastic distinction between love as benevolence and love as concupiscence Descartes sees these, as well as the desire to possess the object loved, as effects of love not part of its essence (*Passions*, art. 81, AT XI 388). Considering oneself as united, *de volonté*, with the object loved is thus neither to possess nor to desire to possess it. It is rather like a commitment to treat the object as part of a larger self, and once this emotion or will to love is in place, benevolence follows.

Before considering the love of God, which in the passage quoted is said to be the greatest and most perfect of all loves, we need to look at the account Descartes gives of the other kind of love mentioned at the beginning, love as a passion. The latter depends on the body (the animal spirits) and the senses, and this passion of love is said to commonly accompany the rational love described above, as long as the soul is joined to a body. While the rational love is a “clear and distinct” thought – a judgment really, love of this latter, sensuous kind, “is nothing but a confused thought, aroused in the soul by some motion of the nerves”, and just like other passions, these motions and the “confused thought” caused by it also affect the soul or its will, making “it disposed to have the other, clearer, thought which constitutes rational love.”(AT IV 602–603, CSMK 306.)

How can a confused thought dispose the soul to a clearer thought? According to Descartes, a strong passion of love inclines the will (the motive power of the intellect as it were) to judge that the object causing it is a good, so worth uniting oneself to. These two kinds of love, the rational love, which depends on a distinct perception and judgment of the soul or mind alone, and the passion, which depends on the organs and affections of the body, commonly occur together in this present life. It is not just that the bodily movements cause a confused thought which is a sentiment or sensation, say of heat around the heart, but moreover, the latter affects the will disposing it to consider itself as united to whatever good presents itself to it, or is associated in thought to this sentiment. Passions thus are complex psycho-physical phenomena: there is, on the one hand, always a thought or an intellectual element even in the obscurest passion caused through the body, and, on the other hand, rational love never occurs without some passion or bodily

affection. The two kinds of love normally occur together: as soon as “the soul judges an object worthy of it”, this makes “the heart disposed to the motions which excite the passion of love; and when the heart is similarly disposed by other causes, that makes the soul imagine lovable qualities in objects in which, at another time, it would see nothing but faults.” (AT IV 603; CSMK 307.) So were it not for the special disposition of the heart, that is, the feeling of warmth occurring in love, an object, no matter how good it is, might not be perceived as worth loving at all by the soul. This connection or “association between each of its thoughts and certain motions or conditions of this body” is natural, something that Descartes goes on to explain with great detail, speculating about the first emotions of love and joy experienced by the soul of the foetus in the womb of the mother when receiving suitable nourishment:

I consider it probable that the soul felt joy at the first moment of its union with the body, and immediately felt love, then perhaps also hatred, and sadness; and that the same bodily conditions which then caused those passions have ever since naturally accompanied the corresponding thoughts. I think that the soul’s first passion was joy, because it is not credible that the soul was put into the body at a time when the body was not in a good condition; and a good condition of the body naturally gives us joy. I say that love followed because the matter of our body being in constant flux (*s’écoulant sans cesse*) like the water in a stream, there is always need for new matter to replace it, so that it is scarcely likely that the body would have been in good condition unless there were nearby some matter suitable for food. The soul, uniting itself willingly to that new matter, felt love for it; and later, if the food happened to be lacking, it felt sadness. And if its place were taken by some other matter unsuitable as food for the body, it felt hatred. (AT IV 605; CSMK 307–308.)

Being united to the body, the soul loves, i.e. joins itself by will to the body and to whatever nourishes it. What is good for the preservation and wellbeing of the body will necessarily make it feel love and joy, and what hurts it, for instance, the lack of what it needs to do well and grow, makes it feel sadness (and perhaps anger). Here, in the letter to Chanut, Descartes derives from this

first sentiment of love the five basic passions (love, hatred, joy, sadness and desire), which are part of his list of what in the *Passions of the Soul* are called primitive passions, and from whose various combinations all other passions are formed. To begin with, they were all “only sensations or very confused thoughts”, because “the soul was so attached to matter that it could not yet do anything else except receive various impressions from it.” (AT IV 605, CSMK 308.)

The derivation parallels that of the rational emotions from the movement of will in intellectual love, which is given at the beginning of the same letter. Thus, once the soul deems an object worth loving, and this object is present, that is, when the soul is joined to it not only by will (in thought that is) but also in fact or actually, then the movement of the will caused by the knowledge of being united to the good constitutes its joy. On the contrary, when the good is absent, the movement of the will, which follows from knowing that it does not possess it, is its sadness. Finally “the movement of the will which accompanies the knowledge that it would be a good thing to acquire it is desire.” All these emotions, love, joy, sadness and desire, qua rational thoughts, are mere movements of the will – the will that for Descartes is part of intellect or reason, so “could exist in our soul even if it has no body”, while the corresponding passions are confused thoughts caused by movements in the body.⁴

The will, as Descartes explains elsewhere has these two operations or acts as it were: assenting and denying on the one hand, and pursuing or avoiding on the other (Fourth Meditation, AT VII 56–57). It is of the nature of the will to assent to the true and to pursue (tend or move toward) the good. As soon as a good is distinctly perceived the will unites itself to it, treating it as a part of a larger self. To Chanut he writes:

[I]f the soul perceived that there are many fine things to be known about nature, its will would be infallibly impelled to love the knowledge of those things, that is, to consider it as belonging to itself. And if it was aware of not having that knowledge, it would have sadness, and if it

⁴ One should be careful with the term motion or movement of will here. What Descartes calls movements of the will are purely rational acts of the soul, i.e., clear distinct judgments about the good instantly joining the soul to it, and have nothing in common with the movements causing and accompanying the passions which occur mechanically in the body.

thought it would be a good thing to acquire it, it would have desire. There is nothing in all these movements of its will which would be obscure to it, or anything of which it could fail to be perfectly aware, provided it reflected on its own thoughts. (AT IV 602; CSMK 306.)

I will hereafter speak of these movements of will as ‘intellectual emotions’, by contrast to the confused affections or ‘passions’, which are caused by movements in the body and are psycho-physical states. The former depend on the soul and are active or express its activity, the latter depend on the body so they are passive: they are ‘passions’ in Descartes’s strict sense of the word. It is important to remember that although Descartes makes a point of distinguishing between the two, and this distinction is crucial for his ethics, they occur in reality and fact together. It is also noteworthy that love, from which the others are derived, is essentially an intellectual emotion. This judgment by which the soul considers itself united to the object of love is mentioned also as part of the definition of love as a passion. The latter, no matter how confused, seems thus to always include a judgment, an intellectual element.

4. Descartes on the Love of God

There are two sorts of reasons, Descartes notes, which make it doubtful that one could love God without the help of grace. First of all God, or his attributes, are so high above us that it is hard to see how they can be fitting for us, which is what would make us willing to unite ourselves to God. Secondly, “nothing about God can be visualized by the imagination”, so even if one could have an intellectual love for God, one would not have any accompanying passion to support it. This is why some think that we need Christian religion and the mystery of incarnation to make us love God. Descartes does not share these doubts, however. He argues, revealing his heresy, that it is possible to truly love God by the “sole power of our nature”, leaving it to theologians to determine whether that would count as a merit without the assistance of divine grace... He declares, “that with regard to the present life it [intellectual love of God] is the most delightful and useful passion possible; and *it can even be the strongest, though only if we meditate very attentively, since we are continually distracted by the presence of other objects.*” (AT IV

608; CMSK 309; italics added.) Consider carefully what kind of meditations he recommends to produce in us the love of God. For sake of clarity I list them as separate points. We should consider:

- (1) That God “is a spirit (*esprit*) or thinking thing”;
- (2) that the nature of our soul resembles his and is “an emanation of his supreme intelligence, & *divinae quasi particula aurae*”;
- (3) that our knowledge can increase gradually, it seems, *ad infinitum*;
- (4) that God’s knowledge is infinite and is the end to which our knowledge aims;
- (5) that if we stopped here (at the consideration of (1)–(4)) we might succumb to the extravagant wish to be gods, and by a very great mistake love only divinity instead of loving God (a mistake Descartes might have charged Spinoza for). The following considerations save us from this hubris:
 - (6) that God’s power by which he has created all things is infinite;
 - (7) that we are only the tiniest part of these things;
 - (8) that God’s providence by which he sees in a glance “all that has been, that is, that will be, and that could be” is unlimited;
 - (9) that God’s decrees are infallible and immutable (AT IV 608); and finally,
- (10) that we must weigh our smallness against the immensity (and infinity)⁵ of all things created, observing how they depend on God and regarding them in a manner proper to his (infinite) omnipotence – instead of “enclosing them in a ball (*une boule*)” as those who insist that the world is finite (the Aristotelians).

He concludes in exalted terms that could have been approved by Spinoza if not to the letter, at least in spirit:

- (11) “When understanding all these things properly, meditating on them fills a man with extreme joy...the knowledge with which God has honored him is enough by itself to make his life worth living, and

⁵ Although ‘infinity’ is not explicitly mentioned here, God’s infinite power is mentioned above in (6).

joining himself by will entirely to God, he loves him so perfectly that he desires nothing at all except that his will should be done. Because he knows that nothing can befall him that has not been decreed by God, he no longer fears death, pain or disgrace. He so loves this divine decree, deems it so just and so necessary, and knows that he must be so completely subject to it that even when he expects it to bring death or some other evil, he would not will to change it even if, *per impossibile*, he could do so. He does not shun evils and affliction...still less does he shun the permissible goods or pleasures he may enjoy in this life, since they too come from God. He accepts them with joy, without fear of evils, and his love makes him perfectly happy.” (AT IV 609; CSMK 309–310.)

If Descartes in this letter seems to agree with Spinoza that our highest bliss is the intellectual love of God and the joy that causes, their ways to this goal are not the same. For Spinoza, it goes through adequate knowledge alone, which in itself causes the most perfect active joy – Spinoza’s counterpart of Descartes’s intellectual joy, but which because it is a pure (self-caused) intellectual activity excludes, by its nature, the passion of joy that Spinoza calls ‘passive affect’ and must be caused by external things affecting the body (see below). For Descartes too the intellectual emotion is an ‘activity’ in so far as it is caused by the soul, but is normally accompanied by a passion that supports it, even though as we saw passions may also be distracting when too confused and violent.⁶ Not only is it hard for the soul to detach itself entirely from the senses in order to represent the truths arousing this perfect intellectual love, it is moreover not possible to communicate it, merely by willing it, to the imagination “so as to make it a passion”. Imagination needs aid from the senses and its images of corporeal things to work, and the passions need imagination to be sustained. Indeed towards the end of his letter to Chanut, Descartes confesses that though he does not think such metaphysical thoughts “hold any difficulty for your [Chanut’s] mind”, his own mind tires easily by them, “and the presence of sensible objects does not allow me to dwell on such things for too long.” (AT IV 613; CSMK 331.)

⁶ The natural function of passions being to support and strengthen the thoughts causing them and that are good or useful for it or for the mind-body union. See *Passions*, AT XI 385; CSM I 354.

Before making this blunt admission, he gave Chanut this glowing account of how the proper passion could nevertheless be aroused:

For if we cannot imagine anything in God, who is the object of our love, we can imagine our love itself, which consists in willing to unite ourselves to some object. That is, we can consider ourselves in relation to God as a minute part of all the immensity of the created universe...*and the idea of such a union by itself is sufficient to produce heat around the heart and cause a violent passion.* (AT IV 610; CSMK 310; italics added.)

The passion, *nota bene*, depends on the consideration of our self as a minute part of the infinite universe! From the point of view of orthodoxy, this pronouncement would be problematic for more reasons than one. Just before, as we saw, Descartes had committed to the infinity of the created universe – a heresy in itself. And now we are talking of uniting ourselves not to God, but to his creation, almost as if, like for Spinoza, they were one and the same thing. (Descartes will have a lot of explaining to do, both to Chanut, to Christina and to other Christian philosophers who take him to task about the infinity of the created universe...) Note also the intense affect or passion that such thoughts are supposed to produce: thinking of being part of the immense universe created by God by itself causes the movements in the brain and heart of the passion of love. Spinoza agrees on the intensity of this affect that he would argue is produced by reason or intellect alone. But since he denies mind-body interaction, he could not agree on Descartes' explanation of it. He assumes instead, without explaining it, that what is an activity in the mind is an activity in the body, so the active affect of love produced by the idea of God, in his story too, has to be matched, by parallelism or identity theory, by corresponding bodily effects, i.e., an increase of bodily activity or force.

5. Spinoza on the Intellectual Love of God

Spinoza's God has no ties to Christian divinity, but is an immanent and infinite self-causing power identified with nature. Spinoza radically breaks with the tradition in declaring that since everything happens of necessity in nature – nature's laws being everywhere the same –, there can be nothing good or bad in nature. If Spinoza's doctrine has many striking similarities to

that of the ancient Stoics, it departs from theirs through his strong identification of God and nature and the rejection of any kind of divine providence (see Long 2003). What we consider good depends on desire, and we do not desire a thing because it is good, on the contrary: it is good because we desire it (3p9s). There is, however, one line of thought that Spinoza retains from the tradition (and from Descartes), namely, the ancient identification of being or reality with perfection as unimpeded activity, and with it the notion of degrees of reality or perfection understood as degrees of activity. In his universe of necessity there is no room for changing things, yet there is, remarkably, room for improvement in the sense of increased perfection. His *Ethics* is after all a book of therapy and moreover a guide to salvation – showing the way to the highest good and happiness, the latter arising only through the increasing perfection that the right use of reason and true self-esteem or contentment, *acquiescentia in se ipso*, can bring about (4p52s).

The perhaps most radical innovation in Spinoza's system is his rethinking of the place and nature of the human being and human agency in the infinite physical nature. We are not substances and though we are rational we are not individual self-moving agents in any traditional sense of the word. We are instead transient modes of the one and only individual substance there is: infinite self-causing nature, and whatever activity we manifest depends on the substance of whose causal power we or our own power to persevere are manifestations.⁷ We perceive the substance under the only two attributes our minds can apprehend: thinking and extension. So we are, at once, modes of thinking and modes of extension: minds and bodies, where our mind is related to our body as an idea to its object. Our mind is an idea (or a system of ideas), which is part of God's infinite intellect, and its object is the system of motion and rest that defines our body – a determinate, finite part of the infinite physical world. Because of the peculiar identity of our mind and body, there is no action or activity in the mind that is not an action or activity in the body, and vice versa. The same holds for passivity: the mind acts and suffers with the body. (3p2s.)

⁷ Olli Koistinen (2009a) offers an interesting reading of this. Cf. Koistinen 2009b.

6. Desire as a Kind of Striving

When Spinoza speaks of singular or particular things (bodies), he speaks of finite systems of motion and rest immersed in and sustained by other finite systems concurring with or opposing the basic striving through which each of them tends to persevere in its own being. This striving (*conatus*) that moves you and defines your being is what Spinoza calls appetite, and when it is “accompanied by consciousness”, desire or will. Desire relates as appetite does to mind and body, will relates in his vocabulary to mind alone, but these are names only and there is no real distinction between striving, appetite, desire and will. (3p9d,s.)⁸

Desire – appetite in so far as we are conscious of it (3p9) – is one of the three “primary” affects (with joy and sadness, 3p11s) and is also the first on Spinoza’s list of passions. It is defined as “man’s very essence, insofar as it is conceived to be determined, from any given affection of it, to do something” (3 Def. Aff. 1). He adds that desire, as said before, “is appetite together with consciousness of it. And appetite is the very essence of man, in so far as it is determined to do whatever promotes his preservation.” Spinoza specifies:

Here ... by the word desire I understand any of mans’s strivings, impulses, appetites and volitions, which vary as man’s constitution varies, and which are not infrequently so opposed to one another that the man is pulled in different directions and knows not where to turn. (3 Def. Aff. 1 Expl.)

The consequence of this definition, as spelled out already in 3p9s, is that we have no special faculty of will or volition directed at the good. We are not moved by ends at all, but by this basic *conatus* force driving us – the changing systems of motion and rest that we are – to seek our self-preservation in whatever we do. This is not, Spinoza scholars emphasize, an end-directed

⁸ I differ on this point from Koistinen (2009b) who sees the will as an (intellectual) affirmation of the body’s power to persist. If the will, which is desire referred to the mind alone, has any distinctive power of its own, it is, I argue, that of affirming adequate ideas. See my “Affects and Ideas” (forthcoming). For interesting very different recent readings see Huebner (forthcoming) and Carriero 2015.

force. It is rather an inertia like force, striving to maintain the system as it is in its being. Spinoza puts it very clearly:

From all this, then, it is clear that we neither strive for, nor will, neither want, nor desire anything because we judge it to be good; on the contrary, we judge something to be good because we strive for it, will it, want it, and desire it. (3p9s.)

You may have thought, as you set out on All Saints Day on a long journey to light a candle on the grave of your beloved deceased (a North European custom), or when (following a North American one) you are planning to fly across the continent to visit your family on Thanksgiving, that you act for considered reasons out of your free decision. Although this may appear so to you, Spinoza teaches that in deciding, or acknowledging that this is something you want to do and choose to do, you are merely registering (becoming conscious of) an appetite that already moves you this way but the causes of which you are not fully aware.

Passions, passive affects as Spinoza calls them, are all determinate modifications of our basic striving. Your striving depends on your bodily constitution and is under constant pressure from surrounding things affecting you, sometimes inhibiting or opposing, sometimes supporting and enhancing your constitution (the ratio of motion and rest determining your constitution). What affects your body affects your mind. (3d3.) Spinoza describes this in terms of diminishing or increasing your power of acting. Affects are transitions in your power of acting and their ideas from greater to lesser or lesser to greater power or force. Ideas of things that increase or diminish the body's power of acting increase or diminish the mind's power of thinking. This yields Spinoza's definition of joy as an affect that "increases the mind's power of thinking", i.e., "by which the mind passes to a greater perfection", and of sadness as "a passion by which it passes to a lesser perfection" (3p11s). Desire, love and joy are thus the three basic or "primary affects" in Spinoza's system of which all the other passions are composed.

7. Spinoza's Transformation of Descartes's Concept of Love

Spinoza defines love simply as “a joy, accompanied by the idea of an external cause [*laetitia concomitante idea causae externae*]” (3 Def. Aff. VI). This definition “explains the essence of love clearly enough.” The will to join oneself with the object of love, Spinoza argues, may be a property of love but does not express its essence. Descartes is not mentioned in this context but that he must be on Spinoza's mind is clear from the explanation he gives of how “property of love” should be understood here.

[W]hen I say it is a property in the lover, that he wills to join himself to the thing loved, I do not understand by will a consent, or a deliberation of the mind, or a free decision (for we have demonstrated that this is fiction in *Iip48*). Nor do I understand a desire of joining oneself to the thing loved when it is absent or continuing in its presence when it is present. For love can be conceived without either of these desires. Rather, *by will I understand a satisfaction (acquiescentia) in the lover on account of the presence of the thing loved, by which the lover's Joy is strengthened or at least encouraged.*” (3 Def. Aff. VI; my italics.)

What Descartes sees as uniting oneself by will to the thing loved, Spinoza sees as the joy or self-satisfaction caused by the ‘presence’ of the thing loved. Spinoza, like Descartes, thinks that love causes different kinds of desires, but insists, like Descartes, that these are not part of its essence. Its essence is the joy or satisfaction (an increase of perfection) at the thought “of the presence of the object loved”. Spinoza's transformation of the Cartesian notion where the essence of love is the volition to form a union with the thing loved, enlarging as it were oneself to the whole now formed with the object loved, is very subtle. Where Descartes speaks of a judgment uniting one by will to a thing (i.e., to consider oneself as part of the whole thus constituted), Spinoza talks of the ‘presence’ of the object which causes joy. But things can be present many ways, physically or spatially, or in imagination or thought. Just thinking of or imagining a thing is, according to Spinoza's psychology, having it present to one – the traces in one's body of the thing thought of being activated is enough for calling it present, for this can affect you just as much as if it were there in front of you. (Thus you can, for instance, rejoice in thinking of people

that are dear to you or of a community you'd like to consider yourself a member of, whether they are near you or distant, perhaps even merely imagined, in a fairy land.)

That Spinoza finds it important to dissociate himself from the Cartesian notion of love has to do, I argue, with his different understanding of the nature of will and desire. Love is essentially joy, an increase of perfection associated with the idea of an external thing. The increase of perfection is itself an augmentation of your striving to persist or appetite (which is called desire when it is conscious). The thing thought of could still be, as in Descartes's account in the letter to Chanut, a flower, a human being, a friend or a sovereign, or a group of distant fighters for a cause you share, or whatever else may give you joy to think of as being part of or as related to you. It could even be God though it does not seem right, all things considered, to call God, who is omnipresent, an external thing for Spinoza,⁹ except in the sense that God is infinitely greater than you are. Yet it would, I surmise, suit Spinoza's view well to describe this as considering oneself as part of a whole of which the thing loved is another part, since this is precisely what finite minds are, ideas and as such parts also of the infinite intellect which encompasses all these other ideas of things, the thinking of which gives finite minds joy or pleasure. Indeed it is the consideration of the (true) greatness of the infinite God nature and my finite mind as a part of it that is supposed to cause the greatest joy there can be for Spinoza. On this last point Spinoza seems to be in agreement with the view expressed in the letter to Chanut quoted above.

8. Human Love and Divine Love

Spinoza, however, is more restrictive than Descartes on the proper object of love, for he seems to think that love if not possessive by essence unfailingly degenerates into possessive love. Thus we read about the dangers of ordinary human love:

It should be noted that sickness of the mind and misfortunes take their origin especially from too much Love toward a thing which is liable to many variations and which we can never fully possess. For no one is

⁹ Cf. Yitchak Melamed who sees a problem here in his "Spinoza's *amor Dei intellectualis*" (forthcoming). I am grateful to the author for letting me read this paper in manuscript.

disturbed or anxious concerning anything unless he loves it, nor do wrongs, suspicions, and enmities arise except from Love for a thing which no one can really fully possess (*quarum nemo potest revera esse compos*). (5p20s.)

There does not seem to be any proper, reliable object of love other than God. Why does Spinoza seem to think that love, to be free from envy or jealousy requires full possession of the object loved? The following line of thought might explain this. Any affect of joy, any increase in the power to persevere, is by nature such that it seeks to perpetuate itself and to grow. (See 3p7–p13.) Naturally, then, no finite, inconstant object can satisfy it but is bound to make us, sooner or later, miserable. We, our minds, require an object of love such that none can take it away from us, an object that is immutable and eternal. Such a thing can only be found through what Spinoza calls the third kind of knowledge, intuitive knowledge (explained in 2p45) “whose foundation is the cognition of God itself” and that is out of reach of ordinary passions which are always based on inadequate cognition. No vices or misfortunes can threaten that love, the continued joy produced by the contemplation of God. (5p20s.)

What then can we, or “the Mind, considered only in itself, ... do against the affects”? (5p20s.) The only cure for the latter depends on the third kind of cognition and Spinoza’s notion of “ordering the mind by the third kind of knowledge”. Spinoza writes:

In so far as the affects are passions, if clear and distinct knowledge does not absolutely remove them (see p3 and 44s), at least it brings it about that they constitute the smallest part of the mind (see p14), and it then begets a love toward a thing immutable and eternal (see p15), which we really fully possess (2p45), and which therefore cannot be tainted by any of the vices which are in ordinary love, but can always be greater and greater (by p15), and occupy the greatest part of the mind (by p16), and affect it extensively with joy. (5p20s.)

The originality and beauty of this doctrine is that this object of love, which is an object of cognition (of theoretical contemplation an Aristotelian might say) is something that we as matter of fact already possess! Our love for it can

therefore only grow – become greater and greater, it can continue to affect our mind with more or greater joy – but it can never be taken away or degenerate into something ugly. For Spinoza, cognition of God is something a human mind already has (whether we know it or not) because it is involved in any idea of any particular thing, and since all things are in God none can be conceived without God. (2p45.) This holds also for the ideas of the affections of the body, and one would think, of any affect or passion as well, envy, hate and misery included. There is nothing good or bad in nature. More to the point, there actually is nothing bad in nature, what is bad is nothing really but a lack of being or perfection. (Thus as soon as we understand God to be the cause of sadness, i.e., understand the true causes of sadness, it is supposed to cease to be sadness and be replaced by the joy of understanding. This I gather should hold for any evils, including the miseries we take to be caused by wars, terror and tyranny. (See 5p18.)

This is hard to understand but this is what we must try to understand if we are to make any sense of Spinoza’s therapy of passions and concept of virtue. For, Spinoza argues, the more a person understands herself and her “affects clearly and distinctly loves God”, and “the more she understands herself and her affects”, the more she loves God (5p14). Love, as we have seen, is joy accompanied by the idea of its (external) cause. Now if we could learn always to refer whatever idea of finite “external” objects we take to cause our passions, including those that diminish our power of acting, to God, who is the only true cause of the effects we perceive, we would love God, i.e., rejoice. The object, I presume, appears as “external” only because of the truncated and inadequate ideas of our imagination, which belong to the first kind of cognition that dominates our confused minds in the ordinary course of nature. Once we learn to see things as they are, clearly and distinctly, we understand that we are these finite transient modes of God, which means that we are in God and God is within us, and we can then marvel at the richness, intricacy and necessity of the causes explaining our present misery, and see how it shrinks and becomes insignificant as compared to the activity and joy that now occupies the greater part of our mind.

9. Love, Activity and Self-Satisfaction

An interesting consequence of this doctrine, which again may seem to bring Spinoza's doctrine closer to Descartes, is that the greatest, true kind of love, the love of God, and the bliss it involves – “the greatest satisfaction of mind there can be” (5p32d) – is actually a kind of self-love. Spinoza, as Melamed (forthcoming) suggests, identifies God's intellectual love¹⁰ with *acquiescentia in se ipso* – a notion he says is close to intellectual love and is borrowed (again!) from Descartes' treatise on the passions. In fact, Spinoza's notion of contentment comes very close to one of Descartes's two senses of self-contentment. In Descartes's account it (*satisfaction de soi-même*) comes in two kinds, one is the “satisfaction of those who steadfastly pursue virtue” and the other “is a habit of their soul which is called ‘tranquillity’ and ‘peace of mind’”. The other he describes as follows:

But the satisfaction we gain when we have just performed an action we think good is a passion—a kind of joy (*une espèce de joie*) which I believe to be the sweetest of all joys, because its cause depends only on ourselves. (AT XI 471; CSM I 396.)

Spinoza, in the same spirit, defines self-satisfaction as “a Joy (*Laetitia*) born of the fact that a man considers himself and his own power of acting” (4p52d). Further on, he identifies self-satisfaction with God's intellectual love, i.e., with the Mind's intellectual love of God, which is part of the infinite love by which God loves himself. (5p36.)

The sweetest and greatest joy there is, according to our philosophers, is the one which we cause ourselves. But note that what is a passion or passive affect for Descartes, is, in so far as it is produced by our (mind's) own activity, an active affect for Spinoza, and this difference may prove important.

While Spinoza takes pride, as it seems, in our rationality and the share we have through our reason in the eternal intellect, Descartes sees the rational use of the will here and now as the only legitimate ground for self-esteem. On the surface these two conceptions may not seem so far apart. When

¹⁰ God's intellectual love is our intellectual love for God, which is nothing but God's love for himself (5p36).

considered more closely and in detail, the differences are significant, but can only be gestured at here.

To be affected by passions is for Descartes a good thing in itself, provided the passions are not excessive but remain controlled by reason. He holds that it is on the passions that all the good and evil of this life depends, and that the persons who can be most deeply moved by the body-dependent passions, “are capable of enjoying the sweetest pleasures in this life” (Art. 212, AT XI 488; CSM I 404). He also thinks it is the passion of joy (accompanying self-satisfaction) that causes the highest bliss, and while it cannot be everlasting, it still depends on one self.¹¹ Descartes’s moral doctrine, moreover, does not aim at removing but at using and enjoying the passions, something that he also thinks anyone with the full use of her reason should be able to do: taking pleasure in being moved by them even when they are painful. (Art. 212.)

Spinoza’s picture of the ordinary human passions is more negative since like the Stoics he sees them as perverting reason: they are primarily truncated and inadequate ideas of imagination which belong to the first kind of cognition that is the realm of confusion and error. Nothing really good can follow from them. Although Spinoza thinks they are unavoidable and that our finite minds will suffer from them in the ordinary course of Nature as long as our bodies last, he clearly holds that we would be better off without them. If the passions cannot be removed, we can at least, Spinoza believes, consider ourselves as removed from them, i.e., we can learn to see that what is our ‘true self’ cannot be troubled by them. We can learn this to the extent we learn to consider ourselves (our minds) as eternal, as parts of the eternal intellect. This is the most perplexing, and in my view, troubling aspect of Spinoza’s doctrine, and raises hard question about the identity of the self that his therapy is supposed to set free.

Spinoza’s universe contains only one individual agent, God or *Natura naturans*. When he speaks of power of action he means power of understanding or intellectual activity, rather than performing actions in any current, worldly sense of actions. “For the eternal part of the Mind (by p23 and p29) is the intellect, through which alone we are said to act (by 3p3).” (5p40c.) What

¹¹ The right use of the will – the highest virtue Descartes calls generosity – which is crucial here is discussed in the longer essay referred to above in note 1.

human beings, including Descartes, consider as actions and agents, are fictions of the imagination. Human beings themselves are only partly real, they are real to the extent only that their minds are conscious and understand many things. The highest or “greatest striving of the mind, and its greatest virtue is understanding by the third kind of knowledge.” Virtue, power, nature, striving are all the same for Spinoza. (5p24.) In striving for this kind of understanding we thereby strive to minimize the force with which the passive, externally caused affects assert themselves. As Spinoza puts it: “We strive...that whatever is related to its [our mind’s] memory or imagination is of hardly any moment in the relation to the intellect.” (5p39s.) Memory and imagination relate to and depend on the body and belong to the part of the mind that perishes with the body, so is “of no moment in relation to what remains” (5p38s). The more a thing is perfect, the more reality it has, the more it acts (and the less it suffers or is acted on). The part of the mind that remains when the body perishes is more perfect than the rest. The eternal part of the Mind is the intellect “through which alone we are said to act” (5p40c). The ‘part of the mind’ through which it acts, it seems, is not really a part of our human mind but is the eternal intellect acting in it. All we do as human minds is rejoicing, the credit for the acting goes to the eternal intellect.

Commentators worry how the intellectual love of God can be love, i.e., joy, if it is eternal (5p33)? Spinoza himself raises this difficulty. What is eternal is unchanging, immutable, and human love and joy were defined in terms of increase of perfection, so are transient, changing states of this particular body’s power to persist. What is eternal moreover has no beginning, but the increase of perfection has a temporal cause. Spinoza argues here that the intellectual love of God in spite of this does have “all the perfections of Love”. The mind has eternally the same perfections – the love and the eternal bliss in which it consists and was, in the preceding propositions (5p32), supposed to originate with the third kind of cognition, actually is enjoyed by the mind eternally. He writes: “If joy, then, consists in the passage to a greater perfection, blessedness must surely consist in the fact that the mind is endowed with perfection itself.” (5p33s.) The mind in understanding loves intellectually, and in understanding is endowed with perfection, is blessed.

I am not sure what to make of this, but here’s an attempt. If Spinoza, following Aristotle on this point, identifies perfection with activity, unimpeded

eternal activity, then a mind that is endowed with perfection itself, takes part in this unimpeded activity, and to this extent, the joy which is here a purely active intellectual joy, cannot be the same as that enjoyed by the ‘presence’ (thought) of some imagined cause, but must be *sui generis*. The oddity of Spinoza’s account here may have to do with the fact that he cannot explain this in terms of actualization of a potentiality at all times present, as other followers of Aristotle would do.

University of Uppsala

Bibliography

- Alanen, Lilli (forthcoming). “Affects and Ideas in Spinoza’s Therapy of Passions”, in Alix Cohen and Robert Stern (eds.), *Thinking About Emotions: A Philosophical History*. Oxford: Oxford University Press.
- Baier, Annette (1994). “Unsafe loves”, reprinted in Annette Baier, *Moral Prejudices. Essays on Ethics*. Cambridge, MA: Harvard University Press, 33–49.
- Carriero, John (2014) “The Ethics in Spinoza’s Ethics”, in Matthew J. Kisner & Andrew Youpa (eds.), *Essays on Spinoza’s Ethical Theory*. Oxford: Oxford University Press, 20–40.
- Carriero, John (2015). “Spinoza, The Will and the Ontology of Power”, in Yitchak Y. Melamed (ed.), *The Young Spinoza: A Metaphysician in the Making*. Oxford: Oxford University Press, 160–182.
- Descartes, René (1964–76). *Oeuvres de Descartes*. 12 volumes, revised edition. Ed. C. Adam & P. Tannery. Paris: Vrin/CNRS. [AT]
- Descartes, René (1985). *The Philosophical Writings of Descartes*. Volumes I–II. Transl. J. Cottingham, R. Stoothoff & D. Murdoch. Cambridge: Cambridge University Press. [CSM I/II]
- Descartes, René (1991). *The Philosophical Writings of Descartes*. Volume III: The Correspondence. Transl. J. Cottingham, R. Stoothoff, D. Murdoch & A. Kenny. Cambridge: Cambridge University Press. [CSMK]
- Huebner, Karolina (forthcoming). “Spinoza’s Unorthodox Metaphysics of Will”, in Michael Della Rocca (ed.), *The Oxford Handbook of Spinoza*. Oxford: Oxford University Press.
- Long, Anthony (2003). “Stoicism in the Philosophical Tradition: Spinoza, Lipsius, Butler”, in Brad Inwood (ed.), *The Cambridge Companion to the Stoics*. Cambridge: Cambridge University Press, 365–392.
- Koistinen, Olli (2009a). “Spinoza’s Eternal Self”, in Jon Miller (ed.), *Topics in Early Modern Philosophy of Mind*. Studies in the History of Philosophy of Mind 9, 151–169.
- Koistinen, Olli (2009b). “Spinoza on Action”, in Olli Koistinen (ed.), *The Cambridge Companion to Spinoza’s Ethics*. Cambridge: Cambridge University Press, 167–187.

Melamed, Yitchak (forthcoming). "Spinoza's *amor Dei intellectualis*", in Noa Naaman & Tom Vinci (eds.), *Freedom and the Passions in Spinoza*. Cambridge: Cambridge University Press.

Spinoza, Benedictus de (1985). *The Collected Works of Spinoza*. Ed. & transl. Edwin Curley. Princeton, NJ: Princeton University Press.

Spinoza ja moderni kosmologia

Juhani Pietarinen

Spinoza tarjoaa *Etiikassaan* metafyyssisen selityksen järjestyneelle maailmankaikkeudelle, kosmokselle. Selitys edustaa rationalistista pyrkimystä lähestyä luonnon järjestystä tieteellisesti hyväksyttävällä tavalla. Spinozan ajoista luonnontieteellinen tutkimus on edennyt huimaa vauhtia ja tuonut esiin hämmästyttäviä asioita maailmankaikkeuden kehityksestä ja toiminnasta. Niistä lukiessa herää aika ajoin kysymys, miltä Spinozan päätelmät mahtaisivat näyttää nykyisen fysikaalisen kosmologian tarjoamien näkymien valossa. Kirjaan tässä joitakin spekulatioita. Pysyttelen hyvin yleisellä ja populaarilla tasolla, olenhan varsinkin fysiikassa täysi harrastelija. Toivon niiden tarjoavan Ollille ja Turun muille rationalisteille lähtökohtia syvemmille spekulatioille.

Spinozan metafysiikka rakentuu opille yhdestä substanssista, joka on ääretön, ikuinen ja jonka olemassaolo on välttämätön totuus. Ääretön substanssi ei rajoitu mihinkään, joten se sisältää kaiken mitä on. Spinozan substanssi on *ens realissimum*. Se on myös ikuinen, ja ikuisuudella Spinoza tarkoittaa välttämätöntä olemassaoloa. Substanssin olemassaolon välttämättömyys tarkoittaa Spinozalla sitä, että olemassaolo seuraa substanssin olemuksesta, siis siitä mitä se itse on. Substanssi on itsensä syy, tekee itse itsensä todelliseksi. On mahdotonta ajatella, ettei substanssia ole tai että se syntyy jostain muusta. Substanssin ikuinen olemassaolo on luonteeltaan ajatonta, se ei ole sidoksissa aikaan eikä paikkaan.

Spinozalainen substanssi ei ole niin etäällä luonnontieteestä kuin tuon kuvauksen perusteella saattaa näyttää. Spinoza itse identifioi substanssin luontoon: ulotteisena se muodostaa aineellisen tai fysikaalisen todellisuuden perustan. *Etiikan* ensimmäisen luvun lopussa osoitetaan, että substanssi on olemukseltaan kausaalista voimaa. Voimallaan se saa aikaan kaiken mitä on, kaiken todellisen, niin substanssin oman olemassaolon kuin kaiken mikä sen olemuksesta seuraa. Kysymys on luonnon perusvoimasta. Luonnon olemuksen muodostaa siis yksi ainoa kausaalinen voima, jota ei mikään rajoita ja joka tuottaa itsensä ja kaiken, mitä se voi luonnon lakien mukaan toteuttaa.

Spinoza puhuu erikseen ”synnyttävästä luonnosta” (*natura naturans*) ja ”synnytetystä luonnosta” (*natura naturata*). Synnyttävän voiman luonteeseen kuuluu, että se ”purkautuu” välttämättä eli toteuttaa olemustaan, jos mikään ei estä, ja koska luonto-substanssia ei mikään rajaa, sen olemassaolo on välttämätöntä. Näin kaikki, mikä on luonnon perusvoiman mukaan mahdollista, toteutuu välttämättä. Tämä sulkee pois mahdollisuuden, että jotain luonnonlakien mukaista jäisi toteutumatta, samoin teologisen näkemyksen, että Jumala olisi voinut jättää koko maailman luomatta – Jumalahan on Spinozalle sama kuin substanssi eli luonto. Monistisesta voiman metafysiikasta seuraa edelleen, että kaikki luonnossa ilmaisee tavalla tai toisella yhtä ja samaa perusvoimaa. Jokainen olio, jokainen ilmiö ja tapahtuma on olemukseltaan tuota voimaa. Voima on kausaalista, ja Spinoza ymmärtää voiman sillä tavoin aktiiviseksi, että sen ansiosta kaikki luonnossa pyrkii toteuttamaan omat potentiaalinsa niin täysimääräisesti kuin kulloinkin on mahdollista.

Spinozan mukaan substanssi, ja siten luonnon perusvoima, on luonteeltaan ajatonta. Meille luonto näyttäytyy kuitenkin ajallisena ja alati muuttuvina ilmiöinä. Ei pidä kuitenkaan ymmärtää niin, että Spinoza ajattelisi ajattoman perusvoiman kuuluvan eri todellisuuteen kuin muuttuvan luonnon – kaikki todellinenhan on yhtä ja samaa perusvoimaa. Kysymys on siitä yksinkertaisesta asiasta, että ajallisesta näkökulmasta katsoen luonnon perusvoima on läsnä aina ja kaikkialla. Se ohjaa kaikkia ilmiöitä ja tapahtumia olemuksellisten lakiensa mukaisesti. Jokainen ajallinen tapahtuma ja vaihe universumissa seuraa kausaalisesti edellisestä vaiheesta, eikä tuossa ketjussa ole alkua eikä loppua.

Miten Spinozan metafysiikka sopii nykyisen fysikaalisen kosmologian käsityksiin universumista? Onko niillä lainkaan yhtymäkohtia? Onko Spinozan rakennelmaa ylipäätään mahdollista vertailla järkevästi nykyiseen tieteelliseen kosmologiaan? Jotain kiinnostavaa voi ehkä löytää, ainakin voi yrittää.

Nykyinen fysikaalinen kosmologia perustaa käsityksensä tähtitieteen havaintoihin, Einsteinin suhteellisuusteoriaan ja hiukkasfysiikan teoreettisiin ja empiirisiin löydöksiin. Yleisimmin kannatetun teorian mukaan nykyinen universumi on syntynyt äärimmäisen tiheään pakatun aineen äkillisestä laajenemisesta, niin sanotusta ”alkuräjähdyksestä” noin 13,7 miljardia vuotta

sitten. Tälle standarditeorialle on tarjolla myös vaihtoehtoja: on teorioita universumin pysyvistä vakaasta tilasta, ehdotuksia kehityksen syklisyydestä sekä visioita useista eri universumeista. Aloitan standardikäsitteestä.

1. Alkuräjähdysteoria

Empiiriset havainnot osoittavat, että universumi laajenee kiihtyvällä nopeudella. Jos ilmiöön sovelletaan Einsteinin yleistä suhteellisuusteoriaa, voidaan teoreettisesti laskea, että aika-avaruudessa ajan parametrilla on absoluuttinen alaraja, raja missä aika häviää. Kun tuota rajaa lähestytään ajassa taaksepäin, avaruus käyristyy ja sen sisältämä aine puristuu yhä pienempään tilaan, kunnes päädytään yhteen matemaattiseen pisteeseen. Pistettä kutsutaan alkusingulariteetiksi.

Singulariteetti on suhteellisuusteorian yhtälöistä johdettu tila, jossa nykyinen aika-avaruudellinen universumi ei ole voinut koskaan olla. Matemaattisella pisteellä ei ole ulottuvuutta, joten aika-avaruus häviää siinä olemattomiin. Alkusingulariteetti on ajaton ja paikaton. Teoreettisten laskelmien mukaan aineen tiheys olisi singulariteetipisteessä ääretön, mikä merkitsee ääretöntä kuumuutta ja siten ääretöntä energiaa. Tuollaisesta singulariteetista muodostui standarditeorian mukaan nykyinen universumi suunnattoman nopeana ja rajuna laajenemisena, inflaationa tai ”alkuräjähdyksenä”. Näin ymmärrettynä alkuräjähdyks tarkoittaa siirtymistä ajattomasta ja paikattomasta tilasta nykyiseen aika-avaruuteen, joka laajenee edelleen kaiken aikaa kiihtyvällä vauhdilla.

Singulariteetin tulkinta on monella tavalla ongelmallinen. Singulaarisen tilan tulisi olla jotain, mikä kuuluu fyysikaaliseen todellisuuteen, selitetäänhän sillä luonnon tapahtumia. Miten sellainen, mikä on ajan ja paikan ulkopuolella, voi kuulua luontoon ja selittää luonnon ilmiöitä? Miten jostain ajattomasta ja paikattomasta voi syntyä mitään aikaan ja paikkaan sidottua? Eikä luonnossa pitäisi esiintyä mitään ääretöntä, ei ääretöntä tiheyttä eikä ääretöntä energiaa. Mitä sellaiset voisivat olla? Kysymys singulariteettien olemassaolosta ja tulkinnasta tuottaa suurta päänvaivaa niin fyysikoille kuin filosofeille.

Rajoitun tässä yhteen ongelmaan, universumin alun käsitteeseen. Singulariteetti ei voi olla universumin alku, koska alkaminen voi tapahtua vain ajassa. Jos jokin X alkaa hetkellä t , se tarkoittaa, että X ei ole ennen

hetkeä t , mutta se on sillä hetkellä ja sen jälkeen. Singulariteetissa ei ole aikaa eikä siten sellaista ajallista kohtaa, jota ennen universumia ei ole mutta josta lähtien se on olemassa. Ne avaruuden ilmiöt, joita tähtitiede ja fysiikka tutkivat, tapahtuvat kaikki aika-avaruudessa ja edellyttävät aikaa, mutta aika tulee teorian mukaan vasta alkuräjähdyksen myötä. Mikä silloin on universumin ajallinen alku? Sellaista ei ole: voimme kulkea kohti singulariteettia kuinka lähelle tahansa ajan lopullisen häviämisen hetken tulematta vastaan. Vaikka maailmállamme olisikin ajallisesti rajallinen menneisyys, alkuräjähdysteorian mukaiset vajaat 14 miljardia vuotta, singulariteetista nykyaikaan mahtuu ääretön määrä tapahtumia.

Alkuräjähdysteoria herättää periaatteellisia ontologisia ja metafysisiä kysymyksiä. Mitä siitä voi sanoa Spinozan metafysiikan kannalta?

Alkusingulariteetti muistuttaa eräiltä keskeisiltä ominaisuuksiltaan yllättävän läheisesti spinozalaista substanssia ulotteisuuden näkökulmasta. Tietenkin näyttää absurdilta ajatella ulotteisuutta geometrisen pisteen kokoisena, sehän kutistuisi silloin olemattomiin. Mutta mitä ulotteisuus on Spinozan metafysiikassa? Se on alussa esitetyn mukaan ääretöntä ja ajatonta voimaa, josta ajallinen kaikkeus muuttuvine ilmiöineen seuraa. Voiman äärettömyys tarkoittaa Spinozalla rajoittavan voiman puuttumista, joten substanssin voima sisältää universumin kaiken voiman. Samat asiat voi todeta myös singulariteetista: se on ääretöntä voimaa aika-avaruuden ulkopuolella. Substanssin ulotteinen voima ilmenee Spinozan mukaan ”liikkeen-ja-levon” muodossa, toisin sanoen voimana, joka tuottaa kaiken liikkeen universumissa. Nykyisin sanoisimme ehkä, että se on rajatonta energiaa ja sellaiseksi singulariteetin ääretön voima kaikeksi on käsitettävä.

Mutta miten voi olla jotain, mikä ei esiinny aika-avaruudessa, mikä ei ole aikaan eikä paikkaan sidottua – sellaista kuin Spinozan substanssi ja alkuräjähdysteorian singulariteetti? Mikä voi saada sellaisen aikaan? Spinozan vastaus on: se tekee itse itsensä. Toisin sanoen substanssin olemassaolo on sillä tavoin välttämätöntä, ettei mitään muuta mahdollisuutta ole. Substanssin ja sen voiman, luonnon perusvoiman, olemassaoloa ei voi epäillä yhtään enempää kuin sitä, että tasokolmion kulmat summautuvat oikokulmaksi, Spinoza sanoo. Kumpikin on luonteeltaan ajaton totuus. Spinozasta hurmaantunut fyysikko voisi sanoa singulariteetista, että se on ikään kuin substanssi, joka tuottaa itse oman voimansa ja on siten olemassa ajattomasti.

Ei ole mitään ulkopuolista syytä, joka saa aikaan sen sisältämän energian, eikä sen olemassaololle tarvita enää muuta selitystä, ei yliluonnollista Jumalaa eikä luonnontieteen selvittävää muuta syytä. Edeltävää syytähän ei voi ollakaan, koska mitään singulariteettia edeltävää ei ole.

Kuinka hyvä tällainen vastaus on? Ongelman muodostaa tietysti ajatus itsensä syystä sillä tavoin kuin Spinoza esittää: että jonkin olemassaolo on välttämätöntä. Voimme ilman suurempaa vastahankaa hyväksyä sen, että kolmion kulmien summa on välttämättä oikokulma ja että niin on ajattomasti, mutta voiko mikään luonnossa olla tuolla tavalla välttämätöntä – voiko mikään luonnossa olla itsensä syy?

Luonnontieteilijän on epäilemättä vaikea hyväksyä ajatusta. Hänhän etsii jokaiselle ilmiölle syytä tai selitystä jostakin muusta kuin ilmiöstä itsestään, ja ellei sitä löydy, on kysymykseen ”miksi näin on?” tyydyttävä vastaamaan: ”Emme tiedä, niin vain on.” Mitä alkusingulariteetti on ja mistä se on tullut? Miksi alkuräjähdyks tapahtui? ”Emme tiedä, emmekä ehkä koskaan saa tietääkään. Niin vain on ollut, niin on tapahtunut.” Mutta Spinoza ei voinut sanoa niin; hänen rationalisminsa esti tyytymästä tuollaisiin ”raaka tosiasia” -vastauksiin. Rationalistin mukaan kaikille asioille pitää löytää riittävä syy, niin myös universumin olemassaololle. Tuo ohjenuora johti Spinozan käsitykseen substanssista, joka on olemassa välttämättä ja josta kaikki muu syntyy.

Nykyisin filosofit hylkäävät yleisesti ajatuksen, että välttämätön olemassaolo voisi kuulua muille kuin abstrakteille kohteille. Tuskin ajatus houkuttelee luonnontieteilijöitäkään – siitäkään huolimatta, että singulariteetin ja siitä lähtevän aika-avaruuden selittäminen saattaa houkutellessa ”itsensä syy” -tyyppisiin selityksiin. Niinpä on ainakin toistaiseksi tyydyttävä hyväksymään alkuräjähdyks raakana tosiasiana. Toiseenkin mahdollisuuteen voi tietenkin tarttua. Ehkä yleinen suhteellisuusteoria ei pidäkään kaikin puolin paikkaansa eikä singulariteettia ja siitä lähtenyttä alkuräjähdystä tarvitse pitää uskottavana asiana? Tämä mahdollisuus tulee kohta esille.

Spinozalaisen substanssin oleminen ei ole aikaan sidottua, eikä myöskään singulariteetin omalla tavallaan. Miten ne liittyvät ajallisesti ikääntyvään kaikkeuteen? Kumpikin on olemukseltaan voimaa, ja voimaa voi sanoa kyvyksi tai taipumukseksi tuottaa vaikutuksia. Voiman luonteeseen kuuluu myös, että se toteutuu välttämättä silloin kun mikään ei estä.

Aikaisemmin oli esillä, kuinka Spinozan substanssin ulotteinen voima, ”liike- ja-lepo”, toteutuu ajallisesti kaikkena, mitä luonnossa tapahtuu. Se sisältää tietyllä tavalla mahdollisuutena kaiken sen, mikä ylipäättään voi toteutua, ja kun substanssin voimaa ei mikään estä (koska mitään muuta ei ole), nuo mahdollisuudet toteutuvat välttämättä. Sama ajattelu tekee singulariteetistakin ”synnyttävän luonnon”, joka äärettömän rajulla ponnistuksella vapauttaa kohdussaan piilevän aika-avaruuden. Synnytetty luonto jatkaa ajallista vaellustaan sillä elinvoimalla ja niiden lakien mukaisesti, jotka luontoemon olemuksesta on siihen virrannut.

Spinozan mukaan substanssin koko voima toteutuu luonnon järjestyksessä aina, joka hetki ikuisuudesta ikuisuuteen. Hän ajatteli ajalleen ominaisesti kaikkeuden pysyvän tilana, joka loputtomiin vaihtuvina äärellisine tapahtumineen säilyy kokonaisuudessaan samana aina ja kaikkialla, vailla äärellistä alkua ja loppua. Se ajatus ei näytä sopivan alkuräjähdysteorian tarjoamaan kuvaan avaruuden ajallisesta alusta ja laajenemisesta. Asiaa kannattaa kuitenkin pohtia.

Myös nykyisessä fysikaalisessa kosmologiassa on puolustettu näkemystä, että universumi pysyy ajasta toiseen suhteellisen vakaana. Kun relativistinen kosmologia pääättelee universumin laajenemiselle äärellisen iän, vakaan tilan teoria olettaa laajenemisen edenneen äärettömän ajan ja jatkuvan tulevaisuuteen ainaisesti, olettaahan se universumin pysyvän aina samanlaisena. Mikä pysyy samana? Eikö kaikkeuden tulisi laajentuessaan jotenkin oheta, aineen tiheyden vähentyä? Vakaan tilan puolustajien mukaan aineen tai paremminkin aine-energian keskimääräinen tiheys ei muutu, koska universumiin tulee tuntemattomasta lähteestä uutta ainetta korvaamaan menetystä. Selitys jää teoreettisesti jotenkin ontoksi, *ad hoc* -tyyppiseksi, ja kun empiiriset löydöksetkin puhuvat vakaan tilan oletusta vastaan, teorian hylkääminen näyttää väistämättömältä nykyisessä kosmologiassa. On kuitenkin muistutettava, että tietomme avaruuden sisältämästä energiasta ja energiatoiminnoista on puutteellista ja tulee todennäköisesti muuttumaan radikaalilla tavalla.

Spinozan metafysiikkaa ja vakaan tilan kosmologiaa sitoo yhteen ajatus samana pysyvistä universumista. Mutta samuuden selitykset eroavat. Spinozan mukaan samana pysyy universumin synnyttävä ja sitä ylläpitävä voima. Substanssin energiaa ei häviä eikä sitä tule mistään lisää. Voima

organisoi tuu kyllä lukemattomiksi erilaisiksi liikkeiden järjestelmiksi, jotka muuttuvat alituisen, mutta itse voima pysyy muuttumattomana. Toisin sanoen voiman määrä pysyy, mutta sen jakaantuminen muuttuu. Edellä kuvattu vakaan tilan teoria näkee asian toisin: sen mukaan energiaa häviää avaruudessa ja tulee lisää. Sellainen on vastoin energian säilymisen lakia, ja tätä seikkaa onkin käytetty yhtenä perusteena vakaan tilan teorian hylkäämiselle. Sama arvostelu ei pure Spinozan metafysiikkaan, joka on säilymisen laille kuuliainen. Se ei kuitenkaan välttämättä tee Spinozan opista fysiikan kannalta vakaan tilan kosmologiaa uskottavampaa, sillä energian säilyminen suhteellisuusteorian mukaisessa aika-avaruudessa ei ole ainakaan kaikkien fyysikkojen mukaan mitenkään selvä asia, etupäässä energian ja voiman käsitteisiin liittyvien tulkintaongelmien takia.

Spinozan metafysiikan tutkijoita on askarruttanut paljon kysymys, miten on ymmärrettävä ajattoman substanssin ja ajallisen maailman suhde. Miksi ajallinen todellisuus on ylipäätään syntynyt? Miten substanssi, joka on ikuinen ja välttämätön, voi tuottaa sellaisen kaikkeuden, jossa ilmiöt seuraavat toisiaan kausaalisesti ajankohdasta toiseen, vieläpä niin, että vain yksi luonnon järjestys on mahdollinen?

Spinozan vastaus ensimmäiseen kysymykseen näyttää lyhyesti seuraavalta. Täydellisesti todellisella substanssilla täytyy olla kaikki erilaiset attribuutit, toisin sanoen sen täytyy olla olemassa kaikilla mahdollisilla eri tavoilla. Todellisuus ilmenee meille ulotteisena (kappaleina) ja mentaalisenä (ajatuksina), ja koska kaikki todellisuudessa ilmentää substanssin attribuutteja, substanssilla täytyy olla ulotteisuuden ja ajattelun attribuutit. Substanssin olemus on voimaa, ja ulotteisuuden kannalta se tarkoittaa liikkeitä muodostavaa voimaa, siis kykyä tuottaa kaikkea mahdollista liikettä. Koska substanssin voimaa ei mikään estä, se toteutuu välttämättä erilaisina liikejärjestelminä, joiksi Spinoza ymmärsi aikansa fysiikan mukaisesti fysikaaliset kappaleet ja niiden interaktiot. Ajallisesti etenevä kausaaliketjujen todellisuus syntyi näin välttämättä ulotteisesta substanssista, siis luonnon äärettömästä perusvoimasta.

Tämä selitys ei kuitenkaan vastaa siihen kysymykseen, miten ajattomasta voimasta voi ylipäätään syntyä ajallisia ilmiöitä. Spinoza ei anna muuta vastausta kuin että niin on täytynyt tapahtua. Alkuräjhdysteoria nostaa saman ongelman: miten singulariteetista, joka on ajaton, voi syntyä

ajallinen kosmos? Tiede ei anna vastausta – muuta kuin että niin on täytynyt tapahtua. Joka tapauksessa Spinozan mukaan substanssin voima vaikuttaa nyt, tässä meidän ajallisessa luonnossamme aina ja kaikkialla. Alkuräjähdysteoria antaa universumille äärellisen iän, eikä ole syytä arvella tuollaisen mahdollisuuden koskaan edes juolahtaneen Spinozan mieleen – vielä vähemmän sen, että ajallakin olisi äärellinen alku. Mutta voisiko Spinozan käsityksessä ja alkuräjähdysteoriassa nähdä silti jotain yhteistä? Tuon teorian mukaanhan aika syntyi singulariteetista alkuräjähdyksessä, joten nykyisen kosmoksen on täytynyt olla *aina*, joka hetki niin kauan kuin aika on ollut olemassa, ja se tulee olemaan niin kauan kuin aikaa on. Kosmoksella ei ole ajassa määriteltävää ensimmäistä eikä viimeistä hetkeä. Ja juuri niin Spinozan metafysiikka väittääkin.

Alkuräjähdysteoriaa ei suinkaan hyväksytä fysiikassa varauksetta. On esitetty perusteita epäillä sen mukaisia laskelmia aika-avaruuden synnystä. Teoria perustuu yleiseen suhteellisuusteoriaan, mutta kvanttiteoreettisten tulosten mukaan suhteellisuusteoria ei päde olosuhteissa, missä energiatiheys on suunnattoman korkea, niin kuin sen on täytynyt olla hyvin lähellä singulariteettia. Siksi on lähdetty rakentamaan kosmologioita yhdistämällä kvanttifysiikkaa ja suhteellisuusteoriaa.

Tunnetuimpia kvanttikosmologioista ovat silmukkateoria ja säieteoria. Molempien tulokset viittaavat siihen, että nykyinen universumi on ollut olemassa ennen alkuräjähdyksessä tapahtunutta äkillistä laajentumista (inflaatiota), joten aika ei olisikaan lähtenyt alku-räjähdyksestä. Säieteorian pohjalta on hahmoteltu kuvaa universumin pitkästä kehityksestä: ennen nykyistä laajenevaa avaruutta on tapahtunut aika-avaruuden hidasta kutistumista, joka tietyssä äärikohtassa, ennen singulariteetin saavuttamista, kääntyi äkillisesti nykyiseksi laajenemiseksi. Teorian mukaan supistunut kosmos voi olla äärettömän vanha, samoin nyt laajeneva voi jatkaa loputtomiin. On myös kehitetty versioita, joissa universumi kehittyy supistumisten ja laajentumisten sykleinä loputtomiin. Näiden teorioiden mukaan aktuaalinen kosmos on olemassa ikuisuudesta ikuisuuteen, kuten Spinoza esitti. Kvanttiperusteiset kosmologiat ovat kuitenkin vielä kehitysvaiheessa ja pitkälle spekulatiivisia.

2. Monen universumin teorit

Spinozan substanssi on *ens realissimum*, kaikkein todellisin sillä tavoin, että se käsittää kaiken mikä on olemassa. Spinozan mukaan se tarkoittaa, että substanssilla on kaikki attribuutit, mitä vain voi olla. Attribuutilla hän tarkoittaa erilaisia olemisen tapoja. Substanssi voi olla olemassa äärettömän monilla tavoilla, myös meille tuntemattomilla. Me tunnemme niistä vain kaksi, nimittäin ulotteisuuden eli fysikaalinen todellisuuden ja ajattelun eli mentaalisen todellisuuden, ja ne ovat Spinozan mukaan perimmiltään sama asia. Muita mahdollisia substanssin olemisen tapoja meillä ei ole mitään mahdollisuutta tuntea. Kun substanssi ymmärretään luonnoksi, se tarkoittaa, että rajaton määrä luonnon perusvoiman mahdollisista ilmenemistavoista on meille tuntematonta.

Tässä Spinozan opissa lukuisista ja enimmäkseen tuntemattomista attribuuteista on nähty yhtymäkohtia fysikaalisen kosmologian ehdotuksiin useista erilaisista toisistaan riippumattomista universumeista. Avaruus ei olisikaan vain yksi universumi vaan monen eri universumin ”multiversumi”. Se universumi, jossa me elämme ja josta teemme havaintoja, olisi vain yksi suunnattomasta määrästä hyvin erilaisia muita maailmankaikkeuksia. Säieteoreettisten laskelmien perusteella voidaan johtaa kuvauksia sellaisista universumeista, joissa fysiikan tietyt perusparametrit saavat toisia arvoja kuin meillä ja joissa kaikki toimii siksi täysin eri tavalla. Kysymys on puhtaasti laskennallisista mahdollisuuksista, mutta multiversumi-hypoteesi lähtee siitä, että kaikki tuollaiset mahdollisuudet toteutuvat ja fysikaalinen todellisuus muodostuu rajattomasta määrästä toisistaan erillisiä maailmankaikkeuksia. Hypoteesi on pitkälle kehitetyn fysiikan teorian mukainen ja selittää potentiaalisesti monia avoimia kysymyksiä hiukkasfysiikassa ja kosmologiassa. Mutta vain potentiaalisesti, sillä hypoteesia on mahdotonta konfirmoida kokeellisesti, koska emme voi tehdä havaintoja sellaisista universumeista, joilla ei ole mitään yhteyttä kotikaikkeuteemme – muutoinhan olisi kysymys vain omaan kaikkeuteemme kuuluvista osista.

Onko multiversumi-hypoteesilla mitään kiinnostavaa yhteyttä Spinozan ajatukseen tuntemattomista attribuuteista? Heti ensiksi on huomattava, että kaikki hypoteesin oletamat avaruudet ovat spinozalaisittain ajatellen ulotteisia, sillä fysiikka tutkii vain ulotteista todellisuutta. Kysymys ei siten ole spinozalaisittain tuntemattomien attribuuttien alaan kuuluvista kohteista.

Kaikkien universumien oletetaan noudattavan fysiikan lakeja, koska jo oletus niiden olemassaolosta perustuu fysiikan kehittämiin teorioihin. Emme vain tunne niitä kaikkia, ja Spinozalle olisi täysin luonnollista todeta, että emme vielä tunne kaikkia substanssin voimaa ohjaavia lakeja. Hän ei välttämättä torjuisi ajatusta monista erilaisista kaikkeuksista. Oleellista hänen kannaltaan on kuitenkin, että kaikki ne mahdollisuudet, jotka substanssin voimaa ohjaavat lait sallivat, toteutuvat ulotteisessa todellisuudessa – siis kaikki substanssin lakien mukaiset universumit, ovatpa lait kuinka poikkeavia tahansa meidän tuntemistamme liike- ja muista laeista. Mikään universeista ei voisi jäädä Spinozalle pelkäksi mahdollisuudeksi, eikä niiden toteutuminen voisi noudattaa muuta kuin yhtä järjestystä. Siis vain yksi multiversumi olisi hänelle mahdollinen, ja se toteutuisi välttämättä. Se olisi ajallisesti rajaton molempiin suuntiin, kuten monet multiversumi-ajatuksen kannattajat ehdottavatkin.

Ongelmaksi muodostuu kuitenkin kysymys: Voiko Spinoza hyväksyä sen, ettemme voi edes periaatteessa saada tietoa muista universeista? Vastaus riippuu kysymyksestä, mitä ”periaatteessa” tässä tarkoittaa. Spinoza puolusti rationalistien tapaan ihmisen tiedollisen kyvyn mahdollisuuksia. Hän korostaa *Etiikassa* alinomaa, kuinka pystymme lisäämään tietoa substanssista. Mitään etukäteisiä rajoituksia ei tietämisellemme ole mahdollista asettaa. Siksi hän voisi hyväksyä ajatuksen tuntemattomien lakien mukaan toimivista ulotteisista kaikkeuksista, mutta lisäisi, että niitäkin on meille mahdollista oppia ajan mittaan tuntemaan. Entä jos lakien määrä onkin rajaton? Eikö silloin astu kuvaan äärellisen ihmisjärjen ja äärettömän intellektin ero: jos fysiikan lakeja on äärettömästi, emme voi koskaan saavuttaa äärettömän intellektin koko tietoa ulotteisesta maailmankaikkeudesta – vaikka ihmisäly säilyisi äärettömiin?

3. Älyllisen suunnittelun kielto

Yhteen kysymykseen, joka on herännyt monen universumin kosmologian yhteydessä, Spinoza ottaa hyvin selvän kannan. Se on niin sanotun älyllisen suunnittelun periaate.

Teoreettisen fysiikan tutkimus on tuonut esille, että universumi, jossa elämä siinä muodossa kuin sitä esiintyy maapallolla, on edellyttänyt tiettyjen fysikaalisten perusparametrien käsittämättömän tarkkaa asettumista kohdilleen. Ne on ikään kuin äärimmäisellä taidolla hienosäädetty elämän

syntymiselle otolliseksi: vety-ytimien synty alkuräjähdyksessä, hiili- ja happiatomien muodostuminen alkutähdissä, niiden kulkeutuminen supernovien kautta aurinkokuntaan ja siellä yhteen pieneen planeettaan, jossa vedestä ja hapesta riippuvan hiilipohjaisen elämän kehittyminen on ollut mahdollista – kaikki tuo edellyttää valtavan pitkää makroskooppista kehitystä ja samalla mikroskooppisen tarkkoja asetuksia alkeishiukkasten muodostumiselle. Mikä voisi selittää tuollaisen äärimmäisen tarkan hienosäädön, joka näyttää palvelevan elämän ja erityisesti älyllisen elämän syntymistä kaikkeudessamme?

Vastausten perusteella kosmologit voi jakaa kahteen ryhmään. Ensimmäiseen kuuluvat hyväksyvät ajatuksen, että universumin kehitys on jollakin tavalla määrätty ihmiselämän tai yleensä elämän muodostumiselle suotuisaksi – että sitä ohjaa ”antrooppinen periaate”, kuten ajatusta kutsutaan kosmologisessa keskustelussa. Periaatetta on pyritty täsmentämään monella tavalla, mutta tässä riittää kannattajien jako kahteen ryhmään: uskonnollisiin, joiden mukaan kehitystä ohjaa yliluonnollinen voima, Jumala tai Luoja, ja tieteellisiin, jotka olettavat, että fysikaaliseen maailmaan itseensä sisältyy mekanismeja, joiden tarkoituksena on varmistaa ihmisille sija kaikkeudessa. Uskonnollisen kannan tueksi on otettu perinteisen ”suunnitteluargumentin” muunnos: universumin antrooppinen hienosäätö osoittaa, ettei ”sokea” fysikaalinen kehitys tai peräti pelkkä sattuma voi tuottaa elävän luonnon ja ihmiskunnan kaltaisia ihmeellisyyksiä, vaan kaiken taakse täytyy olettaa älyllinen voima. Maailma on tulos älyllisestä suunnittelusta, hienoimpana päämääränä älykäs ihminen.

Spinozan kanta suunnitteluargumenttiin ja ylipäänsä päämäärään vetoaviin selityksiin on hyvin selvä. ”Luonnolla ei ole mitään sille ennalta-asetettuja päämääriä, eikä yksikään päämääräsyy ole muuta kuin ihmisen kuvittelua,” hän kirjoittaa *Etiikan* ensimmäisen osan liitteessä. Spinoza ajatteli kyllä, että luonnon kulku seuraa tietyllä tavalla ”älykästä” ohjailua: se seuraa luonnon lakeja, ja lait edustavat älyllistä ajattelua, koska ne voidaan muotoilla matemaattisesti. Substanssi on olemukseltaan ääretön intellekti. Spinoza kyllä yhdisti substanssin Jumalaan, mutta ei Jumalaan yliluonnollisena maailmankaikkeuden luojana vaan luonnon perusvoimana, joka noudattaa tarkkoja matemaattisia lakeja. ”Kaikki tapahtuu luonnossa tietyllä ikuisella välttämättömyydellä,” hän lisää edelliseen samassa liitteessä.

Maailmankaikkeuden synty ja toiminta luonnon ulkopuolisen voiman älykkäänä suunnitteluna on hänestä virheellistä kuvittelua.

Spinozan tuomio ei kohdistu vain älykkään suunnittelun argumenttiin, vaan hän hylkää kaiken päämääriin vetoavan luonnon tapahtumien selittämisen. Luonnolla ei ole päämääriä, kaikki tapahtuu välttämättä luonnonlakien mukaisesti. Antrooppista periaatetta on yritetty käyttää fyysikassa tekemään uskomattoman tuntuisia hienosäätöjä ja yhteensattumia ymmärrettäviksi, jopa ennustamaan lisää vastaavia. Toisaalta antrooppisia selityksiä on arvosteltu perusteettomiksi, tarpeettomiksi ja haitallisiksi: ihmisen elämän olemassaolon nostaminen etusijalle universumissa on perusteetonta, kaikki ilmiöt, myös elämän muodostuminen, voidaan selittää ilman antropia-oletusta, ja oletukseen takertuminen johtaa pois todellisten fysikaalisten syiden etsinnästä. Sellaiseen kritiikkiin Spinoza epäilemättä yhtyi.

Monet fyysikot ovat nähneet multiversumi-hypoteesissa mahdollisuuden tehdä universumissa ilmenevä hienosäätö ymmärrettäväksi. Jos universumin luonnetta säätelevien perusparametrien arvot voivat vaihdella vapaasti, silloin toisistaan poikkeavia kaikkeuksia voi syntyä myriadeittain, kukaties rajaton määrä. Suurimmassa osassa ei otaksuttavasti olisi elämälle suotuisia olosuhteita, mutta joukkoon voi odottaa hyvin suurella todennäköisyydellä mahtuvan myös niitä, joissa parametrit ovat älyllisen elämän kehittymisen kannalta kohdallaan. Oma universumimme on sellainen, eikä siinä havaittavaa hienosäätöä tarvitse ihmetellä: kysymyshän on vain siitä, että älyllisen elämän syntyminen edellyttää tiettyjä olosuhteita, ja multiversumissa odotusarvo sellaisten kehittymiselle jossakin lähes rajattomasti varioivista universumeista on äärimmäisen korkea.

Tämäkään yritys vastata kysymykseen, miksi universumistamme on muodostunut ihmiselle suotuisa, ei voisi tyydyttää Spinozaa, koska multiversumi-hypoteesi ei tarjoa riittävää perustetta hienosäädölle. Sehän vetoaa vain tilastolliseen sattumaan: on uskottavaa, että hyvin suuressa tapausten määrässä kehittyä ennen pitkää myös meidän universumimme kaltainen järjestelmä. Ehdottoman välttämätöntä sellainen ei kuitenkaan ole, ja moni muukin kysymys jää vaille sellaista riittävää perustelua, jota Spinoza vaati. Miksi pitäisi uskoa poikkeavien kaikkeuksien olemassaoloon? Eihän niistä ole mitään todisteita, eikä ilmeisesti ole mahdollista saadaakaan. Mikä määrää perusparametrien arvon ja siten sen, millainen universumi kulloinkin

kehitty? Pelkästään siitä, että parametrit voivat jonkin nykyisen teorian perusteella saada erilaisia arvoja, ei voi vakuuttavasti päätellä, että kaikki mahdolliset arvot myös toteutuvat – mahdollisuus ei merkitse välttämättömyyttä, eikä multiversumin mahdollisuuskaan ole selvää säieteorian kiistanalaisuuden vuoksi. Spinozan mukaan kaikki mikä tapahtuu, tapahtuu ehdottoman välttämättä eikä jätä vaihtoehdolle sijaa.

Jos hylätään multiversumin idea ja antrooppinen ”ihmistä varten” -ajattelu, miten havaintoihin hienosäädöstä tulisi suhtautua? Tieteellisen rationaalisuuden kannalta luonnollisinta on ottaa havainnot havaintoina ja pitää hienosäädöksi tulkittuja ilmiöitä tosiasioina muiden tosiasioiden joukossa. Hienosäädöllä ei ole sen kummempaa tarkoitusta kuin millä tahansa muulla fysikaalisella ilmiöllä. Universumistamme on tullut se mikä se on, sen perusparametrit ovat saaneet ne arvot mitä niillä on. Jos ei vielä pystytäkunnolla selittämään, miksi se on juuri tuollainen tai saanut juuri nuo arvot, on tyydyttävä ottamaan ne toistaiseksi raakoina tosiasioina ja etsittävä uusia selittäviä teorioita – kenties fysiikassa löytyy pätevä ”kaiken teoria”, joka selittää tyydyttävästi, miksi universumin ominaisuuksilla on juuri ne arvot mitkä havaitaan.

Tämä kanta vastaisi parhaiten Spinozan asennetta. Hän meni kuitenkin vielä pidemmälle selittämisen ehdottomuudessa kuin tieteessä on aiheellista olettaa. Hänen metafysiikkansa valossa nykyinen kaikkeus on *välttämättä* sellainen kuin on, sen vakiot eivät voi olla muita, mitään muuta kaikkeutta ei olisi voinut kehittyä. Tämä johtuu siitä, ettei substanssin olemus, luonnon perusvoima, voi toteutua muutoin kuin yhdellä ainoalla tavalla. Kuinka moni on valmis hyväksymään tuon maailmanselityksen lähtökohdaksi?

Turun yliopisto

Kirjallisuus

- Ellis, George F. R. (2014). ”On the Philosophy of Cosmology”, *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 46, 5–23.
- Enqvist, Kari (2003). *Kosmoksen hahmo*. Helsinki: WSOY.
- Enqvist, Kari (2011). *Olemisen porteilla*. Helsinki : WSOY.
- Spinoza, Baruch (1992 [1677]). *Ethics*. Engl. Samuel Shirley. Indianapolis: Hackett.
- Spinoza, Benedict de (1994 [1677]). *Etiikka*. Suom. Vesa Oittinen. Jyväskylä: Gummerus.

Spinozan tulkintaoppi

Seppo Sajama

1. Taustaa

Spinozan *Tractatus theologico-politicus* (josta jatkossa käytän lyhennettä ”TTP”¹) pidetään usein kummallisena teoksena, jossa yhdistetään oudosti politiikan teoriaa ja raamatuntulkintaa ja -historiaa. Kirjan kahden osan yhteys oli kuitenkin kirjoittajalleen päivänselvä: Eurooppaa kiusaavista uskonsodista päästään eroon vain suhtautumalla uskontoon rationaalisesti. Siksi järkevän raamatuntulkintaopin omaksuminen edistäisi rauhaa ja sopusointua.

Spinozan vastustajia olivat hänen oman aikansa fundamentalistit, jotka väittivät, että Raamatun ”jokainen kohta on tosi ja jumalallinen” (TTP Esipuhe, v). Spinozan mielestä tuo oletus pitäisi ensin todistaa eikä vain olettaa tutkimuksen lähtökohdaksi (TTP Esipuhe, v). Sitä kukaan ei ollut edes yrittänyt tehdä. Ja kun protestantit olivat antaneet jokaiselle yksilölle luvan tulkita Pyhää Kirjaa, tilanne oli Martti Lutherin kuvaama:

On melkein yhtä monta lahkoa ja uskomusta kuin on ihmistä ... Ei ole yhtäkään ihmistä, olkoon hän millainen pelle tahansa, joka ei väittäisi olevansa Pyhän Hengen innoittama ja joka ei esittäisi omia uniaan ja houereitaan profetioina. (Ray 1997, takaviite 62.)²

Koska jokainen pelle oli valmis taistelemaan ase-in oikean uskonsa puolesta, tilanne vaati Spinozan mielestä Raamatun luonteen ymmärtämistä. Esimerkiksi ”se, joka yrittää etsiä profettojen kirjoista viisautta ja tietoa luonnollisista ja henkisistä asioista, on täysin eksyksissä” (TTP II, 15). Raamattu ei ole luonnontieteen oppikirja, eivätkä Vanhan testamentin kertomukset ole tosiasioiden kuvauksia. Esimerkiksi kertomus Aatamista,

¹ Sivunumerot ovat WBG-editiosta, jossa noudatetaan alkuperäisen *editio princepsin* sivunumerointia.

² Protestantit ja katoliset kiistelevät tänä päivänäkin tämän lainauksen aitoudesta ja oikeasta tulkinnasta.

Eevasta ja omenasta on pelkkä allegoria, ei tosiasioiden kuvaus (TTP IV, 52).

Spinoza tarjoaa TTP:n seitsemännessä luvussa tulkintametodin, jolla Raamatun sanoma kirkastuu. TTP VII merkitsi teologiassa historiallis-kriittisen koulukunnan syntyä, ja TTP:n loppuosa oli yhteiskuntafilosofiassa väkevä liberalismiin aatteiden puolustus jo ennen John Lockeaa.

2. Tulkinnan metodi ei eroa luonnontieteen metodista

Spinozan mukaan Raamatun tulkitsemiseen ei tarvita mitään erityistä Pyhän Hengen innoitusta, vaan se sujuu aivan maallisilla keinoilla: ”Raamatun tulkitsemisen metodi eroaa tuskin luonnon tulkitsemisen metodista, vaan on sen kanssa yhdenpitävä [dico methodum interpretandi Scripturam haud differre a methodo interpretandi naturam, sed cum ea prorsus convenire]” (TTP VII, 84). Siksi Raamattua pitää tulkita samalla asenteella kuin luontoakin. Mitään ”ilmoituksen” ja ”kokemuksen” tai ”ymmärtämisen” ja ”selittämisen” vastakkainasettelua on Spinozalta turha etsiä. Luonnontiede on naturalistista siinä mielessä, ettei se salli yliluonnollisia selityksiä. Jos tulkintaoppi haluaa olla tiede, senkin on oltava omalla tavallaan ”naturalistista”: mielikuvituksellisia allegorisia tulkintoja ei sallita. Sama eetos, kaksi eri ilmenemismuotoa.

Spinoza selittää metodinsa kantavaa ideaa – luonnontieteen ja hermeneutiikan lähisukulaisuutta – seuraavasti:

Nam sicuti methodus interpretandi naturam in hoc potissimum consistit, in concinnanda scilicet historia naturae, ex qua, utpote ex certis datis, rerum naturalium definitiones concludimus: sic etiam ad Scripturam interpretandam necesse est ejus sinceram historiam adornare, & ex ea tanquam ex certis datis & principiis mentem authorum Scripturae legitimis consequentiis concludere. (TTP VII, 84.)

Tätä keskeistä tekstinkohtaa on lähes mahdotonta suomentaa ”oikein”. En edes yritä kääntää sitä sana sanalta, vaan esitän vain parafraasin, joka ilmaisee Spinozan idean – ja joka vaatii vielä runsaasti selittäviä kommentteja:

Luonnon tulkitsemisen metodin ydin on (i) luonnonhistorian konstruointi ja (ii) luonnonolioiden määritelmien johtaminen siitä samalla tavalla kuin geometriassa teoreemat johdetaan aksioomista.

Aivan samalla tavalla Raamatun tulkitsemiseen tarvitaan (i) raamatunhistorian konstruointi ja (ii) Raamatun kirjoittajien näkemysten johtaminen pätevillä päättelyillä tuosta historiasta samalla tavalla kuin geometriassa teoreemat johdetaan aksiomista.

Tämä katkelma herättää useita kysymyksiä. (1) Mikä on se ”luonnonhistoria” (*historia naturae*), josta Spinoza puhuu? Sanan merkityksen kannalta on relevanttia se, että antiikissa Plinius Vanhempi kirjoitti moniosaisen teoksen nimeltä *Naturalis historia*. Tämä kaikkien ensyklopedioiden äiti kattoi osan siitä, mitä nykyisin kutsuisimme luonnontieteiksi – etnografiasta gemmologian kautta astrologiaan. Nykyisin luonnonhistorian ymmärretään olevan elävän ja elottoman luonnon olioiden – erityisesti niiden alkuperän, kehityksen ja keskinäisten suhteiden – tutkimusta, kertoo *The American Heritage Science Dictionary*.

On siis turvallista olettaa, että ”luonnonhistoria” tarkoitti Spinozalle suurin piirtein samaa kuin ”luonnontiede” meille. Jos luonnehdintaa tarvitsisi tarkentaa, voisi lisätä, että luonnonhistoria on pikemminkin empiiristä ja kuvailevaa kuin teoreettista ja spekuloiavaa luontoa koskevaa tietoa. ”Tekstihistoria” olisi vastaava nimitys sille tekstiä, sen kirjoittajaa ja kirjoitustilannetta koskevalle tiedolle, jonka avulla ymmärrämme tekstin.

(2) Spinozan puhe ”luonnonolioiden määritelmien johtamisesta luonnonhistoriasta” osoittaa, että hänellä oli varsin aristoteelinen käsitys tieteestä. Aristoteleen mukaanhan tiede pyrkii määritelmiin, jotka ovat löytämisjärjestyksessä tieteen ”viimeisiä” saavutuksia, vaikka perustelujärjestyksessä ne ovat ”ensimmäisiä” lähtökohtia. Tämä pätee ainakin luonnontieteestä, mutta myös raamatuntulkinnassa vastaavia tutkimuksen päätepiteitä ovat eri Raamatun kirjojen kirjoittajien ajatukset ja tarkoitukset (*mentes auctorum*), ja tutkimuksen lähtökohtia ovat puolestaan (i) hermeneutiikan ”yleiset opit” ja (ii) tulkittavaa tekstiä koskevat faktatiedot, joista voisi käyttää yhteisnimeä ”tekstihistoria”.

(3) Spinozan luonnontieteen ja tulkintataidon metodien luonnehdinnat ovat samaa loogista muotoa:

Asian X tulkinta koostuu (i) X:n *historian* laatimisesta ja (ii) X:n loogisesta johtamisesta siitä.

Spinozan (seuraavissa jaksoissa tarkemmin esiteltävän) oman tulkintaopin mukaan on toisarvoinen kysymys, ovatko kirjoittajan näkemykset tutkimuskohteestaan tosia. Tärkeitä ovat sen sijaan seuraavat kysymykset: (i) Miten kirjoittaja itse ymmärsi käsittelemänsä asian? (ii) Miten hänen siitä sanomansa asiat sopivat yhteen hänen muiden näkemystensä kanssa? Nämä kysymykset vaivaavat jokaista tulkintatieteen harjoittajaa.

3. Merkitys edeltää totuutta

Spinoza sanoo edellä lainatussa tulkintametodia koskevassa kohdassa, että tulkinnan tarkoituksena on kirjoittajan aikomuksen (*mens authoris*) rekonstruktio. Kirjoittajan aikomus johdetaan siis loogisesti (1) tekstin ”historiasta” sanan laajassa merkityksessä ja (2) tietyistä yksittäisistä tiedoista ja yleisistä periaatteista. Tässä prosessissa on olennaista tehdä Descartesin metodista epäilyä tai Husserlin *epokhêta* (havaintokokemukseen sisältyvän totuusvaateen neutralointia) muistuttava operaatio: huomiota ei pidä kiinnittää tutkittavan tekstin totuuteen vaan sen merkitykseen. Kuten Spinoza sanoo: ”Meitä kiinnostaa vain tekstien merkitys, ei niiden totuus [De solo enim sensu orationum, non autem de earum veritate laboramus].” (TTP VII, 86.)

Heti kirjansa esipuheessa Spinoza valittaa aikansa ”fundamentalistien” epäkriittistä lähestymistapaa. He lähtevät liikkeelle siitä oletuksesta, että Raamatun jokainen sana on tosi. Se on heille aksioma, jota ei pidä epäillä. Spinozan mukaan se on kuitenkin vain yksi teoreema tai hypoteesi muiden joukossa, ja siksi sitä pitäisi ankarasti koetella ennen hyväksymistä: ”plerique tanquam fundamentum supponunt ... [Scripturam] ubique veracem, et divinam esse; id nempe ipsum, quod ex ejusdem intellectione, et severo examine demum deberet constare” (TTP Esipuhe, v).

Spinoza päättää itse ryhtyä tähän tehtävään eli Raamatun luotettavuuden arviointiin ja sen kriteerien kehittämiseen. Hänen ohjelmanjulistuksensa tuo mieleen Descartesin aikomuksen selvittää asiat kerran elämässään:

... sedulo statui, Scripturam de novo integro et libero animo examinare, et nihil de eadem affirmare, nihilque tanquam ejus doctrinam

admittere, quod ab eadem clarissime non edocerer. (TTP Esipuhe, v)

... päätin tutkia Raamattua uudestaan, tarkasti ja puolueettomasti, niin etten väittäisi siitä mitään enkä myöskään hyväksyisi sen opiksi mitään, mitä en selvästi näkisi sen sanovan.

Tällainen epistemologisesti pidättyväinen projekti oli uusi ja vallankumouksellinen teologiassa. Spinoza oli tietoinen hankkeensa vaarallisuudesta ja siksi hän ei julkaissut TTP:tä omalla nimellään. Willem van Blijenbergh, Spinozan kirjeenvaihtokumppani, sanoi kirjassaan *Kristinuskon totuus* (1674), että ”Tractatus on täynnä kiehtovia, mutta kammottavia löydöksiä, joiden täytyy olla helvetistä kotoisin. Jokaisen kristityn, ja jokaisen järkevän ihmisen, tulisi kammoksua tällaista kirjaa. Sen tekijä yrittää hylätä kristinuskon ja järkyttää siihen perustuvaa toivoamme.” (Colerus 1999, 579–580.)

Spinoza oivalsi, että ennen kuin Raamatun *totuutta* voidaan arvioida, on ymmärrettävä sen *merkitys* eli tiedettävä, mitä se sanoo. Siksi oli tärkeää paitsi erottaa kysymykset tekstin merkityksestä ja sen totuudesta, myös pohtia perusteellisesti merkityksen luonnetta.

4. Sanan merkitys on sen käyttö

Kysymykset sanojen *merkityksestä* on ratkaistava tutkimalla niiden *käyttöä*:

... ne verum sensum cum rerum veritate confundamus, ille ex solo linguae usu erit investigandus, vel ex ratiocinio, quod nullum aliud fundamentum agnoscit, quam Scripturam. (TTP VII, 86.)

Jottemme sekoittaisi lauseen oikeaa merkitystä sen totuuden kanssa, merkitystä on etsittävä vain kielenkäytöstä tai sellaisista päättelyistä, jotka eivät tunnusta muuta perustaa kuin Raamatun.

Spinozan teosten, erityisesti *Etiikan* ja *Tractatus de intellectus emendationen*, lukijan on äärettömän helppo sortua ajattelemaan, että Spinozan mielestä sanojen merkitykset ovat jonkinlaisia ikuisia platonisia entiteettejä. Toki TTP:stäkin löytyy ”platonistisia” oppeja, kuten esimerkiksi se, että on olemassa ei-

kielellinen tapa käsittää asioita ja että se on kielellistä korkeampi: ”asia ymmärretään, kun se havaitaan vain mielellä ilman sanoja ja kuvia [tum enim res intelligitur, cum ipsa pura mente ex verba et imagines percipitur]”) (TTP IV, 50–51).

Tässä ei ole onneksi tarpeen lähteä pohtimaan näin syvällisiä tietopillisiä kysymyksiä, koska TTP:ssä tutkimuskohteena on tulkinta, jossa (i) on aina kysymys *merkityksen* etsimisestä ja löytämisestä ja jossa (ii) etsitty merkitys liittyy aina *merkkeihin* – Raamatun tapauksessa paperilla oleviin merkkeihin.

Sitä paitsi monia TTP:n kohtia olisi hyvin vaikea sovittaa yhteen merkitysplatonismin kanssa. Spinoza sanoo esimerkiksi, että Raamatun tulkinta on vaikeaa siksi, että sen käyttämien sanojen merkitykset ovat unohtuneet (ks. TTP VII, 92). Miten ikuisia ja muuttumattomia merkityksiä voisi unohtaa?³ Spinoza joutuu myös toteamaan, että ”Raamatussa on monia selittämättömiä kohtia [multae orationes inexplicabiles in S. Scriptis occurrunt]” (TTP VII, 95). Näin ei varmaan olisi, jos sanoilla olisi ikuiset platoniset merkitykset, jotka nähtäisiin intellektuaalisella intuitiolla. Tässä epätäydellisessä maailmassamme ainoa tapa löytää ilmaisun merkitys on vertailla sen erilaisia käyttötapoja eri yhteyksissä. Tämäkään metodi ei välttämättä johda hyväksyttävään tulokseen, koska esimerkiksi yhden profetan sanoja ei automaattisesti voida käyttää toisen profetan sanojen selittämiseen. (TTP VII, 95.)

TTP:n XII luvussa Spinoza ilmaisee selvimmin käsityksensä kielellisestä merkityksestä: ”Verba ex solo usu certam habent significationem” (TTP XII, 146) eli ”sanat saavat tarkan merkityksensä vain käytöstä”. Tämän selvemmin merkityksen käyttöteoriaa ei juuri voisi ilmaista. Samanlaiseen muotoiluun päätyi myös teorian isänä pidetty Wittgenstein, jonka mukaan ”Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache” (1953, § 43). Erona näiden kahden filosofin välillä on kolmen vuosisadan lisäksi vain se, että Spinozan formulaatiossa on varaus tai tarkennus ”certam” eli ”tarkan”.⁴ Varauksen vuoksi olisi periaatteessa mahdollista väittää, että Wittgensteinilla sanan *koko* merkitys tulee käytöstä, kun taas Spinozalla vain osa –

³ On tietysti teoriassa mahdollista väittää, että Spinoza hyväksyy sekä muuttumattomat merkitykset että alituisesti muuttuvat sanojen käyttötavat, esimerkiksi siksi, että hän uskoo, että on olemassa jokin ajattelun kieli, jonka merkitykset eivät muutu. Paljon taloudellisempaa on kuitenkin tulkita Spinoza merkityksen käyttöteorian edustajaksi.

⁴ Wittgensteinillakin oli varaus: hänen teesinsä pätee vain ”useimmista” sanoista.

ydinmerkitystä tarkentava osa – merkityksestä tulee käytöstä. Tämä näkemys ei tunnu uskottavalta, koska sille ei ole helppo löytää tukea TTP:stä.

Ongelma on samanlainen kuin tulkittaessa Fregen kontekstiperiaatetta: ”Nur im Zusammenhange eines Satzes bedeuten die Wörter etwas” (Frege 1884, § 62).⁵ Tarkoittaako Frege sitä, että sanalla ei ole *mitään* merkitystä ennen kuin se asetetaan lauseeseen, vai sitä, että sana saa *tarkan* merkityksensä vasta jonkin lauseen osana? Eikö esimerkiksi sanalla ”kuusi” todellakaan ole minkäänlaista merkitystä ennen kuin lausekontekstista selviää, onko kyseessä havupuu vai kokonaisluku? Vai onko tulkitsijan peräti oletettava ennen tutkimuksen alkamista, että sana ”kuusi” voi viitata aivan mihin tahansa – vaikkapa keltaisuuteen tai vihaan – ja että hän on ylen hämmästynyt, kun sillä viitataan havupuuhun? Koko ajatus on järjetön, sillä mikään tulkinta ei ikinä pääsisi alkuun, ellei tulkitsija ”aina jo” käsittäisi sanoilla olevan *jonkinlaisen* merkityksen, joka tulkinnassa sitten tarkentuu. Miten yksikielinen sanakirjakaan olisi ylipäänsä mahdollinen, jos sanat saisivat merkityksensä *vain* lausekontekstista?⁶

5. Sanan merkitys ja puhujan tarkoitus

Spinozan käsitys merkityksestä tuntuu hieman paradoksaaliselta. Luonnehtiessaan tulkintaa hän sanoo, että tulkinnan tehtävänä on *kirjoittajan tarkoituksen (mens authoris)* paljastaminen (TTP VII, 84). Mutta myöhemmin hän vaatii useaan kertaan *pysymään tekstin sisällä* ja olemaan lisäämättä siihen omiaan (esim. TTP Esipuhe, v). Onko tehtävänä siis etsiä kirjoittajan intentio, mutta pysyä silti koko ajan tekstin sisällä? Miten tulkinnassa tarvittavan ”tekstihistorian” laatiminen onnistuu, jos tulkitsija ei saa mennä tulkittavan tekstin ulkopuolelle?

Ongelman selvittämisen voi aloittaa erottamalla merkityksen ja tarkoituksen. Apostoli Paavali on ollut kuolleena lähes kaksi tuhatta vuotta,

⁵ Fregellä on toinenkin, selvästi varovaisempi kontekstiperiaatteen muotoilu: ”Nur in [einem vollständigen Satz] haben die Wörter eigentlich eine Bedeutung” (Frege 1884, § 60). Sana ”eigentlich” muuttaa kaiken. Sen vaikutus on sama kuin Spinozan sanalla ”certam”.

⁶ Ei muuten ole lainkaan selvää, kuka on merkityksen käyttöteorian keksijä. Spinozan lisäksi Wittgensteinilla on ainakin kaksi muutakin kovaa kilpailijaa: Erich Danz ja Antoine Meillet. Danz (1912, § 4.ii) sanoo: ”Unter der Bedeutung eines gesprochenes Wortes versteht man die Vorstellung, welche die Mitglieder derselben Sprachgemeinschaft ... zu einer gewissen Zeit mit einem Lautgebilde verbinden.” Meillet (1918, 177) sanoo vielä selvemmin: ”le sens d’un mot ne se laisse définir que par une moyenne entre [ses] emplois linguistiques”.

joten on vaikeaa ymmärtää hänen antamansa normin ”Nainen vaietkoon seurakunnassa” sisältö. Voimme kyllä tietää – tai ainakin saada selville – hänen käyttämiensä sanojen *merkityksen*, mutta on aivan eri asia saada selville hänen *tarkoituksensa*. Paavalin 1 Kor. 14:34:ssa käyttämät tarkat sanat olivat: ”olkoot vaimot vaiti teidänkin seurakunnankokouksissanne [αἱ γυναῖκες ὑμῶν ἐν ταῖς ἐκκλησιαῖς σιγατωσαν]”. Niiden *merkitys* on selvä, mutta ei ole selvää, mitä hän *tarkoitti* niillä, eli minkä vaikutuksen hän halusi niillä saada aikaan. Joidenkin teologioiden mukaan hän halusi kieltää ikiajoiksi naisten puhumisen seurakunnan tilaisuuksissa – ja siksi myös naispappeuden. Toisten mukaan hän halusi vain hiljentää joukon korinttilaisia naisia, jotka olivat käyttäytyneet häiritsevästi joissakin tietyissä seurakunnan tilaisuuksissa pari tuhatta vuotta sitten, joten sanoilla ei ole mitään yhteyttä naispappeuteen.

Merkityksen ja tarkoituksen käsitteet menevät pahasti päällekkäin. Englannissa niistä molemmista voidaan käyttää sanaa ”meaning”, mutta on niillä syvällisempikin yhteys: sanalla ei voisi olla merkitystä, ellei olisi ihmisiä, jotka tarkoittavat tuolla sanalla jotain. Ja noita ihmisiä pitää olla paljon. Yksityiskielen ideassa itsessään ei ole mitään ristiriitaista, sillä Tolkien kehitti useita sellaisia. Yksityiskieltä ei kuitenkaan voi käyttää ihmisten väliseen kommunikaatioon, koska onnistunut kommunikaatio edellyttää sanojen käyttöä niiden *tavanomaisessa* merkityksessä. Siksi yksittäisen puhujan kannattaa käyttää sanoja samalla tavalla kuin muut, jos haluaa tulla ymmärretyksi. Puhuminen on siis sosiaalinen teko, koska sen aikomus sisältää viittauksen muiden ihmisten asenteisiin ja uskomuksiin.

Lisäksi ihmisellä ei voisi edes olla tarkoitusta ilmaista sanoillaan jotain, ellei noilla sanoilla olisi jo vakiintuneita merkityksiä – siis muita ihmisiä, jotka tarkoittavat noilla sanoilla samaa kuin muutkin ihmiset. Siksi ei ole mielekäästä kysyä, mikä on sanan ”kotka” merkitys, ellei samalla kerro, minkä kieliyhteisön sanaa tarkoittaa. Suomen sanalla ”kotka” viitataan isoon petolintuun, mutta bulgarian samoin ääntyvä sana ”котка” merkitsee kissaa.

Vaikka sanan merkitys ei olekaan sama asia kuin yksittäisen puhujan tarkoitus, sanan merkitystä ei voi erottaa koko *kieliyhteisön* tarkoituksista – siis heidän tavastaan käyttää kieltä. Koska ”kieliyhteisön yhteinen käyttötapa” on sama asia kuin ”merkitys”, yksilön kannattaa sovittaa oman tarkoituksensa yhteen sanojen merkityksien eli kieliyhteisön tapojen kanssa.

Juristeilla on tapana erottaa laki- ja sopimustekstin tulkinnessa

subjektiivinen ja objektiivinen tarkoitus. Subjektiivinen tarkoitus on se yksityinen mielentila, joka tekstin laatijalla oli tekstin laatimisen hetkellä. Objektiivinen tarkoitus on puolestaan se tarkoitus, jonka tosiasioita perillä oleva, objektiivinen, puolueeton ja hyvässä uskossa toimiva tarkkailija omistaa tekstin laatijalle. Karkeasti sanottuna se on mielentila, joka *kenellä tahansa* normaalijärkisellä, asioista perillä olevalla ihmisellä *on täytynyt* olla tuollaista tekstiä laatiessaan.

On selvää, että Paavalin subjektiivinen aikomus – yksittäisenä ja yksilöllisenä mentaalisenä tilana – on kadonnut lopullisesti tietämättömyyden hämärään. Yleisemminkin voi sanoa, että ellei subjektiivinen mielentila ole jättänyt minkäänlaisia objektiivisia *jälkiä* (paperille, filmille, ääninauhalle, muiden muistiin tms.), on aivan kuin sitä ei olisi koskaan ollutkaan.

Tekstin tulkinta näyttää siis olevan pikemmin objektiivisen kuin subjektiivisen aikomuksen paljastamista, ja siksi se näyttää välttämättä vaativan kolmannen persoonan näkökulman omaksumisen. Spinozan kannalta kolmannen persoonan näkökulma on ongelmallinen. Toisaalta hän nimittäin sanoo, että tulkinnassa tekstiin ei pidä lukea sisälle asioita, joita siellä ei ole, eikä poistaa asioita, jotka siellä ovat. Mutta toisaalta hän myös sanoo epäsuorasti, että kirjoittajan aikomusta selvitetessä voidaan turvautua myös kolmannen persoonan näkökulmaan ja sen myötä tekstin- ja mielenulkoisiin seikkoihin. Näin hän tekee esimerkiksi seuraavassa kohdassa:

[N]ostras cogitationes ita determina[re debemus], ut nullo praejudicio praeoccupemur, ne scilicet auctori, vel ei, in cuius gratiam auctor scripsit, plus minusve justò tribuamus, & *ne de ullis aliis rebus cogitemus, quam de iis, quas auctor in mente habere potuerit, vel quas tempus & occasio exegerit.* (TTP VII, 96; kursivointi lisätty.)

[Meidän pitää esittää tulkintamme niin,] ettemme omista kirjoittajalle emmekä sille, jonka puolesta hän kirjoittaa, enempiä eikä vähempää kuin mitä hän aikoi, ja *keskitymme vain niihin asioihin, jotka kirjoittajalla olisi voinut olla mielessään tai joita aika ja tilanne vaativat.*

Spinoza mainitsee tässä kaksi *hypoteettista* seikkaa, joita käytetään tekstin tulkinnassa: (i) asiat, jotka kirjoittajalla *olisi voinut olla* mielessään, ja (ii) asiat,

joita *aika ja tilanne vaativat*. Kumpikaan niistä ei ole aktuaalinen subjektiivinen mentaalinen tila.

Se, ”mitä aika ja tilanne vaativat”, luetaan *teon* tulkinnassa automaattisesti tekijän mielessä olleiksi asioiksi. Palavasta talosta ulos juosseen ihmisen voidaan huoletta tulkita uskoneen, että talo palaa ja että taloon jääminen merkitsisi kuolemaa, ja halunneen sekä päästä talosta ulos että välttää kuoleman. Kenellä tahansa samassa tilanteessa olevalla on samat halut ja uskomukset. Tai ainakin niin on järkevää olettaa. (On helppo kuvitella tilanne, jossa tuo järkevä oletus osoittautuu vääräksi: Mies juoksee ulos palavasta talosta kalsareissaan, mutta hänellä ei ole aavistustakaan siitä, että talo on liekeissä. Paon syy oli se, että paholainen oli tullut hänen asuntoonsa eteisen peilin läpi.)

Sanottu pätee myös tekstien tulkinnasta. Se on vähintäänkin yhtä vaikeaa kuin tekojen tulkinta. (Itse asiassa se on vaikeampaa, koska tekstien tulkintaan sisältyvät kaikki tekojen tulkinnan vaikeudet – koska teksti on aina teon tulos – ja lisäksi joukko uusia vaikeuksia.) Emme koskaan pääse sisälle kirjoittajan mieleen, mutta voimme (usein) päätellä – *sub specie familiaritatis* – että hän tarkoittaa juuri sitä, mitä hänen asemassaan olevan ihmisen voisi olettaa tarkoittavan. Voimme esimerkiksi olla varmoja, että ihminen, joka jättää keittiön pöydälle lapun ”Menin käymään kaupassa”, erittäin todennäköisesti joko (1) todella meni kauppaan, tai ainakin (2) aikoi mennä kauppaan, tai vähintäänkin (3) aikoi herättää viestinsä lukijoissa uskomuksen, että hän meni kauppaan, vaikka tosiasiallisesti hän meni jonnekin muualle ja vaikka hän ei koskaan aikonutkaan mennä kauppaan.

Kun jossakin raamatunkohdassa on epäselvyyttä, voimme yrittää poistaa sen vertaamalla sitä saman luvun tai kirjan muihin jakeisiin. Emme voi kuitenkaan ilman muuta päätellä, että *kaikkien* Raamatun kirjojen kirjoittajat käyttävät sanoja samalla tavalla (TTP VII, 90).⁷

Kolmannen persoonan apuun turvautuminenkaan ei aina auta. Tietämättömyys kirjoituskontekstista (tekstihistoriasta) voi tehdä tekstin tulkinnan mahdottomaksi: ”Ellemme tunne näitä asioista [kirjoittajaa, kirjoitustilannetta jne.], emme tiedä myöskään, mitä kirjoittaja aikoi tai *saattoi aikoa* [quid author intenderit aut *intendere potuerit*]” (TTP VII, 95; kursivointi

⁷ Tätä oppia ei Luther eivätkä nykyiset fundamentalistit olisi hyväksyneet. He yhdistelevät surutta Uutta ja Vanhaa testamenttia, koska kaikessa kuuluu sama Pyhän Hengen ääni.

lisätty).

Spinozan kirjoittajan intentio (*mens auctoris*) näyttää olevan samanlainen hyvänlaatuinen fiktio kuin juristien käyttämä *lainsäätäjän tahto*. Sen avulla he selittävät, mitä jokin hämärä lainkohta merkitsee. Oikeusfilosofi Gustav Radbruch suhtautuu koko käsitteeseen epäilevästi. Hänen mukaansa lainsäätäjän tahto on vasta tulkinnan lopputulos ja siksi sillä saattaa olla sisältö, jota lainsäätäjä itse ei ollut aikonut:

Lainsäätäjän tahto ei siis ole tulkintaväline, vaan tulkinnan tavoite ja lopputulos ... Siksi on mahdollista väittää lainsäätäjän tahdoksi sellaista, mitä lainlaatija ei koskaan ollut tietoisesti ajatellut. (Radbruch 2003, 107.)

6. Spinozan tulkintasäännöt käytännössä

Lopuksi on syytä katsoa, miten Spinoza itse tulkitsi Raamattua ja miten hänen tulkintasääntönsä poikkesivat keskiajan johtavan juutalaisen ajattelijan Moses Maimonideen tulkintasäännöistä ja -käytännöstä.

Spinoza toistaa TTP VII:ssä useita kertoja ”raamatuntulkinnan yleissäännön” (*Regula universalis interpretandi Scripturam*), jonka mukaan Raamatun tulkinnassa on pidättäydyttävä siihen, mitä Raamattu itse sanoo. Tässä on tämän periaatteen yksi muotoilu:

Tota itaque Scripturae cognitio ab ipsa sola peti debet. ... Regula igitur universalis interpretandi Scripturam est, nihil Scripturae tanquam ejus documentum tribuere, quod ex ipsius historia quam maxime perspectum non habeamus. (TTP VII, 85.)

Kaiken Raamattua koskevan tiedon on siis tultava siitä itsestään. ... Raamatun tulkinnan yleissääntö on siis se, ettei Raamatulle pidä omistaa mitään sellaista oppia, jota emme selvästi voi johtaa sen historiasta.

Spinozan tulkintaperiaate muistuttaa hyvin paljon Lutherin *Scriptura sui ipsius interpretres* -teesiä; ne molemmat sanovat, että Raamattua pitää tulkita Raamatun avulla. Käytännössä Spinoza on kuitenkin paljon Lutheria radikaalimpi

teesinsä ”merkitys edeltää totuutta” vuoksi. Luther ei olisi voinut kuvitellakaan tällä tavalla kyseenalaistavansa Raamatun totuutta – ei edes metodologisenä keinona. Spinoza taas sanoo, ettei Raamatun lauseita saa pitää tosina ilman perusteellista tutkimusta ja että lauseiden merkitys pitää saada selville ennen kuin niiden totuutta voidaan lähteä arvioimaan.

TTP VII:n loppupuolella Spinoza kritisoi ankarasti Maimonideen näkemystä, että vaikka Raamattu on kirjoitettu tavallisille ihmisille heidän ymmärryskykynsä mukaisesti, se sisältää silti syvällisiä filosofisia totuuksia, jotka vain hyvä tulkitsija pystyy löytämään sieltä, jos hän on valmis näkemään vaivaa. Kun Raamatusta esimerkiksi sanotaan, että joku profetta ”näki” Jumalan (2. Moos. 24:10, 4. Moos. 12:8, Jes. 6:1–3, Hes. 1:26–29), ei suinkaan tarkoiteta tavallista aistimellista näkemistä, vaan sellaista intellektuaalista näkemistä, jolla ”nähdään” geometrisen todistuksen pätevyys.

Kaikki Raamatun sisältämät totuudet eivät kuitenkaan ole ilmeisiä, vaan ne pitää kaivaa esille tulkinnan avulla. Mikä tahansa tulkinta ei kuitenkaan kelpaa edes Maimonideelle. Jos tulkinta on järjenvastainen, se on hylättävä. Esimerkiksi jokainen tulkinta, josta seuraisi, että Jumalalla on fyysinen ruumis, on hylättävä suoralta kädeltä, koska Jumala ei ole fyysinen olio. Maimonides tietysti myöntää, että Raamattu puhuu monessa kohdassa *ikään kuin* Jumalalla olisi ihmisruumiin kaltainen ruumis, mutta hänen mukaansa nämä kohdat on ymmärrettävä allegorisiksi. Raamatusta on siis kohtia, joita ei voi tulkita kirjaimellisesti vaan jotka pitää ymmärtää allegoriseksi puheeksi, joka yrittää välittää tavalliselle ihmiselle syvällisiä filosofisia totuuksia.

Spinoza kehitti edellä käsitellyn periaatteen ”merkitys edeltää totuutta” nimenomaan kritisoidakseen Maimonidesta. Tämän näkemyksestään käytännössä seuraa, että *raamatunkohdan tulkinta riippuu sen totuusarvosta*: jos kohta esittää toden väitteen, se on tulkittava kirjaimellisesti; jos se esittää epätoden väitteen, se on tulkittava allegorisesti. Maimonideen käsityksen ongelma on se, että se asettaa kärryt hevosen eteen, kuten englantilaiset sanovat. Miten voisimme tietää, onko hämärä väite tosi, ennen kuin tiedämme, mitä se tarkoittaa, eli ennen kuin se on tulkittu?

Spinoza tyrmää täysin myös Maimonideen ajatuksen, että Raamattu sisältää syvällisiä filosofisia totuuksia. ”Raamatun oppi ei sisällä vaikeita spekulatioita eikä filosofisia järkeilyjä, vaan hyvin yksinkertaisia asioita, jotka voi käsittää hidasjärkisinkin” (TTP 13, 153). Kun olemme ymmärtäneet

nämä yksinkertaiset asiat eli perustavimmat moraaliset opetukset, meidän ei enää tarvitse välittää noista muista opetuksista (TTP VII, 97).

Spinozalta löytyy hyvä esimerkki siitä, mitä periaate ”merkitys edeltää totuutta” merkitsee käytännön tulkinnassa. Miten pitäisi tulkita ne raamatunkohdat, joissa sanotaan, että Jumala on tuli? Tällainen on esimerkiksi 5. Moos. 4:24. Uskoiko Mooses todella, että Jumala on tuli? Maimonideen ratkaisu on suoraviivainen: koska järki sanoo, että tuli on jotain materiaalista ja että Jumala ei ole materiaallinen, Jumala ei voi olla tuli. Siksi allegorinen tulkinta on tarpeen tässä.

Spinoza on samaa mieltä siitä, että tällaiset raamatunkohdat on tulkittava allegorisesti, mutta hänen perustelunsa on toinen: ”Kysymystä, uskoiko Mooses Jumalan olevan tuli, ei pidä missään tapauksessa ratkaista sillä perusteella, onko käsitys järkevä vai ei, vaan ainoastaan Mooseksen muiden sanomisten perusteella” (TTP VII, 86–87). Ja Mooses sanookin monissa muissa kohdissa, ettei Jumala muistuta mitään näkyviä olioita. Nyt tulkitsijalla on kaksi mahdollisuutta: hän tulkitsee allegorisesti joko väitteen, että Jumala on tuli, tai kaikki nuo muut väitteet (TTP VII, 87). Koska allegoriseen tulkintaan tulisi turvautua mahdollisimman harvoin, on järkevää soveltaa sitä yhteen kohtaan mieluummin kuin moneen. Seuraavaksi Spinoza kysyy, löytyykö Raamatusta muita kohtia, joissa sanaa ”tuli” on käytetty metaforisesti. Sellainen löytyy Jobin kirjasta (Job 31:12), jossa ”tuli” tarkoittaa vihaa ja kateutta.⁸

Kaikki tulkintaongelmat eivät kuitenkaan ratkea näin onnellisesti. Jos allegorinen tulkintakaan ei onnistu, tulkitsijan on vilpittömästi myönnettävä tappionsa:

Ellei [tekstinkohtien harmonisointi] onnistu kielenkäytön perusteella (ex usu linguae), silloin ne ovat yhteen sopimattomia, ja tulkitsija tulee pidättyä esittämästä arvostelmaa niistä. (TTP VII, 87.)

Näin Spinoza on selittänyt, mitä periaate ”Kaikki Raamattua koskeva tieto on johdettava siitä itsestään” (TTP VII, 87) käytännössä tarkoittaa. Se tarkoittaa siis muun muassa sitä, että ensin on yritettävä kirjaimellista tulkintaa, mutta

⁸ Spinozan esimerkkitaupaus ei kestä tarkempaa tutkimusta. Hänen ideansa tulee silti täysin selväksi.

ellei se onnistu, myös allegorinen tulkinta on sallittu, mutta vain toiseksi viimeisenä mahdollisuutena. Viimeinen on pyyhkeen heittäminen kehään. Sekin vaihtoehto on joskus valittava.

7. Yhteenveto teesien muodossa

(1) Spinozan tulkintateorian lähtökohta on sama kuin protestanteilla: Raamattua on tulkittava Raamatun avulla. Spinoza on kuitenkin radikaalimpi kuin Luther: hän ei epäkriittisesti *oletta* Raamatun jumalallisuutta ja sen lauseiden totuutta. Ne ovat hypoteeseja, joita pitää koetella ankarasti.

(2) Hermeneutiikalla ja luonnontieteellä on olennaisesti sama metodi: ensin rakennetaan tutkittavan asian ”historia” (eli sitä koskevan tiedon systemaattinen esitys), josta sitten johdetaan loogisesti kaikki sitä koskevat yksittäiset tiedot.

(3) Keskeinen metodologinen idea on merkitys- ja totuuskysymysten erottaminen. Emme voi perustellusti pitää tekstin lauseita tosina (tai epätosina) ennen kuin olemme selvittäneet, mitä ne merkitsevät (vai merkitsevätkö ne mitään).

(4) Kysymykset Raamatun sanojen merkityksistä on ratkaistava Raamatun avulla, vertailemalla saman sanan erilaisia käyttötapoja eri konteksteissa. Sanan merkitys on niiden käyttö (tai vähintäänkin paljastuu sen käytöstä).

(5) Tulkinta on kirjoittajan intention (*mens auctoris*) paljastamista. ”Kirjoittajan intentio” on kuitenkin vain nimi sille rekonstruktioille, jonka tulkitsija tekee *sub specie familiaritatis* tulkittavan tekstin takana *mahdollisesti* (eli kaiken evidenssin valossa *luultavasti*) olevasta aikomuksesta.

(6) Maimonideen tulkintateoria on monella tapaa virheellinen. (i) Hän ei erota merkitys- ja totuuskysymyksiä, ja siksi hänellä vastaus totuuskysymykseen määrää vastauksen merkityskysymykseen. (ii) Hän olettaa, että jokaisella Raamatun lauseella on syvälinen metafyyminen merkitys, joka löytyy allegorisella tulkinnalla. – Nämä oletukset vievät hänet harhaan.

(7) Spinozan omia tulkintasääntöjä ovat mm. seuraavat: (i) Tulkinnessa on lähdettävä kirjaimellisesta tulkinnasta eli tavanomaisen käyttötavan mukaisesta tulkinnasta. (ii) Allegoriseen tulkintaan saa turvautua vain siinä tapauksessa, että kirjaimellinen tulkinta johtaisi ristiriitaan muiden *saman kirjoittajan* laatimien tekstinkohtien kanssa. (iii) Myös allegorista tulkintaa

käytettäessä on pyrittävä poikkeamaan mahdollisimman vähän kirjaimellisesta tulkinnasta. (iv) On myönnettävä, ettei kaikille tekstinkohdille löydy uskottavaa tulkintaa.

Itä-Suomen yliopisto

Kirjallisuus

- Colerus, Jean (1999). *Vie de B. Spinoza*. Teoksessa Spinoza, *Éthique*. Paris: Éditions du Seuil 1999, 547–600.
- Danz, Erich (1912). *Einführung in die Rechtsprechung – Anleitung für junge Juristen*. Jena: Verlag von Gustav Fischer.
- Frege, Gottlob (1884). *Die Grundlagen der Arithmetik*. URL = <https://archive.org/details/diegrundlagende00freggoog>
- Meillet, Antoine (1918). Kirja-arvio Felix Restrepon teoksesta *Diseño de semántica general*, *Bulletin de la Société de linguistique de Paris* 21. URL = <https://archive.org/stream/bulletin2122sociuoft#page/n181/mode/2up>
- Radbruch, Gustav (2003). *Rechtsphilosophie*. 2. Aufl. München: C. F. Müller.
- Ray, Stephen K. (1997). *Crossing the Tiber: Evangelical Protestants Discover the Historical Church*. San Francisco: Ignatius Press.
- Spinoza, Baruch (1979). *Tractatus theologico-politicus / Theologisch-politische Traktat*. Opera / Werke vol. I. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Spinoza, Baruch (1978). *Ethica / Ethik & Tractatus de intellectus emendatione / Abhandlung über die Berichtigung des Verstandes*. Opera / Werke vol. II. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Wittgenstein, Ludwig (1953). *Philosophische Untersuchungen*. URL = http://www.geocities.jp/mickindex/wittgenstein/witt_pu_gm.html

Spinoza on Self-Contentment¹

Daryl de Bruyn

Self-contentment is pleasure arising from a man's contemplation of himself and his power of activity. (Def. of the Emotions 25.)

Spinoza sets out “to enquire whether there existed a true good, one which was capable of communicating itself and could alone affect the mind to the exclusion of all else, whether, in fact, there was something whose discovery and acquisition would afford a continuous and supreme joy to all eternity” (TdIE§1). In my view, his philosophical quest culminates in the discovery of true self-contentment, which is the highest good we as humans can hope for (4p52sch). I argue in this brief paper that Spinoza's *acquiescentia in se ipso* is an active emotion, a pleasure that co-occurs with the love of God, which ensues when the mind certainly affirms the adequate idea of God or Nature. I also claim that his notion of self-contentment is closely related to the highest form of knowing, i.e. the intuitive knowledge of the third kind.

In Spinoza, reaching true spiritual contentment is not easy (5p42sch) and requires a disciplined lifestyle, combined with a proper philosophical method. His rationalist method stipulates that an object of philosophical enquiry must be conceived either through the adequate idea of its own nature or through the true idea of its cause (TdIE§92).² In Spinoza's system humans are perceived as dependent things or modes or affections of substance. To adequately understand the nature of our self-contentment we must then conceive our self through the clear and distinct or adequate idea of our proximate cause (5p28def). The all-important point is that we cannot understand ourselves adequately in isolation from our cause, i.e. God or Nature, and should not think of our self as something independent. Spinoza's self-contentment arises from understanding and accepting that we are

¹ I gladly dedicate this little paper to Olli Koistinen, who first introduced to me the important and interesting notion of the self in Spinoza's thought.

² For more on this see De Bruyn 2014.

inextricably part of something else on which we totally depend, i.e. God or Nature.

Spinoza assigns the highest degree of intellectual pleasure or self-contentment to God:

God loves himself with an infinite intellectual love. (5p35.)

God is absolutely infinite (1def6); that is (2def6), God's nature enjoys infinite perfection, accompanied (2p3) by the idea of itself, that is (1p11 and 1def1), by the idea of its own cause; and that is what, in 5p32cor, we declared to be intellectual love. (5p35def.)

It is interesting to compare the above citation to the following one, which deals with our human self-contentment:

We take pleasure in whatever we understand by the third kind of knowledge, and this is accompanied by the idea of God as cause. (5p32.)

From this kind of knowledge there arises the highest possible contentment of mind (5p27), that is (Def. of Emotions 25), the highest possible pleasure, and this is accompanied by the idea of oneself, and consequently (5p30) also by the idea of God, as cause. (5p32def.)

From the third kind of knowledge there necessarily arises the intellectual love of God (*amor Dei intellectualis*). (5p32cor.)

Note in the above, firstly, that Spinoza refers to both God's love of himself and to our love of God, as the intellectual love of God. In the former, God's self-contentment follows from the idea of himself, accompanied by the idea of being self-caused. In the latter, our self-contentment follows from the idea of our self, accompanied by the idea of God, as cause. Secondly, Spinoza associates the intellectual love of God with intuitive knowledge of the third

kind.³ I argue that both self-contentment and the intellectual love of God, arise from intuitive knowledge and amount to the same thing. Spinoza's self-contentment is the active emotion or pleasure that co-occurs with the adequate idea of myself (the idea of the eternal essence of the mind and the body), accompanied by the idea of God, as cause. Spinoza's third kind of knowing surpasses abstract conceptual knowledge and is akin to an intuitive *feeling* of wellbeing that arises from knowing certainly that whatever is, is contained *in* and follows from God or Nature (1p15). This intellectual delight co-occurs with the simultaneous conscious comprehension of oneself and other things with the idea of God, as proximate cause.⁴ This is what Spinoza seems to be getting at in the closing lines of his Ethics (my emphasis):

The wise man, insofar as he is considered as such, suffers scarcely any disturbance of spirit, but being conscious, by virtue of a certain eternal necessity, *of himself, of God and of things*, never ceases to be, but always possesses true spiritual contentment. (5p42sch.)

By contrast, the “ignorant man, besides being driven hither and thither by external causes, never possessing true contentment of spirit, lives as if he were unconscious of himself, God and things” (Ibid.). The following citation from his *Theological-Political Treatise* gives a clearer idea of what Spinoza is getting at when he speaks of intuitively relating all aspects of the idea of oneself and other things to the idea of God, as cause:

Again, since the power of Nature in its entirety is nothing other than the power of God through which alone all things happen and are determined, it follows that whatever man – who is also part of Nature – acquires for himself to help preserve his own being, or whatever Nature provides for him without any effort on his part, all this is provided for

³ In my view it is the active emotion of pleasure, which accompanies intuitive knowledge of the third kind that is its distinguishing characteristic. In comparison, conceptual reason or knowledge of the second kind is rather abstract and therefore a tad less convincing (5p36sch).

⁴ Towards the end of the fifth meditation Descartes seems to be thinking along the same lines (my emphasis): “Now, however, I have perceived that God exists, *and at the same time* I have understood that everything else depends on him, and that he is no deceiver, and I have drawn the conclusion that everything which I clearly and distinctly perceive is of necessity true.” (Descartes 1988, 109.) For more good discussion of the fifth meditation see Nolan 2015 and Koistinen 2014.

him solely by the divine power, acting either through human nature or externally to human nature. Therefore whatever human nature can effect solely by its own power to preserve its own being can rightly be called God's internal help, and whatever falls to a man's advantage from the power of external causes can rightly be called God's external help. And from this, too, can readily be deduced what must be meant by God's choosing, for since no one acts except by the predetermined order of Nature – that is, from God's eternal direction and decree – it follows that no one chooses a way of life for himself or accomplishes anything except by the special vocation of God, who has chosen one man before others for a particular work or a particular way of life. Finally, by fortune I mean simply God's direction insofar as he directs human affairs through causes that are external and unforeseen. (TTP3 S: 417.)⁵

Spinoza's intuitive knowledge of the third kind, in my opinion, is the conscious action of the mind by which we continuously associate the idea of our self and other things with God, as direct cause. Self-contentment and pleasure follows inasmuch as such an intuitive understanding of things becomes *second nature* to us, so to speak.⁶ For Spinoza true spiritual contentment depends on ridding the mind of its inadequate ideas, which are the chief cause of its discontent. However, the human condition is such that it is impossible to evade external affections from other things and to avoid having obscure and confused ideas and the ensuing passions. Spinoza's cure for our mental agitation comes in the form of a *medicina mentis* in which the mind actively relates *whatever* mode of thought it has, to God or Nature as closest cause.

⁵ The theory of concurrence is clearly at work here. This theory is found throughout Spinoza's writings (CM2/11; 1p24cor; 2p45sch; 5p36sch). This notion suggests the two elements (*Natura naturans*) and (*Natura naturata*) ineffably "working" together or co-occurring.

⁶ In my view, reason's arguments and proofs are heuristic devices that assist in establishing the adequate idea of God. Once this is clearly known, the mind's affirmation of God's existence follows naturally. Reason acts like the stairs to board an aircraft. Once we have climbed onboard they can be removed. Once reason has lead us to the idea of God, the mind is fully united with it and no longer depends on rational argumentation. Other ideas contained in the true idea of God can now be directly inferred or intuited. See Nolan 2015 for a very good discussion of a similar method in Descartes.

We take pleasure in whatever we understand by the third kind of knowledge and this is accompanied by the idea of God, as cause. (5p32.)

For Spinoza the perfected or emended mind is able to conceive any idea or thing as an expression of the eternal and infinite essence of God (2p45), i.e. as in God and conceived through God. To the extent that we associate *all* aspects of our lives (both pleasure and pain) to God, we are active and consequently content – since no discontent can co-occur with the idea of God (5p18). The following is a brief outline of the main elements of his theory of ‘remedies for the emotions’ as found in the preface and first section of E5.⁷

In Spinoza pleasure arises from adequate or active ideas, whereas pain follows from inadequate or passive ideas. Pleasure is an active emotion that co-occurs with the idea of transition from a state of less perfection to a state of greater perfection. Pain or discontent, on the other hand, accompanies the idea of a shift from a level of greater perfection to a lower level. The remedy for discontent is for the mind to take action – to detach the particular form of agitation or anxiety (the passive emotion) from the unclear thought of an external cause and to attach it to an adequate thought, i.e. the idea of God as cause. According to Spinoza, the force of a passion – be it love, hatred, hope, despair or any other, will then be removed and be replaced by an active emotion (5p2). A powerful intellectual tool of the mind, to resist and remove the passions, is the idea of necessity. In reflecting on the idea of a most perfect being, the mind discovers the very nature of its dependent existence, that in essence, itself and its body is a part of Nature whose order must be followed, that “all things are from the necessity of the divine nature determined to exist and to act in a definite way” (1p29). The more the mind understands and positively affirms this idea, the greater its resistance will be towards the passive emotions that cause its discomfort (5p6). For Spinoza, our human nature is to think and “insofar as we understand, we can desire nothing but that which must be, nor, in an absolute sense, can we find contentment in anything but truth” (4app32). The nature of the mind is knowledge and consequently it will

⁷ Underlying this method is the doctrine of parallelism. Passive emotions are painful and accompany confused ideas, which parallel complex states of the body, caused by external forces. If one relates the confused idea or emotion to a clear and distinct internal cause (the idea of God, as cause), the emotion becomes active and pleasurable, which parallels a changed state in the body as well. See also 5ax2.

naturally affirm any idea that it clearly and distinctly understands, such as the idea of necessity. Spinoza argues that the pleasure arising from the clear and distinct idea of necessity is more powerful than a passion that co-occurs with the obscure and confused thought of external causes (5p8, 9). Unfortunately, it seems that our natural disposition is to be determined externally (4app1), to passively perceive things by the common order of nature and to attribute our trials and tribulations to the fortuitous run of circumstances (2p29sch). However, the idea of necessity is clearly understood to belong to the nature of God or Nature itself. When we attribute things to necessity we are internally conditioned, in that the mind is able (by actively conceiving things under a form of eternity) to attach all its modes of thought *directly* to the idea of God or Nature itself, as cause.

The mind can bring it about that all the affections of the body – i.e. images of things – be related to the idea of God. (5p14.)⁸

However, does the acceptance of necessity or determination not undermine our own self or activity? In Spinoza, human virtue seems to depend on being able “to bring about that which can be understood solely through the laws of his own nature” (4def8). For me, Spinoza gives the best clarification of his concept of freedom in one of his letters to Willem van Blyenbergh.⁹ He writes:

Our freedom lies not in a kind of contingency nor in a kind of indifference, but in the mode of affirmation and denial, so that the less indifference there is in our affirmation or denial, the more we are free. For instance, if God’s nature is known to us, the affirmation of God’s existence follows from our nature with the same necessity as it results from the nature of a triangle that its three angles are equal to two right

⁸ To make steady progress in the pursuit of self-contentment it is important to introduce some rational discipline and order to the mind. This can be done by arranging the vacillating affections of the body according to the order of the intellect (5p10). Examples are planning a right method of living or some fixed rules of life, to which we commit and apply as best as we can (5p10sch). Another task is to clearly understand our emotions. The mind is able to form an adequate idea of any bodily affection, because it is something common to all things (5p4def).

⁹ Spinoza’s correspondence with W. van Blyenbergh (Ep18–24) is very interesting. The discussion represents the confrontation found in early modern thought between the traditional monotheistic idea of God and the new philosophical idea of God’s nature as absolutely infinite.

angles. Yet, we are never so free as when we make an affirmation in this way, despite the fact that we are acting necessarily and from God's decree. (Ep21 S: 825.)

Inasmuch as it is possible for us to act solely in accordance with our own power, the intellectual affirmation of the love of God represents the highest degree of human action. In Spinoza we are not deemed mere spectators or victims of God or Nature's work. Neither are we told to passively accept things with an attitude of Stoic forbearance. The *amor Dei intellectualis* is the action of the intellect, to clearly understand and positively affirm the true idea of God or Nature – *that whatever is, is in God or Nature*. When adequately conceived in this way, all of life, in essence, is seen to follow directly from the infinite essence of God or Nature and to be perfect, i.e. to contain and express all degrees of reality or perfection.¹⁰ Conceiving things in this way and living accordingly is to be determined by the Truth, which is to be fully self-content *in* what must be (4app32).

For blessedness is nothing other than that self-contentment that arises from the intuitive knowledge of God. (4app4.)¹¹

¹⁰ Almog (2014, 95–100) describes Spinoza's intellectual love of God aptly as being 'at home in the universe' and captures the very heart of the matter.

¹¹ Spinoza's theory of self-contentment is based on the arising Cartesian conception of God's nature as absolutely infinite. Spinoza conceives the divine nature as the all-inclusive *base* in which the being of all things whatsoever is contained. For Spinoza there is no privation in God or Nature and perfection and reality amount to the same thing. God is not seen as a person or a judge, but is understood to be perfect only in the sense of grounding all possible expressions of perfection, from the lowest to the highest degree (1app). However, it seems rather difficult to comprehend the absolutely infinite as 'something' at all. Jung (1955) suggests that this ineffability of an infinite being comes close to the Chinese idea of Tao:

Because the eye gazes but can catch no glimpse,
It is called elusive.

Because the ear listens but cannot hear it,
It is called rarefied.

Because the hand feels for it, but cannot find it,
It is called infinitesimal ...

These are called the shapeless shapes,
Forms without form.

Vague semblances.

Go towards them, and you see no front;
Go after it and you see no rear.

Bibliography

- Almog, Joseph (2014). *Everything in its Right Place: Spinoza and Life by the Light of Nature*. New York: Oxford University Press.
- De Bruyn, Daryl (2014). *Spinoza's Concept of Emending the Intellect*. Turku: University of Turku.
- Descartes, René (1988). *Descartes: Selected Philosophical Writings*. Transl. J. Cottingham, R. Stoothoff & D. Murdoch. New York: Cambridge University Press.
- Jung, C. G. (1955). *Synchronicity: An Acausal Connecting Principal*. New York: Routledge.
- Koistinen, Olli (2009a). "Spinoza's Eternal Self", in Jon Miller (ed.), *Topics in Early Modern Philosophy. Studies in the History of Philosophy of Mind* 9, 151–169.
- Koistinen, Olli (2009b). "Spinoza on Action", in Olli Koistinen (ed.), *The Cambridge Companion to Spinoza's Ethics*. New York: Cambridge University Press, 167–187.
- Koistinen, Olli (2014). "The Fifth Meditation: externality and true and immutable natures", in David Cunning (ed.), *The Cambridge Companion to Descartes' Meditations*. New York: Cambridge University Press, 223–239.
- Nolan, Lawrence (2015). "Descartes' Ontological Argument", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL = <http://plato.stanford.edu/archives/fall2015/entries/descartes-ontological/>
- Spinoza, Benedictus de (2002). *Spinoza: Complete Works*. Transl. Samuel Shirley. Ed. Michael L. Morgan. Indianapolis: Hackett.

Remarks on Cognition in Spinoza: Understanding, Sensation, and Belief

John Carriero

In this paper, I sketch some of the ways in which Spinoza's thinking about cognition differs from those prevalent today.

1. Essence and Understanding

Let's begin with understanding. Spinoza, like others before him, thinks of understanding as a matter of grasping a thing's fundamental character or structure—its essence—and seeing what follows from, or flows from, that essence.¹ This is a recognizable idea. Even today, one might think that to understand water, for example, is to appreciate that its essence or basic structure is H₂O, and to be able to see what follows from this essence. Similarly, one might think that to understand a triangle is to grasp its essence, which, in the tradition, amounted to knowing a Euclidean procedure for constructing a triangle, and being able to see what properties follow from being the result of such a construction procedure (see, e.g., TdIE §§ 95–96). Since there is not much distance between a thing and its fundamental structure—for example, there is not much distance between the water and its being H₂O or between the triangle and its being the result of such and such a construction procedure²—to get the fundamental structure right is to understand the thing.

Grasping an essence involves more than being able to recite a formula (e.g., “water is H₂O”) and knowing the meaning of its terms. Further, one can have greater or less understanding. Think of the difference between how an

¹ I have in mind Spinoza's second and third kinds of cognition (2p40s2). For their connection with essence see items three and four in TdIE, §19. The second kind of cognition seems to be *a posteriori* (in the pre-Kantian sense), roughly a movement from *propria* to essence, and the third kind of cognition seems to be *a priori* (again, in the pre-Kantian sense), a movement from essence to *propria*.

² More precisely, *this* water is *this* H₂O, and *this* triangle is the result of *this* application of a Euclidean construction procedure.

experienced chemist and a novice student operate with the knowledge that water is H_2O . This would seem to be in large part a matter of being able to discern more or less in the essence. Other factors, no doubt, come into play as well.

The preceding may sound innocent (or empty) enough. It seems to be the matter of a certain philosophical grammar of understanding. A thing's essence is what you've grasped when you've understood it; you've understood a thing when you've grasped its essence. So far, it might seem, we've only set out certain labels. But these labels, I think, harbor a certain outlook that differs from a contemporary viewpoint.

To begin with, as a thing's intelligible structure, an essence bridges the world and cognition in an interesting way. As a *thing's* intelligible structure, it is grounded in the world; as the thing's *intelligible* structure, it looks toward the mind and understanding. So, the assumption that things have essences is the idea that things are capable of being understood. If things lacked a structure—if they were the simple property collectors that blockish empiricist *substrata* can seem to be—then there would be nothing to understand, as opposed to, perhaps, catalogue. That water has a structure—an essence—makes it possible for someone (perhaps not us, but some more advanced intellect) to understand why water behaves as it does. That a triangle has a structure—an essence—is what makes it possible for someone to understand why the triangle has the further properties that it does (e.g., having an area equal to one half of the product of its base and height).

Further, for Spinoza and Descartes, understanding is primarily a relation between a cognizer, on the one hand, and a thing and its structure, on the other. For Spinoza, to grasp an essence of a thing is to have what he calls a true idea of a thing (see, e.g., TdIE § 70).³ To have such an idea is to have the thing in your mind completely enough that it contains the thing's essential structure, as opposed to merely certain aspects of the thing. And, for Descartes, distinct perception is supposed to be what makes available to the perceiver the essential structure of what is perceived.

Now, it is reasonable to suppose that someone who has perspicuous cognition of a thing's basic structure, so that she grasps the thing's essence,

³ Although this article is about a true thought, I think everything Spinoza says there applies to true ideas.

will often be able to provide an explication of that thing's essence. In the tradition an explication or account of a thing's essence (i.e., a *logos*) was called a (real) definition.⁴ But it is not clear that someone who has seen to the bottom of things will always be able to provide a real definition. And it is also unclear to what extent a real definition captures the richness of the essence. Here we might think of an idiot savant, who seems to be onto something that she is unable to explain to others, or, again, of a being so advanced in its understanding that it is unable to explain the things that it understands to lesser intellects. Grasping an essence, which involves an activity of mind that may or may not be expressible discursively, is part of what Spinoza has in view when he speaks of intuitive cognition.

Nowadays, it seems that understanding is thought of as fundamentally discursive, as opposed to intuitive. Providing a theory is often thought of as a matter of providing some set of basic propositions or axioms that characterize some subject matter, and understanding is thought of as a matter of determining the deductive consequences of the propositions, or of tracing back the consequences to the axioms. In the case of non-empirical subject matters, sometimes the subject matter is felt to be exhausted by the axioms. Group theory comes to mind here. There is nothing more to the subject matter, it seems, than what the axioms say there is; the axioms exhaust the nature of groups. In the case of other non-empirical subject matters, the situation is less clear. Some seem to think that there is nothing more to sets than what the axioms of ZFC imply; others disagree. Empirical subject matters tend to be more complicated: what counts as chemical as opposed to (more basic) physical structure is determined by some interplay between theory and subject matter. A sign of this is that it seems that future empirical discoveries could, at least in principle, lead us to revise our views about the chemical. This attitude differs from that of Spinoza and Descartes, for whom a thing and its essence, rather than an axiomatization of a subject matter, hold first place in the enterprise of understanding. For them, to understand is to get onto a thing's essence (or nature). The articulation of that grasp into a theory—perhaps some system of axioms from which subsequent truths can be deduced—comes later, to the extent that such a thing is possible at all.

⁴ Spinoza shows interest in such definitions in the TdIE. See, especially, § 91 to the end.

This is important, for example, for making sense of Spinoza's conception of God's cognition. According to Spinoza, God has a grasp of his own essence and can intuitively see everything that follows from that essence: God's essence is so rich that a sufficiently powerful intellect (such as God's) can see unlimited realities following from it, which realities include everything that exists and everything that happens in the world. This picture does not commit Spinoza to the thesis that there is some set of axioms from which deductively follow (in the sense of a predicate calculus) all the truths of the universe. And the apparent impossibility of the latter does not count against the coherence of the former.

To be sure, Spinoza thinks it is possible demonstratively, through some process of discursive reasoning (and here I am thinking of what Spinoza calls the second kind of cognition), to make clear that certain things follow from God's essence. Presumably this is what is going on in much of the first part of the *Ethics*. But there is no reason to believe that Spinoza would think that what can be shown through discursive reasoning *exhausts* what can be seen in God's essence. And there is no reason to think Spinoza thinks that all the things that can be shown discursively can be combined into some single, quasi-formal system. Here, I see an analogy between Spinoza's picture of divine cognition and Gödel's (as I understand it) thought about human cognition of arithmetic. Gödel didn't think his incompleteness theorems showed that any arithmetical truths lay in principle beyond the capacity of the human being. Rather, he thought that what is open to us to know in this domain cannot be fit into a single, all-embracing formal system but is a more piecemeal affair.

Relatedly, in Spinoza's view, intuitive cognition enjoys a certain priority over discursive cognition, which he regards as second best. While discursive cognition may help lead one to an intuitive purchase on an essence and what follows from it, then, intuitive cognition does not consist in discursive cognition (which is not to deny that intuitive cognition is structured in some other way). One way in which we might try to put this last point is that while the grasp of the essence (and what that involves) may be *taught*, it is not the sort of thing that can in general be *communicated*. (And here we may be arriving at one of the pressures toward nativism in early modern rationalism.)

Philosophically, the fact that there is a certain interplay between intuitive cognition and discursive cognition is important; it keeps intuitive cognition from seeming like mysticism. The sort of thing that Spinoza has in mind by intuitive cognition seems familiar enough. It does not seem far-fetched to think you might find yourself in a situation, perhaps as a teacher, where you've seen something that you are unable to get across, perhaps to your students. There's a lot you can do discursively, related to what you have grasped and your students haven't yet got—again, that's what keeps the whole thing from seeming like mysticism (though sometimes it feels like that)—but all the same it's natural to think that what you've grasped transcends and is, in a way, prior to what you can manage discursively. Spinoza's views on intuitive cognition require that we take such experiences at face value.

As perhaps is clear by now, the more traditional idea of essence as providing a thing's intelligible structure differs from the contemporary idea of essence that one sometimes sees as the basis of a thing's modal profile.⁵ Here the example of water and H_2O is helpful. The discovery that water is H_2O is first and foremost an accomplishment of scientific understanding. When we learned that water is H_2O , we understood why water behaves as it does. And this is what scientists were interested in when they sought the chemical structure of water. They were not concerned with questions like “Could something be water and not have this or that characteristic?” It may be true that having learned that water is H_2O , certain modal conclusions follow: for example, there is no possible world in which there is water but only one hydrogen atom. But modality is at best a secondary concern here, a consequence of getting one's account of water right.

We should note—and this is a point of some consequence for Spinoza—that a formula like the “essence of water is H_2O ” is shorthand. It does not make sense by itself but only in the context of a more comprehensive chemical and physical theory. Similarly, the characterization of a triangle as the result of a certain construction procedure tacitly draws on a background Euclidean structure. Spinoza was well aware of the role played by global structure. This is why he thought finite beings had to be understood (must be “conceived through,” is the way he puts it) through the attributes of God or

⁵ For related discussion, see Føllesdal 1986, 97–113.

substance, that is, the infinite invariant structure of the universe. It would be impossible to understand, say, a structure like the human body, apart from the laws of physics, that is, the global structure that pervades the physical universe.

This creates an interesting difficulty for him: if a finite thing cannot be understood without reference to God or substance's invariant structure (here we might think of the pervasive laws of nature), then shouldn't God or substance be included in a finite thing's essence? In Part Two of the *Ethics*, Spinoza tries to solve this difficulty by introducing a technical modification to the idea of essence intended to keep God out of the essence of finite things.⁶

2. The Human Mind

The primary cognition in Spinoza's universe is God's cognition of his own essence and of the following of subsequent things (i.e., modes) from that essence (cf. 1p16 and 2p3). For a partial model, it helps to consider Spinoza's views on the physical order. According to Spinoza, God (necessarily) produces within the plenum a maximal order of reality, an overall pattern of matter—what Spinoza calls the face of the universe—which contains (in some sense) every possible (physical) reality. Understanding what God does, and why, resembles beginning with Euclidean space and solving a fantastically complex construction problem, with a result that is infinitely rich in detail, more than it does starting with universals, say, *human* or *equine* (as Aristotelians did), or with concepts and their determinations, say, *red* and *crimson* (as Kant later will), and reasoning about them. (In fact, according to Spinoza, such universals show up late in the game, and are not of much use for understanding, as opposed to getting around in the world.) Although Spinoza did not have the resources to put the point this way, in his universe it is no easier to figure out what's what and why things are the way they are, than it is to solve families of differential equations.

This cognitive activity structures the attribute of thought, yielding what Spinoza calls the “infinite intellect of God” (see 2p4). The infinite intellect of God is the way Thinking Being becomes structured through God's cognitive activity, just as the face of the whole universe is the way Extended Being

⁶ See 2d2 and Spinoza's explanation of it in 2p10cs.

becomes structured through God's motion-and-rest activity.⁷ All other cognitive activity, including the cognitive activity of finite beings, is to be understood in terms of this *ur* activity, as parts of it: "the human mind is a part of the infinite intellect of God" (2p11c). According to Spinoza, just as all bodies find their place within the face of the whole universe, an infinite determinate mode of Extended Being, so too all finite cognition finds its place within the infinite intellect of God, an infinite determinate mode of Thinking Being.

The idea of God is all encompassing: that is to say, each thing exists objectively (or cognitively) within it. Another way to put this is that God has an idea of each thing, which we might think of as a sort of sub-idea within the all-embracing idea that God has of his essence and everything that flows from it. In particular, God has an idea of my body.

The idea of my body—the region of God's cognition where my body exists objectively—is, according to Spinoza, my mind. This is his attempt to balance the claims of the sameness of mind and body with those of their difference. There is a way in which the mind, as God or the universe's cognition of the body, is the same thing as the body: my mind is my body as it exists objectively within the idea of God. But there is a way in which they differ. As Spinoza reminds us in TdIE, § 33, the idea of the circle is not a circle; the idea does not, for example, have a circumference.

My mind, then, in Spinoza's view, is a partial or incomplete cognition of reality. To the extent that our corner of God's cognition is ordered according to "the order of the intellect" as opposed to "the order and connection of the affections of the human body," we perceive "things through their first causes" (2p18s), i.e., we understand. This is a function of what is and is not found in our partial view of things, and so of the extent to which what is missing renders our cognition "fragmented" or "mutilated." (Notice that Spinoza is trying to explain our intellectual and nonintellectual cognition in a way that

⁷ Cf. Letter 32: "So you see in what way and why I hold that the human body is part of Nature. As regards the human mind, I maintain that it, too, is a part of Nature; for I hold that in Nature there also exists an infinite power of thinking which, insofar as it is infinite, contains within itself the whole of Nature objectively [*in se continet totam naturam obiective*], and whose thoughts proceed in the same manner as does Nature, which is in fact the object of its thought."

does not appeal to powers or faculties: the difference is a matter of the way in which our thought is ordered.)

3. Sensory and Imaginative Cognition as Atrophied Understanding

All of this makes puzzling the place of nonintellectual cognition for Spinoza. After all, if our minds are fragmentary patches within God's cognition of reality, and God's cognition is intellectual, how do nonintellectual forms of cognition—sensation, imagination, and memory—arise, on Spinoza's view?

Spinoza's general strategy is to locate sensation and imagination in the fact of the mind's finiteness, that is, in the fact that, as God's idea of the human body, the human mind is only a partial cognition of the universe. In other words, the finiteness of the human mind can be ultimately traced back to the finiteness of the human body.

Now, a sensory or imaginative idea is, on Spinoza's telling, the idea of some goings on within the human body, which goings on the body cannot fully (causally) account for.⁸ Since the body cannot causally explain these goings on, the idea of the body—that is, the body's mind—does not suffice of itself to make these goings on intelligible. Inasmuch as these goings on are not intelligible to the mind, the idea it has of them counts as sensation (or, depending on the details of the case, imagination or memory), as opposed to understanding.

For example, a tulip affects some light, and the light affects my eyes and optical nerves, causing a pattern of motions in my brain. This motive pattern—the brain image—exists objectively in my mind, as an idea. According to Spinoza, the idea of these motions “represents” the tulip “as present” (2p17cs).⁹ The idea of the brain motions is incomplete in me because the cause of the motions, especially the tulip, does not exist in my mind

⁸ I am limiting my discussion to minds that are the objective or cognitive existence of bodies. Presumably, the objective existence of modes comprehended by God's other attributes also are minds, and these minds would have their own partial and incomplete ideas, and something analogous to sensation and imagination.

⁹ “Next, to retain customary words, the affections of the human bodies whose ideas present [*repraesentant*] external bodies as present to us [*velut nobis praesentia*], we shall call images of things” (2p17cs). Spinoza does not try to explain why it is the distal cause, rather than some other intermediate cause, that the idea represents as present. Perhaps he simply thinks that it is obvious that this happens, and a further understanding of how and why awaits scientific progress.

objectively. With respect to me, the idea of the motions is an example of sensation (and, subsequently, imagination, which includes memory). Spinoza famously compares such ideas to “conclusions without premises [*consequentiae absque praemissis*]” (2p28dem). In a certain sense, I just seem to be landed with them; they seem to come from nowhere. By way of contrast, *God’s* cognition of the brain motions is not incomplete: The pattern of brain motions, the tulip, and whatever else is involved in the causal path from the tulip to my brain, all exist objectively in the idea of God. In this way, what’s involved in my sensory cognition of the tulip is very different from what’s involved in God’s cognition of the tulip. Although the tulip exists objectively in God’s cognition, it does not exist objectively in mine; only the brain motions, whose idea “represents” the tulip “as present,” do.¹⁰

It is evident, in this way of theorizing about the difference between sensory and intellectual cognition, that Spinoza is not much moved by the problem of *qualia*. His attitude here seems to be similar to that of Leibniz, that our experience of green is very complex and what puzzlement we find surrounding it would sort itself out if only the complexity of the cognition were rendered distinct. Leibniz thinks we get an inkling of this when some apparently green powder turns out to be, upon closer inspection, a mixture of blue and yellow powders.¹¹ Whether Spinoza and Leibniz are oblivious to some sort of intractable mystery surrounding the “what it is like” of experiencing green is a question I won’t try to delve into here.

Spinoza is offering, then, what might be called an “understanding first” picture of cognition, where sensation and imagination are construed as sorts of atrophied forms of understanding. In this respect, his attitude seems similar

¹⁰ See note 9 above. As far as I am aware, Spinoza does not describe sensory or imaginative ideas as bringing it about that the things they represent as present to mind, should objectively exist in the mind. In the *Ethics*, he reserves the term *obiective* for the way the objects of thought exist within Thinking Being (2p7c), as he does in Letter 32. It comes up frequently in the TdIE (see, e.g., §§ 33–36 and 41), where it appears in the context of intellectual cognition. Spinoza may be signaling the obliqueness of imaginative cognition by declining to characterize it as involving the objective existence of that which is imagined. I think it is unlikely that sensory cognition, of itself, imports enough of the tulip’s structure into a human mind for it to count as a “true idea” in the sense of §§ 33–34, i.e., the objective existence of the tulip’s essence. Eventually, with the help of science (and by coming to have further ideas so that one’s cognition of the tulip becomes less “fragment and mutilated”), a human mind may come to possess the tulip’s essence objectively, that is, may understand its essence. (I am grateful to Lilli Alanen for discussion of this point.)

¹¹ Leibniz employs this example in “Meditations on Cognition, Truth, and Ideas” (AG 27) and the *New Essays*, II.2.1 and IV.6.7. Spinoza touches on perceptual relativity in Iapp.

to that of Leibniz and (I would argue) Descartes, according to which sensation is a confused form of intellection. What brings these rationalists to such a view, I think, is their desire for an integrated theory of cognition. To be sure, high scholastic theories of cognition, too, sought to provide an integrated picture. For them, understanding is seen as a matter of extracting intelligible structure—say, the universal *caninehood*—immanently contained in the experience (the phantasms) of particulars—say, Spot, Rover, and Fido. But Spinoza (along with, I believe, Descartes and Leibniz) thought that this picture was not well suited to the mathematical structure of the physical order. If the sensory somehow immanently enveloped an intelligible structure, this, they thought, had to work along radically different lines. Ultimately, this led them to the view that sensory and imaginative cognition are more like understanding than one might pre-theoretically have suspected. For them, there's a lot more going on in, say, my experience of pain than I might think.

The rationalist picture of sensation runs against the grain of modern sensibilities. For many today, as perhaps for Locke and Hume, it seems more natural to begin with more primitive forms of cognition, such as sensation and imagination, and see understanding as in some way built up from or a development of these more primitive forms of cognition. Often this approach is motivated by the thought that human understanding must in some way be based on experience, which is a certain interplay between sensation, imagination, and memory. Whether it is possible, starting with such materials, to reach understanding is hard to say. Locke, for what it is worth, had his doubts, and Hume thought we needed to reconceive what understanding is.

Wherever one's sympathies lie here—with the rationalists or with the empiricists—it is worth keeping in mind that achieving a more balanced view of the respective contribution of understanding and sensation, while providing some account of the understanding of the physical world that it seems we obviously have, has proven to be a difficult problem that has engaged much subsequent philosophy. This makes it less surprising perhaps that the first philosophers to confront it felt pressed to interpret sensation in terms of understanding or understanding in terms of sensation.

4. Belief

For Spinoza, ideas, which are things (the “objects” of the ideas) existing objectively (or cognitively), are the coin of the cognitive realm. By way of contrast, for many philosophers today, beliefs and propositional attitudes are the common currency for a theory of the mind.

Some have felt that Spinoza’s priming of ideas in his theory of cognition leaves the place of belief uneasy in this theory. It is not that Spinoza has nothing to say in the direction of belief—he talks about affirming and denying ideas—but he has a lot less to say in the vicinity of belief than about ideas, and what he does say that bears on belief has struck some as problematic. Ideas, taken at face value, don’t seem to be sort of entities that one can affirm or deny. My idea of the sun is not a proposition for me to believe or disbelieve. Confronted with this difficulty, some readers have concluded, appearances notwithstanding, that ideas, for Spinoza, must be entities with propositional structure.¹²

I think it is worth trying to understand Spinoza on his own terms here. Converting his ideas into propositions runs the risk of missing something valuable and interesting about how his philosophical starting point might be different from what is prevalent today. Some of this difference is already implicit in the point made earlier that for Spinoza the primary intellectual moment is the grasping of an essence and not, say, the establishment of a set of axioms that implicitly define a theory. This is not to deny that one can think discursively, by way of propositions, as it were. But, for Spinoza, intuitive thought is prior to discursive thought and, as far as I can tell, not supposed to be exhausted by it.

This being so, how, then, ought we to think about the relationship between an idea and the beliefs it engenders? Spinoza takes ideas to be psychological realities that have affective tendencies. Among an idea’s affective tendencies is “affirmation and negation” (see 2p49 and 2p49s).¹³ In his view, a mind never simply “has” an idea. An idea is not “something mute,

¹² See, e.g., Curley 1969, 119–126, and Bennett 1984, 128, 162–167. It is worth keeping in mind that similar issues come up in Descartes and Hume. For discussion of Descartes, see Carriero 2009, 228–230.

¹³ It is not clear to me whether to think of this affective tendency as a *proprium*, that is, a necessary concomitant of the idea, or as a part of the idea’s essence.

like a picture on a tablet” (2p43s). Ideas make a difference to my cognitive economy in various ways—to my overall cognitive orientation—and in particular to what I think or believe, as we might put it colloquially. Depending on the adequacy or distinctness of my idea—that is, depending on how well I understand the triangle—the idea’s tendencies may express themselves in my overall cognitive orientation in different ways over time, in my beliefs, as we might put it. For example, initially my idea might not generate a belief in its three angles’ being equal to a straight line, but eventually, as I understand the triangle better, I may come to a point where this tendency becomes a part of my overall cognitive orientation.

Thus far, there is a certain looseness over exactly which aspects of the essence lead to beliefs. Perhaps Spinoza would agree that as one tries to articulate some of what is implicit in my grasp of the triangle’s essence notions like propositional structure and propositional content would prove useful. Perhaps not. As far as I can see, nothing in his system precludes this, although it must be said that explicating the fine structure of beliefs is not a project that he shows a great deal of interest in. What Spinoza is committed to, I take it, is the denial that such items are found at the fundamental level of mental reality.

It is not so difficult, I think, to make sense of Spinoza’s position in situations where understanding is involved. After all, the view that there is an intimate connection between understanding a thing and having beliefs about it seems natural, even if the exact nature of the relationship between the understanding and the attendant beliefs is controversial. But what about other sorts of beliefs, say, my belief that there is a tulip in front of me? How does Spinoza think about this?

As we saw earlier, a sensory or imaginative idea is, for Spinoza, a pattern of brain motions that exist objectively in the human mind. The idea of this pattern represents the tulip as present. This idea is the basic psychological reality. The presence of this idea in the mind makes a difference to the cognitive system. It makes a contribution to the mind’s overall affective orientation.

Some of that affective orientation concerns what we would naturally think of as beliefs. As with the idea of the triangle we considered earlier, it is not straightforward to translate what is involved here into some propositional

attitude or, perhaps, some set of propositional attitudes. The problem is not with the “attitude” side. Ideas of brain images are not generally causally neutral; they do make a difference to the rest of the cognitive system, including what we would call its beliefs. My idea of the brain image produced by the tulip carries with it a commitment to the tulip’s existence—what Spinoza calls an “affirmation” of the existence of the tulip; typically this affirmation continues even after I am no longer seeing the tulip.¹⁴ But exactly what such “affirmation” of the tulip’s existence involves is left open. It seems reasonable to characterize this affirmation as a belief (or perhaps a constellation of beliefs) that is perhaps expressible in language as “The tulip exists” or “That tulip exists” or maybe just “That exists.” It is reasonable to do so, as long as we are not too fussy about what a belief is and as long as we recognize that there may be some gap between the psychological reality (the idea and its affective tendencies) and our attempts to describe it.

It is also reasonable to think, with the same caveats, that the difference that the idea of the tulip image makes to the mind’s affective orientation, when taken in conjunction with other ideas, involves affirmations of its being next to the daffodil, its being red, its being taller than the rock, and so forth. Spinoza does not have a great deal to say about the division of labor between what the idea of the tulip does and what other ideas in the cognitive system do (which ideas, as it were, provide collateral information). He does, though, note that other ideas may block the affirmation of the tulip’s existence, so even this rather basic affective aspect is contingent on what other ideas there are in the system. He also does not say much about exactly what affirmations accompany ideas and to what extent these affirmations can be made articulate by being put into language. The fact that Spinoza does not worry about such matters may be explained by the fact that he thinks that the words “tulip” and “daffodil” stand for an idea of a blurred brain image (see 2p40s1),

¹⁴ See 2p17cs and Spinoza’s treatment of the example of the boy and the idea of the winged horse in 2p49s.

rather than a classical universal (abstracted shared, intelligible structure) or a modern concept (a function-like rule).¹⁵

University of California, Los Angeles

Bibliography

- Bennett, Jonathan (1984). *A Study of Spinoza's "Ethics"*. Indianapolis: Hackett.
- Carriero, John (2009). *Between Two Worlds: A Reading of Descartes's "Meditations"*. Princeton: Princeton University Press.
- Curley, Edwin M. (1969). *Spinoza's Metaphysics: An Essay in Interpretation*. Cambridge: Harvard University Press.
- Føllesdal, Dagfinn (1986). "Essentialism and Reference", in Lewis Edwin Hahn (ed.), *The Philosophy of W. V. Quine*. La Salle: Open Court, 97–113.
- Leibniz, Gottfried Wilhelm (1989). "Meditations on Cognition, Truth, and Ideas", in G. W. Leibniz, *Philosophical Essays*. Ed. & transl. Roger Ariew & Daniel Garber. Indianapolis: Hackett, 23–27. [AG]
- Leibniz, Gottfried Wilhelm (1981). *New Essays on Human Understanding*. Ed. & transl. Peter Remnant & Jonathan Bennett. Cambridge: Cambridge University Press.
- Spinoza, Benedictus de (1985). *The Collected Works of Spinoza*. Ed. & transl. Edwin Curley. Princeton: Princeton University Press.

¹⁵ I am grateful to the editors of this volume, Hemmo Laiho and Arto Repo, for their suggestions, and to Don Garrett and Karolina Hübner for their helpful comments. And special thanks go to Olli, who has taught me much about Spinoza over many years of writings, conversation, and friendship.

Leibniz on Primitive Concepts and Conceiving Reality

Peter Myrdal & Arto Repo

In this paper, we consider what is commonly referred to as Leibniz's argument for primitive concepts. After presenting and criticizing (in sections 1 and 2) one recent rather straightforward way of interpreting this argument, by Paul Lodge and Stephen Puryear, which takes the argument to be merely about the structure of concepts, we offer an alternative way of looking at the argument. We think it is best seen as being fundamentally about the relation between thought and reality. In order to prepare the ground for our reconstruction (which we present in section 5), we have to introduce his view of ideas or concepts (section 3), as well as some metaphysical principles concerning reality-dependence (section 4).

1. The Argument

In his early piece, *Of an Organum or Ars Magna of Thinking*, Leibniz offers the following argument for why, in his own terms, insofar as we conceive anything, there needs to be something that is conceived through itself:

Whatever is thought by us is either conceived through itself [*per se*], or involves the concept of another.

Whatever is involved in the concept of another is again either conceived through itself or involves the concept of another; and so on.

So one must either proceed to infinity, or all thoughts are resolved into those which are conceived through themselves.

If nothing is conceived through itself, nothing will be conceived at all. For what is conceived only through others will be conceived only in so far as those others are conceived, and so on; so that we may only be

said to conceive something in actuality when we arrive at those things which are conceived through themselves.¹ (A VI.iv 156/MP 1–2.)

The expressions “conceived through itself” and “conceived through something else” may at first sight appear a bit obscure. One initial worry is that there seems to be some ambiguity in Leibniz’s formulation. He can be taken to be talking about what we are thinking about, about *things* conceived, on the one hand, or about our *thinking* or *conceiving* of those things, on the other.

A natural way to try to make Leibniz’s claims clearer is to take the argument to be about concepts, i.e. about something in our minds, about the constituents of our thoughts, and to understand the argument as trying to show that there has to be primitive or simple concepts in order for there to be complex concepts. As textual evidence for this kind of reading one can refer to another of Leibniz’s early texts, *An Introduction to a Secret Encyclopaedia*, where he writes that “we can have no derivative concepts except by the aid of a primitive concept.” (A VI.iv 529/MP 7.) On this view, for a concept to be complex (or derivative) is for it to have constituent concepts into which it can be analyzed. For a concept to be simple or primitive (we will use these terms interchangeably here) is for it to lack such constituents. For example, the concept *man* is complex because it can be analyzed into the concepts *animal* and *rational*. Similarly, the concept *animal* is complex because it can be analyzed into, say, the concepts *corporeal* and *living*. The point of the argument would be that relatively more complex concepts (such as *man*) result from a combination of relatively less complex concepts (such as *animal*), which in turn

¹ It may be useful to have the full quote in the original Latin:

Quicquid cogitatur a nobis aut per se concipitur, aut alterius conceptum involvit.

Quicquid in alterius conceptu involvitur id rursus vel per se concipitur vel alterius conceptum involvit. Et ita porro.

Itaque vel eundem est in infinitum, vel cogitationes omnes resolvuntur in eas quae per se concipiuntur.

Si nihil per se concipitur, nihil omnino concipietur. Nam quod non nisi per alia concipitur, in tantum concipietur in quantum alia illa concipiuntur et hoc rursus ita: ac proinde tum demum actu ipso aliquid concipere dicemur, cum in ea quae per se concipiuntur incidemus.

result from other relatively less complex concepts, until some simple concepts are reached, which are not the result of any combination.

Such a reading is offered by Paul Lodge and Stephen Puryear, who propose to clarify Leibniz's argument by substituting "primitive concept" for "conceived through itself" and "complex concept" for "conceived through something else," resulting in the following reconstruction of the *Ars Magna* argument:²

- (1) Every concept is either primitive or complex, i.e., composed of other concepts (assumption).
- (2) Every concept that composes a complex concept is itself either primitive or complex (from (1)).
- (3) Complex concepts are either composed of other concepts to infinity, or ultimately composed of primitives (from (2)).
- (4) Complex concepts are conceived only insofar as their constituents are conceived (assumption).
- (5) A complex concept is conceived only if it is ultimately composed of primitives (from (4)).
- (6) If a complex concept is composed of other concepts to infinity, then it will not be conceived (from 5).

This only establishes the conditional conclusion (6) and not the existence of primitive concepts. It seems, however, that Leibniz does intend the argument to show the existence of primitive concepts. As Lodge and Puryear (henceforth L&P) point out, he seems to rely on the following additional premise:

- (7) Some complex concepts are conceived (unstated assumption).

This premise allows us to draw the intended conclusion:

- (8) Some complex concepts are not composed of other concepts to infinity (from (6), (7)).
- (9) Some concepts are ultimately composed of primitives (from (3), (8)).

² Lodge & Puryear 2006/7, 178.

(10) There are primitive concepts (from (9)).

L&P's reconstruction also serves to bring out some difficulties with the argument, which we will consider next.

2. Some Problems with the Argument

The first problem has to do with the fact that the argument does not only show that there are primitive concepts, but also that insofar as we conceive a complex concept we also conceive its primitive constituents. (This follows from (4).) It has been suggested that for Leibniz conceiving x implies awareness of x .³ If this is true, then it would further follow that in conceiving a complex concept, we are aware of all of its constituents, down to the primitives. To conceive a complex concept would mean that one is able to provide the full analysis of it. The problem is that Leibniz explicitly denies the last claim. He holds for example that we are aware of the concepts of colors without being aware of what they contain (this is distinctive of what he calls clear and confused concepts (e.g. A VI.iv 585f./AG 23f.)). In fact, Leibniz is rather skeptical of the possibility that we would ever be capable of analyzing our complex concepts into primitives. As he explains in *An Introduction to a Secret Encyclopaedia*: “A concept is primitive when it cannot be analysed into others [...] But it can be doubted whether any concept of this kind appears distinctly to men.” (A VI.iv 529/MP 7; cf. A VI.iv 590/AG 26.)

L&P's main aim is to show that this alleged problem is not a problem. The reason is that it is incorrect to ascribe to Leibniz the thesis that conceiving x implies awareness of x .⁴ In that case, there seems to be no conflict between, on the one hand, claiming that in order to conceive a complex concept we need to conceive all of its constituents, and, on the other hand, denying that we are capable of analyzing all complex concepts. At a general level the notion of unconscious conceiving fits nicely with Leibniz's thesis that much of our mental life is unconscious. More particularly, L&P

³ See Plaisted 2003, 338. He relies on *Discourse on Metaphysics* 27: “the expressions in our soul, whether we conceive them or not, can be called ideas, but those we conceive or form can be called notions, concepts [*conceptus*].” (A VI.iv 1572/AG 59.)

⁴ They persuasively argue that *Discourse* 27 does not constitute textual evidence for this thesis (Lodge & Puryear 2006/7, 193).

point to the fact we can *have* and *use* concepts without being aware of them, in line with how Leibniz in the *New Essays* claims that we use the principle of contradiction without being aware of the principle or its constituent concepts (*impossibility, being, etc.*) (NE 76, 83–84). L&P do not go so far as claiming that conceiving a concept is just having or using it, but it is not quite clear what, in their view, is involved in conceiving a concept in addition to merely having or using it. One might worry that this additional aspect of unconscious conceiving will come close to having the analysis of a concept, which (as we have seen) would be problematic. While we agree with L&P that conceiving for Leibniz does imply awareness, their reading leaves the nature of conceiving somewhat obscure.

Another difficulty with the argument, also noted by L&P, concerns the crucial step from (4) to (5). There are in fact two related problems here. The first has to do with the motivation for (4): why should it be the case that in order to conceive a complex concept, I need to conceive all of the constituents of that concept? Couldn't I conceive, for example, *gold* without conceiving all of its constituents? The second problem is that even if we accept (4) this does not rule out the possibility of infinitely complex concepts. Hence we cannot from (4) conclude that (5) – that a complex concept is conceived only if it is ultimately composed of primitives.

In order to motivate assumption (4) as well as to understand the move from (4) to (5), L&P suggest that Leibniz accepts the following further thesis concerning concepts, which they call the *Inheritance Thesis*:

(IT) Complex concepts inherit (or borrow) their content from the concepts that compose them. (Lodge & Puryear 2006/7, 189.)

They claim that if we accept (IT), then (4) follows: “If complex concepts must (continually) borrow their content from the constituents, then it would stand to reason that we cannot conceive a complex concept without in the process conceiving its constituents. For in grasping the (borrowed) content of a complex concept, we would in effect be grasping the content of its components.” (Lodge & Puryear 2006/7, 189.) Support for attributing (IT) to Leibniz is supposed to be found in a simile he offers in *Ars magna* as an illustration of the argument:

I will illustrate this by a simile. I give you a hundred crowns, to be received from Titus; Titus will send you to Caius, Caius to Maevius; but if you are perpetually sent on in this way you will never be said to have received anything. (A VI.iv 156/MP 2.)

According to L&P this continual deferring of the payment can be characterized as a matter of the payment's being borrowed or inherited. The simile, they claim, also allows us to see how (IT) helps to explain the move to (5), that is, how the regress can be blocked: "Just as nothing will be received at all unless one of the people to whom we are sent actually pays [...] so also nothing will be conceived at all unless the complex concept ultimately resolves into concepts that do not inherit their content from others, since conceiving requires that there be a content to conceive." (Lodge & Puryear 2006/7, 189.)

There appears, however, to be something wrong with the suggestion that the simile straightforwardly illustrates (IT). L&P overlook what seems to be an important *disanalogy* between the two. In the simile I am given 100 crowns, but in order to receive the money I am perpetually sent onto new people. This means that I haven't received anything until there is someone who actually pays. (IT), on the other hand, assumes that complex concepts (such as *gold* or *triangle*) *have* content, even if that content is borrowed or inherited from its constituents. If complex concepts have content, something – some content – seems to be received already at the first stage, unlike in the simile, where nothing is received. While it may be true that the chain in the simile must terminate – if some money is given, there has to be some money to be collected – this is not so in the case of (IT). Given that there is content at the first stage, it is hard to see why the fact that this content is borrowed from the constituents would rule out a situation where those constituents in turn borrow their content from their constituents, and so on to infinity.

This means that that even if we accept that (IT) helps to motivate (4), it does not explain the move to (5), contrary to what L&P suggest. For in order to use (IT) to infer that a complex concept is conceived only if it is ultimately composed of primitives, the thesis would have to entail that there is no content at all – nothing received – until we reach some primitive concept, which, as we just saw, does not appear to be the case. The disanalogy between the simile and (IT) also means that we are left without a satisfactory

understanding of what Leibniz takes the point of the simile to be: how exactly is it intended to illustrate the argument? We believe that L&P's appeal to the notion of borrowing does in fact contain an important insight, but that it needs to be developed in a different direction (as we will argue in section 4 below) in order for it to help us with Leibniz's argument.

There is a third, and perhaps even deeper worry, about L&P's reconstruction. The conclusion of the argument reads: "we may only be said to conceive something in actuality when we arrive at those things which are conceived through themselves." For purposes of clarification L&P substitute (as we have seen) Leibniz's original formulation with (10): "There are primitive concepts." The problem is that these two formulations do not seem to be equivalent. The latter is a claim about the mereological structure of concepts: complex concepts have to be composed of simple or primitive concepts. The former is a claim about the connection between cognition and external reality: in order for us to conceive anything there has to be some *thing* that is conceived through itself.

It is true that in the argument Leibniz does not explicitly use the Latin *res*, but rather the pronouns *quicquid*, *alterius*, *ea*, etc. However, it seems clear that these are intended to refer to some external reality (things), rather than to something mental (concepts): in *Ars magna* Leibniz goes on to argue that that which is conceived through itself is "God himself [*Deus ipse*]" (A VI.iv 158/MP 2); in *An Introduction to a Secret Encyclopaedia* he talks explicitly about "the *thing* which is conceived through itself [*rei quae per se concipitur*], namely the supreme substance, that is, God" (A VI.iv 528–9/MP 7; our emphasis). The claim is then that in order to conceive anything, we need to conceive some (or rather *the*) basic reality.⁵ We will attempt to make some more sense of this idea in section 3 below. For now we are simply interested in bringing out the significance of Leibniz's emphasis on conceiving a *thing* through itself. (It is also worth noticing that the idea of there being a single primitive (God) does not fit very well with reading the argument as a matter of the mereology of concepts. For if the role of primitives is to serve as the basic building-blocks of

⁵ There seems to be an interesting connection to an idea that Olli Koistinen has brought out in Descartes (as well as in Spinoza and Kant): "For something to be possible, there must already be some existent things on which the possibility depends, and through which the possibility can be understood." (Koistinen 2014, 237.)

complex concepts, then there would have to be several (maybe infinitely many) primitives.)⁶

At this point it may be suggested that we are making too much of the contrast between, on the one hand, talk of concepts and, on the other hand, of things. After all, L&P do not mean to deny that Leibniz is concerned with conceiving *things*. In fact, they insist that the locution “conceiving (of) a concept” is “shorthand for something like ‘conceiving a *thing* through a concept’” (Lodge & Puryear 2006/7, 178 fn. 5). Still, it is hard to see that the idea of conceiving things has any substantial role to play in their reconstruction. Their presentation of the steps of the argument seems entirely consistent with staying neutral with respect to the question of whether there is any relationship between concepts and things. The argument appears, as already noted, simply to concern the structure of concepts.

This seems to leave Leibniz’s position vulnerable to Kant’s well-known objection in the *Critique of Pure Reason* that Leibniz created a mere “intellectual system of the world” (A 270/B 326). Famously, Kant claimed that the problem is that Leibniz thought of himself as able to arrive at objects merely through considering their concepts, failing to see that there is a gap between concept and object – that concepts alone do not allow us to reach reality. This is not the place to go further into the details of Kant’s criticism. Our point is simply that the contrast we have drawn attention to between concept-talk and thing-talk, cannot simply be overcome by asserting that we conceive things through concepts. At first sight, the claim that we conceive a primitive concept does not in itself license the inference to there being some *thing* that is conceived through that concept. And what seems important in Leibniz’s conclusion is precisely that we arrive at things.

We believe that the difficulty just described can be addressed by taking a closer look at the central terms used in the argument. This will also provide help with the other problems noted (the nature of conceiving, as well as the step from (4) to (5)).

⁶ In later writings Leibniz tends to talk about the primitives as the various attributes of God (e.g. A VI.iv 590/AG 26), which may perhaps be more easily fitted into a picture of Leibniz as engaging the question of the structure of concepts. To us it seems, however, more plausible to read the later texts in light of the earlier, so “attributes” refer to things or reality rather than to concepts. Indeed, this would be a natural reading from the perspective of Scholastic theology, which understood God’s attributes as following from – or even as aspects of – the divine essence.

3. Concepts, Ideas, and Expression

While it is natural to read the argument as concerned with concepts, and Leibniz himself sometimes formulates the argument in terms of concepts, this can easily lead astray. We need to pay attention to some deep differences between what is for us a familiar way of taking ‘concept’ and Leibniz’s way of understanding that term. Nowadays philosophers tend to think (in some respects following Kant) that concepts themselves offer us only a very “thin” cognitive relation to reality – they are mainly instruments we use in our thinking in order to pick up objects for our thoughts. We may be able to arrive at some analytic judgments on the basis of having the concept of, say, a cat – maybe that cats are animals and some other information of that sort. But in order to know more about cats, or in order to arrive at synthetic judgments about cats, we would have to go “outside” our concept, and rely on experience of the objects we pick up with the help of the concept.

Now, even if Leibniz is well aware of the need for experience in our knowledge of the world, his view of concepts is fundamentally different from many later views. In order to see this we can consider Leibniz’s attempts to define the term ‘idea’ (a term that is closely related to the term ‘concept’). One such attempt is in a short paper with the title “What is an idea?”, which Leibniz wrote in 1678. Leibniz starts there by saying that “by the term idea we understand something which is in our mind” (A VI.iv 1370/L 207). But not everything in our mind is an idea: there are also “thoughts, perceptions, and affections”. The latter are occurrent states, but ideas are not: “an idea consists, not in some act, but in the faculty of thinking”. To have an idea of something is to be able to think about it.

Leibniz argues, however, that this is not enough to define what it is to have ideas. One must still introduce the notion of *expression*: an idea of a thing must express the thing. The expression relation, as Leibniz explains it in “What is an idea?”, seems to be a very general cognitive relation: x expresses y if and only if “we can pass from a consideration of the relations in the expression [x] to a knowledge of the corresponding properties of the thing expressed” (A VI.iv 1370/L 207). This generality means that Leibniz can apply the notion to many, quite different cases: the model of a machine expresses the machine itself, an algebraic equation expresses a circle or some other figure, speech expresses thoughts, characters express numbers.

Ideas, however, are in certain respect quite peculiar as expressions.⁷ The most important peculiarity is that the expression relation between an idea and a thing is a necessary and internal relation. It may be possible to use a model of a machine as a model of some other machines as well. An idea of x , however, is essentially, in virtue of what it is, an idea of x and not of anything else. For Leibniz, this means that there is a connection between the individuation of ideas and the individuation of objects: it is not possible to individuate the idea of x independently of x . We could call this dependence *individivative expressivism*.⁸

For Leibniz this notion of idea has far reaching implications:

That the ideas of things are in us means therefore nothing but that God, the creator alike of the things and of the mind, has impressed a faculty of thinking, so that from its own operation it can derive what perfectly corresponds [*respondeant*] to what follows from things [*sequuntur ex rebus*]. Although, therefore, the idea of a circle is not similar to the circle, truths can be derived from it which would be confirmed beyond doubt by investigating a real circle. (A VI.iv 1371/L 208.)

One may at first wonder, whether Leibniz, in claiming that there is a “perfect correspondence” between ideas and the nature of things, assumes some kind of epistemic optimism based on his view of God. However, what he says is really meant to follow from his individivative expressivism concerning ideas. In fact, it seems rather trivial that if an idea of x is individuated by x , it is individuated by the nature of x (what x is).

At the same time, we can surely ask what this talk of *things* in individivative expressivism involves. If it is not possible to individuate an idea without a thing, does it follow that we can only have ideas of what actually exist? Here it is helpful to consider *Discourse on Metaphysics* §26 where Leibniz explains that “this quality of our soul, insofar as it [our soul] expresses some nature, form, or essence, is properly the idea of the thing” (A VI.iv 1570/AG

⁷ Leibniz’s notion of expression is often explained in terms of isomorphism (see. e.g. Swoyer 1995). But it seems to us that isomorphism is not enough to account for ideas as expressions, although we are not able to pursue this point in detail here.

⁸ ‘Expressivism’ as we use it here has of course nothing do with the term as used in contemporary philosophy of language and metaethics.

58). The emphasis on natures, forms, or essences is actually more adequate than the earlier talk of things. For it is not Leibniz's view that in order for us to have an idea of x , x needs to actually exist. We can have ideas of merely possible entities, but it does follow that even these should be what we could call *real possibilities*. The important point for our purposes here is that for Leibniz the being of ideas necessarily involves reality. (It is worth noticing, however, that he further holds that realities need to have a ground in some ultimate actual existent (God).⁹ As will we see, this grounding of realities in something ultimate plays an important role in the argument of *Ars Magna*.)

These points about ideas are true of concepts as well. Concepts are, as Leibniz says in § 27 of the *Discourse*, “those [expressions in our soul] we conceive or form” (A VI.iv 1572/AG 59). To have an idea is to have “in one's soul” something that expresses the nature of some real possibility; the actualization, as we might say, of this idea in thinking is a concept, or a conceiving.

The way of thinking introduced in this section concerning ideas and concepts is going to motivate our reconstruction of the *Ars Magna* argument. Before going into details some general points can be made:

First, it should now be more understandable that Leibniz seems to move rather freely from talk of “concepts” to talk of “conceiving things” to talk of “things conceived” in his formulation of the argument. This need not be a sign of confusion or ambiguity but rather just a sign that what we have called individuating expressivism is somehow involved in the argument. Thereby we can also arrive at a different understanding of Leibniz's implicit starting point. Rather than (as in L&P's (7)) take him to set out from our conceiving some *concept*, we take him to begin with our having a cognitive relation to some *thing* or *reality*.¹⁰ This would serve to mitigate Kantian qualms about the possibility to reach reality by starting with mere concepts. (At the same time, one might wonder whether such a move gives rise to another problem. We will try to say something about that at the end of section 5 below.)

Second, we can now also understand better why, even if conceiving something complex requires conceiving its ultimate constituents, there is no

⁹ In addition to the passages from *Ars Magna* discussed in section 4 below, see also a later writing such as the *Monadology* par. 44 (GP VI 614).

¹⁰ Koistinen (2009, 168) develops a somewhat similar idea in connection to Spinoza. For a helpful discussion of related issues in the Aristotelian tradition and Descartes see also Carriero 2009, 17ff.

implication of complete analysis – that is, we do not face the sort of worries about the nature of conceiving, which seem to arise on L&P’s reconstruction. In the case of conceiving a circle, what we conceive is not the concept of a circle, but the circle itself – that thing or reality. The relevant constituents are not concepts, but further things conceived, i.e. other realities on which conceiving the circle depends (for example space). Conceiving, we could perhaps say, is not a matter of exercising purely conceptual capacities, but involves reality conceived. We hope to clarify this line of thought in the course of next section, where we will focus on how Leibniz’s individuating expressivism can help us understand how the regress is supposed to be blocked. The conception of cognition as essentially involved with reality or things will allow us to develop in a different direction L&P’s insight that the notion of “borrowing” is important to the argument. For it is as a metaphysical notion, rather than as a conceptual one – as in L&P’s *Inheritance Thesis* – that borrowing has a key role to play.

4. Conceiving and Reality-Dependence

The thesis that some things or realities depend upon other things or realities for their reality – for what they are – is central to Leibniz’s metaphysics. He elaborates on the thesis in connection with the question of the nature of aggregates in a well-known passage from a letter to Arnauld: “every being by aggregation presupposes beings endowed with real unity, because every being derives its reality only from the reality of those of which it is composed.” (A II.ii 184/AG 85.) A herd of sheep (to use one of Leibniz’s examples) is an aggregation of sheep, and so the reality of the herd will depend on the reality of the sheep. There is, however, another case of reality-dependence – which has received somewhat less attention in the literature – but which seems more significant for the *Ars Magna* argument. In this case the dependence runs in the opposite direction: rather than one reality depending on a plurality of constituent realities, a plurality of realities (finite things) depend on a single reality (God). The thought is that we start with God’s being, reality or perfection, and that finite things are, as Leibniz explains in *Ars Magna*,

generated through negation or privation (limitation) of that reality.¹¹ Perhaps we can think of this relationship between a plurality of realities and a single basic reality somewhat along the lines of the way in which there can be a plurality of determinations of a single determinable. In *Ars Magna* Leibniz uses the case of figures in space as an analogy for how a multitude of things can be generated from a few: “granted space, body, a straight line, and continuous motion, one could also demonstrate the possibility of a circle; further, even a straight line can be demonstrated, granted space, body and continuous motion.” (A VI.iv 159/MP 3.)

Leibniz characterizes this relation of reality-dependence in terms of one thing’s *borrowing* reality from another (see e.g. LDV 301). It is important to notice that the notion of reality-borrowing is very strong. If x depends on y for its reality, then the very identity of x – what x is – depends on y . It is not only that x could not exist without y , but the specification of what x is, what its reality or being is, includes y (what y is): not only does the present existence of the herd depend on the sheep, but the specification of the reality of the herd includes the sheep (e.g. what the herd does involves the doings of the sheep). Similarly, the specification of the reality of determinates will include the determinable – a figure, as a description in space, includes space, and the specification of the reality of a finite thing includes divine reality.

The metaphysical thesis of reality-borrowing has important implications with respect to the nature of cognition, given Leibniz’s individuating expressivism. Let us consider an idea I that expresses reality x , and suppose that x borrows its reality from reality y . As I is individuated by the reality which it expresses, and reality-borrowing is a matter of the identity of x depending on y , it follows that in order for x to be expressed by I , y also needs to be expressed, and if the reality of y depends on z , z also needs to be expressed by I , and so on. Conversely, if I express x through the idea of y , then the reality of x must depend on y , for otherwise I would not be expressing x , but some other reality. The order of ideas must follow the order of realities. We can sum up this thesis – which we may call the *Borrowing Reality-Conceiving Principle* – in the terminology of *Ars Magna* by substituting ‘conceiving a thing’ for ‘idea’:

¹¹ In this connection, he uses a somewhat bewildering analogy with binary numbers (A VI.iv 156/MP 2).

(BC) If, and only if, some thing x borrows its reality from some other thing y , then if x is conceived, then y is conceived.

It is easy to see the connection to L&P's *Inheritance Thesis*, but also an important difference. (IT) is about conceptual content, whereas (BC) involves a claim about metaphysical dependence. Indeed, as we just saw, the motivation for (BC) lies in the notion of borrowing reality, together with individuating expressivism about ideas.

Does (BC) then help us to block the regress? This is perhaps not quite obvious. For at first sight (BC) may seem to be consistent with there being an infinite chain of reality-borrowers. The case of the herd of sheep appears to offer a good example. The reality of the herd borrows its reality from the sheep, but in some sense the sheep themselves are composite beings, and so also aggregates, and thus depend for their reality on their parts. Why would such a chain of reality-borrowing have to terminate?

Here we need to introduce a further idea, which Leibniz typically presents as a corollary of the notion of reality-borrowing. The idea is clearly formulated in a letter to De Volder: "And where there is no reality except that which is borrowed, there will never be reality, since it ultimately must be proper to some subject [*subjecto propria*]." (LDV 301.) Why must reality be proper to a subject? At one level it is perhaps rather trivial that if A borrows, say, a book, then the book must be owned by someone else, who hasn't herself borrowed it. The owner may either be B – the person from which A borrowed the book – or someone from whom B has borrowed it, and so on. It is, as it were, in the logic of borrowing that the chain of borrowers must terminate in an owner. This also helps us to understand how Leibniz can intend the case of a deferred gift of some money work as a simile illustrating the argument. The point of the simile is that no money is given unless there is somewhere some money to be collected. Analogously, if there is no owner, there is nothing to be borrowed.

The real question then is not whether such a chain has to terminate, but rather how to make sense of reality-dependence in terms of borrowing. For if x borrows its reality from y , not only does the identity of x depend on y , but the reality of x is somehow the very same reality as is found in y , and ultimately in a "reality owner," something to which that reality is "proper" –

in the same way as the book that I borrow is the very same book that is owned by someone else (although in the case of reality it seems as if the idea is that it is simultaneously present in both the borrower and the owner).

In order to make this notion of reality-borrowing seem less abstract, it may help to consider the focal case of *Ars Magna*, where a plurality of things are conceived through (and results as limitations or determinations from) God. If finite things are determinations of unlimited reality (whatever that exactly comes to), it does not seem so far-fetched that it is the same reality that is present in each determination. The determinations do not constitute new levels of reality, but are just variations of divine reality, just as figures can be seen as simply ways in which the very same space is bounded.

These considerations are not meant to in any way exhaust the nature of reality-dependence – there are, for example, important problems about the relationship of determinations (finite things) and determinable reality. Our aim has merely been to give a sense of the picture that motivates Leibniz to think of reality-dependence in terms of borrowing. For it is precisely the connection just outlined between borrowing and owning reality that seems to us to be the key to seeing how (BC) can help to block the regress in the *Ars Magna* argument. The following *Borrowing-Owning Reality Principle* is meant to capture that connection:

(BO) If there is some thing x that borrows its reality, there must be some other thing y that owns that reality, and y is an owner of that reality if, and only if, y is not a borrower of that reality (*Borrowing-Owning Reality Principle*).

While (BO) is, as already noted, not supposed to be an independent principle, but rather something that follows trivially from (BC), it will be helpful to introduce it separately for purposes of reconstructing the argument, which is what we will do next.¹²

¹² One might perhaps wonder whether this consideration would not be available to L&P: couldn't they too rely on something like (BO), but with respect to conceptual content? This would indeed be possible. One advantage with our reading is, however, that the metaphysical principles we rely on are to be found in Leibniz's writings (we do not introduce any new principles about concepts). In addition, it seems to us that the other problems with L&P's reading of the argument (discussed

5. The Argument Reconsidered

We suggested above (see section 3) that there is no ambiguity or confusion in Leibniz's tendency to move between talk of concepts, conceiving of things, and things conceived in different versions of the argument – provided we take individuating expressivism into consideration. However, for sake of clarity we will choose a formulation of the argument in terms of conceiving things, as this most lucidly conveys what we take to be Leibniz's view. Thus, in the following reconstruction of the argument, we have replaced Leibniz's 'concept' with 'conceiving a thing'. We have also chosen to render explicit the reference to *things* by replacing 'whatever is thought' (*quicquid cogitatur*) or 'another' (*aliud*) with 'if some thing is thought' and 'another thing'.

Let us then turn to the reconstruction itself:

- (1*) If some thing is thought by us, then it is either conceived through itself [*per se*], or conceived through some other thing (assumption).
- (2*) If some thing is conceived through some other thing, then that other thing is again either conceived through itself or conceived through some other thing (from (1*)).
- (3*) One must either proceed in conceiving one thing through another thing to infinity, or all thoughts are resolved into those things which are conceived through themselves (from (2*); not a premise in the argument).
- (4*) If some thing is conceived only through other things, then it will be conceived only insofar as those other things are conceived (only a reformulation of 2*).
- (4b) If, and only if, some thing x borrows its reality from some other thing y , then if x is conceived, then y is conceived (*Borrowing Reality-Conceiving Principle*).
- (4c) If there is some thing x that borrows its reality, there must be some other thing y that owns that reality, and y is an owner of that reality if, and only if, y is not a borrower of that reality (*Borrowing-Owning Reality Principle*).

in section 2 above) would remain, for example that it seems to fail to deliver the intended conclusion.

- (4d) If x owns reality, then x is conceived through itself (from 2* and (4b) and (4c)).
- (4e) If there is some thing x that borrows its reality from some other thing y , then there is some other thing than x that is conceived through itself (from (4d) and (4c)).
- (4f) If some thing x is conceived only through some other thing y , then there is some other thing than x that is conceived through itself (from (4e) and (4b)).
- (5*) If we conceive some thing, then we conceive some thing through itself (from (2*) and (4f)).¹³
- (6*) If nothing is conceived through itself, nothing will be conceived at all (from 5*).¹⁴
- (7*) We conceive some things (unstated assumption).
- (8*) There is some thing conceived through itself (from (6*), (7*)).

The numbering of the steps is meant to correspond to L&P's reconstruction, (1*) corresponding to their (1) etc., except our (8*), which corresponds to their (10). Steps (4b)–(4f) are steps added in our reconstruction. It is worth noticing that, even if it may look a bit complicated, when we formulate the argument in terms of conceiving things it actually becomes logically simpler than on L&P's reconstruction:

Firstly, we do not need anything corresponding to their steps (8) and (9), but can move directly from (7*) to the conclusion (8*).

Secondly, (3*) in our version does not figure as a premise in the argument, while L&P need to appeal to (3) in order to arrive at the conclusion. Our version seems more in keeping with how Leibniz himself presents the consideration about infinity, which we have expressed as (3*).

Thirdly, L&P need to add (4) – the claim that complex concepts are conceived only insofar as their constituents are conceived – as a new assumption, an assumption that needs to be motivated (a motivation that

¹³ In *Ars Magna* Leibniz here writes “conceive something in actuality,” but for purposes of clarity we have left out “in actuality,” since we do not think it is meant to add anything substantial, but merely works as emphasizing that conceiving is *actual* conceiving. We also take Leibniz's expression “we arrive at those things which are conceived through themselves” to be equivalent to “we conceive some thing through itself”.

¹⁴ This step is of course not necessary, but we include it, since it is included in *Ars Magna*.

raises some difficulties for L&P). Yet when we cast the argument in terms of conceiving things it turns out that the corresponding step (4*) is not a new assumption, but identical to (2*). It is thus not (4*) that needs to be motivated, but the step from (4*) to (5*). That motivation involves appeal to the *Borrowing Reality-Conceiving Principle* and the *Borrowing-Owning Reality Principle*. We have also chosen to articulate in some detail (steps (4b)–(4f)) how these principles help to deduce (5*).

For expository purposes we have left (3*) and (4*) as steps in the argument, even though logically speaking they are not needed.

A possible worry about our reconstruction is that it relies too heavily on metaphysical principles ((4b) and (4c)). Another worry has to do with what is, as it were, the starting-point of the argument – the claim that we do conceive some things (7*). While this formulation avoids what we suggested is a problem with L&P’s version – namely that it leaves the step from concepts to reality unaccounted for – it is easy to get the feeling that it does so by in some sense begging the question.

With respect to the first worry one could begin by saying that by introducing the metaphysical principles, the major gap in the argument is filled without otherwise introducing any significant changes in Leibniz’s formulation of it. Furthermore, this more metaphysical reconstruction stays much closer to the thrust of Leibniz’s argument, expressed in the conclusion that there has to be *things* conceived through themselves – it is an argument for the existence of ultimate reality. From a more general perspective, our interpretation suggests that Leibniz’s theory of cognition forms a unity with his metaphysics. Very few scholars nowadays would agree with Russell’s and Couturat’s famous claim that Leibniz’s whole philosophy is based in his logic (see Couturat 1901 and Russell 1937). However, recent interpretations still tend to retain a logical approach to Leibniz’s theory of concepts (this tendency is implicit in L&P’s reading). In contrast, our reading indicates that even the theory of concepts cannot be understood apart from his metaphysics.

The second worry raises some large issues that we cannot go into here. The basic question is whether Leibniz’s argument can speak to a more general worry (found in, for example, Kant) about the possibility of cognition of reality. The focus here would be on step (7*) in our reconstruction, and crucial question is how strong that assumption must be. It seems to us that the

argument might work even on quite a weak reading of (7*) – i.e. a weak understanding of the relevant notion of reality – but this is something that will have to be explored on another occasion.¹⁵

University of Turku

Bibliography

- Carriero, John (2009). *Between Two Worlds: A Reading of Descartes's Meditations*. Princeton, NJ and Oxford: Princeton University Press.
- Couturat, Louis (1901). *La logique de Leibniz*. Paris: F. Alcan.
- Kant, Immanuel (1998). *Critique of Pure Reason*. Transl. and ed. Paul Guyer & Allen W. Wood. Cambridge: Cambridge University Press. [A/B]
- Koistinen, Olli (2009). "Spinoza on Action", in Olli Koistinen (ed.): *The Cambridge Companion to Spinoza's Ethics*. Cambridge: Cambridge University Press.
- Koistinen, Olli (2014). "The Fifth Meditation: externality and true and immutable natures", in David Cunniff (ed.): *The Cambridge Companion to Descartes' Meditations*. Cambridge: Cambridge University Press.
- Leibniz, Gottfried Wilhelm (1923–). *Sämtliche Schriften und Briefe*. Ed. Deutsche Akademie der Wissenschaften. Darmstadt/Leipzig/Berlin: Akademie Verlag. Cited by series, volume and page. [A]
- Leibniz, Gottfried Wilhelm (1989). *Philosophical Essays*. Ed. & transl. Roger Ariew & Daniel Garber. Indianapolis: Hackett. [AG]
- Leibniz, Gottfried Wilhelm (1989 [1969]). *Philosophical Papers and Letters*. Ed. & transl. Leroy E. Loemker. 2nd ed. Dordrecht: Reidel [Kluwer]. [L]
- Leibniz, Gottfried Wilhelm (2013). *The Leibniz-De Volder Correspondence*. Transl. and ed. Paul Lodge. New Haven and London: Yale University Press. [LDV]
- Leibniz, Gottfried Wilhelm (1973). *Philosophical Writings*. Ed. G. H. R. Parkinson. Transl. Mary Morris and G. H. R. Parkinson. London: Dent. [MP]
- Leibniz, Gottfried Wilhelm (1996). *New Essays on Human Understanding*. Transl. Peter Remnant & Jonathan Bennett. Cambridge: Cambridge University Press. Cited by pages from A VI, vi., which is also the pagination in the English translation. [NE]
- Lodge, Paul & Puryear, Stephen (2006/7). "Unconscious Conceiving and Leibniz's Argument for Primitive Concepts", *Studia Leibnitiana* 38 (2), 177–196.
- Plaisted, Dennis (2003). "Leibniz's Argument for Primitive Concepts", *Journal of the History of Philosophy* 41 (3), 329–341.
- Russell, Bertrand (1937). *A Critical Exposition of the Philosophy of Leibniz: With an Appendix of Leading Passages*. London: George Allen & Uwin.
- Swoyer, Chris (1995). "Leibnizian Expression", *Journal of the History of Philosophy* 33(1), 65–99.

¹⁵ For discussion of the issues taken up in this paper, we would like to thank participants in our Fall 2015 class on Leibniz at the University of Turku. We would also like to thank Olli for many fruitful conversations about related questions. Myrdal's research has been financially supported by the Academy of Finland (project number 275652).

Kantin maailmat

Ilkka Niiniluoto

Immanuel Kantin ajattelu muodostaa vedenjakajan, johon kaikkien itseään kunnioittavien filosofien on jollain tavoin otettava kantaa. Uuskantilainen Wilhelm Windelband jakoikin filosofian historian kolmeen osaan: ennen Kantia, Kant ja Kantin jälkeen. Suomessa Kantia on tutkittu innokkaasti yli kahden vuosisadan ajan (Oittinen 1999). Eino Kaila, jonka oma suhde Kantiin oli varsin kriittinen, testasi edistyneitä opiskelijoita teoreettisen filosofian laudaturtentissä vaativan teoksen *Kritik der reinen Vernunft* avulla. J. E. Salomaa käänsi Kantin siveysopilliset pääteokset 1928 ja Vesa Oittinen *Prolegomena*-teoksen 1997, mutta vasta professori Olli Koistisen työryhmä Turussa on huolellisen valmistelun jälkeen 2013 saanut aikaan käännöksen teokselle *Puhtaan järjen kritiikki*. Koistisen monografia Kantin teoksesta ehti ilmestyä jo viisi vuotta aikaisemmin (Koistinen 2008). Näin Olli Koistinen on kohottanut suomalaisen Kant-tutkimuksen uudelle tasolle, mutta myöskin antanut meille, jotka eivät ole professionaalisia suuren königsbergiläisen tutkijoita, tervetulleen mahdollisuuden täsmentää ja syventää omaa suhdettamme Kantin ajatteluun.

Tässä pienessä kirjoituksessa tarkastelen kiistakysymystä siitä, olettaako Kant filosofisessa järjestelmässään kaksi erilaista olioiden tyyppiä, jotka vastaavasti muodostavat kaksi ontologisesti erillistä maailmaa. Esitän myös kriittisen tieteellisen realismin pohjalta muodostetun oman näkemyksen siitä, miten tähän kysymykseen todellisuuden tasoista pitäisi vastata.¹

1. Kaksi maailmaa?

Selvittääkseen perinteisen metafysiikan virheellisyyden ja oikeanlaisen metafysiikan mahdollisuuden tieteenä Kant asetti tehtäväkseen osoittaa synteettis-aprioristen arvostelmien mahdollisuuden. Hänen ratkaisunsa apriorisen tiedon ongelmaan seuraa tekijän tiedon mallia (vrt. Hintikka 1974):

¹ Aikaisempia yrityksiäni tässä suunnassa löytyy kirjoituksista Niiniluoto 1984 ja 1990 sekä teoksesta Niiniluoto 1999, luvut 4.3. ja 4.4. Ensimmäisen englannin kielisen käännöksen Kantin puhtaan järjen kritiikistä ostin ja luin 1969.

”tiedämme olioista *a priori* vain sen, minkä itse niihin asetamme” (B XVIII). Kantin ”kopernikaanisen kumouksen” mukaan kognition kohteiden on mukauduttava intuitiokykymme² rakenteeseen, joka liittää näihin kohteisiin ajan ja avaruuden havainnon muotoina sekä puhtaan ymmärryksen kategoriat (kuten kausalityetti). Kognitiomme alkaa kokemuksesta niin, että objektit³ ”vaikuttavat aistehimme”, joten kokemuksemme on ”yhdistelmä siitä, minkä otamme vastaan vaikutelmien kautta, ja siitä, minkä oma kognitiokykymme tuottaa itsestään” (B 1). Näin ollen ”meillä ei voi olla kognitiota mistään kohteesta oliona itsessään vaan ainoastaan sikäli kuin se on aistimellisen intuition objekti, siis ilmentymänä” (B XXVI). Kant lisäsi varauksen, että ”vaikkemme voikaan *tuntea* näitä kohteita olioina itsessään, täytyy meidän voida ainakin *ajatella* niitä sellaisina”, sillä muuten tästä seuraisi ”järjetön lause, että olisi ilmentymiä ilman jotakin, mikä niissä ilmenee” (B XXVII). Ei siis ole mieleöntä ajatella olioita sinänsä tai puhua niistä, vaikka emme tiedä niistä muuta kuin niiden olemassaolon ja kyvyn aiheuttaa meissä vaikutelmia.

Kantin transsendentaalisen idealismin mukaan kaikki, mikä intuoidaan avaruudessa tai ajassa, siis kaikki mahdollisen kokemuksen kohteet, ovat ”pelkkiä ilmentymiä”, eikä niillä ole ajatustemme ulkopuolella mitään itseensä perustuvaa olemassaoloa (A 491). Kant yhdisti tähän oppiin empiirisen realismin, jonka mukaan ajassa ja avaruudessa esiintyvät ilmiöt ovat todellisia (A 28).

Näin päädyimme Kantin tietokritiikin pohjalle rakentaman järjestelmän standarditulkintaan, jonka mukaan siihen sisältyy kahdenlaisia olioita: olioita sinänsä, jotka muodostavat noumenaalisen maailman, ja ilmentymiä (olioita meille), joiden fenomenaalinen maailma on muodoltaan transsendentaalisen egon konstituoma (A 249). Kant itse käytti latinan termejä *mundus phaenomenon* (ilmiömaailma) ja *mundus intelligibilis* (ymmärryksen maailma). Koistisen (2008) ontologinen lukutapa Kantista metafysiikkona tukee tätä kahden maailman tulkintaa. Oliot sinänsä ovat olennainen osa tätä järjestelmää, sillä ne

² Intuitio on käänös saksan sanalle *Anschauung*, jonka Oittinen puolestaan kääntää sanalla ”havainnointi”. Vaikka intuitio, eli subjektin suuntautuminen yksittäiseen oloon, on ihmiselle aina aistimellista, Kant ei sulje pois mahdollisesti Jumalalle kuuluvaa ei-aistimellista intellektuaalista intuitiota.

³ Tässä ”objektit” on mielestäni parempi käänös saksan termille *Gegenstände* kuin ”kohteet”, jolla tavallisesti viitataan ilmentymiin.

kausallisesti aiheuttavat ilmentymät. Tämä tulkinta näkyy Koistisen tavassa kääntää *Erscheinung* sanalla ”ilmentymä” ja *Vorstellung* sanalla ”representaatio”.

Kahden maailman opilla on tärkeä merkitys myös Kantin moraalifilosofian kannalta. Puhtaan järjen kolmannessa antinomiassa Kant esitti teesin, jonka mukaan maailman ilmentymiä ei voi johtaa pelkästään luonnon lakien mukaisesta kausaliteetista, vaan niiden selittämiseksi on välttämätöntä olettaa vielä ”vapauden kausaliteetti” (A 445). Tällä hän viittaa ihmisen kykyyn panna alkuun tapahtumasarjoja (esimerkiksi nousen vapaasti tuoliltani), jotka ilmenevät fenomeenien maailmassa (A 450), jolloin vapaa tekijä ”ensimmäisenä syynä” on ihminen oliona sinänsä (Koistinen 2008, 135).⁴ Teoksessa *Grundlegung zur Metaphysik der Sitten*, joka ilmestyi 1785 *Puhtaan järjen kritiikin* ensimmäisen painoksen (1781) ja toisen painoksen (1787) välissä, Kant jatkoi kompatibilismina tunnettua ratkaisuaan vapaan tahdon ongelmaan. Aistimaailmaan kuuluvana ”ilmiömaailman olentona” ihminen on kausaalisten luonnonlakien alainen, mutta ”järjen maailmaan” kuuluvana ”ajatusoliona” tai ”olentona sinänsä” hän on vapaa (ks. Kant 1990, 153). Tämä erottelu ilmaistaan usein sanomalla, että ihmisen ”empiirinen minä” kuuluu fenomeenien maailmaan, kun taas ihminen järkiolentona on olio sinänsä (vrt. Koistinen 2008, 14, 129). Tätä tulkintaa tukee *Puhtaan järjen kritiikin* toisen painoksen esipuhe, jonka mukaan ”tahto ilmentymässä” on ”luonnonlakien mukainen” ja sikäli ei vapaa, kun taas ”olioon sinänsä kuuluvana se ei ole näiden lakien alainen” (B XXVIII). Tämä teema on keskeisesti esillä myös seuraavassa teoksessa *Kritik der praktischen Vernunft* (1788).

Kantin kahden maailman oppiin liittyy kuitenkin vaikeuksia, joista monet koskevat olion sinänsä käsitettä. Ensimmäisessä painoksessa Kant määritteli noumeenit (*intelligibilia*) positiivisesti olioina, jotka ovat pelkästään ymmärryksen tai ei-aistimellisen intellektuaalisen intuition kohteita (A 249). Toisessa painoksessa hän suosi negatiivista määrittelyä, jonka mukaan noumenon on olio, joka ei ole aistimellisen intuition kohde (B 307). Kyse on ”rajakäsitteestä” (*Grenzbegriff*), joka ilmaisee rajaa, johon tietomme ei voi ulottua (B 311). Mutta mitä oliosta jää jäljelle, jos siitä poistetaan kaikki ilmentymiin kuuluva aines? Sillä ei ole ajallis-avaruudellisia ulottuvuuksia,

⁴ Vaikka Kant ajoittain puhuukin ihmisestä oliona sinänsä, Pihlström (2009) muistuttaa, että Kantille mieli ei ole kartesiolainen substanssi. Turussa vaikuttanut G. I. Hartman esitti teoksessaan *Kunskapslära I-II* (1807–1808), että itseään tiedostava Minä on olio sinänsä (Oittinen 1999).

eikä siihen voi soveltaa ymmärryksen kategorioita. *Prolegomena*-teoksessa Kant sanoi selkeästi, että Locken sekundaaristen kvaliteettien (kuten lämpö, väri ja maku) lisäksi myös primääriset kvaliteetit (kuten ulottuvuus, hahmo ja läpäisemättömyys) kuuluvat pelkästään ilmiöihin eikä olioihin sinänsä (Kant 1997, 88). Tämä on johtanut kommentaattorit kysymään, onko tällainen ”paljas olio” vailla ominaisuuksia ylipäänsä mitään. Erik Stenius (1963) pitääkin negatiivista noumeenin käsitettä virheenä.⁵

Vielä vakavampi ongelma on siinä, että kantilainen kahden maailman oppi on sisäisesti ristiriitainen. Ensinnäkin olemassaolo on yksi kategorioista (A 80), joten sitä ei voi soveltaa ilmiömaailman ulkopuolisiin olioihin sinänsä. Toiseksi syyn ja vaikutuksen suhde myös kuuluu kategorioihin, joten sitä ei voi soveltaa olioiden sinänsä ja ilmentymien väliseen suhteeseen.⁶ Thiodolf Rein, joka itse kannatti lotzelaista emergenttiä idealismia, arvioi juuri jälkimmäisen ristiriidan johtaneen Fichten hylkäämään oliot sinänsä – ja näin aloittaneen pitkän idealismin kauden saksalaisessa filosofiassa (Rein 1908, 292). Samaa ajatusta seuraa Peter Strawson (1966), joka hyväksyy kahden maailman tulkinnan siitä, mitä Kant todella tarkoitti, mutta päätyy sen ristiriitaisuudesta väitteeseen, ettei Kantin oppia voi erottaa Berkeleyyn idealismista tai fenomenalismista.⁷

2. Vai sittenkin vain yksi maailma?

Henry E. Allisonin 1983 ilmestynyt teos (toinen laajennettu painos 2004) Kantin transsendentaalisesta idealismista on tunnetuin vaihtoehto ontologiselle kahden maailman tulkinnalle. Koistisen Kant-kirjan

⁵ Kaila arvioi teoksen *Hahmottuva maailma* postuumissa luonnoksessa (vuodelta 1958), että Kant oli strukturaalinen realisti, jonka mukaan aineella on ominaisuuksien sijaan vain relaatioita. Kaila oli jo teoksessaan *Inhimillinen tieto* (1939) esittänyt, että ”Kant selvästi ajattelee, että ilmiöt ja oliot sinänsä ovat isomorfisessa suhteessa toisiinsa”, joten ”ilmiömaailma esittää olioita sinänsä; molemmilla on sama strukturi” (Kaila 1939, 28).

⁶ Myös *Prolegomena*-teoksessa Kant toisti eksplisiittisesti käsityksen, jonka mukaan ulkopuolellemme olevat kappaleet ”vaikuttavat aisteihimme” ja näin aikaansaavat meissä mielteitä (Kant 1997, 88). Ajatus olioiden sinänsä vaikutuksista ilmiömaailmaan sisältyy myös oppiin ”vapauden kausaliteetista”.

⁷ Oikeastaan Machin fenomenalismi saadaan Kantin järjestelmästä vasta poistamalla olioiden sinänsä lisäksi transsendentaalinen subjekti, jolloin jäljelle jää vain ilmiöitä (niiden joukossa empiirinen minä elämysten kimppuna) (vrt. Niiniluoto 1990). Fenomenologit asettavat ulkomaailman sulkeisiin pyrkien osoittamaan, miten transsendentaalinen ego konstituoii elämismaailman, kun taas monet loogiset positivistit (kuten nuori Rudolf Carnap) ja jotkut pragmatistit ovat pitäneet puhetta tajunnasta riippumattomasta ulkomaailmasta mielettömänä.

arvostelijoista Toni Kannisto (2009) ja Sami Pihlström (2009) viittaavatkin myönteisesti Allisonin ajatuksiin, joita Pihlström pienin varauksin kannatti jo väitöskirjassaan (Pihlström 1996, 222–248). Allisonin ”kaksoisaspektitulkinnan” mukaan oliot sinänsä ja ilmiöt ovat vain kaksi eri tapaa puhua samoista olioista, joten samalla on aihetta hylätä ajatus noumeenien ja fenomeenien muodostamista erillisistä maailmoista.

Tekstievidenssiä Allisonin epistemologiselle yhden maailman tulkinnalle toki löytyy Kantin kirjoituksista. Mutta jos hänen Kant-tulkintansa olisi oikea, olisiko Fichtellä ja muilla idealisteilla ollut enää aihetta kantilaisten olioiden sinänsä hylkäämiseen? Vaikka Allisonin Kant ei olisikaan se historiallinen ajattelija, joka pani vauhtiin 1800-luvun filosofian kehityskulut, hänen antamansa rationaalinen Kant-rekonstruktio on erinomaisen mielenkiintoinen. Itse olen arvellut, että yhden maailman kannattajana Kant olisi kriittinen tieteellinen realisti, joka ei osannut kunnolla muotoilla näkemyksiään (Niiniluoto 1999). Ehkä kuitenkin osuvampi luonnehdinta on Charles Peircen huomautus vuodelta 1905, että ”Kant (jota *enemmän* kuin ihailen) ei ollut muuta kuin jonkin verran sekava pragmatisti”, mikä tulee lähemmäksi Pihlströmin tavoitetta löytää naturalisoidusta transsendentaali-filosofiasta aineksia pragmatistiselle realismille.

Koska ilmiöillä on ominaisuuksia, jotka puuttuvat olioilta sinänsä, kyse ei voi olla näiden oliotyypin identiteetistä missään tavanomaisessa merkityksessä. Niinpä yhden maailman tulkinnassa joudumme kysymään, mitä ovat ne entiteetit, joiden aspekteja nämä oliotyypit ovat. Allisonin mukaan puhuessamme olioista sinänsä oikeastaan puhumme tutuista annetuista spatiotemporaalisista olioista ilman viittausta niitä luonnehtiviin episteemisiin ehtoihin. Tämän mukaan hänelle ensisijaisia olioita ovat kantilaiset ilmiöt, ja noumeenit toissijaisesti luonnehditaan negatiivisella tavalla. Pihlström (2009) myötäilee tätä tulkintaa sanoessaan, että Kantin ilmiöiden ”taakse” ei tarvitse ontologisesti olettaa erillisiä olioita sinänsä, joiden ilmentymistä on kysymys. Allisonin perusolioiden valinta sopii yhteen Pihlströmin pragmatismien kanssa – kuitenkin niin, että jälkimmäiselle eräänlaisen ”moniaspektiteorian” puitteissa erilaiset inhimilliset käytännöt jäsentävät omilla tavoillaan inhimillistä kokemusta ja todellisuutta.

Allisonin kaksoisaspektiteorian ansioita Kant-tulkintana on arvioitava sen uskottavuuden ja koherenssin perusteella. Itse en usko sen ratkaisevan

Kantin järjestelmän sisäisiä ristiriitoja. Ilman lisäoletuksia identiteetistä se ei näytä osoittavan, miksi olioista sinänsä ei voi tietää mitään. Käsitykseni mukaan se ei ratkaise kompatibilismin ongelmaa, vaan jättää ihmisen vapauden liiaksi vain näkökulmasta tai puhetavasta riippuvaksi.⁸ Ehkä kuitenkin vaikein aihe on Kantin teesi olioiden sinänsä kausaalista vaikutuksesta ilmentymiin, mikä nähdäkseni ei mitenkään selity yhden maailman oletuksen pohjalta. Miten kausaalisuhte voisi vallita pelkkään näkökulmien eroon pohjautuvien aspektien välillä?

3. Kriittinen tieteellinen realismi

Idealistien ratkaisuna Kantin järjestelmän ongelmiin oli olioiden sinänsä eliminointi. Toisenlaista ratkaisua ehdottavat suorat realistit, joiden mukaan havainnossa ja kokeissa tietoteoreettinen subjekti voi olla yhteydessä meistä ontologisesti riippumattoman ulkomaailman objektien kanssa. Kantin ajatus, että aistimellinen intuitio suuntautuu ilmentymiin, on verrattavissa empiristien opeihin, joiden mukaan emme näe fyysisiä objekteja vaan sense-dataa. Realistille interaktio ulkoisten objektien kanssa havainnon ja kokeiden muodossa antaa meille informaatiota näistä objekteista itsestään. Kyseinen informaatio ei ole aistimellisen intuition kohde vaan pikemminkin maailmaa koskevan arvostelman sisältö. Tämä sisältö representoi kohdetta siinä mielessä, että se voi olla osittaista ja epätarkkaa, mutta silti – toisin kuin Kantilla – tällainen tieto *a posteriori* koskee ulkomaailman olioita. Näin realistin ehdotus on yhden maailman oppi, jossa kantilaiset ilmentymät verhoina subjektin ja objektin välillä eliminoidaan (Hintikka 1983).⁹ Realistin

⁸ Tämä kysymys liittyy siihen, onko vapauden kausaliteettia koskeva teesi totta (Koistinen 2008, 129) vai pelkästään praktisen järjen postulaatti, kuten Jumalan ja kuolemattomuuden aatteet. Salomaan selitykset käytännöllisen järjen kritiikkiin myötäilevät kahden maailman tulkintaa (ks. Kant 1990, 383–387). Vrt. kuitenkin Kotkavirta 2013.

⁹ Olen itse käyttänyt muotoilua, jonka mukaan tieteellinen realisti eliminoi Kantin fenomeenin (Niiniluoto 1984, 214), mutta tämä voi olla harhaanjohtavaa. Realistinkin ontologiassa esiintyy aistimuksia ulkoisten objektien aiheuttamina mentaalina tiloina. Kant oli oikeassa, että ihminen ei ole tällaisten aistivaikutelmien passiivinen vastaanottaja vaan tulkitsee ja jäsentää niitä käsitejärjestelmien puitteissa. Realisti voi myös hyväksyä visuaaliseen avaruuteen tai muistimaailman aikakokemukseen sijoittuvia havaintoesineitä, mutta nämä ovat tulkittavissa Hintikan perspektiivisinä maailmanviivoina. Kant oli kuitenkin realistin mielestä väärässä hylätessään fyysisen avaruusajan ja olettaessaan visuaalisen avaruuden olevan euklidinen (Kaila 1939; Niiniluoto 1999). Realistille fenomeenit eivät kuitenkaan peitä niiden takana olevaa todellisuutta, vaan esittävät tai sisältävät siitä havaintoon perustuvaa tietoa.

näkökulmasta Kant on myös liian empiristinen väittäessään, että tietomme ei voi koskaan ulottua aistihavainnon tai mahdollisen kokemuksen ulkopuolelle. Hintikka (1983) kutsui tätä ”Kantin aristoteeliseksi virheeksi”. Kant ei ole myöskään oivaltanut, että lisääntynyt ”transsendentaalitieto” tietokykymme toiminnasta samalla lisää tietoa olioista sinänsä.¹⁰ Tieteellisen realismin mukaan selittäviä teorioita muotoilemalla ja empirisesti testaamalla voidaan saavuttaa tietoa, joka korjaa havaintoja ja vie havaittavien ilmiöiden taakse.¹¹

Yleisemmin voimme sanoa, että kriittiselle tieteelliselle realistille noumenaalinen maailma on kausaalisten (osittain indeterminististen) luonnonlakien hallitsema prosessi, josta fysikaaliset objektit erottuvat ajallispaikalliseen jatkuvuuteen perustuvien kriteerien mukaisesti. Siten realisti valitsee maailman perusolioiksi – ilmiöiden sijasta – ihmisen mielestä riippumattomat fysikaaliset entiteetit, joilla on meille vielä tuntemattomia ominaisuuksia ja suhteita. Nämä oliot ja prosessit sinänsä ovat siten ontologisesti rikkaampia kuin niitä koskevat havainnot ja kuvaukset, jotka aina tavoittavat niistä vain joitakin aspekteja. Realistille ne ovat myös tiedon kohteita, jotka kausaalisesti vaikuttavat tutkijayhteisön uskomuksiin tieteellistä menetelmää sovellettaessa. Siten realistik eivät hylkää olion sinänsä käsitettä vaan ajatuksen tiedostamattomista olioista sinänsä. Argumentteja tässä suunnassa Kantin agnostisismia vastaan ovat esittäneet pragmatisti Peirce 1868 (ks. Peirce 2001, 85–86), materialisti Friedrich Engels 1886 ja marxilainen V. I. Lenin 1909. Esimerkiksi Engels vetosi tekijän tiedon hengessä ihmisen käytännölliseen kykyyn manipuloida luonnonilmiöitä ”kokeilussa ja teollisuudessa”, ja Lenin korvasi eron ilmiön ja olion sinänsä välillä ”jo tiedostetun ja vielä tiedostamattoman” erottelulla (Niiniluoto 1990).

Saksassa, jossa Marburgin uskantilaiset korostivat olioiden sinänsä luonnetta pelkkänä ”rajakäsitteenä”, Alois Riehl aloitti 1887 ”kriittisen realismin” koulukunnan, jonka mukaan me voimme saavuttaa tietoa olioista sinänsä. Ohjelmaa jatkoivat mm. Wilhelm Wundt 1896 ja Oswald Külpe 1912 (ks. Neuber 2014). Tyylikkään muotoilun kriittiselle realismille antoi

¹⁰ Hintikka (1974) argumentoi myös, että kantilaisten havainto-objektien sijasta tiedollisen päättelyn lähtökohtina tulisivat olla ”etsimisen ja löytämisen kieliopien” kautta saavutettavat oliot.

¹¹ Pihlström on Heikki Kannistoa seuraten ehdottanut, että Kant voisi hyväksyä teoreettisten entiteettien (kuten fysikaalisen avaruuden) postuloinnin osana ilmiömaailmaa, mutta Kant ei itse tällaista askelta suorita.

Moritz Schlick teoksessaan *Allgemeine Erkenntnislehre* (1918/1925).¹² Hän argumentoi marburgilaista Paul Natorpia vastaan, että mahdottomuus tuntea oliot tyhjentävästi ei kiellä mahdollisuutta tuntea niitä osittain tai likimäärin (Schlick 1985, 362). Nuori Eino Kaila loogisen empirismin muotoilussaan 1926 mainitsi esikuvinaan Külpen ja Schlickin kriittisen realismin. Hintikka (1983) käytti samantapaisia ilmaisuja arvostellessaan Kantin erehdystä ”piilottaa oliot sinänsä aistihavainnon taakse”: olioiden sinänsä tyhjentymättömyydestä ei seuraa niiden saavuttamattomuutta. Oliot sinänsä eivät enää ”piile havainnon esiripun takana, eivätkä muutenkaan muodosta erillistä olioluokkaa, vaan ne ovat meidän lähestyttävissämme”. Eksplikaation tälle totuuden lähestymisen idealle voi antaa totuudenkaltaisuuden (*truthlikeness*) käsitteen avulla (Niiniluoto 1999).

Näin muotoiltu kriittinen tieteellinen realismi on Kantin kahden maailman opin vaihtoehto, joka myös selkeästi poikkeaa Allisonin yhden maailman opista. Havainto-objekteilla on oma roolinsa tieteellisiä teorioita testattaessa, ja ne syntyvät aidon kausaalisesti tutkijan ja todellisuuden välisessä vuorovaikutuksessa. Muutoin keskeiseen asemaan nousevat tieteellisten teorioiden hypoteeseina postuloimat teoreettiset entiteetit, joiden kuvaukset totuudenkaltaisella tavalla pyrkivät approksimoimaan tutkijasta riippumatonta todellisuutta.

Helsingin yliopisto

Kirjallisuus

- Allison, Henry E. (2004). *Kant's Transcendental Idealism: An Interpretation and Defense*. Revised and Enlarged Edition. New Haven: Yale University Press.
- Hintikka, Jaakko (1974). *Knowledge and the Known: Historical Perspectives in Epistemology*. Dordrecht: D. Reidel.
- Hintikka, Jaakko (1983). ”Transsendentaalitiedon paradoksi”, *Ajatus* 40, 20–48.
- Kaila, Eino (1939). *Inhimillinen tieto*. Helsinki: Otava.
- Kaila, Eino (1979). ”Filosofian klassillinen käsitys aineellisen ja sielullisen suhteesta”, teoksessa Simo Knuuttila et al. (toim.), *Aate ja maailmankuva*. Porvoo: WSOY, 436–456.
- Kannisto, Toni (2009). ”Puhdasta järkeä omalla äänellä”, *niin & näin* 1/09, 161–163.
- Kant, Immanuel (1990). *Siveysopilliset pääteokset*. 2. painos. Suom. J. E. Salomaa. Porvoo: WSOY.

¹² Tämä oli ennen kuin hän 1930-luvulla kääntyi loogiseksi positivistiksi.

- Kant, Immanuel (1997). *Prolegomena*. Suom. Vesa Oittinen. Helsinki: Gaudeamus.
- Kant, Immanuel (2013). *Puhtaan järjen kritiikki*. Suom. Markus Nikkarla ja Kreetta Ranki. Helsinki: Gaudeamus. [A/B]
- Koistinen, Olli (2008). *Kant ja Puhtaan järjen kritiikki*. Turku: Areopagus.
- Kotkavirta, Jussi (2013). ”Kantin moraalifilosofia”, teoksessa Vesa Oittinen (toim.), *Immanuel Kantin filosofia*. Helsinki: Gaudeamus, 162–190.
- Neuber, Matthias (2014). ”Critical Realism in Perspective: Remarks on a Neglected Current in Neo-Kantian Epistemology”, teoksessa Maria Carla Galavotti et al. (toim.), *New Directions in the Philosophy of Science*. Cham: Springer International Publishing Switzerland, 657–673.
- Niiniluoto, Ilkka (1984). ”Realism, Worldmaking, and the Social Sciences”, teoksessa *Is Science Progressive?* Dordrecht: D. Reidel, 211–225.
- Niiniluoto, Ilkka (1990). ”Realismi, praksismi ja relativismi”, teoksessa *Maailma, minä ja kulttuuri*. Helsinki: Otava, 43–87.
- Niiniluoto, Ilkka (1999). *Critical Scientific Realism*. Oxford: Oxford University Press.
- Oittinen, Vesa (1999). ”Suomen ensimmäiset kantilaiset”, *Ajatus* 56, 45–87.
- Peirce, Charles S. (2001). *Johdatus tieteen logikkaan ja muita kirjoituksia*. Suom. Markus Lång. Tampere: Vastapaino.
- Pihlström, Sami (1996). *Structuring the World: The Issue of Realism and the Nature of Ontological Problems in Classical and Contemporary Pragmatism*. Acta Philosophica Fennica 59. Helsinki: Societas Philosophica Fennica.
- Pihlström, Sami (2009). ”Analyttisesti Kantin Kritiikistä”, *Tieteessä tapahtuu* 1/09, 59–63.
- Rein, Thiodolf (1908). *Filosofin studium vid Åbo Universitet*. Helsingfors: Svenska Litteratursällskapet i Finland.
- Schlick, Moritz (1985). *General Theory of Knowledge*. Engl. A. E. Blumberg. La Salle, Illinois: Open Court.
- Stenius, Erik (1972). ”On Kant’s Distinction between Phenomena and Noumena”, teoksessa *Critical Essays*. Acta Philosophica Fennica 25. Amsterdam: North-Holland, 231–245.
- Strawson, Peter (1966). *The Bounds of Sense: An Essay on Kant’s Critique of Pure Reason*. London: Methuen.

Kant ja aikajärjestyksen ideaalisuus

Valtteri Arstila

Puhtaan järjen kritiikissä Kant esittää *transsendentaalisen idealismin* oppinsa. Sen mukaan aika ja avaruus ovat aistimellisen intuition muotoja eli havaitsemisen edellytyksiä, jotka mahdollistavat olioiden representaation kokemuksessamme. Tämä tarkoittaa toisaalta sitä, että Kantin mukaan aika ja avaruus ovat tiedettävissä *a priori* ja transsendentaalisesti ideaaleja havaitsevan subjektin havainnon muotoja. Toisaalta taas myös sitä, että aikaa ja avaruutta ei ole olemassa ilman havaitsevaa subjektia. Aika ja avaruus eivät siis ole havaittajasta riippumattomien olioiden ominaisuuksia ja havainnon kohteet eivät ole olemassa havaittajasta riippumattomassa ajassa ja avaruudessa.

Tässä kirjoituksessa tarkastelen Lorne Falkensteinin (2004 [1995]) esittämää kritiikkiä ajan transsendentaaliselle ideaalisuudelle. Se keskittyy tietyn aikajärjestyksen, perättäisyyden, representointiin havainnossa. Tarkoitukseni ei ole osoittaa Falkensteinin tulkinnan Kantista olevan yleisesti väärä, vaan ainoastaan osoittaa tarkastelun kohteena olevan kritiikin ponnettomuuden. Samalla tarkastelen minkälaisia teoreettisia sitoumuksia kritiikkiin vastaamisesta seuraa uudemman filosofisen keskustelun valossa.

1. Argumentti ajan transsendentaalista reaalisuutta vastaan

Falkensteinin kritiikki keskittyy Mendelssohnin, Kantin aikalaisen, vasta-argumenttiin ajan ideaalisuutta vastaan ja Kantin siihen tarjoamaan vastaukseen. Lyhyesti esitettynä Mendelssohn argumentoi perättäisyyden havainnon tarkoittavan sitä, että ensin representoimme A:n ja sen jälkeen representoimme B:n. Tämä edellyttää muutosta ja muutos taas edellyttää ajan reaalisuutta. Kant (A36–37) hyväksyy vasta-argumentin, mutta korostaa ajan olevan reaalista ainoastaan aistimellisen intuition muotona. Kantin mukaan vasta-argumentti osoittaa siis ajan olevan empiirisesti reaalinen, muttei transsendentaalisesti reaalinen.

Falkensteinin mukaan Kantin vastaus ei kuitenkaan riitä Mendelssohnin vasta-argumentin kumoamiseksi. Tämä siksi, että se ei ota huomioon

representaation omaamisen edellyttävän subjektilta representoivassa tilassa olemista. Olettaen Kantin hyväksyvän ajatuksen representoivassa tilassa olemisesta, voimme määrittää henkilön olevan representoivassa tilassa T_a havaitessaan A:n ja representoivassa tilassa T_b havaitessaan B:n. Tällöin Falkensteinin käsitys Mendelssohnin argumentista voidaan konstruoida seuraavasti:¹

- 1) Voimme havaita A:n ja B:n perättäisyyden.
- 2) Voimme havaita A:n ja B:n perättäisyyden joss T_a edeltää T_b :a.
- 3) T_a edeltää T_b :a (seuraus premissistä 1 ja 2).
- 4) Jos T_a edeltää T_b :a, silloin representoivan henkilön tila muuttuu.
- 5) Representoivan henkilön tila muuttuu (seuraus premissistä 3 ja 4).
- 6) Henkilön tilan muuttuminen edellyttää ajan transsendentaalista reaalisuutta.
- 7) Aika on transsendentaalisesti reaalinen (seuraus kohdista 5 ja 6).

Edellä esitetyn argumentin ensimmäinen premissi vastaa Kantin omaa väitettä ajallisen järjestyksen representoinnista havainnossa.² Kant hyväksyisi todennäköisesti myös premissin 6, sillä se on analoginen empiirisen muutoksen ja ajan empiirisen reaalisuuden suhteen kanssa: siinä missä empiirisiin objekteihin liitetty muutos edellyttää Kantin mukaan ajan empiiristä reaalisuutta, on luontevaa ajatella, että Kantin mukaan transsendentaalisesti reaaliin objektiin liitetty muutos edellyttää ajan transsendentaalista reaalisuutta. Kohdat 3, 5 ja 7 ovat seurauksia aiemmista premissistä. Siten vasta-argumentin menestys riippuu ensisijaisesti premissistä 2 ja 4.

Toisen premissin taustalla on yleinen erottelu representaation sisällön (se mitä representoidaan) ja representaation välineen (miten representoidaan) välillä. Olennaista erottelussa on se, että representaation sisällön ja välineen

¹ Tämä versio eroaa uskoakseni Falkensteinin paikoin sekavasta argumentista vain esitystavan systemaattisuuden suhteen. Olennaisesti, argumentin kriittisimmät kohdat (premissit 2 ja 4) ovat sellaisia joihin Falkenstein itse eksplisiittisesti vetoaa.

² Kant (A30) kirjoittaa seuraavasti: ”Samanaikaisuus tai perättäisyys eivät nimittäin itse päätyisi havaintoon, mikäli ajan representaatio ei olisi perustana *a priori*. Se, että jotakin on yhdellä ja samalla hetkellä (samanaikaisesti) tai eri hetkinä (perättäin), voidaan representoida ainoastaan edellyttämällä aika.”

ominaisuuksien ei tarvitse olla samat. Mentaalinen tilamme (representaation väline) voi esimerkiksi representoida sisällön punaisesta kolmiosta, vaikka tila itsessään ei olisi punainen ja kolmiomainen. Toisaalta, vaikka representaation sisältö ja väline voivatkin omata eri ominaisuudet, ei niiden välttämättä tarvitse niin tehdä. Toinen premissi väittääkin, että kokemusten ja ne realisoivien representationaalisten tilojen ajallisen rakenteen suhteen näin ei ole; havaittu perättäisyys edellyttää representaatioiden perättäisyyttä ja päinvastoin. Viitataan tähän representaatioiden ajallisen rakenteen ja representaatioiden sisällön ajallisen rakenteen isomorfiaan *temporaalisen isomorfian teesinä* (Arstila 2016; Mölder 2014).

Neljännän premissin mukaan voimme sanoa henkilön representoivan tilan muuttuneen tilanteessa, jossa henkilön A:ta representoiva tila T_a edeltää B:tä representoivaa tilaa T_b . Yhdessä toinen ja neljäs premissi väittävät siis havaitusta aikajärjestyksestä seuraavan muutos havaitsijan tiloissa. Jos tämä pitää paikkaansa ja muutos edellyttää aikaa (premissi 6), niin havaittu aikajärjestys edellyttää transsendentaalisesti reaalista aikaa. Sen myötä Kantin teoria ajan transsendentaalisesta ideaalisuudesta on epätosi.

Tässä lohduttomassa tilanteessa Falkenstein ehdottaa Kantin voivan pyrkiä puolustamaan teoriaansa kiistämällä toisen premissin eli temporaalisen isomorfian teesin. Tällöin A:n ja B:n havaittu perättäisyys ei edellyttäisi T_a :n ja T_b :n perättäisyyttä, vaan A ja B sekä niiden välinen ajallinen järjestys representoitaisiin yhdessä ja samassa kompleksisessa tilassa (T_k). Koska perättäisyys representoitaisiin muuttumattomassa tilassa, ei perättäisyyden representointi edellytä transsendentaalisesti reaalista aikaa. Falkenstein pitää tätä puolustusta kuitenkin toivottomana, koska (i) se ei sovi yhteen Kantin muun teorian kanssa ja (ii) se on kumottavissa myös Kantin teoriasta itsenäisin perustein. Seuraavassa käsittelen näitä väitteitä tarkemmin.

2. Kantin teoria ja temporaalisen isomorfian teesi

Falkenstein vaikuttaa esittävän kaksi erillistä syytä sille, miksi temporaalisen isomorfian teesin kielto ei sovi yhteen Kantin muun teorian kanssa. Näistä ensimmäinen keskittyy tilaan T_k : kuinka A:n voidaan ajatella edeltävän B:tä, jos molemmat representoidaan samassa tilassa T_k ? Tähän Kant ei Falkensteinin mukaan tarjoa vastausta. Päinvastoin, tiedämme vain, että Kantin mukaan tämä ei voi tapahtua syy-seuraus -suhteen avulla, eikä

myöskään A:n ja B:n sisältöjen kautta (koska käsitys ajasta ei perustu empiiriseen sisältöön). Muita vaihtoehtoja Falkenstein ei pohdi.

Olettaen Falkensteinin väitteen pitävän paikkaansa – so. Kant ei tarjoa vastausta kuinka A:n voidaan ajatella edeltävän B:tä temporaalisen isomorfian teesin ollessa epätosi – ei tästä seuraa vielä Kantin teorian virheellisyys vaan ainoastaan riittämättömyys selittämään perättäisyyden havaitseminen. Siten Kantin teoria olisi ”pelastettavissa” täydentämällä sitä selityksellä miten perättäisyyden representoiminen havainnossa on mahdollista ilman, että sitoudutaan temporaalisen isomorfian teesiin. Tällainen selitys löytyy aikatietoisuuskeskustelusta.

Aikatietoisuutta koskevien teorioiden keskeisin tavoite on tarjota vastaus kysymykseen siitä, millainen kokemuksemme tulee olla, jotta voisimme omata kokemuksia ajallisesti ulotteisilta vaikuttavista ilmiöistä kuten muutos ja perättäisyys. Tätä keskustelua on motivoinut ajatus, että emme voi omata edellä mainittuja kokemuksia, jos kokemuksemme sisällöt rajautuvat pistemäiseen hetkeen tai minimaalisen lyhyeen ajanjaksoon. Jos kokemuksemme sisällöt olisivat tällä tavalla rajattuja, kokisimme ensin esimerkiksi A:n ja myöhemmin B:n. Emme kuitenkaan kokisi B:n seuraavan A:ta, sillä tämä edellyttää A:n olevan jollain tavoin läsnä kokemuksissamme, kun koemme B:n.

Edellä kuvatun päättelyn seurauksena osa filosofiista väittää, että emme omaa kokemuksia ajallisesti ulotteisista vaikuttavista ilmiöistä. Niinpä esimerkiksi Thomas Reid (1850) väittää, että tarkalleen ottaen emme havaitse liikettä. Useimmat filosofit kuitenkin ajattelevat meidän omaavan kokemuksia ajallisesti ulotteisilta vaikuttavista ilmiöistä. He ovat päätyneet kieltämään kokemuksemme sisältöjen olevan rajattuja pistemäiseen hetkeen. Toisin sanoen he väittävät kokemusten sisältöjen vaikuttavan ajallisesti levittäytyneiltä – ensimmäisen persoonan perspektiivistä katsottuna kokemukset kattavat ajallisesti ulotteisen intervallin. William Jamesia (1890) seuraten tätä ideaa kutsutaan *näennäisen nykyhetken doktriiniksi* (the doctrine of the specious present).³ Yksi koettu näennäinen nykyhetki kattaa mitä sillä hetkellä tapahtuu ja mitä juuri on tapahtunut.

³ Teesi koskee *näennäistä* nykyhetkeä, koska teesin mukaan koemme lyhyen ajanjakson, eikä tämä nykyhetki ole siten nykyhetki matemaattisessa mielessä (pistemäinen).

Näennäisen nykyhetken doktriinin taustalla on siis ajatus siitä, että kokemuksemme sisällöt ovat ajallisesti levittäytyneet. Tällöin yhden ”nykyhetken” sisältönä voi olla esimerkiksi kaksi eri aikana tapahtunutta ovenkoputusta. Koska kokemus ensimmäisestä ja toisesta ovenkoputuksesta ovat saman näennäisen nykyhetken osia, olemme tietoisia niistä yhdessä tietoisuudentilassa. Koska näennäinen ajanhetki on ajallisesti ulotteinen, voimme kokea koputusten tapahtuneen eri aikaan. Tämä tarkoittaa sitä, että voimme omata tilan T_k , joka representoi molemmat koputukset sekä niiden välisen perättäisyyden. Tällä tavoin näennäisen nykyhetken doktriini auttaa selittämään, miten voimme omata kokemuksen perättäisyydestä.

Doktriini ei kuitenkaan selitä näennäisen nykyhetken luonnetta sinänsä, vaan sille pitää antaa erillinen selitys. Tarjotut selitykset jakautuvat kahteen. *Intentionalistisen* tai *retentionalistisen mallin* mukaan kokemuksemme sisältö on ajallisesti ulotteinen, mutta kokemusaktilta itseltään puuttuu ajallinen ulotteisuus.⁴ Kokemusaktimme ovat käytännössä pistemäisiä mutta kokemusten sisältö vaikuttaa meille ensimmäisen persoonan perspektiivistä käsin ajallisesti ulotteiselta. *Ekstensionaalinen malli* taas pitää sekä kokemusten sisältöä että kokemusakteja ajallisesti ulotteisina. Sisältö ja kokemusakti kattavat saman ajanjakson ja kokemusaktin ajallinen ulotteisuus selittää osaltaan sitä, miten kokemusten sisällöt voivat olla ajallisesti ulotteisia.

Edellä kuvattu aikatietoisuusteorioita koskeva keskustelu on kahdella tavalla olennainen Falkensteinin argumentoinnille. Ensinnäkin, kahden asian representoinnin peräkkäisyys ei vielä riitä peräkkäisyyden representointiin. Tai kuten James (1890) asian ilmaisee, ”a succession of feelings, in and of itself, is not a feeling of succession.” Vaikka Kant siis hyväksyisi temporaalisen isomorfian teesin, tulee hänen silti tarjota selitys, kuinka perättäisyys representoidaan havainnossa. Siten Falkensteinin kritiikki Kantin mahdollista teesin kieltoa vastaan ei koske teesin totuutta tai epätotuutta, vaan kuvaa ainoastaan Kantin teorian selityksellisen puutteellisuuden.

Toisekseen, intentionalismi – toinen kahdesta näennäisen nykyhetken doktriiniin sitoutuvasta teoriasta – kieltää temporaalisen isomorfian teesin, sillä intentionalismissa representoivan tilan ajalliset ominaisuudet eivät vastaa tilan representoiman sisällön ajallisia ominaisuuksia. Intentionalistiset teoriat

⁴ Osassa teorioista kokemusakti (tai sen realisoiva hermostollinen perusta) voi olla ajallisesti ulotteinen, mutta tämä ulotteisuus ei näytele roolia sisältöjen ajallista ulotteisuutta selitettäessä.

eroavat toisistaan esimerkiksi sen mukaan, kuinka näennäisen nykyhetken ajallinen rakenne niissä toteutetaan. Edmund Husserlin (1991) varhaisempi teoria nojaa retention käsitteeseen, C. D. Broad (1938) taas kehitti teoriaa ajallisista presentaation tavoista (temporal modes of presentation) ja Rick Grush (2007) nojaa kokemuksen sisältöön liitettäviin ajallisiin tunnuksiin (A tapahtui t_1 , B tapahtui t_2 , ...). Huolimatta jokaisen vaihtoehdon ongelmallisuudesta (Arstila 2016), nämä esimerkit osoittavat temporaalisen isomorfian teesin kieltämisen olevan yleisesti kannatettu mahdollisuus.

Falkenstein esittää myös toisen syyn sille, miksi temporaalisen isomorfian teesin kieltö ei sovi yhteen Kantin teorian kanssa. Sen mukaan Kantin teoriassa tietomme ajallisista relaatioista perustuu viime kädessä tosideikkaan, että tietyt representaatiot todella tapahtuvat ja on representoitu intuitiossa ennen toisia representaatioita.⁵ Toisin sanoen, Falkenstein esittää yksikantaisesti Kantin väittäneen (i) representaatioiden tapahtuvan ajallisessa järjestyksessä ja (ii) tietomme ajallisista järjestyksistä edellyttävän representaatioiden ajallista järjestystä.

Molemmat väitteet ovat uskoakseni ongelmallisia, mutta ennen niiden tarkastelua on hyvä korostaa, että Falkenstein ei ole ainoa Kant-oppinut, jonka mukaan Kant on sitoutunut temporaalisen isomorfian teesiin. (Muiden perustelut asialle voivat tosin erota Falkensteinin tarjoamista.) Peter Strawson (1990, 50) esimerkiksi kirjoittaa Kantia ja ajan *a priori* luonnetta pohtiessaan seuraavasti: “they [moments of experiences] are moments none the less, having their place in a temporal procession, followed and preceded by others. ... So the idea of experience in general seems to be truly inseparable from that of a temporal succession of experience.” Toisin sanoen Strawson väittää kokemusten omaavan paikan ajallisessa etenemisessä, olettaen siten kokemusten eli representaation välineiden seuraavan toisiaan. Strawson ei kuitenkaan jatka pohtimalla mitä tästä seuraa kysymykselle ajan transsendentaalisesta reaalisuudesta.

Osa Kantia koskevista kirjoituksista taas voidaan tulkita tavalla, joka vähintäänkin sopii yhteen sen kanssa, että Kantin ajatellaan sitoutuneen

⁵ Falkenstein (2004, 343) kirjoittaa: ”The fact is, as the *Critique* maintains, that our knowledge of temporal relations is ultimately grounded in the fact that certain representations actually occur and are presented in intuition before certain other representations. (This is just what it means for time to be a form of intuition.) This original succession can then be taken up in the unity of apperception and represented in a single, momentary, complex thought.”

temporaalisen isomorfian teesiin. Esimerkiksi (sankarimme) Olli Koistisen (2013, 23–24) mukaan Kant ”huomauttaa sen, että jokin representoidaan ennen toista tai samanaikaisesti kuin toinen tarkoittavan, että näiden ajatellaan valtaavan *ajassa* saman paikan tai eri ajat.” Tässä kohta ”jokin representoidaan ennen toista” on tulkittavissa ajatuksena, että A:n ja B:n representoivat tilat todella tapahtuvat perättäin, eikä vain niin että A ja B näyttäytyvät representaatioissa perättäisinä.

Nämä tulkinnat Kantista voivat olla oikeutettuja Kantin kirjoitusten pohjalta, vaikka ne mahdollisesti johtaisivatkin ajan transsendentaalisen ideaalisuuden kiistämiseen. (Tämä riippuu osaksi myös premissistä 4, jota käsitellen seuraavassa osiossa.) Puhtaan järjen kritiikistä löytyy kuitenkin myös kohtia, joissa Kant vaikuttaa irtisanoutuvan temporaalisen isomorfian teesistä. Mendelssohnin kritiikkiin liittyen Kant (A37, alaviite) esimerkiksi kirjoittaa: ”Voimme kyllä sanoa, että representaationi seuraavat toisiaan, mutta tämä tarkoittaa vain, että olemme niistä tietoisia aikasarjassa eli sisäisen aistin muodon mukaisesti. Aika ei tämän vuoksi ole mitään itsessään eikä myöskään olioihin objektiivisesti liittyvä määre.” Toisin sanoen Kantin mukaan representaatioiden aikajärjestys koskee ainoastaan sitä, miten ne aistimme, eikä representaatioiden aikajärjestystä sinällään. Siten tämän kohdan valossa Kant vaikuttaa kieltävän toisiaan seuraavien representaatioiden transsendentaalisen reaalisuuden.⁶

Väite siitä, että tietomme representaatioiden ajallisista relaatioista edellyttää representaatioiden perättäisyyttä on yhtä lailla ongelmallinen. Tämä siksi, että tietomme representaatioiden aikajärjestyksestä pohjautuu oletettavasti siihen, kuinka aikajärjestyksen havaitsemme. Toisin sanoen tietomme pohjautuu havaintojen sisältöön eikä havaintosisällöt representoivien tilojen ominaisuuksiin sinällään. Tämän vuoksi keskeistä tiedolle representaatioiden ajallisesta järjestyksestä on se, kuinka asiat esitetään, eikä kuinka ne itsessään ovat. Täten meillä voi olla tietoa representaatioiden A ja B perättäisyydestä, vaikka tuon mahdollistaa kompleksinen tila T_k .

⁶ Ilmaus ”toisiaan seuraavat representaatiot” voidaan tulkita kahdella tavalla. Toisaalta ilmaus voi tarkoittaa sitä, että ”seuraaminen” on dynaaminen tapahtuma, joka on riippuvainen ajan etenemisestä. Toisaalta ilmaus voi tarkoittaa sitä, että kaksi representaatiota ovat ajallisessa järjestyksessä (perättäisyys) ilman ajan etenemistä. Seuraavassa osiossa käydyn keskustelun perusteella Kant voisi sitoutua jälkimmäiseen tulkintaan.

Yhteenvedona edellisestä voimme siis todeta temporaalisen isomorfian teesin kiellon olevan mahdollista ylipäätään – kuten useat aikatietoisuusteorioiden tekevätkin – ja olevan mahdollista myös Kantin kirjoitusten valossa.

3. Temporaalisen isomorfian teesin kieltö ei pelasta ajan transsendentaalista ideaalisuutta

Falkensteinin mukaan on olemassa kuitenkin myös Kantin teoriasta riippumattomia syitä hylätä temporaalisen isomorfian teesin kieltoon perustuva ratkaisuehdotus Mendelssohnin argumenttiin ajan transsendentaalista ideaalisuutta vastaan. Falkensteinin argumentti menee seuraavasti: Jos A edeltää B:tä, niin on ollut hetki, jolloin henkilö kykeni representoimaan A:n mutta ei kyennyt representoimaan B:tä. Tällöin henkilö oli tilassa T_a . Myöhemmin henkilö on kompleksisessa tilassa T_k , jossa hän representoi A:n, B:n ja niiden perättäisyyden. Mutta jos henkilö voi jossain vaiheessa olla ainoastaan tilassa T_a ja jos hän on myöhemmin tilassa T_k , silloin henkilön on täytynyt käydä läpi muutos representationaalisessa tilassaan. Siten Mendelssohnin vasta-argumentti pätee.

Aiemman argumentin tapaan Falkensteinin uudempi argumentti nojaa kahden erilaisen representoivan tilan olemassaoloon. Uudemmassa argumentissa pelkästään B:n representoiva tila T_b on kuitenkin korvautunut kompleksisella tilalla T_k . Alkuperäisen ja uudemman argumentin ero ei siis koske argumentin rakennetta, vaan ainoastaan sitä kuinka jälkimmäinen representoiva tila ja sen sisältö tulisi tulkita. Koska uudessa tulkinnassa hylätään temporaalisen isomorfian teesi, voidaan neljäs premissi muotoilla tavalla, jonka Kant voisi hyväksyä (ts. T_b korvataan T_k :lla). Siten vasta-argumentin menestys (ja ajan transsendentaalisen ideaalisuus) riippuu premissistä neljä:

4) Jos T_a edeltää $T_{b/k}$:a, silloin representoivan henkilön tila muuttuu.

Tätä väitettä Kant ei uskoakseni olisi valmis hyväksymään – eikä hänen niin tulisi tehdä, sikäli kuin hän haluaa kannattaa ajan transsendentaalista ideaalisuutta. Kantilla vaikuttaisi olevan kaksi vaihtoehtoa premissin kieltämiseksi. Ensinnäkin, Kant voi kiistää etujäsenen ilmaiseman väitteen mielekkyyden: koska aika on transsendentaalisesti ideaalinen, emme voi sanoa

mitään T_a :n ja $T_{b/k}$:n ajallisesta järjestyksestä. Kant voi toisin sanoen väittää, että kieltäessämme ajan transsendentaalisen reaalisuuden väitämme samalla, ettei tilojen välillä ole ajallista relaatiota ylipäätään. Tällöin emme voi sanoa edeltävätkö kaksi tilaa toisiaan tai ovatko ne samanaikaiset, ja siten etujäsen ei voi olla tosi tai epätosi. (Sama koskee Falkensteinin muotoilua uudesta argumentistaan, jossa hän kirjoittaa A:n edeltävän B:tä, ei siis A:n representaation edeltävän B:n representaatiota.) Tämän ratkaisun ongelmana, tai ainakin mahdollisena ongelmana, on kuinka selittää numeerisesti erillisten mutta sisällöltään identtisten representationaalisten tilojen ero. Jos esimerkiksi omaamme tilan T_a , tilan T_b ja taas tilan T_a , niin mikä erottaa kaksi tilan T_a ilmentymää? Yksi mielekäs vaihtoehto olisi se, että ne tapahtuvat eriaikaan. Kieltäessään neljännen premissin etujäsenen Kant ei voisi kuitenkaan tällä tavoin vastata ja kysymys jäisi vaille vastausta.

Toinen vaihtoehto neljännen premissin kieltämiseksi on hyväksyä etujäsen, mutta kieltää takajäsen. Tämä ratkaisu nojaa siis muutoksen mahdollisuuden kieltoon. Se sopii myös hyvin yhteen Kantin Mendelssohnille antaman vastauksen kanssa. Siinä Kant hyväksyi empiirisesti reaalisuuden muutoksen todellisuuden sekä sen edellyttävän empiirisesti reaalista aikaa. Koska muutos on tällä tavoin alisteinen ajan reaalisuudelle, kieltäessään ajan transsendentaalisen reaalisuuden Kant mitä todennäköisemmin kiistäisi myös muutoksen transsendentaalisen reaalisuuden. Täten hän ei hyväksyisi väitettä, että Falkensteinin esimerkin henkilön representoiva tila muuttuu, vaikka T_a edeltäisikin $T_{b/k}$:a. Sikäli kuin Kant todella vastaisi Falkensteinille tällä jälkimmäisellä tavalla, väittäisi hän siis representoivan tilan voivan olla erilainen eri hetkinä, ilman että voimme sanoa tilan muuttuneen.

Jotta voisimme ymmärtää tätä jälkimmäistä väitettä, ja miksi se ei ole niin epäuskottava kuin miltä se ehkä ensikuulemalta vaikuttaa, tulee meidän tarkastella John Ellis McTaggartin (1908) tekemää erottelua kahdenlaisesta tavasta ymmärtää ajan luonne. Niin sanottu *A-teoria* kuvaa aikaa ja siinä olevia ilmiöitä 'menneisyyden', 'nykyhetken' ja 'tulevaisuuden' termein. A-teoria on dynaaminen teoria, jossa nykyhetki muuttuu alati; nykyhetkessä tapahtuvat ilmiöt ovat seuraavassa hetkessä osa menneisyyttä. *B-teoria* taas kuvaa aikaa ja siinä olevia ilmiöitä relationaalisin termein kuten 'aikaisemmin kuin', 'samanaikainen' ja 'myöhempi kuin'. Näiden ajallisten relaatioiden määrittäessä kaikkien ilmiöiden paikat pysyvästi kaikkina ajankohtina (esim.

jos A edeltää B:tä, niin näin on tarkasteltaessa ilmiöitä minä ajanhetkenä tahansa) on B-teoreettinen aikakäsitys staattinen: aika ei kulu eikä virtaa.

McTaggart argumentoi lisäksi ajan todellisuuden tarkoittavan sitä, että muutos on mahdollista. B-teoria ei kuitenkaan mahdollista muutosta, sillä tässä viitekehyksessä jokaisen tapahtuman ajallinen paikka on ajallisesti määrätty ja siten mikään ei muutu ajan funktiona. Jonkin asian muutoksen suhteen voimme sanoa sillä olevan ensin ominaisuus A ja myöhemmin ominaisuus B, mutta emme voi sanoa asian muuttuneen tai tulleen A:sta B:ksi. Todellinen muutos onkin McTaggartin mukaan mahdollista ainoastaan A-teoriassa, jossa aika virtaa ja asiat voivat ajan edetessä muuttua yhdestä toiseksi. McTaggart pitää kuitenkin A-teoriaa ristiriitaisena, sillä teoriassa ilmiöt omaavat ristiriitaisia ominaisuuksia (mennyt, nykyinen ja tuleva). Siten A-teoria on kestävä kanta eikä todellinen muutos ole mahdollista. Lopputulemana McTaggart totesikin ajan olevan ei-reaalinen (un-real).

McTaggartin esittämä erottelu kahdesta aikakäsityksestä sekä hänen perustelunsa ajan ei-reaalisuudelle ovat leimanneet ajanfilosofiaa viimeisen vuosisadan ajan.⁷ Ottamatta kantaa kummankaan teorian puolesta on todettava B-teorian olevan laajalti kannatettu muun muassa sen vuoksi, että A-teoriaan liittyvää ajatusta ajan virtaamisesta pidetään usein ristiriitaisena Einsteinin suppean suhteellisuusteorian kanssa.

Falkensteinin jälkimmäisen argumentin ja Kantille postuloimani vastauksen kannalta McTaggartin ajanfilosofian keskeinen anti on siinä, että B-teorian viitekehyksessä voimme puhua kahden asian perättäisyydestä ilman, että hyväksymme muutoksen olemassaoloa. Toisin sanoen voimme hyväksyä henkilön representoivan A:n ja jossain myöhemmässä vaiheessa B:n (tai omaavan kompleksisen representationaalisen tilan T_k) ilman, että hyväksymme henkilön representationaalisen tilan tarkkaan ottaen muuttuneen. Tällöin voimme hyväksyä neljännen premissin etujäsenen (T_1 edeltää $T_{2/k;a}$) hyväksymättä sen takajäsentä (joka väittää representoivan henkilön tila muuttuvan) ja siten vastoin Falkensteinia olla hyväksymättä neljättä premissiä.

⁷ Tämä keskustelu on erittäin laajaa ja monipolvista. Koska tarkoituksenani on osoittaa kuinka Kant voisi kiistää neljännen premissin, eikä argumentoida tietyn aikakäsityksen (tai A- ja B-teorian version) puolesta, alla oleva keskustelu seuraa McTaggartin käsitystä ajanfilosofiasta.

McTaggartin tekemä erottelu A-teorian ja B-teorian välillä tuo esille myös Falkensteinin argumentin kehämäisyyden: Falkenstein argumentoi ajan olevan transsendentaalisesti reaalin nojaamalla muutokseen henkilön representoivissa tiloissa. Muutos henkilön representoivissa tiloissa on kuitenkin mahdollista vain, jos hyväksymme jo ajan reaalisuuden (A-teorian). Siten Falkensteinin argumentti olettaa implisiittisesti sen mitä pyrkii todistamaan, ja jos emme lähtökohtaisesti hyväksy ajan transsendentaalista reaalisuutta, ei argumentti ole menestyksenkäs.

4. Yhteenvetoa

Tässä kirjoituksessa olen tarkastellut Falkensteinin kritiikkiä Kantin käsitykselle ajan transsendentaalisesta ideaalisuudesta. Tuo kritiikki pohjautuu Kantin ajatukseen siitä, että voimme representoida havainnoissamme (eli ei vain ajatuksissamme) ajallisia järjestyksiä, kuten samanaikaisuus ja perättäisyys. Falkenstein kiinnittää asiassa huomiota siihen, että havainnot edellyttävät representoivassa tilassa olemista. Tämän seurauksena ajan transsendentaalinen ideaalisuus vaikuttaa tulevan haastetuksi kahdella rintamalla.

Ensimmäinen koskee mielenfilosofian alaa: voiko perättäisyyden kokemisen selittää uskottavasti ja vieläpä tavalla joka sopii yhteen Kantin kirjoitusten kanssa? Perättäisyyden kokemisen selittämisen suhteen olen argumentoinut, vastoin Falkensteinin kantaa, että uskottavan selityksen tarjoaminen (i) ei ole riippuvainen temporaalisen isomorfian teesistä ja (ii) ei ole mahdollista vain olettamalla kokemusten perättäisyys. Esitin myös, että haasteeseen vastaaminen on mahdollista sikäli kuin Kant olisi valmis hyväksymään intentionalistisen mallin aikatietoisuudesta. Ekstensionaalinen malli taas on huomattavasti sovitettavissa yhteen ajan transsendentaalisen ideaalisuuden kanssa, sillä tässä mallissa havaintoakti on todella ulotteinen ajassa.

Toinen haaste koskee metafysiikkaa: miten selittää kahden representoivan tilan ajallinen erillisyyks ilman, että myönnämme representoivan tilan muuttuneen? Myös tässä uudempi filosofinen keskustelu vaikuttaisi tarjoavan keinon vastata haasteeseen. Tämä tarkoittaisi McTaggartin ajan reaalisuutta vastaan esittämän argumentin ja B-teorian hyväksymistä. Olennaista tässä on se, että aikajärjestys (tai ainakin ajallinen

erillisyyttä) on transsendentaalisesti reaallinen, mutta aika itsessään on transsendentaalisesti ideaalinen. Toisaalta Kant voi myös mahdollisesti kiistää koko haasteen ja väittää, ettei puhe representoivien tilojen ajallisesta erillisyydestä ole mielekäästä. Tällöin aikajärjestyksen transsendentaalinen ideaalisuus on seurausta ajan transsendentaalisesta ideaalisuudesta. Tässä tapauksessa Kantin tulisi tarjota selitys sille, mikä erottaa kaksi numeerisesti erillistä, mutta sisällöltään identtistä representationaalista tilaa.⁸

Turun yliopisto

Kirjallisuus

- Arstila, Valtteri (2016). ”The Time of Experience and the Experience of Time”, teoksessa Mölder Bruno, Arstila Valtteri, & Øhrstrøm Peter (toim.), *Philosophy and Psychology of Time*. Dordrecht, The Netherlands: Springer, 163–186.
- Broad, C. D. (1938). *An Examination of McTaggart’s Philosophy*. Cambridge: Cambridge University Press.
- Falkenstein, Lorne (2004 [1995]). *Kant’s Intuitionism: A Commentary on the Transcendental Aesthetic*. Toronto: University of Toronto Press.
- Grush, Rick (2007). ”Time and experience”, teoksessa Müller, Thomas (toim.), *Philosophie der Zeit*. Frankfurt: Klosterman, 1–18.
- Husserl, Edmund (1991). *On the Phenomenology of the Consciousness of Internal Time (1893–1917)*. Dordrecht: Kluwer Academic.
- James, William (1890). *The Principles of Psychology*. New York: Dover.
- Kant, Immanuel (2013). *Puhtaan järjen kritiikki*. Suom. Markus Nikkarla ja Kreetta Ranki. Helsinki: Gaudeamus. [A/B]
- Koistinen, Olli (2013). ”Johdatus Puhtaan järjen kritiikkiin”, teoksessa Immanuel Kant, *Puhtaan järjen kritiikki*. Helsinki: Gaudeamus, 13–46.
- McTaggart, J. E. (1908). ”The Unreality of Time”, *Mind* 17, 456–473.
- Mölder, Bruno (2014). ”Constructing Time: Dennett and Grush on Temporal Representation”, teoksessa Valtteri Arstila & Dan Lloyd (toim.), *Subjective Time: The Philosophy, Psychology, and Neuroscience of Temporality*. Cambridge, MA: The MIT Press, 217–238
- Reid, Thomas (1850). *Essays on the Intellectual Powers of Man*. Cambridge: John Bartlett.
- Strawson, Peter (1990). *The Bounds of Sense: An Essay on Kant’s Critique of Pure Reason*. London: Routledge.

⁸ Haluan kiittää Olli Koistista vuosien varrella käydyistä lukuisista inspiroivista keskusteluista aiheen tiimoilta, sekä Arto Repoa tarkoista ja rakentavista kommentteista kirjoitusta viimeistellessäni.

Kaappikantilaisuutta?

Jussi Haukioja

Tarkastelen tässä artikkelissa lyhyesti yhteneväisyyksiä, joita voidaan nähdä Rae Langtonin Kant-tulkinnan (Langton 1998) ja erään viimeaikaisen käsitteiden luonnetta ja niiden omaksumisen edellytyksiä koskevan teorian välillä. Tulen esittämään, että tämä käsitteiden *globaalina vasteriippuvuutena* tunnettu teoria, jota olen itsekin puolustanut sekä väitöskirjassani (Haukioja 2000) että eräissä myöhemmissä teksteissäni, johtaa yleiseen näkemykseen, joka muistuttaa varsin paljon Langtonin Kantille attribuoimaa näkemystä. Tulen näin ollen ulos kaapista, jos en varauksetta kantilaisena, niin ainakin kantilaisena *tietyiltä osin, niin kuin Langton on Kantia tulkinnut*.

Artikkeli perustuu oleellisilta osin esitelmään, jonka pidin 3.4.2003 Olli Koistisen järjestämässä Kantin filosofiaa käsittelevässä yksipäiväisessä seminaarissa Turun yliopiston filosofian laitoksella. Haluan näin sekä onnitella Ollia syntymäpäivänsä johdosta että jälkikäteen kiittää mahdollisuudesta puhua kyseisessä seminaarissa.

1. Langtonin Kant-tulkinnasta

Kirjassaan *Kantian Humility* (1998) Rae Langton pyrkii esittämään uuden tulkinnan eräistä Kantin metafysiikan ja epistemologian keskeisistä teeseistä. Langtonin päätavoitteena on ratkaista ongelma, joka näyttää syntyvän, kun teemme kaksi Kantiin liitettyä olettamusta: kun väitämme toisaalta, että *olioita sinänsä* on olemassa, ja toisaalta, että emme voi tietää niistä mitään. Langton muotoilee ongelman seuraavasti (Langton 1998, 7):¹

- (K1) On olemassa olioita sinänsä.
- (K2) Oliot sinänsä ovat fenomenaalisten ilmentymien syitä.
- (K3) Meillä ei ole tietoa olioista sinänsä.

¹ Suomennokset ovat omiani; olen tosin pyrkinyt siihen, että käyttämäni terminologia olisi yhtenevää Koistisen (2008) ja hänen johtamansa työryhmän *Puhtaan järjen kritiikki* -käännöksessä (Kant 2013 [1781/1787]) käyttämän kanssa.

Edellä mainittu ongelma on suoraan luettavissa tästä muotoilusta: (K3):sta näyttäisi seuraavan, että emme voi tietää väitteitä (K1) ja (K2). Mikäli Kant siis hyväksyi väitteet (K1)–(K3) – ja väitti meidän voivan tietää ne – näyttää hän syyllistyneen selvään ristiriitaan.

Langton ratkaisee ongelman tulkitsemalla puheen *olioista sinänsä* puheeksi olioiden *sisäisistä* (intrinsicistä) *ominaisuuksista* (Langton 1998, 13). Tällöin väitteet (K1)–(K3) voidaan tulkita seuraavalla tavalla:

- (L1) On olemassa olioita sinänsä, ts. olioita joilla on sisäisiä ominaisuuksia.
- (L2) Oliolla sinänsä on myös relationaalisia ominaisuuksia: nämä ovat kausaalisia voimia jotka konstituivat fenomenaliset ilmentymät.
- (L3) Meillä ei ole tietoa olioiden sisäisistä ominaisuuksista.

Sisäisillä ominaisuuksilla Langton tarkoittaa tässä ominaisuuksia, jotka ovat yhteensopivia olion *yksinäisyyden* kanssa, ts. joita oliolla voi olla vaikka muita olioita ei olisi olemassa; tätä näkemystä sisäisten ja ulkoisten ominaisuuksien välisestä erottelusta Langton on myös puolustanut erikseen yhdessä David Lewisin kanssa (Langton & Lewis 1998).

Langton pyrkii kirjassaan sekä täsmentämään että puolustamaan näkemystä. Keskeiseksi nousevat seuraavat ”kolme kantilaista teesiä”, joita Langton varsin seikkaperäisesti puolustaa sekä Kant-tulkintana että itsenäisinä filosofisina väittäminä:

Erottelu (Distinction): Oliot sinänsä ovat substansseja, joilla on sisäisiä ominaisuuksia; fenomeenit ovat substanssien relationaalisia ominaisuuksia. (Langton 1998, 20; 34–40.)

Reseptiivisyys (Receptivity): (Ihmisen) tieto riippuu aistimuksellisuudesta; aistimuksellisuus on reseptiivistä: voimme saada tietoa oliosta vain sikäli kuin se vaikuttaa meihin. (Langton 1998, 23; 43–47.)

Nöyryys (Humility): Meillä ei ole tietoa substanssien sisäisistä ominaisuuksista. (Langton 1998, 21; 41–43.)

Langtonin näkemyksen mukaan Kant ajatteli, että *Nöyryys* on seuraus *Reseptiivisyydestä*. Keskeiseksi haasteeksi hänen tulkinnalleen muodostuu ymmärtää, miksi näin olisi. Ensi näkemältä ei ole lainkaan selvää, miksi se, että oliota koskeva tieto edellyttää niiden vaikuttavan meihin, estäisi meitä saamasta jollain epäsuoralla tavalla tietoa kyseisten olioiden sisäisistä ominaisuuksista. Langtonin yksi pääteema onkin näyttää, mitä lisäpremissiä tarvitaan, jotta *Reseptiivisyydestä* todella voitaisiin johtaa *Nöyryys*, ja löytää Kantin teksteistä evidenssiä tulkinnan tueksi.

Avainpremissi on Langtonin mukaan seuraava:

Palautumattomuus (Irreducibility): Substanssien välisiä relaatioita ja niiden relationaalisia ominaisuuksia ei voida palauttaa niiden sisäisiin ominaisuuksiin. (Langton 1998, 109.)

Jos olioiden relationaalisia ominaisuuksia – erityisesti niiden kausaalisia voimia – ei voida palauttaa niiden sisäisiin ominaisuuksiin, seuraa Langtonin mukaan vääjäämättömästi, että sisäiset ominaisuudet ovat kausaalisesti inerttejä; tämä huomio on avain siihen, miksi *Nöyryys* seuraa *Reseptiivisyydestä*.

En aio tässä kirjoituksessa tarkastella Langtonin Kant-tulkinnan pätevyyttä – pelkään jo edellä esittämässäni yhteenvedossa tehneeni tarkkaavaiselle (ja Kantinsa tuntevalle) lukijalle piinallisen selväksi, että ylittäisin yrittäessäni oman kompetenssin rajat. Sen sijaan esitän, että eräs viimeaikainen semantiikkaa ja käsitteiden teoriaa koskeva suuntaus, jota itsekin olen puolustanut, rinnastuu mielenkiintoisella tavalla Langtonin Kantiin liittämien ajatusten kanssa.

2. Vasteriippuvuus

Käsitteiden *globaali vasteriippuvuus* on teesi, jonka Philip Pettit johtaa niin sanotun säännön seuraamisen ongelmaan esittämästään ratkaisusta.² En aio tässä yhteydessä mennä sen enempää säännön seuraamisen ongelman kuin

² Tarkemmin ottaen säännön seuraamisen ongelmaan niin kuin Kripke (1982) sen esittää. Pettit on muotoillut sekä kyseisen ratkaisun että vasteriippuvuusteeseinsä monissa eri kirjoituksissaan hieman eri tavoin; keskeisiä tekstejä ovat esim. Pettit 1991 ja Pettit 1993. Olen itse puolustanut Pettitin ratkaisua eräitä vasta-argumentteja vastaan (Haukioja 2005).

vasteriippuvuuden eri muotoilujen yksityiskohtaiseen erittelyyn;³ vasteriippuvuusteetin voi intuitiivisella (ja samalla vähemmän täsmällisemmällä) tasolla nähdä seuraavan tietyistä käsitteiden omaksumista koskevista yleisemmistä olettamuksista.

Pettitin argumentaatiota ja samalla globaalin vasteriippuvuusteetin sisältöä voidaan lähestyä tarkastelemalla kysymystä: miten voimme omaksua käsitteitä? Käytännössä opimme usein uuden käsitteen määritelmän tai muun kielellisen kuvauksen perusteella, mutta on ilmeistä, että *kaikki* käsitteiden omaksuminen ei voi perustua prosessiin, joka olettaa käsitteen omaksujan jo hallitsevan muita käsitteitä. Kutsutaan käsitteitä, jotka voidaan omaksua nojaamatta jo aiemmin hallittuihin käsitteisiin, *semanttisesti perustaviksi* käsitteiksi. Miten siis voimme omaksua semanttisesti perustavia käsitteitä? Vaikuttaa uskottavalta esittää, että käsitteiden oppiminen *esimerkkien* perusteella on ainakin yksi tällainen tapa.⁴

Miten sitten voimme omaksua käsitteen esimerkeistä? Erityisesti: miten voimme *äärellisen* esimerkkijoukon perusteella omaksua käsitteen, jonka ala on määräytynyt (ainakin potentiaalisesti) *äärettömässä* joukossa mahdollisia yksittäistapauksia? (Tässä kohtaa yhteys säännön seuraamisen ongelmaan tulee ilmeiseksi.) Pettitin mukaan yksi ilmeinen edellytys on, että kyseisen esimerkkijoukon jäsenten täytyy *näyttäytyä toistensa kanssa samankaltaisina* käsitteen omaksujalle. Toisin sanoen joukon jäsenillä täytyy olla dispositio vaikuttaa toistensa kanssa samankaltaisina käsitteen omaksujille, tai vastaavasti: käsitteen omaksujalla täytyy olla dispositio kokea esimerkkijoukon jäsenet keskenään samankaltaisina. Omaksuessaan uuden käsitteen henkilö kiinnittää huomiota tähän samankaltaisuuteen ja alkaa *ekstrapoloida* alkuperäisen esimerkkijoukon ulkopuolelle: myös uudet yksittäistapaukset alkavat näyttäytyä joko relevantisti samankaltaisina tai erilaisina.

Tämä ei kuitenkaan vielä yksistään riitä käsitteen hallintaan. Käsitteen käytölle on olemassa *korrektiusehdot*. Pelkät ekstrapolaatio-dispositiot voivat suoraan kertoa meille sen, mitkä yksittäistapaukset *näyttävät* käsitteen käyttäjistä kuuluvan käsitteen alaan, mutta eivät sitä, mitkä *oikeasti* kuuluvat

³ Jälkimmäisestä ks. Haukioja 2013.

⁴ En aio tässä yhteydessä yrittää todistaa, että esimerkkeihin perustuva käsitteiden omaksuminen olisi *ainoa* mahdollinen tapa, joka ei edellytä omaksujan jo hallitsevan käsitteitä, mutta pidän varsin uskottavana, että ainakaan *meidän* kaltaisillemme olioille (ts. ihmisille aktuaalisen maailman kykyineen ja rajoituksineen) muita tapoja ei ole.

sen alaan: joskus käsitteen käyttäjällä on dispositio luokitella *väärin*. Tässä kohdassa dispositioteoretikko nojaa yleensä *ideaaleihin, normaaleihin* tai *suotuisiin* olosuhteisiin: käsitteen alaan kuuluvat ne yksittäistapaukset, joiden tuon käsitteen (kompetentti) käyttäjä *suotuisissa* oloissa arvioisi kuuluvan sen alaan. Ratkaiseva kysymys nyt on, miten suotuisat olosuhteet määritellään.⁵ Tähän kysymykseen en voi tässä yhteydessä tarttua: nyt mielenkiintoni kohdistuu siihen, mitä seuraa, *jos* tähän kysymykseen voidaan antaa tyydyttävä vastaus.

Globaalin vasteriippuvuusteesin tässä yhteydessä oleellinen piirre tuli jo esille edellisessä kappaleessa. Jos kaikki semanttisesti perustavat käsitteemme ovat vasteriippuvia, ts. jos niiden ala määräytyy edellä luonnostellulla tavalla, pätee seuraava periaate: jotta tietty käsitteemme voisi viitata johonkin ominaisuuteen tai olioluokkaan, sen alaan kuuluvilla olioilla tulee olla dispositio näyttäytyä meille samankaltaisena suotuisissa olosuhteissa, ja nimenomaan tämä dispositio määrää, mitkä oliot kuuluvat käsitteen alaan ja mitkä eivät. Ne erottelut, joita käsitteillämme teemme, perustuvat siis pohjimmiltaan *kokonaan* käsitteen alaan kuuluvien olioiden yhteisiin *relatiionaalisiin* ominaisuuksiin.⁶

Nyt yhteys Langtonin Kant-tulkintaan alkaa hahmottua. Voimme itse asiassa johtaa edellä hahmotellusta seuraavat, Langtonin *Reseptiivisyyttä* ja *Nöyryyttä* vastaavat teesit:

Reseptiivisyys?: (Ihmisen perustavien) käsitteiden ala määräytyy aistimusten avulla; aistimukset ovat reseptiivisiä: voimme tehdä (perustavilla) käsitteillämme erotteluja vain sikäli kuin eri käsitteiden alaan kuuluvilla olioilla on meihin erilaisia vaikutuksia.

Nöyryys?: Emme voi käsitteillämme viitata substanssien sisäisiin ominaisuuksiin.

⁵ Kripke (1982) argumentoi tunnetusti, että suotuisia oloja *ei* voida määritellä ilman kehämäisyyttä, monet, esim. Pettit (1993) ovat eri mieltä. Olen puolustanut Pettitin tapaa määritellä suotuisat olosuhteet toisaalla (Haukioja 2007; Haukioja 2013).

⁶ On huomattava, että voimme perustavia käsitteitämme käyttäen määritellä monimutkaisempia käsitteitä, joiden alaan kuuluvilla olioilla ei välttämättä ole suoraan meille havaittavissa olevia yhteisiä ominaisuuksia (ks. Haukioja 2013).

Nämä teesit olettavat Langtonin *Erottelen*. Mikäli edelleen Langtonin *Palautumattomuus* on pätevä, näyttää *Nöyryys*' seuraavan *Reseptiivisyydestä*' vastaavalla tavalla kuin *Nöyryys* seuraa *Reseptiivisyydestä*.

Onko *Nöyryys*' uskottava periaate? Ensi näkemällä se voi vaikuttaa täysin järjenvastaiselta: sanoohan se, että voimme käsitteillämme viitata vain olioiden relationaaliin ominaisuuksiin. Tuntuu kuitenkin hyvin luontevalta ajatella, että viittaamme puheessamme ja ajattelussamme jatkuvasti sekä sisäisiin että ulkoisiin ominaisuuksiin, puhuessamme vaikkapa toisaalta perunan primääreistä ominaisuuksista, kuten sen muodosta, ja toisaalta sen sekundäärisistä ominaisuuksista, kuten mausta. Edelleen, filosofiassa kiistellään usein siitä, pitäisikö joku ominaisuus tai ominaisuuksien luokka (esimerkiksi värit) ymmärtää primääreinä vai sekundäärisinä ominaisuutena – eikö *Nöyryys*' vie kokonaan pohjan näiltä keskusteluilta?

Uskoakseni näin ei ole. Luontevalta tuntuva erottelu primäärien ja sekundääristen ominaisuuksien välillä voidaan tulkita uudelleen ja hyväksyä, vaikka sitoutuisimmekin *Nöyryyteen*' (ja *Nöyryyteen*). Intuitiivisesti sekundäärisiltä vaikuttavat ominaisuudet ovat tyypillisesti ominaisuuksia, jotka voidaan ymmärtää vain viittaamalla tiettyyn, tai tietyn tyyppiseen, havaitisijaan. Esimerkiksi makuominaisuudet voidaan ymmärtää vain relaatiossa tietyn tyyppiseen maistajaan. Muodon ja massan kaltaiset ominaisuudet, jotka intuitiivisesti vaikuttavat primääreiltä, taas näyttäytyvät kyseisen olion kausaalisissa vuorovaikutuksissa, paitsi suhteessa havaitisijoihin, myös suhteessa kaikkiin muihin olioihin (ks. Pettit 1998 ja Jackson & Pettit 2002). Mikäli hyväksymme *Nöyryyden* ja *Nöyryyden*', tulee meidän siis ymmärtää myös tämän kaltaiset primäärit ominaisuudet relationaalisina ominaisuuksina. Näin ymmärrettynä tämä näkemys ei ehkä johdakaan millään tavalla mystiseen noumenalismiin, vaan näkemykseen, jonka mukaan ominaisuudet yleensä – tai ainakin ne ominaisuudet, joista meillä voi olla tietoa ja joihin voimme viitata – pitäisi ymmärtää dispositionaalisina (vrt. esim Shoemaker 1980). Kysymyksen siitä, missä määrin tätä näkemystä voidaan tai tulisi pitää kantilaisena, joudun jättämään johonkin toiseen ajankohtaan (ja todennäköisesti myös jonkun toisen filosofin käsittelyyn).

Kirjallisuus

- Haukioja, Jussi (2000). *Rule-Following, Response-Dependence and Realism*. Turku: University of Turku.
- Haukioja, Jussi (2005). "Hindriks on Rule-Following", *Philosophical Studies* 126, 219–239.
- Haukioja, Jussi (2007). "How (Not) to Specify Normal Conditions for Response-Dependent Concepts", *Australasian Journal of Philosophy* 85, 325–331.
- Haukioja, Jussi (2013). "Different Notions of Response-Dependence", teoksessa Miguel Hoeltje, Benjamin Schnieder & Alex Steinberg (toim.), *Varieties of Dependence*. Munich: Philosophia Verlag, 167–190.
- Jackson, Frank & Pettit, Philip (2002). "Response-Dependence Without Tears", *Noûs* 36, 97–117.
- Kant, Immanuel (2013 [1781/1787]). *Puhtaan järjen kritiikki*. Suom. Markus Nikkarla & Kreetta Ranki. Helsinki: Gaudeamus.
- Koistinen, Olli (2008). *Kant ja puhtaan järjen kritiikki*. Turku: Areopagus.
- Kripke, Saul (1982). *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.
- Langton, Rae (1998). *Kantian Humility. Our Ignorance of Things in Themselves*. Oxford: Oxford University Press.
- Langton, Rae & Lewis, David (1998). "Defining 'Intrinsic'", *Philosophy and Phenomenological Research* 58, 333–345.
- Pettit, Philip (1991). "Realism and Response-Dependence", *Mind* 100, 587–626.
- Pettit, Philip (1993). *The Common Mind. An Essay on Psychology, Society, and Politics*. Oxford: Oxford University Press.
- Pettit, Philip (1998). "Noumenalism and Response-Dependence", *The Monist* 81, 112–132.
- Shoemaker, Sidney (1980). "Causality and Properties", teoksessa Peter van Inwagen (toim.), *Time and Cause: Essays Presented to Richard Taylor*. Dordrecht: Reidel, 109–135.

Kant ja aistimuksen objektiivisuuden ongelma

Matti Saarni

Puhtaan järjen kritiikin (*Kritik der Reinen Vernunft*, 1781/1781) alussa Kant kirjoittaa aistimuksen käsitteestä seuraavasti: ”Sikäli kuin kohteet affektoivat meitä, kohteen vaikutus representaatiokykyyn on aistimus.” (A 20/B 34.)¹ Teoksen loppupuolella Kant määrittelee aistimuksen käsitettä vielä tarkemmin, tällä kertaa suhteessa kognitioon:

Perseptio, joka viittaa pelkästään subjektiin tämän tilan modifikaationa, on *aistimus* (*sensatio*) ja objektiivinen havainto on *kognitio* (*kognitio*). Tämä on joko *intuitio* tai *käsite* (*intuitus vel conceptus*). Edellinen kohdistuu välittömästi kohteeseen ja on singulaarinen; jälkimmäinen kohdistuu siihen välillisesti, jonkin tuntomerkin välityksellä, joka voi olla usealle oliolle yhteinen. (A 320/B 376–377.)

Näiden määritelmien perusteella vaikuttaa selvältä, että aistimus tarkoittaa Kantille ensinnäkin jotain sellaista, jonka kohteet (tai objektit) saavat aikaan meissä (subjekteissa). Toiseksi Kant sanoo aistimuksen kuitenkin olevan jotain sellaista, joka ei viittaa kohteeseen, vaan ainoastaan subjektiin (tämän tilan modifikaationa). Juuri tässä merkityksessä Kant tuntuu käyttävän termejä objektiivinen ja subjektiivinen; objektiivinen on jotain (representaatio), joka viittaa tai kohdistuu objektiin/kohteeseen, kun taas subjektiivinen viittaa/kohdistuu subjektiin.

On huomionarvoista, että puhuessaan kohteiden vaikutuksesta meihin, Kant ei puhu ensisijaisesti esimerkiksi fysikaalisista, mahdollisesti mitattavista vaikutuksista kehoomme, vaan kohteiden vaikutuksesta tietoisuuteemme. Kohteet vaikuttavat meihin meidän representaatiokykymme kautta. Tämän representaatiokykymme eräänlainen pohja taas on *aistimellisuus*. Hieman yksinkertaistaen voidaan ajatella aistimellisuuden tarkoittavan vain sitä, että meillä on jonkinlaiset, tosin tietyllä ja määrättyllä tavalla toimivat aistit, jotka

¹ Viitataan Kantin teoksiin lyhenteillä (ks. lähdeluettelo) ja standardinmukaisilla *Akademie-Ausgabe* -sivumerkinnöillä.

kykenevät vastaanottamaan kohteiden vaikutuksen. Juuri tämä kohteiden tuottama vaikutus subjektissa on se, jota Kant kutsuu nimellä aistimus. Kysessä on siis sellainen kohteen synnyttämä vaikutus subjektiin, joka on vain subjektin tilan modifikaatio, eikä näin ollen mitään kohteeseen tai kohteisiin viittaavaa.

Ylläolevan perusteella lienee selvää miksi voidaan pitää hämmästyttävänä sitä, että sekä *Arvestelmakryvyn kritikissä* (*Kritik der Urteilskraft*, 1790) että metafysiikan luennoilla² Kant yllättäen puhuukin *objektiivisesta aistimuksesta*. Mitä Kant tarkoittaa ”objektiivisellä aistimuksella”, ja miten aistimus voisi olla objektiivinen, jos se määritelmällisesti on kohteen aiheuttama subjektin tilojen modifikaatio, joka ei viittaa lainkaan kohteeseen? Kant kirjoittaa:

Niityn vihreä väri kuuluu *objektiiviseen* aistimukseen, havaintona aistien kohteesta; mutta sen miellyttävyys kuuluu *subjektiiviseen* aistimukseen, jonka kautta mitään kohdetta ei representoida, toisin sanottuna tunteeseen, jonka kautta kohde käsitetään mielihyvän kohteena (joka ei ole kognitio) (*KU* 5:206).³

Tässä tekstissä käsitelen Kantin aistimuksen käsitettä keskittyen erityisesti objektiivisen aistimuksen käsitteeseen ja siihen mahdollisesti liittyvään ongelmallisuuteen. *Aistimus* on nähdäkseni keskeisessä tai ainakin hyvin merkittävässä asemassa puhuttaessa niin Kantin havaintoteoriasta kuin siitä, mitä objektiivisuus ylipäänsä on ja mitä se edellyttää. Tähän nähden voidaankin pitää ongelmallisena sitä, miten Kant yhtäältä puhuu aistimuksesta jonakin puhtaan subjektiivisena (A 320), mutta toisaalta puhuu siitä, miten nimenomaan kohteet (objektit) saavat meissä aikaan aistimuksia, ja jopa siitä, miten jotkin aistimukset itse asiassa ovatkin objektiivisia aistimuksia (*KU* 5:206).⁴

² Luennot *Met-L1* (AA 28) ja *Met-Mron* (AA 29).

³ Oma suomennokseni.

⁴ Jankowiak (2014) tosin esittää tulkinnan, jonka mukaan aistimukset voivat olla sekä subjektiivisia että objektiivisia, mutta tämä vaatii sen eksplikoimista, että Kant käyttää termejä objektiivinen ja subjektiivinen monissa eri merkityksissä, joiden tulkitseminen jää lukijan/tutkijan tehtäväksi. Jankowiakin mukaan tässä yhteydessä on tarpeen tehdä erottelu esteettisesti objektiivisen/subjektiivisen ja representationaalisesti objektiivisen/subjektiivisen välillä. Tällöin voitaisiin hänen mukaansa sanoa, että kaikki aistimukset ovat esteettisesti subjektiivisia, mutta

1. Objektiivisen aistimuksen ongelmallisuus

Toivon kyenneeni yllä alustavasti esittämään miksi ja missä mielessä on ongelmallista ylipäänsä puhua aistimuksen objektiivisuudesta. Kun Kant kuitenkin tekee niin, on syytä kysyä joitakin lisäkysymyksiä. Mitä Kant tarkoittaa aistimuksen objektiivisuudella? Miten aistimus voi olla objektiivinen, kun Kant toisaalta väittää, että objektiivista on nimenomaan kognitio, ja aistimus taas on nimenomaan jotakin subjektiivista? Mitä tarkoitetaan objektiivisella ja subjektiivisella? Perustuuko tämä erottelu puhtaasti representaatioiden kohteeseen-viittaavuuteen ja ei-viittaavuuteen?

Kantin filosofiaa tuntevalle on ehkä päivänselvääkin se, että Kantin näkemyksessä aistimellisuus on merkittävässä roolissa objektiivisuudesta puhuttaessa. Nimenomaan aistimellisuuden avulla objektit voivat tulla meille annetuiksi, ja aistimellisuuden avulla muodostamme intuitioita (A 19/B 33). Kaiken empiriisen kognition edellytys on aistimellisuus, vaikka aistimellisuus ei yksinään riitäkään objektiiviseen kognitioon. Ajatukseni onkin, että objektiivisen aistimuksen käsitteen ongelmallisuus on sikäli näennäistä, että Kant ei lopulta tarkoita pelkän aistimuksen, oli sitten kyse mistä aistimuksesta tahansa, riittävän varsinaiseen objektiiviseen kohdesuhteeseen. Toisin sanottuna pelkkä aistimus ei yksinään voi olla kognitiivisessa mielessä objektiivinen.⁵ Aistimuksen ja objektiivisuuden yhteys onkin nähdäkseni siinä, että aistimus on välttämätön osa kognitiota, joka on varsinainen objektisuhde. Kant siis saattaa nimittää aistimusta siinä mielessä objektiiviseksi, että aistimus on välttämätön osa objektiivista kognitiota. Mutta ovatko kaikki kognition osana olevat aistimukset sitten tässä mielessä objektiivisia, ja ovatko ne yhtä objektiivisia?

Palataan vielä hieman taaksepäin, kysymykseen aistimuksen objektiivisuuden käsitteen ongelmallisuudesta. On hyvinkin luultavaa, että kyseessä ei ole todellinen vaan ainoastaan näennäinen ristiriita, joka johtuu siitä, että Kant käyttää termejä objektiivinen ja subjektiivinen hieman eri merkityksissä riippuen asiayhteydestä. Puhuessaan objektiivisesta

vain jotkin aistimukset ovat tämän lisäksi representationaalisesti objektiivisia, kun taas toiset ovat siis sekä esteettisesti että representationaalisesti subjektiivisia.

⁵ Esimerkiksi Westphalin (1997) mukaan Kantin ”sensationismi” eli näkemys, jonka mukaan aistimukset eivät itsessään kykene viittaamaan objekteihin, on keskeistä Kantin kriittiselle filosofialle. Tämän näkemyksen mukaan vasta ymmärrys ja kyky tehdä arvostelmia kykenevät synnyttämään objektiivisen suhteen kohteeseen.

aistimuksesta Kant ei tarkoita, että aistimus sinänsä olisi kognitio, toisin sanoen, että aistimus sinänsä ilman ymmärrystä synnyttäisi objektiivisen suhteen kohteeseen. Sen sijaan hän ajattelee, että kohteita havaittaessa aistimus on välttämätön elementti, mutta aistimukset eroavat toisistaan siinä, missä määrin ne (osana kognitiota) kertovat jotakin kohteistaan. Vaikka kaikki *aistimukset* siis *sellaisenaan* ovat puhtaita subjektin modifikaatioita, jotkin aistimukset todella ovat, osana intuitiota, tiedollisesti merkittäviä kohteen suhteen, kun taas jotkin aistimukset kertovat vain subjektin tiloista. Lainaamassani *Arvostelmakyvyn kritiikin* kohdassa objektiivinen aistimus asettuukin nimenomaan vasten subjektiivista aistimusta, joka on tunne, ja ”jonka kautta mitään kohdetta ei representoida” (KU 5:206). Toisin sanottuna Kant ei väitä, että objektiivinen aistimus sinänsä olisi kognitio (vaikka siltä tekstin perusteella kyllä saattaa näyttää), vaan että objektiivinen aistimus osallistuu kognitioon, toisin kuin subjektiivinen aistimus, joka koskee vain subjektin tilaa.

2. Erilaiset objektiivisuudet ja objektiivisuuden asteet

Ovatko kaikki aistimukset sitten aivan yhtä objektiivisiä, jos ne vain ovat osa kognitiota? Kant ei vaikuta ajattelevan aivan näin, vaan aistimusten objektiivisuudessa on eroja toisiin verrattaessa. Mitä nämä erot ovat ja mistä ne johtuvat? Vaikuttaa siltä, että perusmerkityksessään objektiivisuus tarkoittaa Kantilla kohteeseen viittaamista. Mitä silloin voi tarkoittaa se, että objektiivisuudessa tai sen määrässä on eroja, kuten aistimuksen objektiivisuuden tapauksessa? Voiko jokin representaatio, aistimus tai aisti viitata toista enemmän (samaa) kohteeseen? Mitä tämä viittaamisen määrän ero voisi tarkoittaa?

Jossain vaiheessa Kantilla vaikuttaakin tapahtuvan käänne pois siitä, että objektiivisuus tarkoittaisi pelkkää kohteeseen viittavuutta, kohti sitä, että objektiivisuus tarkoittaa itse asiassa instruktiivisuutta, opettavuutta tai paljastavuutta kohteen suhteen. Kant kirjoittaa eri aistien subjektiivisuudesta ja objektiivisuudesta metafysiikan luennoillaan:

Jotkin näistä aisteista ovat objektiivisiä, toiset subjektiivisiä. Objektiiviset aistit ovat samanaikaisesti yhteydessä subjektiivisiin; täten objektiiviset aistit eivät ole vain objektiivisiä, vaan myöskin subjektiivisiä.

Aistimuksissa joko objektiivisuus on suurempaa kuin subjektiivisuus, tai toisin päin. Esimerkiksi näkemisessä objektiivisuus on suurempaa kuin subjektiivisuus, ja korviaraastavan kovassa äänessä subjektiivisuus on suurempaa. Mutta mikäli emme tarkastele aistien voimakkuutta vaan niiden laatua, huomaamme, että näkeminen, kuuleminen ja tunteminen ovat pikemminkin objektiivisia kuin subjektiivisia aisteja, kun taas haistaminen ja maistaminen ovat pikemminkin subjektiivisia kuin objektiivisia. (*Met-L1* 28:231.)⁶

Huolimatta Kantin lukuisista esimerkeistä, yllä olevassa ei tule selväksi mitä Kant tässä yhteydessä objektiivisuudella ja subjektiivisuudella tarkoittaa. Tekstin jatko valaisee asiaa:

Subjektiiviset aistit ovat nautintoaisteja, kun taas objektiiviset aistit ovat opettavia aisteja. [...] Näkö ja kosketus ovat täydellisen objektiivisia representaatioita. Mutta kosketus on objektiivisista representaatioista perustavampi, sillä kosketuksen avulla voin havaita [kohteen] muodot koskettamalla sitä eri sivuilta, näin ollen se on muotojen tulkintataittoa. Näkemisen perusteella tiedostan vain kohteen pinnan. (*Met-L1* 28:231–232.)⁷

Aistit ja aistimukset näyttävät siis yhtäältä voivan olla ensinnäkin joko yksiselitteisen subjektiivisia (kuten niityn vihreän värin herättämät tuntemukset) tai objektiivisia (kuten niittyä koskevat empiiriset havainnot, joissa aistimus on välttämättömänä osana). Tässä merkityksessä objektiivisuus tarkoittaa selvästi kohteeseen-viittaavuutta, ja aistimuksella joko on tai ei ole tällaista ominaisuutta. David Berger kirjoittaa Kantin näkemyksestä metafysiikan luennoilla:

Kant ajattelee toistuvasti, että jotkin aistimodaliteetit, ennen kaikkea haju ja maku, ovat vähintäänkin selvästi vähemmän 'objektiivisia', merkityksessä 'suhteessa kognitioon', kuin kuuleminen ja erityisesti näkeminen. Tämä sama asenne maun ei-kognitiivista statusta kohtaan

⁶ Oma suomennokseni.

⁷ Oma suomennokseni.

on havaittavissa muuallakin, mukaan lukien Kantin makuesimerkit (sokerin makea maku, koiruohon vastenmielinen maku) tämän käsitellessä kokemuksen arvostelmia ja havaintoja *Prolegomenan* pykälässä 19. Kant myöntää, että nämä ovat esimerkkejä havainnon arvostelmista, jotka *eivät voi* tulla kokemuksen arvostelmiksi. ”Kaikki myöntävät”, hän sanoo, että maut ovat ”vain jotain subjektiivista”, että ne ”viittaavat ainoastaan tuntemukseen”, ja että ”ne ilmaisevat vain kahden aistimuksen suhteen samaan subjektin, nimittäin minuun itseeni”. Sen sijaan – huolimatta näkemyksestään transsendentaalisessa estetiikassa – Kant usein mieltää väriä koskevat arvostelmat havainnon arvostelmiksi, jotka voivat tulla kokemuksen arvostelmiksi, eli jotka voivat olla suhteessa kohteisiin. Ne eivät ilmaise ainoastaan aistimusten suhdetta subjektiiin; ne artikuloivat objektiivista, arvosteltavaa sisältöä. (Berger 2009, 78.)⁸

Berger käyttää tässä ilmaisua ’suhteessa kognitioon’ osittain samassa merkityksessä kuin itse käytän termiä ’kohteeseen-viittaavuus’. Kuitenkin olen Bergerin kanssa hieman eri mieltä siitä, mieltääkö Kant värit arvosteltavaksi sisällöksi. Pikemminkin kyse voisi olla siitä, että värin aistimus on osa sitä aistimusta, jonka avulla kognitio kohdistuu varsinaiseen kohteeseen. Väri ei silti itsessään ole Kantille objektiivinen ominaisuus.

Tällaisen jokseenkin yksiselitteisen ”joko–tai” -objektiivisuuden lisäksi aistimukset voivat olla eri määrissä objektiivisia siinä mielessä, että ne *kertovat* jotain *enemmän* tai *vähemmän itse empirisestä kohteesta*. Tällöin kyse ei vaikuttaisi olevan niinkään kohteeseen-viittaavuudesta tai ei-viittaavuudesta, vaan pikemminkin siitä, millä tavoin ja missä määrin kohteeseen viitataan tai missä määrin siitä saadaan tietoa. Kyse ei siis ole tällöin puhtaasti siitä, onko jokin representaatio objektiivinen vai ei, vaan siitä, miten objektiivinen se on. Mutta mikä tarkalleen ottaen määrittää sen, miten objektiivinen jokin aistimus on, ja minkä suhteen aistimuksen tulisi olla opettava tai instruktiivinen ollakseen objektiivinen?

⁸ Oma suomennokseni.

3. Objektiivisuus, viittaavuus ja kohteiden ominaisuudet

Kant käyttää siis objektiivisuuden käsitettä ensinnäkin siten, että (1) objektiivisuus tarkoittaa kohteeseen-viittaavuutta. Ilman tätä ominaisuutta aistimus ei voi olla objektiivinen. Edelleen, aistimus ei siltikään voi yksinään olla objektiivinen, vaan vain osana kognitiota, mutta tässä roolissa se voi joko viitata kohteeseen tai olla viittaamatta, eli se voi olla joko objektiivinen tai subjektiivinen. Toisaalta Kantille (2) objektiivisuus voi olla myös määrällistä. Jokin aistimus voi siis olla ensinnäkin objektiivinen, eli viitata kognitiossa kohteeseen, mutta myös enemmän tai vähemmän objektiivinen riippuen siitä, miten ja minkä aistin avulla se kohteeseen viittaa, sekä siitä *mihin se kohteessa oikeastaan viittaa*. Voimmekin ehkä ajatella tämän toisenkin objektiivisuuden merkityksen tarkoittavan tietynlaista viittaussuhdetta, nimittäin *viittaamista empiirisesti todelliseen kohteeseen (sinänsä)*, ja nimenomaisesti tämän tiettyihin, *a priori* primaarisiin ominaisuuksiin, jotka perustuvat olion ajallis-avaruudellisuuteen. Selvennän tätä väitettä ja ajatusta kohteiden *a priori* primaarisista ominaisuuksista lisää alla. Aluksi on olennaista huomata, että mikäli Kantin toinenkin merkitys objektiivisuudelle tarkoittaa tietynlaista viittaussuhdetta, se tarkoittaa tulkintani mukaan representaation kohdistumista *empiiriseen olioon sinänsä*, joka määrittyy nimenomaan olion empiirisesti *primaaristen ominaisuuksien* mukaan. Kant kirjoittaa:

[...] esimerkiksi värejä tai makua ei tule pitää olioiden ominaisuuksina vaan pelkkinä subjektimme muutoksina, jotka voivat vieläpä olla eri ihmisillä erilaisia. Tässä tapauksessa sitä, mikä itse alun perin on pelkkä ilmentymä, esimerkiksi ruusu, näet pidetään *empiirisessä mielessä oliona itsessään*, joka kuitenkin voi ilmetä eri silmille väriltään erilaisena. (A 29–30/B 45; kursivointi lisätty.)

Näin ollen tämän toisen objektiivisuuden merkitys olisi nimenomaan siinä, että aistimukset kohdistuvat juuri empiirisen olion varsinaisiin ominaisuuksiin silloin, kun on kyse objektiivisuudesta.

Voiko aistimuksen objektiivisuudessa sitten ollakin kyse vain siitä, mikä aisti on kyseessä? Kanthan pitää joitakin aisteja objektiivisempina kuin toisia sillä perusteella, että ne kertovat toisia enemmän jotakin kohteesta, toisin sanoen ovat kohteiden suhteen instruktiivisempia. Tässäkin ratkaisussa jää

ainakin nimelliseksi ongelmalliseksi se, miten aisti ylipäänsä voi olla objektiivinen, kun vasta kognition pitäisi olla! Ratkaisu tähän liittyyneeseen siihen, että aistimus voi olla objektiivinen vain osana kognitiota, eikä kognitio itse asiassa voi olla objektiivinen ilman aistimusta⁹, joka on ainakin osittain objektiivinen eli objektia koskeva tai kohteeseen kohdistuva. Toisin sanottuna kaikkeen objektiivisuuteen tarvitaan aistimusta, ja vieläpä sellaista aistimusta, joka todella kertoo jotain (empiirisestä) kohteesta itsestään! Mitkä aistit sitten kertovat, ja mikä on empiirinen kohde itsessään?

Parhaiten kohteista näyttävät Kantiin mukaan kertovan tunto- ja näköaisti, sillä näillä pääsee parhaiten ”käsiksi” empiirisen kohteen todellisiin ominaisuuksiin. Mitkä sitten ovat näitä todellisia ominaisuuksia? Tulkintani mukaan näitä ovat intuition muotoihin, siis käytännössä avaruuteen ja aikaan *a priori* perustuvat välttämättömät ominaisuudet, kuten muoto ja koko. Lyhyesti sanottuna kyse on kohteiden primaarisista ominaisuuksista, jotka nimenomaan määrittävät kohteen. Kuten tiedämme, Kantiin aika ja avaruus ovat inhimillisen intuition muotoja ja tällaisina välttämättömiä elementtejä kaikessa inhimillisessä kohteiden kokemisessa ja jopa näiden kohteiden olemassaolossa. Inhimillisen kokemuksen kohteet ovat välttämättä olemassa ajallis-avaruudellisesti. Elleivät ne olisi, emme me voisi niitä lainkaan kokea, emmekä muodostaa niistä kognitiota, sillä aistimellisuutemme rajoittaa meitä kokemaan kohteita, jotka ovat olemassa nimenomaan ajallis-tilallisesti. Juuri tämä ajallis-avaruudellisuuden välttämätön *a priori* -luonne, joka kaikilla kokemuksemme kohteilla eli ilmentymillä on, takaa sen, että kaikilla empiirisillä kohteilla/olioilla on välttämättä joitakin ominaisuuksia, joita ilman ne eivät voisi olla (ajallis-tilallisesti) olemassa. Väitteeni on, että juuri näitä välttämättömiä ominaisuuksia Kant pitää empiirisessä mielessä kohteiden primaarisina ominaisuuksina.

Vaikka Kant ei missään esitä selvää luetteloa primaarisista ominaisuuksista, on nähdäkseni perusteltua väittää, että hän pitää esimerkiksi muotoa ja kokoa tällaisina ominaisuuksina, kun taas hän ei pidä esimerkiksi hajua ja väriä tällaisina ominaisuuksina. (A 29–30/B 45.) Kant tosin esittää *Prolegomenassa* ajattelevansa, että kaikki ilmentymien ominaisuudet ovat sekundaarisia siinä mielessä, että mikään niistä ei ole olemassa täysin riippumatta havaitsemisen tai kokemisen ehdoista:

⁹ Tämä koskee ainakin empiiristä kognitiota.

Jo kauan ennen Locken aikoja, mutta etenkin hänen jälkeensä on yleisesti otaksuttu ja myönnetty, että ulkoisten olioiden lukuisista predikaateista voidaan olioiden olemassaololle haittaa tekemättä sanoa, etteivät ne kuulu näihin olioihin sinänsä vaan ainoastaan niiden ilmiöihin, eikä niillä ole mieltämistemme ulkopuolella mitään omaa olemassaoloa. Näihin predikaatteihin kuuluvat lämpö, väri, maku ja niin edelleen. Johtuu täysin kiistämättömistä perusteista, että minä tärkeistä syistä luen pelkkiin ilmiöihin näiden lisäksi vielä muutkin kappaleiden laatumääreet – ne, joita kutsutaan ensisijaisiksi (*qualitates primarias*) kuten ulottuvuus, paikka ja yleensä avaruus sekä kaikki, mikä on siitä riippuvaista (läpäisemättömyys eli materiaalisuus, hahmo jne.). [...] [O]len todennut vielä useampien, *jopa kaikkien kappaleiden havainnoinnin muodostavien ominaisuuksien* kuuluvan pelkkiin ilmiöihin, sillä toisin kuin todellisessa idealismissa tässä ei kumota sen olion olemassaoloa, joka ilmenee. Osoitan ainoastaan, ettemme aistien avulla lainkaan voi tiedostaa oliota sellaisena kuin se sinänsä on. (*Prol* 4:289; Kant 2005, 88–89.)

Kantin termin sanottuna mitkään havaitsemamme ilmentymien ominaisuudet eivät ole olioiden itsensä ominaisuuksia. Tässä mielessä kaikki ilmentymien ominaisuudet ovatkin sekundaarisia ominaisuuksia, eli jossain määrin riippuvaisia (transsendentaalisesti) subjektiivisista kokemisen ehdoista.

Kuitenkin siinä mielessä kuin Kant esittää primaaristen ja sekundaaristen ominaisuuksien erottelun *Puhtaan järjen kritiikissä*, kyseinen erottelu näyttää liittyvän vahvasti fysiikkaan ja fysikaaliseen selittävyYTEEN. On tärkeää huomata, että tällöin erottelu koskee vain ilmentymiä, ja näin ollen asiat sinänsä rajautuvat koko tämän keskustelun ulkopuolelle. Gary Hatfield (2011) esittääkin Kantin toisen, ”varsinaisen” primaaristen ja sekundaaristen ominaisuuksien erottelun ennen kaikkea fysiikkaan liittyvänä. Tällöin erottelussa (joka siis sijoittuu ilmentymien eikä olioiden sinänsä alueelle) on lähinnä kyse siitä, että tiettyjen ominaisuuksien avulla, joita kutsutaan ensisijaisiksi eli primaarisiksi, on mahdollista selittää toisten, sekundaaristen ominaisuuksien olemassaolo. Sama ei kuitenkaan ole mahdollista toisinpäin, eli sekundaarisilla ominaisuuksilla ei voida selittää primaaristen olemassaoloa. Tarkkaan ottaen Hatfieldin tulkinnassa Kantin primaariset ominaisuudet ovat

tilallistuneita voimia sellaisina kuin ne ilmenevät ajassa ja tilassa, ja näin ne muodostavat kohteita, joilla on primaarisia ominaisuuksia sekä näistä johtuvia sekundaarisia ominaisuuksia. Nämä ”toissijaiset” ominaisuudet eivät varsinaisesti ottaen edes ole ominaisuuksia, vaan ainoastaan kohteiden (primaaristen ominaisuuksien) vaikutuksia havaitsijassa.

Onkin selvää, että fysikaalinen selittävyys on ollut keskeinen tekijä koko primaaristen ja sekundaaristen ominaisuuksien erottelulle uudella ajalla, linkittyen ”uuteen mekanistiseen luonnontieteeseen” 1600-luvulta lähtien (ks. Nolan 2011 tai Langton 1998). Kant ei vaikuta olevan tässä suhteessa poikkeus, vaan päinvastoin tahtoi saattaa filosofisen systeeminsä vähintäänkin yhteensopivaksi newtonilaisen fysiikan kanssa.¹⁰ Esimerkiksi Ilmari Jauhiainen kirjoittaa, että ”[l]uonnontieteiden ja erityisesti newtonilaisen fysiikan merkitystä Kantin filosofialle sen kaikissa eri vaiheissa ei voi painottaa liikaa” (Jauhiainen 2013, 110). Näin ollen ei ole yllättävää, että Kantin erottelu muistuttaa esimerkiksi Locken kuuluisaa erottelua, toki sillä erolla, että Kantin erottelu ei koske asioita sinänsä, vaan ainoastaan ilmentymiä.¹¹

Jos fysikaalinen selittävyys onkin olennainen tekijä haettaessa niitä ominaisuuksia, jotka ovat erityisen olennaisia kohteille, väitteeni kuitenkin on, että Kantin transsendentaalisesta/formaalisesta idealismista löytyy toinenkin, fysikaalista selittävyttä jopa perustavampi syy sille, miksi jotkin ominaisuudet ovat toisia olennaisempia. Tätä syytä olen kuvannut jo yllä ja se liittyy Kantin näkemykseen intuition *a priori* -muodoista, jotka ovat objektien kokemisen ja olemassaolon välttämättömiä ehtoja ja joiden lähde on transsendentaalisessa subjektiivisuudessa. Tulkintani mukaan Kant ajatteli transsendentaalisen subjektin¹² olemassaolon takaavan ilmiöille ajalliset ja avaruudelliset ominaisuudet. Kant kirjoittaa:

¹⁰ Kantin suhteesta Newtonin fysiikkaan ks. esim. Westphal 2004.

¹¹ Locken teoriasta ks. Locke 2008, II, VIII, §§ 8–15.

¹² Seuraan tässä T. K. Seungin (2007) näkemystä, jolloin tarkoitan transsendentaalisella subjektilla niitä subjektiivisuuden elementtejä, joita Kant pitää transsendentaalisina. Nämä ovat ensisijaisesti intuition muodot ja ymmärryksen kategoriat, jotka Kantin mukaan kaikki ihmiset jakavat ja jotka ovat välttämättömiä kohteiden kokemisen ehtoja. Empiirinen subjekti taas viittaa niihin subjektiivisuuden elementteihin, jotka ovat luonteeltaan empirisiä, eli *a posteriori*, eli perustuvat kokemukseen. Nämä elementit eivät ole universaaleja kokemuksen välttämättömiä ennakkoehtoja, vaan pikemminkin seurausta kokemuksesta.

Jopa se koko tässä jaksossa selostettu pääperiaate, että yleiset luonnonlait ovat tiedostettavissa *a priori*, johtaa jo itsestään väittämään, joka kuuluu: Korkeimman luontoa koskevan lainsäädännön täytyy piillä meissä itsessämme, toisin sanoen meidän ymmärryksessämme. Ja meidän ei pidä etsiä luonnon yleisiä lakeja luonnosta kokemuksen avulla, vaan päinvastoin meidän on etsittävä luontoa yleisessä lainmukaisuudessaan yksistään aistimellisuudessamme ja ymmärryksessämme sijaitsevien kokemuksen mahdollisuuden edellytyksistä. (*Prolog* 4:319; Kant 2005, 133–134.)

Kant viittaa tässä käsitteellä ”luonto” koko ilmentymämaailmaan, jolloin luonnon yleisten lakien voidaan tulkita tarkoittavan juuri todellisia ja perustavia fysiikan lakeja, jotka Kantin mukaan ovat lähtöisin transsendentaaliselta subjektilta. Koska aika ja avaruus ovat kaiken luonnon kokemisen eli ilmentymämaailman *a priori* -muotoja, ajallisuuteen ja avaruudellisuuteen perustuvat ominaisuudet, kuten muoto, koko ja liike, ovat ilmentymillä välttämättömiä ja primaarisia. Sen sijaan värillisuus ja makuisuus eivät ole *a priori* -muotoja, eivätkä lainkaan peräisin transsendentaaliselta subjektilta, vaan empiiriseltä subjektilta, jota ilman näitä ominaisuuksia ei edes olisi. On Kantin mukaan tiedostettavissa *a priori*, että kohtaamillamme kohteilla on välttämättä ajallisuudesta ja avaruudellisuudesta seuraavia ominaisuuksia, joita ilman kohteet eivät voisi olla tiedostettavissa, eivätkä täten kohteina edes olemassa. Sekundaariset ominaisuudet taas eivät ole tällaisia, vaan niiden olemassaolo on luonteeltaan *a posteriori*, mikä edellyttää todellista empiiristä kokemusta.

4. Primaariset ominaisuudet ja kohteiden aistittavuus

Tulkintani mukaan erottelu primaarisiin ja sekundaarisiin ominaisuuksiin on kiinteästi yhteydessä koko transsendentaalisen idealismin perusasetteluun. Ajallisuus, avaruudellisuus ja kategoriat ovat Kantille *a priori*, mutta värillisuus ja makuisuus ovat vain subjektiivisia aistikokemuksia, joskin toki kohteiden aiheuttamia sellaisia. Voisimme jopa sanoa primaaristen ominaisuuksien perustuvan *ilmentymän puhtaaseen muotoon* (avaruus ja aika). Näin ollen voisimme nimittää näitä ominaisuuksia myös *muodollisiksi* tai *formaaleiksi* ominaisuuksiksi. Kant kirjoittaa:

apriorisen havainnoinnin kyky koskee ainoastaan ilmiön muotoa, aikaa ja avaruutta; se ei koske ilmiön materiaa, toisin sanoen sitä, mikä siinä on aistimusta, sillä empiirinen koostuu tästä jälkimmäisestä. (*Prol* 4:284; Kant 2005, 80.)

[...] käy päinsä, että tiedetään *a priori* ja siis ennen mitään olioiden kanssa tehtyä tuttavuutta – nimittäin ennen kuin ne ovat meille annettuja – minkälatauista niiden havainnoinnin täytyy olla. Näinhän on avaruuden ja ajan kohdalla. (*Ibid.*)

Koska aika ja avaruus *a priori* -muotoina ovat kaiken kokemuksen ja kaikkien ilmiöiden ehtoja, on oltava niin, että kaikki kohtaamamme ilmiöt tulevat välttämättä olemaan ajassa ja avaruudessa. Tämän vuoksi niillä on välttämättä oltava myös koko, sijainti ja muoto, toisin sanottuna avaruudelliset ominaisuudet. Kantilaisin termein voisimme siis sanoa, että ilmiöiden (1) primaariset ominaisuudet ovat transsendentaalisesti ideaalisia, mutta empiirisesti reaalisia, kun taas (2) sekundaariset ominaisuudet ovat sekä transsendentaalisesti että empiirisesti ideaalisia.

Tämä tarkoittaa, että primaaristen ominaisuuksien olemassaolo riippuu kaiken kokemuksen mahdollistavasta ja *a priori* jäsentävästä transsendentaalisesta subjektista. Primaaristen ominaisuuksien olemassaolo ei kuitenkaan riipu niiden subjektiivisesta havaitsemisesta eli empiirisestä subjektista. Kant kirjoittaa:

Väitämme avaruuden siis olevan empiirisesti reaalin (kaiken mahdollisen ulkoisen kokemuksen suhteen) joskin samalla transsendentaalisesti ideaalinen, eli että se ei ole mitään, kun luovumme kaiken kokemuksen mahdollisuuden ehdosta ja pidämme sitä jonakin, joka on olioiden itsensä perustana. (A 28/B 44.)

Voimme huomata, että ilmiöiden primaarisia ominaisuuksia koskevat tässä täsmälleen samat määritelmät kuin ilmentymiä yleensäkin. Molemmat ovat siis transsendentaalisesti ideaalisia ja empiirisesti reaalisia. Sekundaariset ominaisuudet eroavat ilmentymästä yleensä siinä, että ne eivät ole reaalisia eli riippumattomia subjekt(e)ista. Tämä tarkoittaa yksinkertaisesti sitä, että

Kantin transsendentaalisessa idealismissa ilmiöiden todellinen luonne on juuri niiden primaarisissa ominaisuuksissa.

Yhteenvedona tähänastisesta voidaan sanoa, että Kantin mukaan voidaksemme saada mielteitä tai kognitioita kohteista meidän on voitava intuitoida kohteet. Intuitioilla taas on *a priori* ajan ja avaruuden muoto. Täten tiedämme *a priori*, että kaikilla intuition kohteilla on ominaisuuksia, jotka seuraavat olemassaolosta ajassa ja avaruudessa. Kantin näkemyksessä olemisesta ajassa ja avaruudessa sen sijaan ei seuraa esimerkiksi se, että kohteella on oltava jokin tuoksu tai väri. Kant kirjoittaa:

Värit eivät ole niiden kappaleiden ominaisuuksia, joita koskevaan intuitioon ne liittyvät, vaan nekin ovat vain näköaistin modifikaatioita valon vaikuttaessa siihen tietyllä tavalla. Avaruus sitä vastoin kuuluu ulkoisten objektien ehtona välttämättä niiden ilmentymään tai intuitioon. Maku ja värit eivät ole välttämättömiä ehtoja, joiden yksinomaisessa alaisuudessa kohteet voivat tulla meille aistien objekteiksi. Ne ovat yhdistyneinä ilmentymiin vain tietynlaisen rakenteen kontingentisti liittäminä vaikutuksina. Siten ne eivät myöskään ole *a priori*-representaatioita, vaan ne perustuvat aistimukseen, miellyttävä maku vieläpä (halun tai haluttomuuden) tuntemukseen aistimuksen vaikutuksena. Kenelläkään ei voi myöskään olla värin tai jonkin maun *a priori*-representaatiota. Avaruudessa sen sijaan on kyse vain intuition puhtaasta muodosta, eikä avaruus siis sisällä mitään aistimusta (mitään empiiristä), ja avaruuden kaikki lajit ja määreet voidaan ja ne täytyy voida representoida myös *a priori*, jotta avaruudellisten muotojen ja relaatioiden käsitteet voisivat syntyä. Ainoastaan avaruuden myötä on mahdollista, että oliot ovat meille ulkoisia kohteita. (A 28–29.)¹³

Keskeistä on, että olioiden on voitava olla meidän aistittavissamme, jotta me voimme muodostaa niihin objektiivisen suhteen. Tämän aistittavuuden perusluonteen määrittää aistimellisuutemme, johon kohteiden on voitava vaikuttaa. Aistimellisuutemme perusmuotojen ollessa aika ja avaruus on kohteiden välttämättä oltava olemassa avaruudellisesti ja ajallisesti, sillä muutoin ne eivät voisi vaikuttaa aistimellisuuteemme. Kohteilta siis vaaditaan

¹³ Ero Nikkarlan ja Rankin käännökseen on tarkoituksellinen.

sellaista olemista, joka vaikuttaa aistimellisuuteemme. Kohteen olemassaolo tällaisella tavalla, ja näin ollen vaikuttaminen aistimellisuuteemme, on edellytys kohteen aistimukselle. Värit ja maku (tai värisyys ja makuisuus) taas eivät ole edellytyksiä kohteen aistimukselle. Onkin pikemminkin niin, että aistimuksen olemassaolo on edellytys värille ja maulle. Aistimus on siis edellytys niin sekundaarisille ominaisuuksille kuin kohteiden intuomiselle ylipäänsä. Sekundaariset ominaisuudetkin ovat osa aistimusta, joka on edellytys intuitiolle kohteesta ja joka kohdistuu kohteeseen. Tulkintani mukaan se, mikä aistimuksessa kohdistuu kohteiden varsinaisiin avaruudellis-ajallisiin ominaisuuksiin, on objektiivista, kun taas se osa aistimuksesta, joka ei kohdistu näihin ominaisuuksiin, on subjektiivista.

5. Lopuksi

Kantin ajatus on, että vain aistimellisuuden avulla, toisin sanottuna *a posteriori*, voimme kokea kohteita ja saada tietoa niistä. Kuitenkin se, mitä voimme kokea ja mitä voimme kohteista edes mahdollisesti tietää, määräytyy siinä mielessä *a priori*, että kokemuksellamme on mahdollisuuden ehdot, jotka ovat samalla kohteiden olemassaolon ehdot. Tiedämme ikään kuin jo etukäteen mitä ominaisuuksia kohtaamillamme kohteilla ainakin tulee olemaan, vaikkakaan emme tiedä mitään niiden tarkasta laadusta ennen tätä kohtaamista. Ne ominaisuudet, joita kohtaamillamme kohteilla välttämättä on, ovat primaarisia ominaisuuksia. Nämä ominaisuudet ovat ne, jotka määrittävät varsinaisen empiirisen kohteen sinänsä. Sekundaariset ominaisuudet taas syntyvät meissä kun kohtaamme kohteita, mutteivät ole lainkaan välttämättömiä kohteen olemassaololle.

Kun kohtaamme kohteita, ne synnyttävät meissä aistimuksia. Väitteeni on, että Kantin mukaan ne aistimukset (tai osat aistimusta), jotka kohdistuvat kohteiden primaarisiin ominaisuuksiin, ovat objektiivisia. Vastaavasti ne aistimukset (tai osat aistimusta), jotka eivät kohdistu kohteiden varsinaisiin (primaarisiin) avaruudellis-ajallisiin ominaisuuksiin, ovat subjektiivisia. Tässä mielessä sekundaariset ominaisuudet eivät ole objektien ominaisuuksia lainkaan, vaan subjektin modifikaatioita. Se, mitkä ominaisuudet ovat kohteiden varsinaisia (primaarisia) avaruudellis-ajallisia ominaisuuksia määrittäyty *a priori* aistimellisuuden ehtojen perusteella.

Kirjallisuus

- Berger, David (2009). *Kant's Aesthetic Theory. The Beautiful and Agreeable*. London: Continuum.
- Hatfield, Gary (2011). "Kant and Helmholtz on Primary and Secondary Qualities", teoksessa Lawrence Nolan (toim.), *Primary and Secondary Qualities*. New York: Oxford University Press.
- Jauhainen, Ilmari (2013). "Kant ja fysiikka", teoksessa Vesa Oittinen (toim.), *Immanuel Kantin filosofia*. Helsinki: Gaudeamus.
- Jankowiak, Tim (2014). "Sensations as Representations in Kant", *British Journal for the History of Philosophy*, 22 (3), 492–513. (DOI:10.1080/09608788.2014.913550)
- Kant, Immanuel (1900–). *Gesammelte Schriften I–XXIX*. Toim. Deutsche (aiemmin Königliche Preußische) Akademie der Wissenschaften. Berlin: Walter de Gruyter. [AA]
- Kant, Immanuel (1997). *Metaphysik L₁*. Teoksessa Immanuel Kant, *Lectures on Metaphysics*. Käänt. ja toim. Karl Ameriks & Steve Naragon. New York: Cambridge University Press. [Met-L1]
- Kant, Immanuel (1997). *Metaphysik Mrongovius*. Teoksessa Immanuel Kant, *Lectures on Metaphysics*. Käänt. ja toim. Karl Ameriks & Steve Naragon. New York: Cambridge University Press. [Met-Mron]
- Kant, Immanuel (2000 [1790]). *Critique of the Power of Judgment*. Toim. Paul Guyer. Käänt. Paul Guyer ja Eric Matthews. New York: Cambridge University Press. [KU]
- Kant, Immanuel (2005 [1783]). *Prolegomena [Prolegomena zu einer jeden künftigen Metaphysik die als Wissenschaft wird auftreten können]*. Käänt. & toim. Vesa Oittinen. Helsinki: Gaudeamus. [Prol]
- Kant, Immanuel (2013 [1781/1787]). *Puhtaan järjen kritiikki [Kritik der Reinen Vernunft]*. Käänt. Markus Nikkarla ja Kreeta Ranki. Työryhmän johtaja: Olli Koistinen. Helsinki: Gaudeamus. [A/B]
- Locke, John (2008 [1689]). *An Essay concerning Human Understanding*. Oxford: Oxford University Press.
- Seung, T. K. (2007). *Kant: A Guide for the Perplexed*. London: Continuum.
- Westphal, Kenneth R. (1997). "Noumenal Causality Reconsidered: Affection, Agency, and Meaning in Kant", *Canadian Journal of Philosophy* 27 (2), 209–245.
- Westphal, Kenneth R. (2004). *Kant's Transcendental Proof of Realism*. Cambridge: Cambridge University Press.

Maksiimeista ja itsetietoisuudesta

Kantin filosofiassa

Markus Nikkarla

Christian Wolff (1679–1754) oli Kantia edeltävän saksalaisfilosofian mahtihahmo, ja wolffilainen rationalismi oli Immanuel Kantin (1724–1804) nuoruuden henkinen koti. Vaikka Kant joutui myöhemmin kaatamaan wolffilaisuuden keskeisiä tukipilareita, hänen kriittistä filosofiaansa on hyödyllistä tarkastella tätä taustaa vasten. Kantin rakennelma on sen verran sokkeloinen ja sen käytävät niin hämärit, että selvittämällä tarkasti, mitkä asiat oli muutettava ja mitkä saivat jäädä ennalleen, voimme löytää kiintopisteitä, jotka auttavat hänen filosofiansa ymmärtämisessä.

Yksi tällaisista kiintopisteistä on Kantin käsitys maksiimeista, ja tässä kirjoituksessa käsittelem maksiimien suhdetta itsetietoisuuteen. Maksiimien keskeisyys Kantin käytännöllisessä filosofiassa näkyy jo siinä, että hän muotoilee kategorisen imperatiivinsa eli ylimmän moraaliperiaatteen niihin nojaten. Kant esittää kategorisesta imperatiivista useita muotoiluja, mutta olennaista tässä periaatteessa on subjektiivisten maksiimien yleistettävyyden objektiiviseksi laiksi. *Moraalin metafysiikan perustuksessa* Kant muotoilee kategorisen imperatiivin näin:

[T]oimi vain sen maksiimin mukaan, joka sallii sinun samanaikaisesti tahtoa, että siitä tulisi yleinen laki. (Kant 2014, 63.)

Maksiimin Kant on aikaisemmin samassa kirjassa määritellyt tahtomisen subjektiiviseksi periaatteeksi (Kant 2014, 42). Tahto taas on käytännöllinen järki, joten tämän perusteella maksiimit ovat sellaisia järjen periaatteita, joita toimija käyttää oman toimintansa ohjaamiseen, ja kategorisen imperatiivin tehtävänä on, nimensä mukaisesti, käskää toimijaa rajoittamaan omia maksiimejaan niin, että hän voisi tahtoa kenen tahansa omaksuvan nämä samat maksiimit.

Maksiimit näyttäisivät siis kuuluvan Kantilla järjen toimintaan. Tämä onkin luonteva tulkinta, koska Wolffin filosofiassa maksiimit ovat praktisten syllogismien yläpremissiä. Myös Kant antaa niille tämän selityksen metafysiikan luennoissaan (Ak. 28: 678),¹ joten voimme todeta kategorisen imperatiivin käskävän toimijaa syllogismin tai syllogismien sarjan välityksellä joko tekemään tietyn teon tai jättämään sen tekemättä. Itse imperatiivi taas on järjen itsensä antama sääntö, joten mikäli toimija on täysin rationaalinen, järjen sääntö määrää hänen toimintansa. Koska maksiimitkin toisaalta ovat järjen sääntöjä, myös moraalin vastainen toiminta perustuu tämän selityksen perusteella järkeen. Ainakin ensi näkemältä tällainen kuva toiminnan perusteista vaikuttaa kuitenkin vääristyneeltä. Tokihan me ihmiset olemme juuri järjellä varustettuja eläimiä ja järki siis tosiaan ohjaa toimintaamme, mutta meissä on myös eläimellinen puoli. Kantkaan ei kiellä tätä vaan hänen mukaansa järjen käsky kohtaa meissä aistimellisten halujen vastarinnan. Jotta moraalista toimintaamme vastustavat halut saataisiin kuvaan mukaan, Kantin tarjoama selitystä toiminnan perusteista on täydennettävä jollakin tavalla, ja koska Kant tuntuu olettavan, että lukija kykenee itse suoriutumaan tästä tehtävästä, on luontevaa etsiä siihen eväitä samasta lähteestä, mistä maksiimitkin ovat peräisin, eli Wolffin filosofiasta.

Wolffin rationalismiin kuului ajatus siitä, että kaikelle on löydettävissä riittävä peruste, ja niinpä sellainen löytyy aina toiminnallekin, määräytyipä se sitten halujen tai tahdon kautta. Tämä määräytyminen johtuu aina hyvän representoimisesta (hyvä on Wolffin mukaan täydellisyyttä), mutta halujen kohdalla mieltymyksemme kohdistuu asiaan, josta meillä on hyvän epätarkka representaatio (Wolff 1747, § 434). Tahdon määrätessä toimintaa maksiimien rooli on helposti ymmärrettävissä, koska silloin mieli representoi maksiimin sääntöä itsetietoisesti, mutta Wolffin mukaan maksiimeja tarvitaan myös aivan eläimellisessä, affektien vallassa määräytyvässä toiminnassa (Wolff 1743, § 192). Nimittäin kun toiminta määräytyy silloinkin hyvän representaatiosta, mieli tarvitsee yleisen säännön, jonka perusteella se päättyy toimintaan. Toiminnan määräytymisen kannalta keskeinen ero halujen ja tahdon välillä on siis maksiimien suhde itsetietoisuuteen. Eläimellisessä toiminnassa

¹ ”Ak.” viittaa Kantin koottuihin teoksiin (*Gesammelte Schriften*) niteen ja sivun mukaan. *Puhtaan järjen kritiikkiin* viitataan vakiintuneeseen tapaan ensimmäisen (A) ja toisen (B) laitoksen sivunumeroinnin mukaan.

itsetietoisuutta ei ole mutta toiminnan määräytymiseen tarvitaan silti yleinen sääntö eli maksimi, josta toiminta ikään kuin päätellään tai johdetaan. Wolff ajattelikin, että eläimillä on järjen kaltainen kyky tai järjen alin aste (Wolff 1747, § 872).

Tämä käsitys pohjautuu Wolffin käsitykseen kognitiokyvystä. Wolffin mukaan representaatioiden selvyys erottaa ymmärryksen eli ylemmän kognitiokyvyn aistimellisuudesta eli alemmasta kognitiokyvystä. Kantin mielestä tämä näkemys alemman ja ylemmän kognitiokyvyn erosta oli kuitenkin onneton virhe, joka oli johtanut filosofian aivan väärille poluille. Selvyys tai epäselvyys ei Kantin mukaan kerro mitään siitä, kuuluuko representaatio aistimellisuuteen vai ymmärrykseen, vaan näitä kykyjä tutkittaessa huomio on kiinnitettävä representaatioiden alkuperään ja sisältöön. (A 44/B 61–62.) Wolff oli laiminlyönyt tällaisen tutkimuksen ja kognitiokyvyn kohdalla Kantin olikin purettava vanhoja rakenteita perustuksia myöten.

Jotta ymmärtäisimme paremmin, minkälaista teoriaa Kant lähti uudistamaan, tarkastelkaamme ensin Wolffin käsitystä mielen kyvyistä. Wolff ajatteli, että mielen fakulteetit, kuten aisti, kuvittelukyky, muisti, tarkkaavaisuus, halu ja tahto, ovat mielen potentiaalisen aktiivisuuden kykyjä, jotka toimiakseen tarvitsevat voiman (Wolff 1734, § 54). Pohjimmiltaan mieli, tai sielu, on Wolffin mukaan maailman representoimisen voima ja mielen fakulteetit ovat tämän yhden voiman vaikutuksia (Wolff 1747, § 747 ja 753). Fakulteetit voidaan siis tämän käsityksen mukaan palauttaa yhteen voimaan, ja tässä valossa onkin ongelmatonta, että ihminen voi järjen avulla vastustaa halujaan. Järjen tehtävänä on tulla tietoiseksi niistä maksiimeista, jotka ovat toiminnan määräytymisen taustalla.

Yhden perusvoiman olettamisen taustalla on Wolffin käsitys sielusta yksinkertaisena substanssina. Kantin näkökulmasta tällainen käsitys kuitenkin ylittää tietokykymme rajat. Emme nimittäin voi kokemuksessa kohdata mitään yksinkertaista, joten sielun väittäminen yksinkertaiseksi substanssiksi kohdistuu kokemuksen alueen ulkopuolelle, eikä mielen fakulteettien tutkimista voida perustaa tällaiselle väitteelle (A 771–772/B 799–800). Sen sijaan voimme tutkia itse kokemuksen mahdollisuuden ehtoja, ja tämä tutkimus on Kantin teoreettisen filosofian perusta. *Puhtaan järjen kritiikin* ensimmäisessä laitoksessa Kant kertoo (A 94), että

[o]n olemassa kolme alkuperäistä lähdetä (sielun valmiutta tai kykyä), jotka sisältävät kaiken kokemuksen mahdollisuuden ehdot ja joita ei voi johtaa mistään muusta mielen kyvystä, nimittäin *aisti*, *kuittelukyky* ja *apperseptio*.

Corey W. Dyck (2008) on esittänyt, mielestäni vakuuttavasti, että Kant asettuu tässä vastustamaan Wolffin ajatusta yhdestä mielen perusvoimasta, johon fakulteetit voitaisiin palauttaa. Koska sielua ei voi väittää yksinkertaiseksi, myöskään fakulteettien edellyttämien voimien ykseyttä ei voida osoittaa. Kant lähteekin liikkeelle täysin toisesta suunnasta, nimittäin siitä, että on vain yksi aika, yksi avaruus ja yksi kokemus, ja pyrkii tämän avulla tavoittamaan kognitiokyvyn analyysissään sellaisia kokemuksen edellytyksiä, joihin tietoisuutemme ei voi suoraan yltää. Edellisessä lainauksessa mainitut kyvyt nimittäin ilmenevät meille empiirisesti ja olemme sikäli niiden toiminnasta tietoisia, mutta kullakin näistä alkuperäisistä kyvyistä on Kantin mukaan myös transsendentaalinen käyttö, joka on tietoisuutemme ulottumattomissa mutta josta voimme kuitenkin tietää jotakin, koska tämä käyttö on kokemuksen edellytys. Voimme tässä keskittyä tarkastelemaan yhtä näistä alkuperäisistä kyvyistä, nimittäin apperseptiota eli itsetietoisuutta.

Kant päätelee *Puhtaan järjen kritiikissä* esittämässään kategorioiden transsendentaalisessa deduktiossa, että transsendentaalinen apperseptio on yksi itsetietoisuus, joka – paradoksaalista kyllä – ei suoraan näyttäydy tietoisuudellemme muuten kuin vaikutuksessaan eli empiirisessä itsetietoisuudessa. Meidän ei tässä yhteydessä ole tarpeen mennä Kantin ajatuskulun yksityiskohtiin vaan riittää, kun toteamme, että yksinkertaisen substanssin ja yhden perusvoiman sijaan Kant rakentaa filosofiansa yhdelle alkuperäiselle itsetietoisuudelle ja yhdelle avaruudelle ja ajalle (jotka muodostavat aistikyvyn transsendentaalisen käytön). Yhdessä kuvittelukyvyn transsendentaalisen käytön kanssa nämä kyvyt saavat aikaan kokemuksen mutta ne eivät ole, ainakaan sikäli kuin tiedämme, palautettavissa yhteen perusvoimaan vaan ne ovat itsenäisiä kykyjä, joiden tosin täytyy toimia yhdessä saadakseen aikaan kokemuksen.

Ensi silmäykseltä tämä näkemys näyttäisi romuttavan Wolffin pystyttämän rakennelman kokonaan. Wolffin filosofiassa itsetietoisuus ei edes

ole itsenäinen fakulteetti vaan se syntyy tarkkaavaisuudesta mielen erottaessa itsensä niistä kohteista, joista se on tietoinen (Wolff 1747, § 730). Lähempi tarkastelu kuitenkin osoittaa, että Kantin tekemät muutokset koskevat juuri hänen peräänkuuluttamaansa kognition alkuperän ja sisällön tutkimista, jonka Wolff aikalaisineen oli laiminlyönyt, ja että empiirisellä tasolla peruskuva tietoisuudesta ja itsetietoisuudesta on sittenkin Kantin ja Wolffin välillä hyvin samankaltainen. Tämän peruskuvan esiin kaivaminen vaatii Kantin kohdalla ylimääräistä työtä, koska niin kognitiokyvyn kuin haluamiskyvynkin kohdalla hänen kiinnostuksensa kohdistuu itsetietoisuuteen eikä tietoisuuteen yleisesti. Logiikan luennoissaan (Ak. 9:65) Kant kuitenkin kertoo, että eläimet ovat tietoisia, vaikka niiltä puuttuu itsetietoisuus. Teoksessa *Anthropologie in Pragmatischer Hinsicht* (Ak. 7:135) hän kutsuu hämäräksi representaatioiksi niitä representaatioita, joista emme ole tietoisia, ja kertoo, että eläinten representaatiot ovat juuri hämää. Ne representaatiot, joista olemme tietoisia, eli itsetietoiset representaatiot, sen sijaan ovat selviä tai tarkkoja (tarkkoja representaatioita ovat ne, joiden osarepresentaatiot ovat selviä). Tämä vastaa täsmälleen Wolffin käsitystä, ja ainoa Kantin esiin tuoma ero tässä yhteydessä koskee juuri Wolffin virheellistä käsitystä aistimellisuuden ja ymmärryksen eroista. Kant nimittäin jakaa itsetietoiset eli selvät ja tarkat representaatiot intuition ja ajattelun representaatioihin, kun taas Wolffilla intuitiivinen kognitio on aina hämää. Radikaalista uudistustyöstään huolimatta Kant pystyi siis kuvaamaan omaa käsitystään itsetietoisuudesta Wolffin käyttämällä kielellä. Tämä tarjoaa pohjan mahdollisuudelle löytää selitys sille, miten Kant ajattelee tahdon voivan vastustaa haluja.

Wolffin filosofiassa itsetietoisuus ja representaatioiden selvyys ja tarkkuus syntyvät tarkkaavaisuuden kyvyn avulla mielen erottaessa representaatioita toisistaan. Erottaminen tuottaa ensin tietoisuuden kohteista ja tarkkaavaisuuden avulla me ihmiset pystymme erottamaan itsemme näistä kohteista. Elämellisestä intuitiivisesta kognitiosta, josta kaikki kognitio alkaa, tulee näin järjellisen olennon käsitteellistä ajattelua. Koska aistimelliset halumme ovat hyvän hämää representaatioita, siirtyminen niistä itsetietoiseen toiminnan maksimien tarkasteluun on Wolffille ongelmatonta. Kant taas ei voinut enää nojata sielun yksinkertaisuuteen ja yhteen perusvoimaan, ja kun itsetietoisuus on hänen mukaansa alkuperäinen kyky, Kantin kohdalla ei ole yhtä selvää, miten aistimelliset halut voisivat olla

riittävän samankaltaisia tahdon määräytymisperusteiden kanssa, jotta järki voisi maksiimeillaan vastustaa niitä. Kantin muutostöiden jäljiltä ovi maksiimien ja halujen välillä näyttäisi puuttuvan.

Yksi mahdollisuus ratkaista tämä ongelma olisi tulkita Kantin ajatelleen niin, että sekä tietoisuutemme että toimintamme ovat aina jollakin tavalla itsetietoista. Silloin nimittäin halummekin voisivat toimia toiminnan määräytymisperusteina vain itsetietoisina. Tällainen tulkinta olisi kuitenkin ristiriidassa sen kanssa, että Kant ajattelee eläinten olevan tietoisia ja oman itsetietoisuutemme ulottuvan vain selviin ja tarkkoihin representaatioihin. Meillä on siis hyvä syy lähteä tutkimaan, olisiko sittenkin mahdollista löytää yhteys tahdon maksiimien ja itsetietoisuuden ulkopuolella olevien aistimellisten halujen välillä. Tällainen yhteys olisi Kantin teoriassa mahdollinen, mikäli haluihin perustuva toiminnan määräytyminen perustuisi sääntöihin, jotka itsetietoisuuden alkuperäisen kyvyn avulla voivat tulla järjen käsitettäviksi.

Tätä mahdollisuutta voi lähestyä teoreettisen filosofian kautta. Järjen syllogismeista päästään Kantinkin kohdalla vaivatta arvostelmien kautta käsitteisiin, sillä hänen mukaansa niin syllogismien, arvostelmien kuin käsitteen muodostuksenkin takana on sama representaatioiden yhdistämisen funktio, jolla ymmärrys tuo representaatioiden moninaisuuteen ykseyden. Tähän ajatukseen nojautuen Kant pystyi löytämään kategoriat, sillä ymmärrys on hänen mukaansa pohjimmiltaan kyky tehdä arvostelmia, ja loogikot olivat jo löytäneet arvostelmien tekemisen muodot. Arvostelmataulukon avulla on näin ollen mahdollista löytää kategoriat eli puhtaat ymmärryksen a priori -käsitteet.

Kategorioiden transsendentaalisen deduktion tehtävänä on oikeuttaa nämä käsitteet, ja Kant ajattelee tämän onnistuvan osoittamalla, että aistimellisuuden kohteiden representoiminen edellyttää juuri niitä samoja a priori -sääntöjä, jotka ymmärryksen kategoriat ilmaisevat. Kantin teoreettisen filosofian lukija kohtaa transsendentaalisessa deduktiossa samankaltaisen ongelman kuin me olemme tässä kohdanneet: ajatteleeko Kant, että aistimellisuuden säännöt ovat itsetietoisia sääntöjä vai kykeneekö hän suorittamaan todistuksensa sen wolffilaisen empiiristä tietoisuutta koskevan käsityksen puitteissa, jossa hän edelleen ilmaisee pitäytyvänsä? Aiheellamme on siis tiivis kytkös Kantin filosofian perustan luovaan kategorioiden

transsendentaalisen deduktion tulkintaan, jota ei tässä ole kuitenkaan tarpeen pohtia tarkemmin. Aiheemme kannalta olennaisia ovat a priori -sääntöjen sijasta empiiriset säännöt ja tehtävänämme on siis tutkia, voiko Kantin filosofiassa olla empiirisiä sääntöjä, jotka eivät edellytä itsetietoisuutta ja joiden välityksellä halut voisivat määrätä toimintaa.

Wolffin mukaan (1747, § 273) yleisen kognition perustana olevat käsitteet syntyvät samankaltaisuuden ja eroavuuden tunnistamisesta, ja tämän edellytyksenä on kuvittelukyky. Kuvittelukyky tuottaa kuvitelmia, jotka ovat koosteiden representaatioita (Wolff 1747, § 751). Toisin sanoen kuvitelma on kuvittelukyvyyn tuottama kuva, joka on moninaisuuden sisältävä ykseys. Koska tällaiseen kuvaan kuuluu hämäryys, kuva voi vastata useampaa aistien kohdetta, ja tällä tavoin ymmärrykselle avautuu mahdollisuus päästä representoimaan yleisyyttä käsitteiden avulla. Myös Kant ajatteli, että kuvittelukyvyyn on ensin saatettava intuition moninaisuus kuvaksi, jotta käsitteellinen kognitio olisi mahdollista (A 120). Kantin ajattelussa yleisyyden lähde on kuitenkin kuvien perustana olevissa kuvittelukyvyyn tuottamissa säännöissä, joita Kant kutsuu skeemoiksi:

[E]mpiirinen käsite kohdistuu aina välittömästi kuvittelukyvyyn skeemaan intuitiomme määrittämisen sääntönä tietyn yleisen käsitteen mukaisesti. Koiran käsite tarkoittaa sääntöä, jonka mukaan kuvittelukyky voi piirtää yleisesti nelijalkaisen eläimen avaruudellisen muodon olematta minkään yksittäisen kokemuksen minulle tarjoaman erityisen avaruudellisen muodon – eikä myöskään minkään mahdollisen kuvan, jonka voin esittää *in concreto* – rajoittama. (Kant 2013, A 141/B 180.)

Kant tarkastelee skeemoja käsitteellisen a priori -kognition näkökulmasta, mutta esimerkiksi Béatrice Longuenesse (1998, 116) on esittänyt, että ne ovat käsitteitä edeltäviä sääntöjä.² Tätä tulkintaa tukee se, että Kant sanoo niiden olevan pelkän kuvittelukyvyyn tuotteita.

Kuten aikaisemmin totesimme, kuvittelukyky on yksi kognitiokyvyyn alkuperäisistä kyvyistä, ja skematismien perusteella kuvittelukyky voisi Kantin teoriassa tuottaa itsetietoisuudesta riippumatta sääntöjä, joiden avulla halut

² Longuenesse tosin ajattelee, että nämä säännöt ovat itsetietoisia.

pystyisivät määräämään toimintaa. Vielä olisi kuitenkin ratkaistava, mistä nämä säännöt saisivat moninaisuudelleen ykseyden. Wolffin filosofiassahan ykseys tulee sielun yksinkertaisuuden myötä, mutta Kantin kohdalla näin ei voi olla. Itsetietoisien representoimisen suhteen Kantin kantavana ajatuksena on, että alkuperäinen itsetietoisuus on sääntöjen sisältämän moninaisuuden ykseyden perusta, mutta tästä ei luonnollisestikaan ole apua, koska olemme olettaneet käsittelemiemme sääntöjen olevan pelkästään tietoisia, ei itsetietoisia. Kuten edellä huomasimme, alkuperäisten kognitiokykyjen joukosta löytyy kuitenkin myös toisenlainen ykseyden lähde: ajan ja avaruuden singulaarisuus. Aistin transsendentaalinen käyttö voisi siis tarjota kuvittelukyvyyn tuottamille säännöille tarvittavan ykseyden.

Minkälaisia nämä pelkästään tietoiset säännöt sitten olisivat? Miten ne toimisivat ja miten ne eroaisivat itsetietoisesta maksimien kautta määräytyvästä toiminnasta? Wolffin mukaan (1743, § 192) pelkästään tietoisessa (ei siis itsetietoisessa) toiminnan määräytymisessä yläpremissin säännön eli maksimin tuottaa kuvittelukyky ja alapremissin tarjoaa kokemus. Jos siis skeema voisi toimia toiminnan määräytymiseen tarvittavana sääntönä, tilanne olisi Kantilla pääpiirteissään sama kuin Wolffilla. Tarkastelkaamme tätä esimerkin avulla.

Olettakaamme, että kissa ei ole itsetietoisuuteen kykenevä eläin. Kun koiria pelkäävä kissa lähtee koiran nähdessään karkuun, toimintaan määräytymisen yleisenä sääntönä olisi kuvittelukyvyyn tuottama assosiaatio koiran skeeman ja mielipahan välillä. Kissa toteaisi edessään olevan kohteen koiran skeeman tapaukseksi ja tämän myötä yhdistäisi aistien kohteen mielipahaan, mikä johtaisi välttämisreaktioon. Koska kissa ei oletuksemme mukaan kykene käsitteelliseen ajatteluun, jossa sääntöjen ykseyden tarjoaa transsendentaalinen apperseptio, toiminnassa ja sen määräytymisessä tarvittavien sääntöjen täytyy kissan kohdalla saada ykseytensä avaruudesta ja ajasta. Kissa siis kykenisi tavoittelemaan toiminnassaan paikkaa, jossa pystyisi välttymään koiran aiheuttamalta epämiellyttävyydeltä.

Me ihmiset saatamme toimia aivan vastaavalla tavalla, mutta itsetietoisuuden ansiosta kykenemme myös ajattelemaan omaa tilannettamme ja toimintaamme. Siirtyminen eläimellisestä representoimisesta ja toiminnasta ajatteluun ja tahtomiseen ei kuitenkaan näyttäisi Kantin teoriassa edellyttävän muuta kuin ykseyden lisäämisen jo mielessä oleviin kuvittelukyvyyn sääntöihin.

Skeemoista tulee tällöin käsitteitä ja ihminen kykenee irrottautumaan pelkästä lainomaisesta ajallis-avaruudellisesta assosioinnista havaitsemiensa kohteiden tunnistamiseen. Tunnistaminen eli tietoinen samuuden representoiminen ei nimittäin onnistu pelkkien ajallis-avaruudellisten ykseyksien avulla vaan siihen tarvitaan itsetietoisuutta. Kun ihminen on ottanut tämän askeleen, hän pystyy vastustamaan halujaan. Tätä voi havainnollistaa edellä esitetyn esimerkin avulla. Ihminenkin voi pelästyä koiraa ja lähteä pakenemaan sitä. Hän voi toisaalta myös ajatella, onko pakeneminen välttämätöntä tai hyödyllistä. Ehkä hän muistaa kuulleensa, ettei tällaisessa tilanteessa kannata juosta karkuun. Hän saattaa edelleen tuntea pelkoa ja haluta lähteä karkuun, mutta silti pysähtyä, koska hänen tahtonsa vastustaa hänen haluaan.

Skematismi tuntuisi siis tarjoavan kaipaamamme yhteyden järjen ja halujen määräämän toiminnan välillä. Kuten mainitsin, tässä käsitellyillä kuvittelukyvyyn empiirisillä säännöillä on kiinteä yhteys kategorioiden transsendentaaliseen deduktioon. Jos on osoitettavissa, että tämä deduktio perustuu kuvittelukyvyyn transsendentaalisen käytön tuottamien a priori -sääntöjen samuuteen pelkästään tietoisessa aistimellisten kohteiden representoimisessa ja niiden itsetietoisessa ajattelussa, tässä esitetty yhteys itsetietoisien maksiimien ja pelkästään tietoisien halujen määräämän toiminnan välillä voidaan perustella siihen tukeutuen. Toisaalta tässä luonnehdittu maksiimien ja halujen välinen yhteys voisi tukea Colin McLearnin (ilmestyy) esittämää väitettä, että sekä konseptualistiset että nonkonseptualistiset tulkinnat Kantin teoreettisesta filosofiasta ovat laajalti perustuneet virheelliseen tulkintaan Kantin havaintokäsityksestä. Voikin olla, että koko Kant-tutkimus kaipaisi uutta näkökulmaa aistimellisuuden ja itsetietoisuuden eroon. Tällaisen uudistuksen tarkastelu jää kuitenkin suoritettavaksi toisessa yhteydessä.

Turun yliopisto

Kirjallisuus

- Dyck, Corey W. (2008). "The Subjective Deduction and the Search for a Fundamental Force", *Kant-Studien* 99 (2), 152–179.
- Kant, Immanuel (1900–). *Gesammelte Schriften*. Berliini: Georg Reimer & Walter de Gruyter. [Ak.]

- Kant, Immanuel (2013). *Puhtaan järjen kritiikki*. Suom. Markus Nikkarla ja Kreetta Ranki. Helsinki: Gaudeamus. [A/B]
- Kant, Immanuel (2014). *Moraalin metafysiikan perustus*. Suom. Markus Nikkarla. Turku: Areopagus.
- Longuenesse, Béatrice (1998). *Kant and the Capacity to Judge. Sensibility and Discursivity in the Transcendental Analytic of the Critique of Pure Reason*. Princeton: Princeton University Press.
- McLear, Colin (ilmestyy). ”Kant on Perceptual Content”, *Mind*.
- Wolff, Christian (1734). *Psychologia rationalis methodo scientifica pertractata*. Frankfurt.
- Wolff, Christian (1743 [1720]). *Vernünfftige Gedancken von der Menschen Thun und Lassen, zu Beförderung ihrer Glückseligkeit*. Halle.
- Wolff, Christian (1747 [1719]). *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, Auch Allen Dingen überhaupt*. Halle.

Transcendental Anti-Theodicy

Sami Pihlström

1. Introduction

This essay examines the ways in which transcendental arguments or, more generally, transcendental reflection can be put to work in philosophical attempts to understand the conditions for the possibility of appropriately acknowledging human suffering and its moral relevance. In particular, I will suggest that the opposition between *theodicy* (religious or secular) and *anti-theodicy* – i.e., what can be regarded as the main opposition in philosophical inquiries into the problem of evil – ought to be investigated from a transcendental point of view. It will be argued that theodicies should be rejected for transcendental reasons and that *anti-theodicy is a transcendental condition for the possibility of the moral perspective itself*. Thus, the essay joins the tradition of moral criticism of theodicies that can be traced back to Immanuel Kant’s “Theodicy Essay” (1791), with more recent representatives among, e.g., post-Holocaust ethical thinkers developing Jewish responses to the Holocaust (such as Hans Jonas and Richard Bernstein), Wittgensteinian moral philosophers and philosophers of religion (such as D. Z. Phillips), and even pragmatists (especially William James). (See also Kivistö and Pihlström 2016.)

Generally, we may say that theodicies seek a justification, legitimation, and/or excusing of an omnipotent, omniscient and absolutely benevolent God’s allowing the world (His creation) to contain evil. Classical formulations can be found, e.g., in Augustine’s appeal to God’s having created human beings with the freedom of the will as the reason why there is evil, and in G. W. Leibniz’s *Théodicée* (1710), according to which God could not have created any better world than the one he did create; hence we live in the best possible world, and while there is some evil there, it is necessary for the overall good. The mainstream approach to the problem of evil in contemporary Anglo-American (broadly analytic) philosophy of religion is, arguably, strongly *theodicist* in the sense that providing a theodicy is regarded as a requirement of

any acceptable form of theism. Conversely, by *anti-theodicism* we may mean the rejection of the very project of theodicy – typically for moral reasons.

The anti-theodacist philosophers listed above cannot be discussed in any detail in a brief essay. However, it needs to be examined at a general level whether, and in what sense, the ethical arguments against theodicies presented by such post-Kantian philosophers – arguments seeking to determine the moral inappropriateness of theodicies – are actually transcendental arguments. The basic structure of a successful anti-theodacist transcendental argument will therefore have to be sketched.

In addition, the question of what degree of truth can be achieved by employing transcendental arguments will be approached by proposing that in this context the relevant notion of truth should be rearticulated in pragmatist terms. Thus, my essay will, by examining the specific case of transcendental arguments concerning the appropriate acknowledgment of suffering, contribute to the more general re-evaluation of the possibility of transcendental argumentation within an overall pragmatist philosophical framework. I therefore also hope to reach metaphilosophical results concerning what I propose to call the relativity, or perspectivalness, of the distinctions between the transcendental and the empirical on the one hand and between the transcendental and the transcendent on the other hand. We may also speak about the shifting roles of the empirical, the transcendental, and the transcendent. To address these issues, the argument of the paper will unfold in four sections (sections 2–5 below), before a brief concluding section will tie the threads together.

2. In What Sense Are Anti-Theodacist Arguments Transcendental?

Let me first briefly examine the sense in which arguments focusing on the moral inappropriateness of theodicies can be regarded as transcendental arguments. Such arguments could take the following basic form:

- (i) An appropriate acknowledgment of other human beings' experiences (especially their experiences of suffering) is a necessary condition for the possibility of occupying a moral perspective.

- (ii) Theodicies do not enable the moral subject to acknowledge others' experiences of suffering. (More specifically, theodicies are a form of ethical non-acknowledgment.)
- (iii) We are able to occupy a moral perspective (because we at least occasionally actually do so).
- (iv) Therefore, theodicies violate the necessary conditions for the possibility of morality and ought to be rejected for moral as well as transcendental reasons.

This argument takes different more specific forms depending on whether we locate it in the writings of post-Holocaust Jewish philosophers like Bernstein or Jonas, in Wittgensteinians like Phillips, or in the pragmatist tradition, e.g., James (for some more detailed formulations, see, e.g., Pihlström 2013, chapter 5; and 2014). All these different anti-theodicians offer, however, some version of this argument, according to which it is morally problematic to offer a theodicy, or more strongly, a moral duty to reject theodicies. In some thinkers, especially Jonas and James, we may find quite explicitly metaphysical conclusions based on a transcendental argument like this starting from ethical premises: both, for instance, are willing to rethink the concept of God – to defend an idea of a finite God – on the grounds of meaningless suffering (see James 1907; Jonas 1996). James also defends pluralism and attacks monism for ethical reasons, arguing that the latter, unlike the former, leads to an irresolvable theodicy problem.

Accordingly, the fate of transcendental arguments of the kind sketched here depends on the more general question concerning the relation between ethics and metaphysics. I have argued elsewhere that according to James, in particular, an anti-theodacist approach to evil and suffering is a necessary condition for the possibility of adopting the moral point of view in the first place, and thus for the employment of his pragmatic method, which seeks to analyze the proper meaning of metaphysical and other concepts in terms of their ethical relevance (cf. Pihlström 2016). We may even say that for James the pragmatist entanglement of ethics and metaphysics is based on a “framing” of the pragmatic method by the problem of evil.

Now, why doesn't an argument like this generalize to *anything* morally problematic? Imagine that someone argues – perhaps quasi-transcendentally

yet admittedly rather controversially – from there being something morally problematic about some particular way of thinking, say X, to non-X being a necessary condition for the moral perspective itself. This would be parallel to the argument for anti-theodicy as a condition for the possibility of morality. Why doesn't anything immoral turn out to be a violation of the conditions for the possibility of morality? Can we no longer draw a distinction between morality and immorality *within* the moral perspective; is the very possibility of that perspective inevitably at issue as soon as we adopt a transcendental perspective on these questions?

Well, the short answer is that in some sense it is. That's what morality is like – or more precisely what its foundationlessness and overridingness are like. Morality, moreover, is largely (though not exclusively) about dealing with evil and suffering. The problem of evil (as I have suggested in more detail in James's case) frames moral philosophy generally, or even philosophy generally. This is why it is crucially important to investigate it in transcendental terms characterizing the conditions for the possibility of ethically responsible thought in general. I would be willing to extend this claim to the claim that morality is about “negativities” more generally – evil, suffering, death, guilt, and the overall finite, limited, and incomplete condition of the human being. At least we might say (quasi-transcendentally again) that an appropriate acknowledgment of such negativities is a necessary presupposition for us to be able to engage in moral argumentation and serious moral deliberation.

3. The Relativity of the Transcendental vs. Empirical Distinction

Ludwig Wittgenstein famously writes in the *Tractatus* (1921): “Ethics is transcendental.” (This should be read in the context of another statement in the same book, about logic being transcendental – and of course also in the context of the identification of ethics with aesthetics.) Qua transcendental, ethics is constitutive of our relation to the world in general, including ontology or metaphysics – this, indeed, is what the kind of transcendental argument imagined in section 1 above incorporates as its key idea and presupposition. (Furthermore, this idea could be compared to Emmanuel Levinas's well-known position that ethics is a first philosophy, preceding ontology and epistemology.) Now, we can ask the further question(s) regarding the

constitutive or necessary conditions of ethics itself. What makes ethics possible, or what do we have to presuppose in order for ethics to be possible? This question must make sense insofar as it is plausible to claim, as I just did, that the very possibility of the moral point of view presupposes anti-theodicism – and thus taking evil and suffering seriously.

There may be an indefinite number of constitutive criteria or transcendental conditions that ethics itself – or the seriousness of the ethical, or of the moral point of view – requires or presupposes, e.g., the possibility of guilt and remorse, the possibility of forgiveness, or anti-theodicism, the key topic of this inquiry. We may say that ethical seriousness presupposes that we do not adopt a too easy “happy end” view on evil and suffering but take them seriously in a proper ethical sense – which also entails taking them ontologically seriously. Therefore, ethics does not simply precede ontology but is entangled with it in a relation of reciprocal containment or interpenetration. The transcendental thus cuts both ways, or in other words, a condition’s being transcendental (in contrast to its being empirical or being itself “conditioned”) depends on our perspective of inquiry. This is a (paradoxically) fundamentally antifoundationalist view in the sense that there is no final transcendental bedrock for our views on the transcendental of ethics. It is precisely the kind of understanding of the transcendental that we might develop within a pragmatist transcendental philosophy.

The kind of foundationlessness relevant here may be illuminated by another remark we find in Wittgenstein. In *Culture and Value*, he says that one’s life can lead one to believe in God – and he adds that by “life” he means certain experiences of life, such as sufferings of a certain kind. Now, sufferings, or the mere awareness of the evil there is in the world, may also lead us to lose religious faith, though not as a conclusion of an argument. From the Wittgensteinian point of view, the issue is hardly one concerning what the mainstream analytic theodicy discourse calls the “argument *from* evil”; sufferings would not lead us either to God or away from God because of the success (or failure) of such arguments. Yet, suffering could play a transcendental role in relation to religion and ethics.

The metaphilosophical moral of this is, of course, that there is no fundamental unchanging structure of the transcendental. Something (e.g., ethics) is transcendental relative to something else (e.g., ontology), and

something (e.g., anti-theodicism) may in turn be transcendental relative to ethics itself. Everything here depends on our perspective of inquiry. In this sense, transcendental philosophy is here subordinated to a certain kind of perspectival pragmatism allowing a critically self-reflective plurality of perspectives guiding our inquiries.

4. The Transcendentality of the Transcendent

Not only do the roles of the empirical (or factual) and the transcendental shift, as in Wittgenstein's *On Certainty* (1969), where empirical-sounding propositions may take the role of transcendental-sounding grammatical rules, while remaining open to historical transformation and reinterpretation. Also the other key distinction relevant here, the one between the transcendental and the transcendent (highly central in Kant) may become shifting and perspectival.

Consider Franz Kafka's *The Trial*, particularly the parable about the man waiting outside the Law ("Vor dem Gesetz"). He just waits until the door is finally closed for him. We might say that here remaining outside the law amounts to being in a transcendent sphere where law and its immanent norms are not operative but where there is only room for something like transcendental guilt (cf. Pihlström 2011), as distinguished from empirical or factual (e.g., legal) guilt – in other words, from the kind of truth about one's "case" (i.e., what one is actually guilty of) that Josef K. in *The Trial* desperately seeks to find out but inevitably fails. Perhaps there is also room for something like transcendental forgiveness (in the sense of Vladimir Jankélévitch's theory of radical forgiveness irreducible to any right or duty to forgive, as articulated in Jankélévitch 1967).

Now, what is important here is that no law or legal punishment *could* be morally sufficient or appropriate as recognition of the sufferings of the victims of a crime of the magnitude of, say, the Holocaust. Even so, it could still be meaningfully asked whether there is a possibility for forgiveness even in such cases. Such forgiveness reaching for the transcendent (which, by definition, can never actually be reached from the immanent perspective) is not, and cannot be, any kind of moral or political duty for us, and it can even be argued, with Dostoevsky's character Ivan Karamazov, that we have no right to forgive the most horrible crimes; what is at issue here can only be a "mere"

radical transcendent forgiveness in Jankélévitch's sense. In my view, it makes sense to think of the Holocaust as a crime and evil beyond any conceivable laws, norms, principles, punishment, trials, etc. – and hence as something transcendent – but not beyond the possibility of transcendent forgiveness.

The moral of these reflections is that the postulation of the possibility of transcendence may itself be a transcendental condition for a certain understanding of the ethical. The distinction between the transcendental and the transcendent may therefore itself be as perspectival as the one between the transcendental and the empirical.

5. Testimony and the Transcendental Impossibility of Full Acknowledgment

These issues, also invoking the issue of truth and truthfulness in relation to the recognition and acknowledgment of suffering, should be further connected with the (possibility of) testimony, especially *moral testimony* (as analyzed, e.g., in Margalit 2002) and *martyrdom* – which, as has sometimes been remarked, were in some sense rendered impossible in the Holocaust, in which people were murdered anonymously, deprived of any reason to die for anything. The witness, especially the moral witness, testifies of something whose actual (factual) propositional truth and its actual historical occurrence is, though not irrelevant, not the most important issue. The *moral truth* of what happened is what really matters in such cases of interpreting and acknowledging human suffering – and therefore we need to examine the relation between truth and truthfulness, as this kind of moral truth may have more to do with what is known as truthfulness than with the propositional truth of moral statements that philosophers participating in the mainstream metaethical dispute over moral realism and antirealism focus on.

Moreover, the moral truth (truthfulness) relevant here is a matter of developing a general attitude to the world, not confined to facts or propositional truths about any particular historical event. Our adopting an anti-theodistic way of thinking about suffering, and about other people's experiences in general, *changes everything*; nothing regarding our attitude to the world remains unchanged, and it is in this sense that we are here dealing with a transcendental matter.

In any case, even then we still need a notion of truth, even if we reach the limits of testimony, or the impossibility of bearing witness – something that Holocaust writers and philosophers investigating the Holocaust (e.g., Primo Levi, Giorgio Agamben, and others) have emphasized in different ways. Even the non-testifiability and non-martyrdom of Auschwitz will not destroy the concept of truth. On the contrary, the relevant notion of truth needed here is, I would like to suggest, open to a pragmatist analysis in terms of William James (1907). We can, however, fully acknowledge the limits and fragility of testimony only by acknowledging the fundamental insecurity and foundationlessness of the (or any) moral truth(s) that we (can, or could) testify about.

The impossibility of ever fully acknowledging the “whole truth and nothing but the truth” about the non-testifiable suffering of the martyr, or of any suffering individual, or any moral witness, is a crucial part of the anti-theodictist acknowledgment of others’ suffering in general. I would like to suggest, though I won’t be able to develop the idea any further here, that the relevant notion of truth needed here for a proper communication of moral witnessing or testimony, viz., a notion of truth incorporating truthfulness, is the broader pragmatic (Jamesian) notion, not the one narrowly restricted to propositional truth.

However, the impossibility of full acknowledgment does not mean that we would not be able to develop, or would not have the duty to try to develop, a truthful and sincere attitude to the other’s suffering. On the contrary, the key insight of a pragmatist view of truth may be exactly here: it is by giving up (mere) truth (and mere knowing) that we can (only) open the space for acknowledgment and truthfulness. This “can” refers to a transcendental presuppositional condition opening up a certain way of viewing, sharing, and responding to the world in general. As explained above, adopting anti-theodictism changes everything in the way we view ourselves, others, and the world we live in. Nothing remains untouched by the realization that there is no excusing of meaningless suffering, no justification for the absurdity of evil. In particular, everything changes as soon as we acknowledge the impossibility of fully acknowledging the *meaninglessness* of suffering, the duty to avoid constructing apparently meaning-bestowing “happy end” stories to account for human catastrophes. Such a transformed

attitude to our fellow human beings' lives may fundamentally transform our existence as such.

This impossibility of full acknowledgment and its world- and life-changing character may be comparable to the way in which the loss of a person – for example, the death of a loved one, some concrete and unique individual who has been present in one's life but no longer is – changes one's entire world into something different. It is also comparable to the way in which, according to Wittgenstein's *Tractatus*, the world can “wax and wane” as a totality; the good or evil will cannot change the states of affairs the world consists of, but it can change the *limits* of the world, and as these limits *are* the transcendental subject itself (cf. § 5.63), the subject's anti-theodicy attitude of acknowledgment (including the meta-level acknowledgment of the impossibility of full acknowledgment) changes its world, and thus *the* world, in a fundamental manner.

However, there is one reflexive reflection to be added. Is the meta-acknowledgment of the impossibility of full acknowledgment itself sufficient as an ethically appropriate anti-theodicy attitude to suffering? Or is it, in the end, a theodicy by other means? Does it lead to an infinite regress or question-begging: are we trying to acknowledge the others' suffering fully and completely, after all, when acknowledging the impossibility of full acknowledgment? Is there any place to stop here? Perhaps the honest and sincere response here is *no*. We must self-reflectively examine our own practices and processes of acknowledgment (and meta-acknowledgment) themselves, being prepared to revise them continuously, not being blind to those sufferers' experiences that question our attempts to acknowledge as insincere, incomplete, or hubristic in their attempted or alleged modesty. The *failures* of proper acknowledgment should be in our focus, not the successes – and this is a transcendental claim about the possibility of morally responding to others at all.

Perhaps somewhat surprisingly, then, all this is connected with the project of engaging in transcendental argumentation (or, more broadly, transcendental reflection) in moral philosophy. This is, we may say, transcendental moral philosophy pragmatically naturalized through the fundamental ethical need to take empirically real evil and suffering seriously while maintaining the philosophical focus in an analysis of the necessary

conditions for the possibility of a certain kind of human actuality. In a quasi-Wittgensteinian sense, we are here interested in the necessary conditions for the possibility of (a certain kind of) meaningful language-use. It is precisely by emphasizing this that we may come to appreciate the transcendental character of the anti-theodist account of acknowledgment defended here.

This essay has, despite its brevity, addressed several topics, but they are all deeply interconnected as features of an analysis of what it means to engage in ethically adequate (truthful) acknowledging of (others') suffering, or a pragmatic-cum-transcendental analysis of the conditions enabling us to recognize such suffering. We also saw, at the meta-level, that there is something like a transcendental impossibility of *full* acknowledgment. It is part of our ethically adequate acknowledgment of otherness, especially of the suffering other and/or her/his suffering, that we admit that we can never fully acknowledge it, let alone recognize it as something specific. The unavailability of the (full) propositional truth about others' suffering is precisely something that a truthful acknowledgment of otherness needs to emphasize.

6. Conclusion

I started from a sketchy analysis of the way in which the argument seeking to establish the moral inappropriateness of theodicies is, or can be interpreted as, a transcendental argument. I ended up, however, discussing broader transcendental issues bringing these reflections onto a metaphilosophical level. In a sense, this inquiry is inevitably metaphilosophical, as what is at stake is the very possibility of a moral perspective and hence also the possibility of ethical inquiry.

One of my basic conclusions here is, then, that mere transcendental *argumentation* cannot morally appropriately respond to human suffering, or acknowledge the suffering other. Yet, a transcendental *perspective* on ethics is vitally needed. In investigating anti-theodicy as a moral response to suffering and evil we are engaging in a transcendental *reflection* on the constitutive features of moral seriousness itself. Therefore, also the examples of anti-theodist transcendental argumentation identified above can be regarded as instances of post-Kantian transcendental anti-theodicy.

Bibliography

- James, William (1907). *Pragmatism: A New Name for Some Old Ways of Thinking*. New York: Longmans, Green, and Co.
- Burkhardt, F. Bowers & I. K. Skrupskelis. Cambridge, MA & London: Harvard University Press, 1975.
- Jankélévitch, Vladimir (1967). *Forgiveness*. Transl. A. Kelley. Chicago: University of Chicago Press, 2005.
- Jonas, Hans (1996). *Mortality and Morality*. Evanston, IL: Northwestern University Press.
- Kant, Immanuel (1791). “On the Miscarriage of All Philosophical Attempts at a Theodicy”, in Immanuel Kant, *Religion and Rational Theology*. Transl. & ed. Allen W. Wood & George Di Giovanni. Cambridge: Cambridge University Press, 1996, 20–37.
- Kivistö, Sari & Pihlström, Sami (2016). “Kantian Anti-Theodicy and Job’s Sincerity”, forthcoming in *Philosophy and Literature* 40 (2) (October 2016).
- Margalit, Avishai (2002). *The Ethics of Memory*. Cambridge, MA: Harvard University Press.
- Pihlström, Sami (2011). *Transcendental Guilt*. Lanham, MD: Lexington.
- Pihlström, Sami (2013). *Pragmatic Pluralism and the Problem of God*. New York: Fordham University Press.
- Pihlström, Sami (2014). *Taking Evil Seriously*. Basingstoke: Palgrave.
- Wittgenstein, Ludwig (1921). *Tractatus Logico-Philosophicus*. Transl. D. Pears & B. McGuinness. London: Routledge and Kegan Paul, 1974.
- Wittgenstein, Ludwig (1969). *On Certainty*. Eds. G.E.M. Anscombe & G. H. von Wright. Oxford: Basil Blackwell.
- Wittgenstein, Ludwig (1980). *Culture and Value*. Transl. P. Winch. Eds. G. H. von Wright & H. Nyman. Chicago: University of Chicago Press.

Vähäpätöinen kiista

Tapio Korte

1. Aluksi

Gottlob Frege kirjoittaa teoksessa *Grundlagen der Arithmetik* (1884) seuraavasti:

Välttääkseni joutumasta tilanteeseen, jossa vähäpätöisen kiistan takia joudun vastustamaan hahmoa, jota voimme vain kiitollisina ihailia, minun on korostettava myös asioita, joista olen hänen kanssaan samaa mieltä. (GLA, 101.)

Frege mainitsema, ihailumme ansaitseva hahmo, on Immanuel Kant ja kyseinen vähäpätöinen kiista koskee aritmetiikan¹ väitteiden episteemistä luonnetta. Kantin mukaan sekä geometrian että aritmetiikan arvostelmat ovat synteettisiä *a priori* -arvostelmia (B14–17). Frege on Kantin kanssa samaa mieltä, mitä tulee synteettisten *a priori* -arvostelmien olemassaoloon ylipäätään ja erityisesti mitä tulee siihen, että geometrian arvostelmat ovat synteettisiä *a priori* -arvostelmia, mutta hän on eri mieltä aritmetiikan arvostelmien luonteesta. Kuten Kant, Frege uskoo aritmetiikan arvostelmien apriorisuuteen,² mutta hänen mukaansa ne ovat analyttisiä eivätkä synteettisiä (ks. GLA). Frege kanta liittyy läheisesti koko hänen filosofista elämäntyötään hallitsevaan hankkeeseen, logistiseen projektiin, eli pyrkimykseen osoittaa, että aritmetiikan arvostelmien totuus kytetään justifioimaan logiikan sisällä käyttämättä apuna ainoatakaan arvostelmaa, joka ei kuulu puhtaaseen logiikkaan. Frege näkemyksen mukaan arvostelmat, jotka voidaan justifioida kuvatulla tavalla, ovat analyttisiä Kantin tarkoittamassa mielessä (GLA, 4).

¹ Nykyään termillä ”aritmetiikka” tarkoitetaan yleensä erityistä matematiikan haaraa, laskentoa, joka käsittelee pääasiassa peruslaskutoimituksia. Frege aikana ”aritmetiikalla” tarkoitettiin laajempaa matematiikan osa-aluea, jota nykyään kutsutaan ”lukuteoriaksi” tai ”korkeammaksi aritmetiikaksi”.

² Tämä ei ollut mitenkään itsestään selvää. Esim. John Stuart Mill esitti, että aritmetiikan väitteet ovat paitsi synteettisiä niin myös tehtävissä vain *a posteriori* (Mill 1843).

Fregen logistinen projekti ja sen suhde Kantin väitteeseen, että aritmetiikan väitteet ovat analyttisiä, herättää välittömästi monia kysymyksiä. Mitä Kant tarkoitti analyttisillä ja synteettisillä arvostelmilla? Onko Fregen käyttämä analyttisyyden käsite sama kuin Kantin? Miksi Kant ajatteli, että aritmetiikan väitteet ovat synteettisiä? Miksi Frege ajatteli, että ne ovat analyttisiä? Miksi puhtaan logiikan väitteet ovat analyttisiä? Miten puhtaan logiikan väitteet erotetaan muista, synteettisistä, väitteistä? Tämä kirjoitus antaa minun vastaukseni muun muassa näihin kysymyksiin.

2. Analyttiset ja synteettiset arvostelmat

Kant jakaa arvostelmat *analyttisiin* ja *synteettisiin* sillä perusteella, onko arvostelman predikaattikäsite osa subjektikäsitettä vai ei. Analyttisissä arvostelmissa predikaatti on osa subjektia, kun taas synteettisissä se ei ole (A6/B10). Erotteluperusteen taustalla on Kantin logiikan teoria. Perinteisen tavan mukaan hän jakaa logiikan käsiteoppiin, arvostelmaoppiin, päättelyoppiin ja metodioppiin.³ Opit muodostavat hierarkian siten, että jälkimmäisen opin pääaihe koostuu aina sitä välittömästi edeltävän opin aiheesta: arvostelmat koostuvat käsitteistä, päättelyt koostuvat arvostelmista ja menetit päättelyistä. Käsiteopissa Kant hyväksyy perinteisen käsityksen, että käsitteet voivat sisältää toisia käsitteitä eli koostuvat osakäsitteistä, *Merkmaleista*. Kantin arvostelmaoppi puolestaan perustuu niin sanottujen arvostelmamuuotojen oppiin. Sen mukaan arvostelmasta on erotettavissa muoto ja materia. Materia koostuu arvostelman subjekti- ja predikaattikäsitteistä. Arvostelman muodon Kant analysoi jakamalla muodon ensin neljään osaan ja erottelemalla kustakin osasta kolme erilaista tapaa, jolla muoto ilmenee: kvantiteetiltaan arvostelma voi olla universaalinen, partikulaarinen tai singulaarinen, kvaliteetiltaan affirmatiivinen, negatiivinen tai ääretön, relaatioltaan kategorinen, hypoteettinen tai disjunktiiivinen ja modaliteetiltaan problemaattinen, assertorinen tai apodiktinen.⁴ Kantin käyttämä peruste luokitella arvostelma analyttiseksi soveltuu luontevasti vain

³ Niin sanotussa Jäsche-logiikassa Kant jakaa logiikan ensin elementtioppiin ja metodioppiin ja edellisen käsite-, arvostelma- ja päättelyoppiin (ks. *Log*). Perinteiset neljä logiikan osa-aluetta säilyvät tässäkin jaottelussa.

⁴ Ilmeisesti tarkoituksena on, että jokaisella arvostelmalla on jokin kvantiteetti, kvaliteetti, relaatio ja modaliteetti. 1800-luvulla käytiin keskustelua, pitääkö tämä paikkaansa. Voidaanko esimerkiksi hypoteettisista arvostelmista erottaa eri kvantiteetit? Ks. esim. Drobisch 1836, §50.

arvostelmiin, jotka ovat universaaleja, affirmatiivisia ja kategorisia. Tästä tuntuu seuraavan, että minkään muun muotoiset arvostelmat eivät voi olla analyyttisiä.⁵

Kantin tavoin myös Frege näyttää hyväksyvän käsitteiden *Merkmal*-analyysin (ks. esim. KS, 174–175), mutta Kantin analyyttisyyden kriteeri soveltuu silti erittäin huonosti yhteen Fregen käyttämään logiikan järjestelmän, käsittekirjoituksen, *Begriffsschriftin*, kanssa. Kantin logiikka perustuu pitkälti aristoteelisiin, kategorisiin, syllogismeihin, joiden perustana on arvostelmien subjekti/predikaatti-analyysiin tukeutuva arvostelmaoppi. Kantin ehdottama analyyttisen arvostelman kriteeri soveltuu luontevasti käytettäväksi tällaisen logiikan yhteydessä. Fregen kohdalla tilanne on täysin toinen. Ensinnäkin Frege erotteli arvostelmista itse arvostelman tekemisen aktin ja arvostelman sisällön. Koska arvostelman tekemisen akti on aina sama ”todeksi tunnistamisen akti” (BS, §2), Fregen *Begriffsschriftin* yhteydessä sananmukaisesti ei voida puhua erilaisista arvostelmamuodoista lainkaan. Kantin ajatusta arvostelma-muodoista voidaan tietysti muuttaa niin, että arvostelmamuotojen sijaan kyse on arvostelman sisältöjen muodoista, mutta tämäkään ei juuri helpota tilannetta, sillä *Begriffsschriftissä* arvostelmien sisällöistä ei eroteta subjektia ja predikaattia. Kantilaisen logiikan arvostelmamuodoista *Begriffsschriftissä* on jäänteitä enää singulaarisesta, kategorisesta, hypoteettisesta ja negatiivisesta arvostelmamuodosta. *Begriffsschriftissä* käytettävää argumentti/funktio-analyysiä voidaan pitää kategorisen muodon seuraajana,⁶ mutta koska argumentit ovat aina yksilöitä, jokainen *Begriffsschriftin* ”kategorisista arvostelmista” on singulaarinen. Universaalisuus ilmaistaan *Begriffsschriftissä* kvanttorin ja materiaalisen implikaation avulla, ja vaikka materiaalista implikaatiota ehkä voidaankin pitää hypoteettisen arvostelmamuodon perillisenä, Kantin analyyttisyyden kriteerin käyttökelpoisuutta se ei paranna lainkaan.

Kyetäkseen osoittamaan, että aritmetiikan väitteet ovat analyyttisiä, Frege tarvitsi analyyttisyyden kriteerin, joka on käyttökelpoinen *Begriffsschriftissä* mutta joka kuitenkin on riittävän lähellä Kantin kriteeriä,

⁵ Poikkeuksia saattaisivat olla singulaariset ja äärettömät arvostelmat, mutta niistä Kant toteaaakin, että käyttäytyvät puhtaassa logiikassa kuten universaalit ja affirmatiiviset arvostelmat (A71/B96–97).

⁶ Frege itsekin toteaa, että *Begriffsschriftissä* argumentin ja funktion käsitteet korvaavat perinteiset subjektin ja predikaatin käsitteet (BS, XIII).

jotta ylipäätään voidaan puhua samasta analytytisyden käsitteestä. Fregen ehdotus on, että analytytisinä arvostelmina Kantin esittämässä mielessä voidaan pitää sellaisia arvostelmia, joiden totuuden justifiointiin ei tarvita kuin perimmäisiä logiikan lakeja. Ennen kuin selvitän, miten Frege aikoi osoittaa aritmetiikan väitteiden analytytisyden tätä kriteeriä käyttäen, on selvítettävä onko Fregen analytytisyden käsite riittävän lähellä Kantin käyttämää käsitettä, jotta voidaan mielekkäästi sanoa heidän käsityksensä aritmetiikan arvostelmien episteemisestä luonteesta eroavan toisistaan. Mielestäni näin on. Vaikka varsinainen analytytisyden kriteeri Kantilla on predikaatin sisältyminen subjektiin, hän toteaa myös, että analytytisyden prinssiippi on ristiriidan periaate, koska, Kant tuntuu ajattelevan, analytytisen arvostelman totuuden kiistäminen johtaa ristiriitaan (A151/B190).⁷ Tämä pitää paikkaansa myös arvostelmista, jotka ovat analytytisiä Fregen käyttämän kriteerin perusteella. Toisaalta Kant tuntuu ajattelevan, että jokainen ristiriitainen arvostelma on analytytinen, sillä synteettisten, siis ei-analytytisten, arvostelmien totuuden kieltäminen ei ilmeisesti hänen mukaansa ole ristiriitaista.⁸ Hän on siten hyvin lähellä analytytisyden kriteeriä, jonka mukaan arvostelma on analytytinen, jos ja vain jos sen totuuden kiistäminen johtaa ristiriitaan. Tämä on itse asiassa myös Fregen analytytisyden määritelmän ydin: arvostelma on analytytinen, jos ja vain jos sen negaatio johtaa loogiseen ristiriitaan.⁹

Loogista ristiriitaa voidaan käyttää analytytinen/synteettinen-erottelun tekemiseen kuitenkin vain, jos tiedämme millainen on looginen ristiriita. Millä perusteella Kant ajatteli, että väitteen ”jos joku A on B, niin joku B on A” negaatio johtaa loogiseen ristiriitaan, kun taas väitteen ” $1 + 1 = 2$ ” negaatio ei, vaikka äkkiseltään molempien negaatiot näyttävät johtavan jonkinlaiseen ristiriitaan? Miksi edellisessä tapauksessa ristiriita on looginen ja jälkimmäisessä ei? Erottelun taustalla on syvemmälle Kantin filosofiaan

⁷ Teoksessa *Prolegomena* Kant kirjoittaa: ”Koska affirmatiivisen analytytisen arvostelman predikaattia jo etukäteen ajatellaan subjektin käsitteessä, sitä ei voi kieltää subjektista ilman ristiriitaa.” (*Prol*, 267; suomennos kirjoittajan.)

⁸ *Prolegomenassa* Kant esittää, että synteettisten arvostelmien totuus riippuu jostain muusta kuin ristiriidan periaatteesta (*Prol*, 267).

⁹ Myös Quinen mukaan Frege tavoitti Kantin analytytisyden käsitteen ytimen (Quine 1951, 21). Asia on kuitenkin jonkin verran kiistanalainen (ks. esim. Katz 1988, 52ff). Tämän kirjoituksen kanalta on kuitenkin tärkeää vain, uskoiko Frege tavoittaneensa olennaisen Kantin analytytisyden käsitteestä. Uskon, että näin on.

menevä peruste. Kant jakoi ihmisen kognition kahteen tekijään eli aistimellisuuteen, jonka kautta olemme yhteydessä mielenulkoiseen todellisuuteen, ja ymmärrykseen, jonka avulla kykenemme ajattelemaan aistimellisuuden meille antamaa todellisuutta. Molempia kykyjä vastaavat omanlaisensa representaatiot. Aistimellisuuden representaatioita ovat intuitiot, jotka ovat singulaarisia ja välittömässä suhteessa todellisuuteen. Ymmärrykseen puolestaan kuuluvat käsitteet, yleiset representaatiot, joiden avulla intuitioiden tarjoamaa todellisuutta ajatellaan. Intuitiot ja käsitteet voivat olla empiirisiä tai puhtaita riippuen siitä liittyykö niihin aistimuksia vai ei. Puhtaista intuitioista ja käsitteistä Kant toteaa, että ”puhdas intuitio sisältää pelkästään muodon, jonka alla jokin intuoidaan, ja puhdas käsite yksistään kohteen ajatteleminen muodon” (A51). Tämä erottelu antaa mahdollisuuden tarkentaa analyttisyyden määritelmää. Voitaneen sanoa, että arvostelma on analyttinen, jos ja vain jos sen negaatio johtaa ristiriitaan sellaisten arvostelmien kanssa, jotka sisältävät vain puhtaita käsitteitä. Uskon, että tämä analyttisyyden luonnehdinta tavoittaa olennaisen sekä Kantin että Fregen käyttämästä analyttisyyden käsitteestä.

3. Fregen logistinen projekti

Osoittaakseen, että aritmetiikan arvostelmat ovat analyttisiä edellisessä luvussa kuvatussa mielessä, Frege tarvitsi keinon osoittaa, milloin väitteen negaatio on ristiriidassa ainoastaan puhtaita käsitteitä sisältävien arvostelmien kanssa. Fregen hankkeen perusajatus on selkeä: väite on analyttinen, jos ja vain jos on olemassa sellainen validi argumentti, jonka johtopäätös kyseinen väite on ja jonka premissit sisältävät vain puhtaita käsitteitä. Mikäli tällainen argumentti jollekin väitteelle on olemassa, sen negaatio on ristiriidassa argumentin premissien kanssa, eli väitteen analyttisyys on osoitettu. Aritmetiikan väitteiden osoittaminen analyttisiksi tapahtuu valitsemalla joukko analyttisiä väitteitä, joista voidaan muodostaa validit argumentit aritmetiikan väitteille.

Hankkeen toteuttaminen tässä muodossa edellyttää kahden ongelman ratkaisemista: ensinnäkin tarvitaan järjestelmä, joka takaa argumenttien validiuden, ja toiseksi tarvitaan jokin keino osoittaa, että argumentin premissiksi valitut arvostelmat todella sisältävät vain puhtaita käsitteitä. *Begriffsschrift* on järjestelmä, jonka tarkoitus on ratkaista hankkeen molemmat

ongelmat. Se on teoria, jonka perustana on yhdeksän aksiomaa, joista annettuja päättelysääntöjä noudattaen johdetaan teorian teoreemat. Jokainen *Begriffsschrift*issä suoritettu deduktio on validi argumentti, joten mikäli voimme olla vakuuttuneita, että aksiomat eivät sisällä empiirisiä käsitteitä, jokainen teorian teoreema on analyttinen. Jos aritmetiikan väitteet ovat *Begriffsschrift*in teoreemoja, kuten Frege pyrkii osoittamaan, aritmetiikan väitteet ovat analyttisiä.

Frege'n ratkaisu ensimmäiseen ongelmaan on teknisesti nerokas, vaikkakaan se ei ole filosofisesti kovin mielenkiintoinen. Kantin logiikan käyttämien kahdentoista arvostelmamuodon ja subjekti/predikaatti-analyysin sijaan *Begriffsschrift*in väitteet ovat joko funktiosta ja argument(e)ista tai singulaaritermeistä identiteettimerkin avulla muodostettuja atomilauseita tai kahden konnektiivin, implikaation ja negaation, sekä identiteetin¹⁰ ja ensimmäisen tai korkeamman tason universaalikvanttorin avulla muodostettuja molekylaarisia lauseita. Kantin arvostelmamuodoista modaliteettien Frege ei katso lainkaan kuuluvan logiikkaan, sillä, kuten Kantkin toteaa, ne eivät lisää mitään arvostelman sisältöön, vaan kuvaavat pelkästään suhtautumistamme siihen (BS, 13). Ero kategorisen, hypoteettisen ja disjunkttiivisen arvostelman välillä on Frege'n mukaan pelkästään kieliopillinen (BS, 13). Mitä kategoriin arvostelmiin tulee, niin hänellä on selvästi mielessään universaaliset ja partikulaariset kategoriset arvostelmat, sillä kummatkin voidaan ilmaista kvanttorin, implikaation ja negaation avulla: ”Jokainen A on B ” ilmaistaan *Begriffsschrift*issä ” $(\forall x)(Ax \rightarrow Bx)$ ” ja ”joku A on B ” ” $\sim(\forall x)(Ax \rightarrow \sim Bx)$ ”.¹¹ Hypoteettiset ja disjunkttiiviset arvostelmat voidaan muodostaa niin ikään implikaatiomerkin ja negatiomerkin avulla. ”Jos A niin B ” ilmaistaan ” $A \rightarrow B$ ” ja ” A tai B ” ” $\sim A \rightarrow B$ ”. Erottelun kieliopillisuudella Frege ilmeisesti tarkoittaa, että logiikan kannalta nämä arvostelmat eivät edusta erilisiä muototyyppisiä, jotka olisi tarkoituksenmukaista sijoittaa eri muotokategorioihin. *Begriffsschrift*in sallitut päättelysäännöt, *modus ponens*, kaksi

¹⁰ *Begriffsschrift*issä identiteettimerkillä ilmaistaan, että sen sitomilla ilmauksilla on sama käsitteellinen sisältö eli että missä hyvänsä argumentissa ilmaukset voidaan korvata toisillaan muuttamatta argumentin validisuutta. Nykyaikaisesta identiteetinmerkistä poiketen *Begriffsschrift*in identiteettimerkki voi sitoa paitsi singulaaritermejä niin myös kokonaisia lauseita.

¹¹ Frege'n *Begriffsschrift*issä käyttämä kohtalaisen monimutkainen notaatio ei saanut seuraajia. Käytän sen sijaan kanonista notaatiota.

kvanttoreita koskevaa sääntöä ja substituutiosääntö takaavat, että jokainen siinä suoritettu päättely on validi.

Toinen ongelma, *Begriffsschriftin* aksiomien analyyttisyys, on filosofisesti edellistä mielenkiintoisempi. Fregen hanke vaikuttaa olevan vaikeuksissa jo ennen kuin se on alkanutkaan. Kantin mukaan nimittäin kognitioon tarvitaan sekä aistimellisuuden tarjoamia intuitioita että ymmärryksen käsitteitä. Kuten Kantin paljon lainattu lausahdus toteaa: ”Ilman aistimellisuutta yksikään kohde ei tulisi meille annetuksi ja ilman ymmärrystä ei yhtäkään kohdetta ajateltaisi. Ajatukset ilman sisältöä ovat tyhjiä, intuitiot ilman käsitteitä ovat sokeita.” (A51/B75.) Tämä lainaus näyttää tekevän aritmetiikan arvostelmien analyyttisyydestä mahdotonta kahdesta syystä. Ensinnäkin aritmetiikka käsittelee lukuja, ja luvut näyttävät olevan jonkinlaisia objekteja. Jos arvostelma on analyyttinen, sen totuus riippuu pelkästään arvostelmista, jotka koostuvat puhtaista käsitteistä eli sen totuus on intuitioista riippumaton. Aritmetiikan ei siis pitäisi pystyä koskemaan objekteja lainkaan. Toiseksi, aritmetiikan väitteet näyttävät olevan informatiivisia. Tietomme luvuista ja niitä koskevista laeista näyttää jatkuvasti lisääntyvän. Mutta analyyttisten arvostelmien pitäisi olla tyhjiä, koska niiden totuus on riippumaton intuitioista. Kantin mukaan ainoastaan synteettiset arvostelmat ovat ampliativisia eli lisäävät tietoa; analyyttiset arvostelmat ovat eksplikatiivisia (A7/B11).

Frege lähestyy näitä ongelmia eri tavoin. Objektiin viittaamisen mahdollisuudesta hän toteaa ainoastaan, että hänen on pakko kiistää Kantin näkemys, että objektit on annettu meille ainoastaan aistimellisuudessa eli siis intuitioiden välityksellä. Hän vetoaa siihen, että nolla ja yksi ovat objekteja, mutta niitä ei voida aistia (GLA, 101). Tämä ei tietenkään riitä perusteeksi, sillä suhde intuitioihin voi syntyä jo aksiomien tasolla. Tätä mahdollisuutta Frege ei missään sulje pois eikä missään edes epäile koko asiaa. Palaan asiaan seuraavassa luvussa.

Aritmetiikan arvostelmien informatiivisuuden ongelmaan Frege vastaa suoraan. Hänen mukaansa se, että analyyttiset arvostelmat vaikuttavat tyhjiltä ja epäinformatiivisilta, johtuu Kantin käyttämien määritelmien yksinkertaisesta rakenteesta. Jos määritelmissä uusia käsitteitä kootaan pelkästään luettelemalla sen *Merkmale*t, Fregen mukaan tosiaankin näyttää siltä, että laatikosta otetaan ulos se, mikä juuri sinne pantiin. (GLA, 101.)

Fregen omat määritelmät, hän korostaa, eivät ole sellaisia, vaan ne ovat hedelmällisiä; niissä ”jokainen määritelmän elementti on läheisesti, sanoisinpa jopa orgaanisesti, sidottu toinen toiseensa” (GLA,100). Frege epäilemättä tiesi, että hänen hedelmällisen määritelmän luonnehdintansa on kaikkea muuta kuin selvä, sillä hän yritti havainnollistaa asiaa geometrisellä esimerkillä: Käsitteen ekstensioita voidaan esittää alueena tasolla. Kantin käyttämällä määritelmällä määritellyn käsitteen ekstensio on sen kaikkien *Merkmalien* alueiden leikkaus. Frege toteaa, että tällaisessa määritelmässä ei synny mitään uutta, sillä määritellyn käsitteen ekstensio määräytyy jo olemassa olevien aluerajojen mukaan. Hedelmällisissä määritelmässä tilanne on toinen. Niissä alueen (ekstension) rajaamiseen ei samalla tavalla käytetä jo ennestään olemassa olevia aluerajoja. Kun käytämme tällä tavalla määriteltyjä käsitteitä, emme voi etukäteen tietää mihin lopputulokseen päädymme. Siksi analyttisetkin arvostelmat voivat olla ampliativisia. (GLA, 101.)

Sekä Fregen hedelmällisen määritelmän luonnehdinta että sen tueksi annettu vertaus ovat toivottoman kielikuvamaisia, mikä kertoo mielestäni siitä, että Fregellä itselläänkään ei ollut selkeää kuvaa, mitä määritelmässä, ja erityisesti hedelmällisissä määritelmässä, tapahtuu. Luonnehdintojen perusteella on mahdoton sanoa, mitä Fregellä tarkalleen ottaen on mielessään. Olennaista kuitenkin on, että vaikka Frege myöhemmissä kirjoituksissa käsittelee paljonkin määritelmiä, hän ei enää kertaakaan palaa hedelmällisiin määritelmiin eikä yritä sen avulla selittää ongelmaa, miten analyttiset väitteet voivat olla informatiivisia.

Aritmetiikan informatiivisuuden ongelma voidaan esittää toisinkin. Aristoteelisen tieteenihanteen mukaisesti Frege hyväksyy ajatuksen, että teorioissa kaikkea ei voida määritellä. Jokainen teoria, ilmeisesti myös *Begriffsschrift*, sisältää, kuten hän sanoo, ”loogisesti primitiivisiä elementtejä”, joita ei voida määritellä (KS, 288). *Begriffsschrift*issä tällaisia näyttävät olevan muun muassa negaatio, implikaatio, universaalikvanttori, sillä *Begriffsschriftin* sisällä niitä ei määritellä. Frege kyllä selittää, mitä kyseiset ilmaukset merkitsevät, mutta selitykset eivät selvästikään ole osa itse *Begriffsschriftiä*. Aritmetiikan informatiivisuuden ongelma voidaan nyt esittää seuraavasti: Koska validit argumentit ovat eksplikatiivisia, kaikki, mitä premisseistä voidaan johtaa, jo sisältyy tavalla tai toisella premisseihin. Miten on mahdollista, että koko aritmetiikka sisältyy *Begriffsschriftin* yhdeksään

aksiomaan, jotka koostuvat muutamasta loogisesti yksinkertaisesta elementistä?

Ilmausten *Sinnin* ja *Bedeutungin* erottaminen tarjoaa tälle aritmetiikan informatiivisuuden ongelmalle selityksen, jossa ei tarvitse vedota teoreettisesti epäselviin hedelmällisiin määritelmiin. Asian yksityiskohtainen käsitteleminen ei tässä yhteydessä tilan puutteen vuoksi ole mahdollista, joten voin vain luonnostella, millainen selitys voisi olla. Fregen mukaan käsitteet ovat ekstensionaalisia, eli käsitteiden samuuden kriteeri on ekstensioiden samuus (esim. NS, 128). Käsitteiden ekstensionaalisuuden seuraus on niin sanottu *praedicatum subiecto inest* -periaate¹² – käsitteen *Merkmaleja* ovat kaikki käsitteet, jotka voivat olla sellaisen universaalisen arvostelman predikaatteina, joissa se on subjektina. Periaatteen nojalla esimerkiksi ihmisen käsitteeseen sisältyvät kaikki käsitteet, jotka voivat olla arvostelman ”jokainen ihminen on *P*” predikaattina *P*, eli kaikki käsitteet, jotka ovat universaalisti totta ihmisistä. Lopputulos on semantiikka, jossa käsitteet ovat sisällöltään erittäin rikkaita; ne sisältävät paljon sellaista, mitä emme koskaan saa tietää. Tämä puolestaan on mahdollista, koska ymmärtäessämme käsiteilmauksia emme itse asiassa käsitä käsitteitä, jotka ovat niiden *Bedeutungeja*, vaan käsitämme käsiteilmausten *Sinnejä*. Ilmauksilla, joilla on yksinkertainen *Sinn*, voi olla äärimmäisen kompleksinen *Bedeutung*. Yksilönimien kohdalla asiassa ei ole mitään kummallista, sillä vaikuttaa kiistattomalta, että voimme viitata yksilöihin, vaikka emme tietäisi kaikkia niiden ominaisuuksia. Fregen *Sinn/Bedeutung*-semantiikassa tilanne on sama käsiteilmausten kohdalla: voimme ymmärtää käsiteilmauksia, tajuta niiden *Sinnin*, tietämättä kaikkia käsitteiden *Merkmaleja*. Tämä tarjoaa ratkaisun aritmetiikan informatiivisuuden ongelmaan. *Begriffsschriftin* peruskäsitteet ovat *Sinnien* tasolla äärimmäisen yksinkertaisia; ne ovat sen ”loogisesti primitiivisiä elementtejä”, joita teoriassa ei määritellä. *Bedeutungin* tasolla ne kuitenkin ovat erittäin monimutkaisia. Aritmetiikan arvostelmat juontuvat tästä rikkaasta käsitteellisestä todellisuudesta. Aritmetiikka on löytöretki joidenkin käsitteiden monimutkaiseen maailmaan.

¹² Näin on seuraavasta syystä: Jos väite ”jokainen *A* on *B*” on totta, niin silloin käsitteen *A* ekstensio on sama kuin käsitteen *A*, joka on *B*. Jos käsitteet ovat ekstensionaalisia, *A* on sama käsite kuin *A*, joka on *B*, josta näemme että *B* on *A*:n *Merkmal*.

4. Miksi *Begriffsschriftin* aksiomat eivät edellytä intuitioita?

Erikoista kyllä Frege ei käsittele lainkaan kysymystä, ovatko *Begriffsschriftin* aksiomat analyttisiä ja ovatko sen käsitteet loogisesti primitiivisiä, vaan tuntuu pitävän sitä itsestään selvänä. Ainoa kirjoitus, jossa Frege sivuaa asiaa, on vuosina 1903 ja 1906 ilmestynyt David Hilbertin teokseen *Grundlagen der Geometrie* keskittyvä artikkelisarja *Über die Grundlagen der Geometrie* (KS, 262–272; 281–323). Kirjoituksessa Frege hahmottelee menetelmää, jolla teorian aksiomien riippumattomuus toisistaan voitaisiin osoittaa, ja toteaa, että sellaisen menetelmän toteuttaminen edellyttää, että ”pystymme kertomaan, mitä ovat logiikan päättelyt ja mitkä [käsitteet ja relaatiot] ovat osa logiikkaa” (KS, 322–323). Ainakin minä pidän erikoisena, että Frege pitää vielä vuonna 1906 kysymystä ratkaisemattomana, vaikka hän itse oli esittänyt logiikan teorian jo vuonna 1879 ja yrittänyt osoittaa vuonna 1893, että aritmetiikan väitteet voidaan todistaa logiikan aksiomeista ja että siksi ne ovat analyttisiä.

Vastaus ongelmaan saattaa löytyä vertaamalla keskenään Fregen ja Kantin käsityksiä logiikan luonteesta. Kant luokitteli logiikan sen yleisyyden perusteella yleiseen ja erityiseen logiikkaan ja yleisen logiikan edelleen puhtaaseen ja sovellettuun. Logiikka on Kantin mukaan yleistä, kun se ”sisältää ajattelun ehdottoman välttämättömät säännöt, joita ilman mitään ymmärryksen käyttöä ei esiinny, ja se siis käsittelee näitä sääntöjä katsomatta niiden kohteiden eroihin, joihin ymmärrys kohdistuu” (A52/B76). Erityinen logiikka sisältää yleisestä poiketen sääntöjä, jotka riippuvat objekteista, joita arvostelmat koskevat. Yleinen logiikka on puhdasta, mikäli se ei sisällä lainkaan sääntöjä, jotka riippuvat ihmisen muista kyvyistä kuin ymmärryksestä eli on riippumaton aisteista, kuvittelusta, muistia koskevista laeista, ajattelutavoista, taipumuksista jne. (A53/B77). Kant tähdentää, että yleinen ja puhdas logiikka sisältää vain *a priori* -periaatteita, ja kutsuu sitä ymmärryksen ja järjen kaanoniksi (A53/B77).

Pelkkä periaatteiden apriorisuus ei vielä tee logiikasta puhdasta. Kant tähdentää, että ollakseen puhdasta logiikan on oltava muodollista ja täysin riippumatonta arvostelmien sisällöistä (A53/B77). John MacFarlane on äskettäin esittänyt, että logiikan muodollisuus on seurausta Kantin logiikan yleisyydestä ja uudistuksena verrattavissa transsendentaaliseen idealismiin (MacFarlane 2000, 80). Olen eri mieltä. Argumenttien validiuden riippumattomuus arvostelmien sisällöistä on ikaikainen havainto, ja se

mahdollisti muun muassa ensimmäisen tunnetun päättelyjärjestelmän, aristoteelisen syllogistiikan. Sillä, pidetäänkö logiikan lakeja epistemisinä tai metafyyisinä, tai vaikkapa empiirisinä, ei ole tekemistä tämän asian kanssa, sillä muodollisuus tarkoittaa riippumattomuutta arvostelmien sisällöistä eikä riippumattomuutta metafysiikasta, epistemologiasta tai empiirisestä todellisuudesta,¹³ kuten MacFarlane tuntuu ajattelevan. Muodollisuudesta seuraa logiikan yleisyys eikä päinvastoin, sillä ajattelun muoto on, kuten Kant tähdentää, sisällöstä riippumaton.

Muodon ja sisällön erottaminen oli myös perusteena keskiajalla tehtyyn erotteluun kategoremaattisiin eli sisällöllisiin ja synkategoremaattisiin eli muodollisiin termeihin. Muodollisen logiikan termistö on puhtaan synkategoremaattista. Tämän seurauksena muodollisen logiikan arvostelmilla ei ole sisältöä, joten ne eivät myöskään voi olla sanan varsinaisessa merkityksessä tosia tai epätosia. Logiikan väitteet ovat skeemoja, kuten asia nykyään ilmaistaisiin. Toisinaan Kant erottelee toisistaan materiaalisen ja muodollisen totuuden (esim. *Log*, 51). Arvattavasti Kant olisi ollut valmis sanomaan, että logiikan väitteet voivat olla vain muodollisesti tosia.

Vaikka Frege epäilemättä on Kantin kanssa samaa mieltä logiikan muodollisuudesta siinä mielessä, että *Begriffsschrift* sisältää vain synkategoremaattisia termejä, hän yhtä epäilemättä on eri mieltä siitä, että logiikan väitteet eivät olisi tosia tai epätosia ja ettei niillä olisi sisältöä. Frege tähdentää, että hänen looginen järjestelmänsä, *Begriffsschrift*, ei ole pelkkä *calculus* vaan aito *lingua*. Uskon, että *Begriffsschriftin* kieliluonteella Frege viittaa nimenomaan siihen, että sen lauseet eivät ole pelkkiä skeemoja vaan niillä on sisältö ja ne ovat tosia samassa mielessä kuin mitkä hyvänsä sisällöt ovat tosia. Fregen suuri keksintö oli – joka, niin hän uskoi, erotti *Begriffsschriftin* kaikista muista sen ajan logiikan järjestelmistä – miten lauseilla, jotka sisältävät pelkkiä synkategoremaattisia termejä, voi olla sisältö. Esimerkiksi skeemakirjaimilta näyttävät ” p ” ja ” q ” *Begriffsschriftin* lauseessa ” $p \rightarrow (q \rightarrow p)$ ” ovat itse asiassa sidottuja muuttujia – $(\forall p)(\forall q)(p \rightarrow (q \rightarrow p))$ – ja se sanoo, että kaikista objekteista, p ja q , pätee että p implikoi sen että q implikoi p :n. Tämä on väite, jolla on sisältö, vaikka se sisältää ainoastaan synkategoremaattisia termejä!

¹³ Edes Kantin näkemys ei voi olla, että muodollinen logiikka on riippumatonta epistemologiasta. Jos se olisi, niin miten puhtaiden *a priori* -käsitteiden johtaminen arvostelmamuodoista voisi olla mahdollista?

Sama pätee kaikista yhdeksästä *Begriffsschriftin* aksioomasta: niillä kaikilla on sisältö ja ne ovat tosia, vaikka ne eivät sisällä sisällöllisiä, kategoriaattisia termejä. Fregen logistista projektia ajatellen tämä onkin välttämätöntä. Mikäli *Begriffsschriftin* aksioomilla ei olisi sisältöä ja mikäli ne eivät olisi tosia, aritmetiikan johtaminen näistä aksioomista osoittaisi, että aritmetiikan väitteillä ei ole sisältöä ja että ne eivät ole tosia tai epätosia. Tämä on johtopäätös, jonka olisi hyväksynyt matematiikan formalismin kannattaja, mutta ei Frege, joka oli formalismin kiivas vastustaja (esim. GGAI, §104–123).

Käsitykseni *Begriffsschriftin* luonteesta eroaa täysin esimerkiksi Joan Weinerin näkemyksestä. Hänen mukaansa *Begriffsschrift* on teoria, jonka loogisesti primitiivisiä elementtejä ovat muun muassa negaatio, implikaatio ja kvanttori (Weiner 1990, 233). Weiner kyllä näyttää olevan oikeassa, sillä kyseisiä negaation, implikaation ja kvanttorin merkkejä ei määritellä itse *Begriffsschriftissä*, vaan niiden merkitys selitetään luonnollista kieltä käyttäen, ja koska ilmaukset, joita teoriassa ei määritellä, ovat sen loogisesti primitiivisiä elementtejä. Mielestäni hän on kuitenkin väärässä, sillä negaatiomerkki, implikaatiomerkki ja kvanttorimerkki eivät ole kategoriaattisia ilmauksia. Ne eivät ole ilmauksia, joilla olisi sisältö, joten niitä ei myöskään voida määritellä. Alkuperäisessä, vuoden 1879 *Begriffsschriftissä* ei ole olemassa negaatio-, implikaatio- ja kvanttori-funktioita, jotka olisivat kyseisten ilmausten sisältöjä.¹⁴

Tulkintaani, jonka mukaan *Begriffsschriftin* sisältämät termit ovat synkategoriaattisia, sisällyksettömiä ilmauksia, joita sen vuoksi ei määritellä, tukee kaksi seikkaa. Ensinnäkin Frege kyllä mainitsee eri yhteyksissä esimerkkejä loogisesti primitiivisistä elementeistä, kuten objekti (KS, 134), funktio (KS, 279), käsite (KS, 167), arvostelma (KS, 150), ajatus ja totuus (KS, 370n), mutta missään hän ei väitä, että negaatio, implikaatio tai kvanttori olisivat sellaisia. Erityisesti objektin ja funktion käsitteiden primitiivisyys *Begriffsschriftissä* on ilmeistä, sillä kuten edellä esitin, sen lauseet ovat väitteitä, jotka koskevat nimenomaan objekteja ja funktioita. Me emme voisi ymmärtää *Begriffsschriftin* väitteitä, elleimme tietäisi, mitä objektit ja funktiot ovat, mutta,

¹⁴ Teoksessa *Grundgesetze der Arithmetik* (1893) tilanne on eri. Siinä Frege esittelee logiikkansa primitiiviset merkit funktioniminä. Uskon, että hän teki näin ainoastaan virtaviivaistaakseen logiikkaansa. Negaatiota, implikaatiota ja kvanttoria voitaisiin ehkä kutsua ”muodollisiksi funktioiksi”.

yhtä kaikki, objektin ja käsitteen käsitteet ovat primitiivisiä käsitteitä, joita ei voida määritellä *Begriffsschrift*issä mutta ei myöskään missään muussa kielessä. Ne ovat kaiken ajattelun edellytyksiä, eikä niitä näin ollen voida määritellä lainkaan.

Toinen tulkintaani tukeva seikka on se, että Russellin paradoksin paljastuttua vuonna 1902 ja aksioman V paljastuttua syylliseksi, Frege totesi, että hän oli alusta asti epäillyt kyseisen aksioman johtavan vaikeuksiin (GGAI, 253). Jo teoksen *Grundgesetze der Arithmetik* ensimmäisen osan johdanto-osuudessa Frege kirjoittaa seuraavasti:

Kiistaa voi syntyä, sikäli kuin minä voin nähdä, vain aksiomasta, joka koskee arvonkuluja (V) ja jota loogikot kenties eivät ole muotoilleet selkeästi. Se on kuitenkin se, mikä ihmisillä on mielessään, kun he puhuvat käsitteiden ekstensioista (GGAI, vii).

Peruslain V mukaan kahdella käsitteellä on sama arvonkulku, jos ja vain jos ne ovat tosia täsmälleen samoista objekteista. Frege ei missään määrittele arvonkulun käsitettä, vaan toteaa pelkästään, että kahdella funktiolla on sama arvonkulku, jos ja vain jos niillä on aina sama arvo samalla argumentilla.¹⁵ Hän vertaa arvonkulun käsitettä ekstension käsitteeseen ja toteaa, että käsitteiden ollessa kyseessä arvonkulku on se, mitä perinteisesti kutsutaan käsitteen ekstensioksi (GGAI, §3). Tulkintani selittää, miksi Frege epäili peruslakia V jo vuosia ennen kuin sen ongelmat tulivat ilmi. Frege epäilemättä ajatteli, että arvonkulun käsite on samalla tavoin primitiivinen, sanoisinko kategorinen, kuin objektin ja funktion käsitteet ja että peruslaki V on tosi samasta muodollisesta syystä kuin muutkin aksiomat. Frege selvästi epäröi ja hyvästä syystä. Aksioma V oli kuitenkin välttämätön luvun käsitteen määrittelyssä, joten Fregen valitsemasta lähtökohdasta käsin logisismi ei onnistu ilman sitä.

Russellin paradoksin myötä Frege oli pakotettu pohtimaan logiikkansa perusteita. Vuonna 1906, kirjoituksessa ”Über die Grundlagen der Geometrie” (KS, 322), Frege tuntuu myöntävän, että logiikan käsitteiden

¹⁵ Tämä ei kelpaa arvonkulun määritelmäksi, sillä, kuten Frege itse toteaa teoksessa *Grundlagen der Arithmetik* vastaavan muotoisesti suunnan käsitteen määritelmästä (GLA, §66), emme sen perusteella pysty sanomaan mistään mielivaltaisesta objektista, onko se arvonkulku vai ei.

erottaminen kaikista muista käsitteistä on kaikkea muuta kuin selvää, toisin kuin hän oli ajatellut. Mikäli tulkintani on oikea, asia voidaan ilmaista niin, että Frege huomasi, että eroa synkategoremaattisiin ja kategoremaattisiin ilmauksiin ei voida yksiselitteisesti tehdä; ero muodon ja sisällön välillä ei olekaan yksiselitteinen. Ehkä negatiomerkki sittenkin merkitsee yksipaikkaisen ja implikaatio kaksipaikkaisen funktion? Jos näin on, niin miksi nämä funktiot ovat osa logiikkaa, mutta moni muu samankaltainen funktio ei ole? Samaten herää kysymys *Begriffsschriftin* ja luonnollisen kielen välisestä suhteesta. Negaatio ja implikaatio eivät selvästi ole primitiivisiä elementtejä, sillä molemmille voidaan antaa yksiselitteinen määritelmä. Niitä ei kuitenkaan voida määritellä *Begriffsschriftissä* itsessään, joten *Begriffsschrift*, jonka piti olla täydellinen kieli, on perustavalla tavalla riippuvainen luonnollisesta kielestä. Russellin paradoksi ei siis vaikuttanut ainoastaan Fregen logistiseen projektiin, vaan sillä oli seurauksia koko Fregen filosofiaan.

Turun yliopisto

Kirjallisuus

- Drobisch, Moriz (1836). *Neue Darstellung der Logik nach ihren einfachsten Verhältnissen. Nebst einem logisch-mathematischen Anhang*. Leipzig: Verlag von Leopold Voss.
- Frege, Gottlob (1993 [1879]). *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Teoksessa I. Angelelli (toim.), *Begriffsschrift und andere Aufsätze*, toinen laitos. Hildesheim: Georg Olms Verlag.
- Frege, Gottlob (1884). *Grundlagen der Arithmetik. Eine logisch mathematische Untersuchung über den Begriff der Zahl*. Breslau: Verlag von Wilhelm Koebner. [GLA]
- Frege, Gottlob (1893, 1903). *Grundgesetze der Arithmetik I/II*. Jena: Verlag Herman Pohle. [GGAI/GGAII]
- Frege, Gottlob (1967). *Kleine Schriften*. Toim. I. Angelelli. Hildesheim: Olms. [KS]
- Frege, Gottlob (1969). *Nachgelassene Schriften*. Toim. H. Hermes, F. Kambartel & F. Kaulbach. Hamburg: Felix Meiner. [NS]
- Kant, Immanuel (1911 [1783]). *Prolegomena*. Teoksessa Königlich Preussischen Akademie der Wissenschaften (toim.), *Kants gesammelte Schriften*, vol. IV. Berlin: G. Reimer. [Pro]
- Kant, Immanuel (1923 [1800]). *Logik*. Teoksessa Königlich Preussischen Akademie der Wissenschaften (toim.), *Kants gesammelte Schriften*, vol. IX. Berlin: G. Reimer. [Log]
- Kant, Immanuel (1998 [1781, 1787]). *Kritik der reinen Vernunft*. Hamburg: Meiner Verlag. [A/B]
- Kant, Immanuel (2013 [1781, 1787]). *Puhtaan järjen kritiikki*. Suom. M. Nikkarla ja K. Ranki. Helsinki: Gaudeamus. [A/B]
- Katz, Jerrold (1988). *Cogitations*. Oxford: Oxford University Press.

- MacFarlane, John (2000). *What Does It Mean to Say That Logic Is Formal?* Pittsburgh: University of Pittsburgh.
- Mill, J. S. (1974 [1843]). *System of Logic, Ratiocinative and Inductive*. Teoksessa J. M. Robson (toim.), *Collected Works of John Stuart Mill*, vol. 7–8. Toronto: University of Toronto Press.
- Quine, W. V. O. (1951). "Two Dogmas of Empiricism", *The Philosophical Review* 60, 20–43.
- Weiner, Joan (1990). *Frege in Perspective*. Ithaca: Cornell University Press.

Bradley's *Reductio* of Relations and Formal Ontological Relations

Jani Hakkarainen & Markku Keinänen

1. Introduction: Bradley's *Reductio* of Relations

Bradley's most systematic attack on relations occurs in Book 1, Chapter III (*Relation and Quality*) in *Appearance and Reality* (1893):

- (1) Qualities *without* relations are impossible, or at least not fully intelligible (AR, III, 21–5).¹
- (2) Equally with qualities together *with* relations (AR, III, 25–7).
- (3) Relations *without* qualities or relata are “nothing” (AR, III, 27).
- (4) Relations *with* qualities or relata are “unintelligible” (AR, III, 27–8).²

We shall call these intertwined arguments “Bradley's *reductio* of relations”. Their context consists of the problem Bradley poses in the beginning of Chapter II (*Substantive and Adjective*): What is the relation between “a thing” and its “properties”? (AR, III, 16.) In order to solve this problem of monadic inherence, Bradley soon proposes that the thing is its properties *in a relation* (AR, II, 17). His example is that a lump of sugar as a unity would be white, sweet and hard in some relation (AR, II, 16–7). However, then we face the problem how *many* entities (plurality of properties), even in a relation, constitute *one* entity (a unity) over and above the many (addition to being). This unity-in-complexity or complex-unity problem leads to the discussion of properties, or relata, and relations in Chapter III.

Typically, the entire Bradley's *reductio* of relations is ignored or set aside in contemporary metaphysics and only the relation regress occurring in the fourth part is discussed. However, when we are responding to Bradley, it is

¹ References to Bradley's *Appearance and Reality* are to the second edition (1897), hereafter cited as “AR” followed by chapter and page numbers.

² Bradley uses “term” instead of “relatum” but we take it for granted that “relatum” and “relata” are unproblematic here.

not sufficient that we respond to the second strand of the second horn of his dilemma in the fourth argument, which is known as his relation regress or just “Bradley’s regress”. We have to be as certain as possible that our response is not subject to his *reductio* of relations in Chapter III since these arguments are, indeed, intertwined. They consist of several eliminative moves, the most relevant of which for the present purposes are the relation regress and especially *Bradley’s relata regress*. With the relata regress, he intends to show that the relata of any given relation fall apart into an infinitely complex non-terminating tree of alleged relations and their relata. Therefore the initial relata are not unities (see more below). Consequently, since Bradley thinks that relations presuppose relata as unities (AR, III, 27), relations are not fully intelligible.

Bradley’s relation regress and relata regress have the common critical target that any plurality of entities is not fully intelligible since every plurality of entities presupposes holding of relations, at least the holding of numerical distinctness. For the sake of these arguments, Bradley assumes that there is a plurality of entities and then attempts to show that any possible manner of understanding this is not ultimately fully intelligible. Hence, given Bradley’s view that being fully intelligible is a necessary condition for being true and fully real, existence pluralism is not the truth and the world according to existence pluralism is not and cannot be fully real (AR, III, 28).

However, looked at from another angle, one may respond to Bradley by showing that there is an intelligible way of understanding existence pluralism. This is not question-begging against Bradley because it does not assume the full intelligibility of existence pluralism. One thing one has to do in order to defend this response to Bradley is to argue that it is indeed fully intelligible that entities are related and these relations do not fall afoul of either of Bradley’s regresses. Bradley’s eliminative moves in them are not exhaustive of possible views on relations. So we can tell how Bradley’s relation and relata regress and his entire *reductio* of relations, indeed, can be *avoided* from arising, which differs from *solving* the regress problems. This is what we intend to do in this paper. Our focus will be on *formal ontological relations*, some of which at least do not fall prey to Bradley’s *reductio* of relations. So we shall argue.

The paper has a dual character. It includes both reconstructions of Bradley’s regresses and a systematic treatment of them. We have to do this

because our intention is both to convince contemporary metaphysicians of the power of Bradley's arguments and to be fair to Bradley, even if we cannot be fair to *all* of his intentions. This paper is systematic metaphysics, which gains from a scholarly work on Bradley. One also has to make an argument as solid as possible if one intends to answer the argument; there are already too many straw men in philosophy. So we have to dig deeper into Bradley than is usually done in contemporary metaphysics.

We begin in §2 with Bradley's relation regress, which leads to the discussion of how this regress may be avoided (§3). §4 is devoted to Bradley's relata regress and its critique. In §5, we shall show that there may hold formal ontological relations which escape the entire *reductio* of relations. Before the summary (§7), there is a coda (§6), which is a brief remark on relational thought and language.

2. Relation Regress

Arguably, Bradley's problem in his relation regress argument is how relation r is able to stand to two *distinct* qualities or relata, a and b . By his talk of "relating", Bradley apparently means that r is a determinate property of a kind and that r *relationally inheres* in a and b . For example, r is the distance of 2 m, which relationally inheres in a and b . One may compare this with monadic inherence in which a property (e.g. 1 kg mass) inheres in a bearer (object/substance). So r 's relating a and b involves both being a determinate property and relationally inhering in a and b . Secondly, Bradley thinks that r *connects* a and b , which means uniting a , b and r into a single relational complex aRb (Wieland & Betti 2008, 509–11; Betti 2015, 39–41).³

The *explanandum* in Bradley's relation regress is therefore the state of affairs⁴ that (1) r is a determinate property, (2) r relationally inheres in a and b , and (3) r connects a and b . The problem that the regress poses is to give a metaphysical *explanans* for this *explanandum*: which entities account for the

³ Consequently, the relation regress problem is not the same as the complex-unity problem, the special case of which is only part of the former problem.

⁴ Our talk about states of affairs does not presuppose the ontological commitment that they are entities in the category of facts.

obtaining of the state of affairs that r relates a and b by connecting them.⁵ Bradley's conclusion is that finding the *explanans* leads to a vicious infinite regress. Since r is an arbitrary existing relation, the vicious infinite regress has the consequence that there can be no metaphysical explanation for the state of affairs that an existing relation relates its relata by connecting them.⁶

Bradley's relation regress starts by assuming, for the sake of the argument, that the existence of r itself *as a thing* explains the state of affairs that it does relate by connecting a and b . What this must mean at least is that r is a distinct entity: it is an addition to being, a third entity in relation to its relata a and b . It is actually this that generates the relation regress according to Bradley.

The next step in the regress is Bradley's crucial assumption that r not only relates a and b by connecting them, *it itself must also be related and connected to both of a and b* . Hence, the only way for r to relate by connecting a and b is bearing relations and being connected to both a and b . On the assumptions of the argument, this can happen only if there are two additional relating and connecting entities. Thus, there is a new relation r^1 that relates by connecting r and a and a new relation r^2 that relates by connecting r and b . From this, an infinite regress of relations is generated, given that Bradley's argument goes through, by returning to the beginning in the case of r^1 or r^2 .⁷ So the regress has actually the structure of a tree with infinitely many levels:

⁵ In this paper, we do not take any positive stance on the difficult metametaphysical question what metaphysical explanation is. We try to do metaphysical explanation, and characterize it only negatively: it is not causal.

⁶ To be precise, Bradley's ultimate conclusion from the relation regress is that r 's relating a and b by connecting them is not the truth and fully real. This stems from his view that being fully intelligible is a necessary condition for being true and fully real, which does not hold of r 's relating by connecting a and b (AR, III, 28).

⁷ For a closely similar argument, see Bradley's first work *The Principles of Logic* (1883, 96), which was published ten years before *Appearance and Reality* (cf. Allard 2005, 61–66). There is also a version of the relation regress in Chapter II of *Appearance and Reality*. It has the slightly different form that first we need a two-place relation (e.g. instantiation), then a third-place relation and so on *ad infinitum* (the adicity of the relations ascends into infinity vs. infinite number of distinct applications of a multigrade relation). This is the way Armstrong (1997, 114), Cameron (2008) and Gaskin (2008, 314–6), for instance, understand Bradley's relation regress. We focus on the form of the regress advanced by Bradley in the chapter discussing relations (III).

Level 0			aRb		
Level 1		aR ¹ r		rR ² b	
Level 2	aR ³ r ¹		r ¹ R ⁴ r	rR ⁵ r ²	r ² R ⁶ b
...					

Independently from Bradley’s reasons for the infinite regress being vicious, there is a convincing argument to the result that the regress is, indeed, vicious. Rickie Leigh Bliss (2013) has argued that the regress is vicious because on each consecutive level of the tree, the *explanandum* and the components of *explanans* are of the same type (cf. Armstrong 1978, 106). The *explanandum* is that distinct entities are related and connected by a distinct relational entity (*aRb*). The *explanans* is to repeat infinitely that distinct entities are related and connected by a distinct entity (*xRⁿy*). So the *explanans* consists of what is type-identical to the *explanandum*; both are of the type *xRⁿy*. Consequently, nothing at all has been explained metaphysically about distinct entities being related and connected by a distinct entity (*xRⁿy*). This problem is just repeated infinitely. The connecting relations form an infinite tree of entities accounting for *r* relating *a* and *b* by connecting them.⁸ The tree structure does not terminate; the lowest level does not exist. Hence, no step is taken to explain metaphysically *aRb*.

It is quite clear that the relation regress makes two assumptions about relations. If there can be an account of relations that does not have to make those assumptions, it seems that it does not fall foul of this infinite regress (vicious or not). The first (1) of these assumptions is that relations are entities that relate and connect their relata. This presupposition involves two things at least: (1.1) relations are entities (reification of relations), (1.2) every relation is a genuine relating and connecting entity in such a manner that it unites itself, *a* and *b* into relational complex *aRb*. (1.2) entails (1.1) because it involves the commitment that relations are existences. The other assumption (2) is, as was seen above, that necessarily, if a relation relates and connects *a* and *b*, it itself

⁸ The same holds about every *rⁿ* on an infinite number of levels; every connecting by relational entity *rⁿ* unfolds an infinitely complex tree structure.

is connected and related to both *a* and *b*.⁹ (2) presupposes (1.1), since if a relation is itself connected to both of them, the relation must be an entity; non-entity cannot connect distinct entities. Only as constituents of what there is (entities), relations can connect their relata.

3. Could One Avoid Bradley's Trap?

A good start for not falling afoul of Bradley's relation regress is therefore to show that there may be related entities *without there being any reified relation that relates and connects them* (cf. Lowe 2006, 30). Then (1.1), (1.2) and (2) would be rejected since (1.2) and (2) presuppose (1.1). These putative relations would not relate by existing and by being connectors: they do not connect and unite distinct entities into *relational* complexes. Rather than an individual relational complex entity, these related entities form only a *plurality*. In a word, these relations are metaphysically speaking relatednesses of entities without relating and connecting entities. Still these relatednesses may be called relations since we can talk about them by means of relational predicate terms, which is a familiar possibility from the several different views of "internal relations". Neither does it follow that these relations do not mark any difference in being because they may mark a difference in *how* the world is even if they in themselves do not contribute to *what there is in* the world (see more below). As such, Bradley's point above that non-entities cannot connect does not really target them: "And, if you take it [a relation] as a kind of medium of unsubstantial atmosphere, it is a connexion no longer." (AR, III, 28.)

However, this is only a start. It is also necessary to show that the *relata* of these possible relations do not launch any vicious infinite regress. In Chapter III, Bradley considers two ways in which this might happen. The first way is the *relata* regress discussed below that the *relata* are relational in nature (ontologically complex), which leads to an infinitely complex, non-terminating tree of alleged relations and their *relata*. The second way is that *a* or *b* bears some *other* relation than *r* to some entity distinct from *a* or *b*, which re-launches the infinite regress. In the first part of Chapter III, no qualities without relations, Bradley argues that one possibility of this is that the difference in nature between *a* and *b* is due to an entity distinct from them

⁹ We omit the critical discussion of this principle since such a discussion is not needed for the purposes of this paper.

(e.g. an idealist view that the understanding distinguishes them), to which a and b have to bear some other relation than difference in nature (AR, III, 24).

It seems to us that there can hold relations that fall afoul of neither the relation nor the relata regress. This idea is familiar from internal relations (we do not discuss the definition of internal relations since our focus is elsewhere). Let us also be as sure as possible that at least some relations are not subject to Bradley's entire *reductio* of relations. First, it appears to be metaphysically possible that entities are numerically distinct without these entities involving a relational structure, composing an individual relational complex and reifying the relation of numerical distinctness, which would launch an infinite regress. Secondly, it seems to us that these numerically distinct entities can stand in other relations than numerical distinctness without an infinite relation regress ensuing. These metaphysically possible distinct entities would be metaphysically simple: they would have no metaphysical constituents whatsoever. Let us take a schematic example.

Assume that particulars t and u exist. Assume further that both t and u are metaphysically simple: t is identified with some particular nature P and u is identified with some particular nature Q in such a way that P and Q are thin natures in every metaphysically possible respect. On this assumption, t , u , P and Q do not divide into constituents that are distinct in any metaphysical sense. This assumption is part of the standard trope theory, but a realist may also hold it if P and Q are construed as universals.

Suppose that the difference in nature of P and Q consists of nothing but P and Q themselves. In other words, P and Q are different in nature for the fact that they are essentially certain thin natures. But the difference in nature of P and Q is not an additional entity to t and u . Bradley's supposition (1.1) is not true about it. That P and Q are different in nature means nothing but that there holds a certain relation in the world: P and Q are different in nature. Neither does this difference make P and Q different in nature, nor what they are; rather, the obtaining of their difference in nature consists of nothing other than the existence of P and Q . Nothing else than the existence of P and Q is metaphysically necessary or sufficient for their difference in nature. As thin natures, P and Q are intrinsically non-relational so no regress of relations and their relata can be generated *in* them. At least in this respect, our starting assumption seems to represent a metaphysically possible matter of fact.

Bradley would not be happy with this. He would be quick to point out that by the principle of the non-identity of discernibles (the contrapositive of Leibniz's law: the indiscernibility of identicals), it follows that t and u are numerically distinct. So there is another relation holding in our example, the relation of numerical distinctness, and it is that relation that generates an infinite relation regress. However, here we can insist against Bradley that the generation of an infinite relation regress is not necessary in this case. Both the difference in nature of P and Q and the numerical distinctness of t and u consists of nothing but P/t and Q/u themselves, which are non-relational in nature and do not generate relations that would lead to an infinite regress. Numerical distinctness or the plurality of numerically distinct P/t and Q/u is not an entity either.

4. Relata Regress

Against this insistence, Bradley advances a relata-regress argument in the second part of his *reductio* of relations, where he argues that qualities with relations are not fully intelligible. His point is that the difference in nature and numerical distinctness of a simple relatum from the other simple relatum are numerically distinct aspects of the former simple relatum. In our schematic example, t 's difference in nature and numerical distinctness from u , are numerically distinct aspects of t . As such, t is not simple but complex: it consists of two aspects that are related by numerical distinctness. The same holds of these two aspects and so on. Hence, an infinite regress of relations and their relata within t ensues: "We, in brief, are led by a principle of fission which conducts us to no end." (AR, III, 26) The key move in Bradley's argument to this result is that as " a it [t] is the difference on which distinction is based, while as a it is the numerical distinctness that results from connexion." (Ibid.)

Bradley clearly thinks, even if he does not say it, that this infinite regress in t is vicious. He says that the relatum, t in our example, loses "its unity" because t has "a diversity within its own nature", which is "fatal to the internal unity of" t (AR, III, 26–27). It "demands a new relation, and so on without limit. In short, qualities in a relation have turned out as unintelligible as were qualities without one." (AR, III, 27.) These passages tell us not only that Bradley thinks the infinite regress is vicious but that it is vicious since it

turns t into an infinitely complex, non-terminating tree of alleged relations and their relata. If this infinite regress ensues, it is a serious problem on Bradley's plausible view that the relata of any given relation should be unities. This time the tree has the following structure, where a^n is the difference on which the relation of numerical distinctness is based, while as a^n it is the aspect of the numerical distinctness that results from a relation of numerical distinctness r^n :

Level 0	$t = aR\alpha$			
Level 1	$a = a^1R^1\alpha^1$	$\alpha = a^2R^2\alpha^2$		
Level 2	$a^1 = a^3R^3\alpha^3$	$\alpha^1 = a^4R^4\alpha^4$	$a^2 = a^5R^5\alpha^5$	$\alpha^2 = a^6R^6\alpha^6$
...				

Regardless of his reasons for it, Bradley is correct that this relata regress would be vicious if it went through. As in the case of the relation regress, there is type-identity in the *explanandum* and *explanans*. Here the *explanandum* is the unity of a distinct entity t as a relatum of the difference in nature and numerical distinctness between t and u , given t 's ontological structure aRa . This is an instance of the complex-unity problem, which we characterized in the Introduction. The *explanans* is the unity of distinct entities a^n and a^n , which have a similar ontological structure to t (two entities related by numerical distinctness). The *explanandum* and each element of the *explanans* are of the same type; the latter is just repeated in an infinite number of times. The problem with this, as in the case of the relation regress, is that this is a metaphysical explanation failure. The complex-unity of a distinct entity is left without explanation; it does not shed any light on this complex-unity problem to repeat infinitely what the problem is.¹⁰ Consequently, on the assumption that the infinite regress is generated, nothing about the complex unity of any distinct entity in the infinite tree, including t , is explained metaphysically.

It is a different question whether the relata of a relation, a and a for example, have to be unities, which the viciousness of the relata regress clearly presupposes. This is a requirement that is not without its merits. If the relata

¹⁰ Recently, the complex-unity problem is taken up in the discussion on the different answers to van Inwagen's (1990) Special Composition Question. Metaphysicians attempt to lay down the non-mereological conditions that the parts of a complex entity must fulfil in order to constitute a complex unity, i.e., to satisfy the conditions of restricted composition.

were not unities, *what* would be related (e.g. what would be numerically distinct)? At least before anyone can come up with an account, the answer is: no entities. It does not help to point out that then the apparently two-place relation r^n would turn into an n -place relation holding between pluralities instead of two complex unities. These n -places have to be occupied by unities, for which no metaphysical explanation is reached if the infinite regress ensues.¹¹ So if Bradley's argument regarding t went through, we would be in fact dealing with a vicious infinite regress. Both t and u would exchange their intelligibility as unities for an infinitely complex non-terminating tree of alleged relations and their relata in both t and u .

Accordingly, let us judge this argument by Bradley. Then we shall see that his reasoning may be resisted: this relata regress does not get off the ground in the case of t and u in our schematic example. His move that t 's numerical distinctness from u introduces a distinct aspect to t is not easy to understand, but a reasonable reading, put in terms of our example, is the following. Bradley begins with the assumption that t 's nature is different from u . He argues that a relation of numerical distinctness holds between t and u by the principle of non-identity of discernibles and the holding of the relation of difference in nature between t and u . The holding of numerical distinctness, in turn, introduces the aspect of numerical distinctness (from u) into t ; this holding is constitutive of the *aspect* of t that it is numerically distinct from u . Therefore there are two aspects in t and "these different aspects are not each the other, nor again is either" t (AR, III, 26).

It is hard to see how this reasoning by Bradley could be solid. It rests on the premise that the holding of numerical distinctness between t and u is constitutive of the aspect of t that t is distinct from u . However, Bradley does not give us any reason to think that this is so. Furthermore, it is a strange view indeed that a non-entity, such as the relation of numerical distinctness, could make a difference to *what* there is. But Bradley clearly supposes that the relation of numerical distinctness does make a difference to what there is;

¹¹ It does not do to respond to Bradley that the relata a^n and a^n are relations (relations all the way down). This view is endorsed e.g. by Simon Saunders, James Ladyman and Don Ross, who are structural realists (Saunders 2003, 129; Ladyman & Ross 2007, 155). This is not assumed to be ruled out by Bradley's argument here. Rather, the heart of the problem is not the relational or non-relational nature of the relata. It is, to repeat, that there is no metaphysical explanation for the unity of the relata, *relations or not*.

perhaps he thinks it as an entity since he speaks about it as a “connexion” in the previous quote. As this may be reasonably denied, we conclude that Bradley has not succeeded in arguing that t 's difference in nature and numerical distinctness from u make t relational in nature and infinitely ontologically complex. Consequently, his *relata regress*, vicious or not, does not get off the ground.

5. Formal Ontological Relations

Do our replies to Bradley mean that difference in nature and numerical distinctness are properties of their *relata*? They must be something and the only reasonable option seems to be that they are properties of their *relata*, which, in turn, would subject them to Bradley's objection that relations cannot be properties of their *relata*. This objection occurs in his footnote just before the relation regress (AR, III, 27). The relevant part of the objection would be that “the problem of inherence would break out in an aggravated form.” (Ibid.) Is not this exactly the case with difference in nature and numerical distinctness?

Our reply to this possible objection is that difference in nature and numerical distinctness are not entities of any category (constituents of what there is). Rather, they are *ways* of the being of entities (how they exist), in contradistinction to *what exists*: that t exists, for instance, *as* different in nature from u . As such, difference in nature and numerical distinctness are not to be categorised as properties, which their *relata* would bear. So this objection of Bradley just does not arise in their case.

Still Bradley would not be happy (would an existence monist ever be with an argument defending pluralism?). He would insist that t and u are not simple since they have a dual nature, this time because of their numerical *identity* and difference in nature from each other. So there is a third relation in addition to difference in nature and numerical distinctness in our picture: the (logical) relation of numerical identity. “Inside” both t and u “we must distinguish its own quality and its otherness.” (AR, III, 24.) Entity t is identical with P , entity u is identical with Q and t/P is different from u/Q . There are different and hence distinct entities *in* them that constitute their numerical identity (t is t and u is u) and difference in nature (t is different in nature from u) in the same manner as in the case of difference in nature and numerical

distinctness just above. “And if so, then the unsolved problem breaks out inside each quality, and separates each into two qualities in relation.” (Ibid.) To be precise, Bradley does not speak about any infinite regress ensuing, vicious or not, in this context. It is easy to see, however, that he could be arguing for it along the lines of the argument discussed above that difference in nature and numerical distinctness make *relata* lose their needed unity.

Our reply to Bradley here is that he just cannot be seen to advance any argument to the result that the mere joint existence of *t* and *u* as identified with *P* and *Q* in our example would not be metaphysically sufficient for the holding of numerical identity of *t* and *u*, respectively, and their difference in nature from each other. He merely asserts that it would not be sufficient without *t* and *u* dividing into numerically distinct entities. So Bradley has not shown that the numerical identity of *t* and *u* and their difference in nature bestow a dual nature to them. This argument of Bradley may be resisted. No infinite regress is generated, vicious or not.

Actually, difference in nature, numerical distinctness and numerical identity belong to what are called “formal ontological relations” in the contemporary metaphysical literature. Formal ontological relations spell out what E. J. Lowe’s calls “ontological form”, which he distinguishes from “ontological content” (2006, 48; cf. Smith & Grenon 2004).¹² This distinction is analogical to the logical contrast between the non-logical content of a proposition and its logical form (Smith & Murray 1981; Smith & Mulligan 1983; Lowe 2006, 48). It is to be insisted, however, that ontological and logical form should not be conflated because the former is the form of *being* in general, whereas the latter is much more restricted: the form of truth-bearers (if they are entities) (Smith & Mulligan 1983, 73; Smith 2005, ch. 5; Lowe 2006, 48).

It is not an easy task to give a comprehensive account of the distinction between ontological content and ontological form. However, a comprehensive account is not necessary for the present purposes. Suffice it to introduce the distinction by way of examples.¹³ One may consider anything

¹² Husserl’s distinction between ontological content and ontological form was brought into contemporary discussion by Barry Smith, together with David Murray (1981) and Kevin Mulligan (1983).

¹³ In fact, we believe that at least ontological form is a concept that cannot be given an eliminative non-circular definition. It is a primitive notion in formal ontology (or ideology).

one believes to exist. This entity adds to the entities one believes to exist. In one's view, then, the entity is a part of the ontological *content* of the world: what beings there are ("ontology" in Quinean/Lewisian parlance, one makes an "ontological commitment").

For example, if one believes in the existence of Olli Koistinen, one believes that there is an entity to which "Olli Koistinen" refers. One holds a belief about the ontological content of the world. If one also believes that Olli Koistinen exists *as* having numerical identity (he is the same with himself), one holds a belief about the ontological *form* of the world.¹⁴ This is a belief about *how* Olli Koistinen exists as a constituent of the world ("ideology"), that is, as being the same with himself. One does not believe in any entity in addition to Olli Koistinen; rather, the belief concerns Olli Koistinen's *form of being* as being the same with itself. It can be understood as a formal feature of existence, what it is to be the same with itself. Arguably, further examples of formal features of existence are *being one* (unity), *being individual* (having numerical identity and being one) and *being particular* (and *being universal* for the realist). In Lowe's terms, all these are expressed by "formal ontological predicates", which apply but do not refer to entities (Lowe 2006, 193). In our example, "is numerically identical" applies to Olli Koistinen and the sentence "Olli Koistinen is numerically identical" is true about him.¹⁵

Accordingly, formal ontological relations may be characterized as formal relational features of existence. Take numerical distinctness as an example: it should be seen as how an entity exists, that is, as numerically distinct *from* another entity. If this formal ontological relation holds, it is part of the formal structure of the world: the world includes these two entities, which exist as not one and the same entity. In addition to numerical distinctness and difference in nature, possible formal ontological relations include numerical identity, relations between categories in different category systems (e.g. instantiation and characterization in the Neo-Aristotelian ontological square, different ontological dependencies), similarity or resemblance, quantitative relations (e.g. proportion), composition,

¹⁴ In principle, it seems not to be necessary for believing in an entity to hold that it has numerical identity since it appears to be possible to believe in the existence of something as a mere unity (cf. Lowe 2006, 75).

¹⁵ If this sentence is true, there are in principle several possible semantic accounts for its truth. In this paper, we set this semantic question aside.

constitution and mereological relations. Of these, at least dependences spell out the possible ways of existence of entities rather than their actual ways of existence.

Which formal ontological relations hold, what the formal structure of the world is, is dependent on which formal ontological category system is true. Still, formal ontological relations are necessary for every category system according to which there is something in the world, as Lowe argues (2006, 34).¹⁶ Actually, disputes between different ontological category systems are partly disagreements of what the (basic) formal ontological relations are. Whatever formal ontological relations obtain at the end of the day, they are in general good candidates for relations in the world that do not fall afoul of Bradley's *reductio* of relations, as we have argued.¹⁷ First, they are not subject to the relation regress because they are not entities of any category. Secondly, at least in the case of simple entities, they may hold between simple entities without falling prey to Bradley's *reductio* of relata.

6. Coda: Relations in Thought and Language

It seems to us that Bradley's last resort at this point would be to point out that as human beings we have to think and talk about our schematic scenario, formal ontological relations and their relata. This re-launches his relation regress for two reasons. (1) There has to hold some relation between our thinking or talking about the scenario and the scenario (e.g. representation, reference, truth, truth-making). (2) Thinking and talking about the scenario and formal ontological relations involves relations since they involve predications that have relational structure. When we think or talk about the scenario, for instance, we predicate over t and u , P and Q . When we do this,

¹⁶ Arguably, even an existence monist like Bradley should accept the formal ontological relation of numerical identity at least because there is one self-identical entity in the world. It is also quite unclear whether the one could exist without any difference or distinction in it. If it could not, it would involve formal ontological relations.

¹⁷ In some category systems, the holding of formal ontological relations results in the existence of complex unities. Ontological dependences are typically such formal ontological relations. They spell out how entities *can* exist as constituents of the world. It is to be emphasised, however, that these formal ontological relations ought not to be identified with these complex unities. Neither are these complex unities relational complexes. So no holding of formal ontological relation should be identified with the existence of a relational complex unity. The (1.2) assumption of Bradley's relation regress does not hold true of them.

something works as a subject and something as a predicate of the subject. Therefore there must hold relations between predicates and subjects. Bradley believes that “the conclusion to which I am brought [at the end of Chapter III right after his relation regress argument] is that a relational way of thought—any one that moves by the machinery of terms and relations—must give appearance, and not truth.” (AR, III, 28.)

These two points raise extremely difficult questions in epistemology, philosophy of mind, theory of judgement or assertion and philosophy of language, which obviously cannot be discussed in a paper in metaphysics. We can point out, however, as an initial response to Bradley, that our argument has shown that at least some formal ontological relations are, indeed, intelligible. So the possible problems in relational thought and talk are not relevant in this respect. There are two different things here: the possibility of fully intelligible relational thought and talk and the possibility of related *being*. Even if the former were impossible, it does not follow that the latter is not possible, without some further (idealist) premise. Moreover, it is not necessary that predications or judgments are relational complexes: complex entities with a relational ontological structure. First, it can be reasonably denied that predications or judgements are entities. Secondly, according to Allard (2005, ch. 3) and Basile & Candlish (2013, sec. 5), Bradley’s own theory of judgement is monistic.

7. Conclusion

In this paper, we have reconstructed Bradley’s relation and *relata* regress, including their structure, and defended their nature as vicious infinite regress arguments. If they go through, they are vicious infinite regresses since the *explanandum* and *explanans* are type-identical in the regresses and this type identity is repeated in an infinite number of times. Therefore no step is really taken to metaphysically explain the initial *explanandum*.

Our intention has not been to solve the problems that these regress arguments pose. We have limited ourselves to the argument for the result that there are possibly ways of standing in relation in the world that escape both regresses and Bradley’s entire *reductio* of relations. At least difference in nature, numerical distinctness and identity of ontologically simple entities count as such relations. Since they are examples of formal ontological relations, this

suggests that formal ontological relations in general are good candidates for relations that do not fall afoul of either regress.

University of Tampere

University of Helsinki

Bibliography

- Allard, J. (2005). *The Logical Foundations of Bradley's Metaphysics: Judgment, Inference, and Truth*. Cambridge: Cambridge University Press.
- Armstrong, D. M. (1978). *Universals and Scientific Realism. Volume 1: Nominalism and Realism*. Cambridge: Cambridge University Press.
- Armstrong, D. M. (1997). *The World States of Affairs*. Cambridge: Cambridge University Press.
- Betti, A. (2015). *Against Facts*. Cambridge, MA: MIT Press.
- Bliss, R.L. (2013). "Viciousness and the Structure of Reality", *Philosophical Studies* 166 (2), 399–418.
- Bradley, F. H. (1883). *Principles of Logic*. London: Oxford University Press.
- Bradley, F. H. (1897). *Appearance and Reality* (second edition). London: George Allen & Unwin.
- Cameron, R. (2008). "Turtles All The Way Down: Regress, Priority and Fundamentality", *The Philosophical Quarterly* 58 (230), 1–14.
- Candlish, S. & P. Basile (2013). "Francis Herbert Bradley", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL = <http://plato.stanford.edu/archives/spr2013/entries/bradley/>
- Gaskin, R. (2008). *The Unity of the Proposition*. Oxford: Oxford University Press.
- Ladyman, J. & D. Ross (2007). *Every Thing Must Go*. Oxford: Oxford University Press.
- Lowe, E. J. (2006). *The Four-Category Ontology*. Oxford: Oxford University Press.
- Saunders, S. (2003). "Structural Realism, Again", *Synthese* 136, 127–133.
- Smith, B. (2005). "Against Fantology", in J. Marek & E. M. Reicher (eds.), *Experience and Analysis*. Vienna: Austrian Ludwig Wittgenstein Society, 153–170.
- Smith, B. & P. Grenon (2004). "The Cornucopia of Formal-Ontological Relations", *Dialectica* 58 (3), 279–296.
- Smith, B. & K. Mulligan (1983). "Framework for Formal Ontology", *Topoi* 2, 73–85.
- Smith, B. & D. Murray (1981). "Logic, Form and Matter", *Aristotelian Society Supplementary Volume*, 55, 47–74.
- van Inwagen, P. (1990). *Material Beings*. Ithaca: Cornell University Press.
- Wieland, J-W. & Betti, A. (2008). "Relata-specific Relations: A Response to Vallicella", *Dialectica* 62 (4), 509–524.

Emootioteoriat ja toisen emootioita koskeva tieto

Leila Haaparanta

1. Johdanto

Muutaman vuosikymmenen ajan on keskusteltu kahdesta emootioteorian päätyypistä. Toiset filosofit väittävät, että emootiot ovat kognitiivisia. Toiset taas puolustavat kantaa, jonka mukaan emootiot ovat affekteja, tunnevaikutuksia, siis jotakin sellaista, jota käsitteemme eivät voi tavoittaa. Jos emootiot olisivat kognitiivisia arvostelmateoriaksi kutsutun emootioteorian tarkoittamalla tavalla, ne olisivat maailmaa koskevia arvostelmia ja niiden pätevyyttä voitaisiin arvioida tarkastelemalla mahdollisia perustelujen ketjuja, joihin ne kytkeytyvät. Emootioteorioiden päätyypeissä kiinnitetään huomiota yhden persoonan emootioihin, ei persoonien välisiin suhteisiin. Toisenkin keskustelu on ollut meneillään filosofien ja erityistieteilijöiden parissa. Se koskee empatiaa ja sympatiaa ja siten persoonan suhdetta toisen persoonan emootioihin. Näitä kahta kysymystä, siis mitä emootiot ovat ja mikä on suhteemme toisen persoonan emootioihin, ei juuri ole tarkasteltu samanaikaisesti. Tämän kirjoituksen tarkoituksena on yhdistää ne toisiinsa.

Aluksi esitän lyhyen katsauksen emootioteorioihin. Sen jälkeen luonnehdin empatian ja sympatian käsitteitä ja pohdin empatian mahdollisuutta, kun emootiot ymmärretään arvostelmiksi, maailman käsitteellistyksiksi tai affekteiksi. Empatiaa pidetään filosofiassa usein kognitiivisena käsitteenä: siinä on kyse toisen persoonan emootioita koskevasta tiedosta. Puolustan sitä väitettä, että jos empatia ymmärretään tällä tavoin, arvostelmateorian kannattajalle se on tulkinnallista päättelyä; se on hänelle toisen persoonan arvostelmien ja niitä tukeviksi oletettujen argumentaatioketjujen paljastamista. Tarkastelen ongelmia, joita tähän empatiakäsitykseen sisältyy. Myös affektiteorian ja empatian yhdistelmä on ongelmallinen, jos affektiteorian kannattaja pitää empatiaa toisen affekteja koskevana tietona. Artikkelin lopussa ehdotan osittain Edith Steinin

näkemyksiin pohjautuvaa empatian käsitteen analyysia, jolla ongelmat vältettäisiin.

2. Arvostelma- ja affektiteoriat

William Jamesin varhaista emotioteoriaa pidetään yleensä ei-kognitiivisena ja tyypillisenä affektiteorianana. Artikkelissaan ”What is an Emotion?” (1884) James kirjoittaa:

*Our natural way of thinking about these standard emotions is that the mental perception of some fact excites the mental affection called the emotion, and that this latter state of mind gives rise to the bodily expression. My thesis on the contrary is that *the bodily changes follow directly the PERCEPTION of the exciting fact, and that our feeling of the same changes as they occur IS the emotion.* (James 1969, 247.)*

Viime vuosisadan jälkipuoliskolla filosofiassa keskusteltiin paljon kognitivismista, jonka mukaan mentaalisen aluetta voidaan analysoida tarkastelematta ulkoista havaittavaa käyttäytymistä, esimerkiksi ottamalla tietokone ihmismielen malliksi. Emotioiden filosofiassa kognitiivinen teoria taas on näkemys, jonka mukaan emotiot eivät ole ainakaan yksinomaan tunnevaikutuksia, vaan ne liittyvät tavalla tai toisella tiedonhankintaan eli niihin prosesseihin, joista psykologit ja filosofit olivat kiinnostuneet behaviorismin menetettyä vahvan asemansa. Kognitiivisia teorioita esittivät muiden muassa Robert Solomon, johon vaikuttivat Martin Heideggerin ja Jean-Paul Sartren ajatukset, ja Martha Nussbaum, joka lähti liikkeelle antiikin filosofeista. Solomonin mukaan emotiot ovat rationaalisia, tavoitteellisia ja tekojen kaltaisia ja me valitsemme emotiomme samoin kuin valitsemme tekemme (Solomon 1980, 251–252). Solomon väitti myös, että emotiot ovat normatiivisia arvostelmia. Hän esitti, että koska normatiivisia arvostelmia voidaan muuttaa, esimerkiksi argumenttien ja evidenssin avulla tai muuten vaikuttamalla arvostelman esittäjään, ja koska voin itse etsiä vaikutteita, argumentteja ja evidenssiä, olen vastuussa emotioistani, kuten olen vastuussa esittämistäni arvostelmista (ibid., 261). Solomonin mukaan emotiot ovat kuitenkin yleensä hätäisesti esitettyjä arvostelmia (ibid., 262). Stoalaisen emotioteorian kannattajana myös Martha Nussbaum (1990, 1994) on

pitänyt emootioita uskomuksen hyväksymisinä, siis maailmaa koskevana arvostelmina. Hänen mukaansa käytännöllistä viisautta ilmentää kuitenkin se, joka ottaa huomioon emootioiden affektiivisen puolen.

Viime vuosisadan lopulla monet filosofit ehdottivat näkemyksiä, joiden mukaan emootiot sisältävät sekä kognitiivisia että affekti-komponentteja (ks. esim. Gordon 1987, Greenspan 1988 ja Green 1992). Kognitiivisten teorioiden suosion jälkeen muutamat filosofit, muiden muassa Peter Goldie (2000), halusivat jopa puolustaa affektien erityisasemaa emootioissa. Myös uudenlaisia, Solomonin teoriaa maltillisempia, vaikka vahvasti kognitiivisia teorioita on ehdotettu. Näistä esimerkki on Robert C. Robertsin teoria. Siinä emootiot ymmärretään ”näkemisiksi jonakin” (”seeings as”), joita ei voi palauttaa arvostelmiin. Se, että esimerkiksi näen jonkun henkilön pelottavana tai ihailtavana, ei tarkoita samaa kuin se, että esitän arvostelman – vaikka vain itselleni –, että kyseinen henkilö on pelottava tai ihailtava. Tämä seikka selittää ainakin jossain määrin, miksi ennakkoluuloja ei ole helppo muuttaa argumenttien avulla. Asenteellinen henkilö on sidoksissa emootioihinsa, jotka Robertsin teorian mukaan voidaan ymmärtää maailman käsitteellistyksiksi. Käsitteellistyksiä vastaan on hankala argumentoida ja niistä on vaikea irrottautua, vaikka niistä haluttaisiin esimerkiksi moraalisyistä vapautua. Niihin voidaan ehkä vaikuttaa synnyttämällä uusia, aiemmista poikkeavia käsitteellistyksiä. Vaikkei henkilö pääse irti tavoistaan hahmottaa maailmaa, hänet voidaan kuitenkin velvoittaa pidättäytymään arvostelmasta tai hän voi itse haluta pidättäytyä käsitteellistystään vastaavasta arvostelmasta. Riippumatta siitä, kuinka hän toisen ihmisen näkee, hän voi esimerkiksi olla esittämättä arvostelmaa, että toinen, vaikkapa etniseltä taustaltaan itsestä poikkeava ihminen, on pelottava tai halveksittava. Arvostelmasta pidättäytyminen saattaa merkitä myös, että henkilö ei enää argumentoi emotionaalisen asenteensa puolesta; hänen ei tarvitse enää osoittaa, että jotkin perustelujen ketjut tukevat emotion ilmaisevaa arvostelmaa. Joissakin tilanteissa se, että henkilö ei halua esittää emotion ilmaisevaa arvostelmaa julkisesti, saattaa tietenkin olla merkki siitä, että hän haluaa paeta velvollisuutta argumentoida arvostelmansa puolesta.

3. Empatian käsite

Tarkastelen seuraavaksi toisen persoonan emootioiden tuntemisen ongelmaa. Analyttisessä mielenfilosofiassa on keskusteltu paljon simulaatiosta ja projektiosta, joiden ajatellaan olevan tapoja tuntea toisen emootioita (ks. esim. Goldman 2006). Empatian ja sympatian käsitteitä käytetään filosofian perinteessä ja nykykeskustelussa vaihtelevilla tavoilla. Peter Lamarque kuvaa empatiaa ”ymmärtämiseksi elämällä läpi” (”realizing by living through”) ja sen tietämiseksi tai kuvittelemiseksi, millaista on olla toinen (Lamarque 2009, 244), kun taas Suzanne Keen toteaa, että empatia on tuntemista niin kuin toinen tuntee ja sympatia on tunne, joka tukee toisen tunnetta, esimerkiksi sääli (Keen 2007, 4–5). Teoksessaan *Game Theory and the Social Contract I: Playing Fair* (1994), Ken Binmore puolestaan kirjoittaa:

A loan shark might attend a movie and succeed in identifying with an old lady that the story-line plunges into deep distress. But nothing then prevents his wiping the tears from his eyes on leaving the theater and proceeding to use the insights gained in watching the movie in gouging the next elderly widow to come his way. Such empathizing is about as far from sympathetic as it is possible to get. (Binmore 1994, 28.)

Binmoren mukaan sana ”sympatia” soveltuu tilanteisiin, joissa ihmisillä on altruistisia preferenssejä, kun taas empatia on prosessi, jossa kuvittelemme olevamme toisen asemassa ja näemme asiat hänen näkökulmastaan; tuo prosessi on hänen mukaansa puhtaasti kognitiivinen, sillä ei ole vaikutusta omiin preferensseihimme.

Muutamat 1800-luvun lopun ja 1900-luvun alun saksalaiset filosofit, esimerkiksi Edith Stein ja Edmund Husserl, kiinnittivät erityistä huomiota empatiaan. Edith Stein erotti toisistaan sympatian (*Mitfühlen*), esimerkiksi iloitsemisen siitä, mistä toinen iloitsee, tai tunteiden jakamisen, ja empatian (*Einfühlung*), esimerkiksi toisen iloa koskevan tiedon tai toisen ilon tuntemisen, johon saattaa liittyä toisen ilon aiheuttamaa iloa. Hän myös väitti, että empatiassa on kyse kognitiosta, siitä, että kokee toisen sellaisena, jolla on tunteita. Steinille empatia on toisen tietoisuuden sisäisten tunteiden abstraktia ymmärtämistä, ilman että kokee näitä tunteita samanaikaisesti, esimerkiksi erityislaatuista tietoa ystävän surusta ilman, että on omilla tunteilla mukana

suremisessa. Sympatia sen sijaan on hänelle tunteiden jakamista tai vastaavien tunteiden tuntemista. Steinin analyysissä on kaksi vaihetta. Ensiksi empatia on toisen kokemista sellaisena, jolla on kokemuksia, siis myös emootioita. Toiseksi toisen persoonan kokemusten oletetaan olevan osana empatiaa tuntevan intentionaalisissa kokemuksissa, mutta erityisellä tavalla; ne eivät ole samanaikaisesti hänen kokemuksiaan, vaan ne ovat tietoa toisen kokemuksista. Empatia ei Steinin mukaan ole mystinen kokemus, sillä mystisessä kokemuksessa, toisin kuin empatiassa, toinen hallitsee minua. Stein korostaa, että toisen tavoittaminen empaattisesti on kognitiivinen prosessi. Hänen mukaansa emotionaalinen reaktio ei ole sen olennainen piirre. (Ks. Stein 1989, 11 ja 14–17.)

4. Arvostelmateoria, affektiteoria ja empatia

Jos empatiassa ajatellaan olevan kyse jonkinlaisesta tiedosta, kuten Stein ja monet muutkin filosofit näyttävät ajattelevan, millaista tämä tieto on silloin, jos emootiot ovat (1) arvostelmia, (2) maailman käsitteellistyksiä tai (3) affekteja? Jos ajattelemme, että kognitiivinen emootioteoria, joka pitää emootioita arvostelmina, on oikea teoria ja että empatia on toisen emootioita koskevaa tietoa, empatia on ymmärrettävä tulkitseväksi päättelyksi. Se on toisen persoonan emotiivisten arvostelmien ja niitä tukemaan tarkoitettujen argumentaatioketjujen paljastamista. Tulkitsija ehkä luottaa teoreettisen rationaalisuuden malliin ja yrittää rekonstruoida argumentatiivisen rakenteen, johon tuo erityinen arvostelmaksi ymmärretty emootio kuuluu. Filosofisen tekstin tulkitsijan tavoin toista persoonaa tulkitseva saattaa muodostaa argumentatiivista rakennetta koskevia hypoteeseja. Hän saattaa myös pitää praktisen päättelyn mallia oppaanaan ja olla kiinnostunut toisen persoonan päämääristä ja uskomuksista, jotka koskevat sitä, mitkä teot ovat välttämättömiä päämäärien saavuttamiseksi; emootioilla on siis tämän näkemyksen mukaan paikkansa sellaisissa pohdinnoissa.

Jos empatia tulkitaan toisen persoonan emootioita koskevaksi tiedoksi ja toisen emootiot ajatellaan arvostelmiksi ja liitetyiksi perustelujen ketjuihin, empaattiseksi subjektiksi ajatellulla henkilöllä voi sanoa olevan tietoa toisen emootioista, vain jos hän tekee samat johtopäätökset, siis jos hän esittää samat arvostelmat. Jos suljemme pois perusteluketjut ja tarkastelemme vain emootion ilmaisevaa arvostelmaa, on mahdollista ajatella, että tulkitsija

ymmärtää arvostelman propositionaalisen sisällön. Tämä ei kuitenkaan vielä tarkoita, että hän tietää, millaista on esittää kyseinen arvostelma. Jos arvostelmateorian sijasta nojaamme kognitiiviseen teoriaan, joka pitää emootioita maailman käsitteellistyksinä, ”näkemisinä jonakin”, emme ole sen paremmassa asemassa. On epävarmaa, kykenemmekö näkemään maailmaa sellaisena kuin toinen sen näkee, siis voimmeko käsitteellistää maailmaa samoin kuin toinen. Tilanne ei ole myöskään parempi, jos luotamme affektiteoriaan. Affektiteorian kannattaja voi yrittää tarkastella affektien tyyppisiä ja väittää, että voimme tietää millaista on tuntee sen tyyppinen tunne, jonka toinen tuntee. Hän ei kuitenkaan tiedä, millaista on tuntee juuri se emotionin esiintymä, joka toisella on. Mikään ei myöskään takaa, että toinen rakentaa emootiotyypit samoin kuin me.

Jos väitämme, että arvostelman esittämisen aktit, käsitteellistykset ja affektit ovat kaikki empaattista suhdetta tavoittelevan subjektin ulottumattomissa, tämä näyttää tarkoittavan, että empatia on mahdotonta. Edith Steinin teoria kuitenkin antaa välineitä analysoida empatiaa tavalla, joka tekee siitä mahdollista myös edellä esitellyille teorioille. Voimme nimittäin erottaa kaksi eri askelta: on yksi asia nähdä toinen persoona sellaisena, jolla on emootioita, ja on toinen asia tietää, millaista on olla juuri niiden emootioiden ”vallassa”, jotka toisella on. Ensimmäinen askel tarkoittaa, että näemme elävän olennon kaltaisenamme, sellaisena, jolla on emootioita. Miksi ajattelisimme lisäksi, että empatia on mahdollista vain jos meillä on tietoa siitä, kuinka toinen tuntee? Pikemminkin voisi ajatella, että empatia kognitiivisena aktina on mahdollinen vain jos epäröimme ottaa toista askelta. Empaattisesti asennoituvan olettaisi pikemminkin suhtautuvan skeptisesti omaan tietokykyynsä.

Edith Steinin analyysi on lähellä yllä esitettyä. Pääero on siinä, että toisin kuin Stein, olen edellyttänyt, että empaattisesti asennoituva henkilö epäilee, onko hänen mahdollista saavuttaa toisen emootioita koskevaa tietoa. Empatiaa olisi pidettävä erityislaatuisena subjektin kokemuksena: siinä subjekti, minä, näkee toisen persoonana, jolla on emootioita, mutta hän epäilee, tietääkö hän tai onko hänen edes mahdollista tietää, millaista on silloin, kun henkilöllä on niitä emootioita, joita toisella on. Tiedon rajan tunnustaminen on siis tämän analyysin mukaan empatian ehto, ei sen mahdollisuuden kieltö.

Edmund Husserlin, Edith Steinin opettajan, empatian mahdollisuutta puolustava argumentti muistuttaa analogia-argumenttia. Jos kyseessä olisi analogiapäätely, silloin Husserlin voisi sanoa päättelevän toisen mielen olemassaolon oman mielen ja oman ruumiin sekä toisen ruumiin havaitsemisen perusteella. Monet Husserl-tutkijat, esimerkiksi Forest Williams (1989) ja Dan Zahavi (2010), ovat kuitenkin perustellusti kritisoineet tätä tulkintaa. He ovat väittäneet, että sen sijaan, että Husserl esittäisi analogia-argumentin, hän puolustaa näkemystä, jonka mukaan empatia on mahdollista, koska toinen nähdään ruumiillisena mielenä, itsemme kaltaisena. Husserl selvittää empatian mahdollisuuden välttämättömiä ehtoja, kuten Kant selvitti tiedon ja kokemuksen mahdollisuuden välttämättömiä ehtoja. Ensin hän suorittaa reduktion ensimmäisen persoonan alueeseen ja kysyy sitten, mikä rakentaa tien toiseen persoonaan. Hän esittää, että me havaitsemme itsemme psykofyysisinä organismeina (Hua, § 44). Hänen mukaansa me näemme toisen ihmisen elettyinä ruumiina (*Leib*), itsemme kaltaisena. Kyse ei ole siitä, että Husserl päättelisi tai väittäisi meidän päättelevän toisen mielen olemassaolon oman mielen, oman ruumiin (*Körper*) ja toisen ruumiin (*Körper*) olemassaolon kautta.

Husserlin näkemys ei ole ristiriidassa edellä ehdotetun analyysin kanssa. Myös siinä empatia ymmärretään yleiseksi asenteeksi, jossa toinen persoona nähdään itsen kaltaiseksi, tuntevaksi eläväksi olennoksi. Husserl ei oleta, että se, joka tuntee empatiaa, tavoittaa täsmälleen ne emootiot, jotka toisella on. Toisaalta hän ei myöskään esitä, että empatia edellyttäisi skeptistä asennetta toisen emootioita koskevaa tietoa kohtaan. Tässä artikkelissa esitetyssä analyysissä juuri sitä edellytetään, jotta voisimme kutsua tarkasteltavaa mielen aktia tai tilaa empatiaksi.

*Tampereen yliopisto
Centre for Advanced Study, Oslo*

Kirjallisuus

- Binmore, Ken (1994). *Game Theory and the Social Contract, Volume I: Playing Fair*. Cambridge, MA: The MIT Press.
- Goldman, Alvin I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. New York: Oxford University Press.

- Gordon, Robert M. (1987). *The Structure of Emotions: Investigations in Cognitive Philosophy*. Cambridge: Cambridge University Press.
- Green, O. Harvey (1992). *The Emotions*. Dordrecht: Kluwer.
- Greenspan, Patricia S. (1988). *Emotions and Reasons: An Inquiry into Emotional Justification*. New York: Routledge.
- Husserl, Edmund (1960 [1931]). *Cartesianische Meditationen und Pariser Vorträge. Husserliana I*. Toim. S. Strasser. Haag: Martinus Nijhoff. [Hua]
- James, William (1969 [1884]). "What is an Emotion?", teoksessa W. James, *Collected Essays and Reviews*. New York: Russell & Russell, 244–275.
- Keen, Suzanne (2007). *Empathy and the Novel*. New York: Oxford University Press.
- Lamarque, Peter (2009). *Philosophy of Literature*. Oxford: Blackwell.
- Nussbaum, Martha C. (1990). *Love's Knowledge: Essays on Philosophy and Literature*. Oxford: Oxford University Press.
- Nussbaum, Martha C. (1994). *The Therapy of Desire: Theory and Practice in Hellenistic Ethics*. Princeton: Princeton University Press.
- Roberts, Robert C. (2003). *Emotions: An Essay in Aid of Moral Psychology*. Cambridge: Cambridge University Press.
- Solomon, Robert (1980 [1973]). "Emotions and Choice", teoksessa Amélie Rorty (toim.), *Explaining Emotions*. Berkeley: University of California Press, 251–281.
- Solomon, Robert (1983). *The Passions*. Notre Dame: University of Notre Dame Press.
- Stein, Edith (1989 [1917]). *On the Problem of Empathy (Zum Problem der Einfühlung)*. Transl. W. Stein. *The Collected Works of Edith Stein, Vol. Three*, 3rd rev. ed. Washington, D.C.: ICS.
- Williams, Forest (1989). "Intersubjectivity: a Brief Guide", teoksessa Jitendra Nath Mohanty & William R. McKenna (toim.), *Husserl's Phenomenology: A Textbook*. Washington D.C.: Centre for Advanced Research in Phenomenology and University Press of America.
- Zahavi, Dan (2010). "Empathy, Embodiment and Interpersonal Understanding: From Lipps to Schutz", *Inquiry* 53 (3), 285–306.

Habituaalisen toiminnan haaste vapaalle toiminnalle

Susanne Uusitalo

Vapaan toiminnan on tyypillisesti ajateltu muodostuvan rationaalisesta toiminnasta, joka on toimijan hallinnassa. Tämä itsessään ei kuulosta liian vaativalta ehdolta inhimilliselle toiminnalle ja sen voidaan nähdä sopivan yhteen arkisenkin elämän kanssa. Lounasajan lähestyessä voin päättää syödä puuroa ja toteuttaa tämän päätökseni asianmukaisin resurssein. Toimintani on vapaata siinä määrin, kun se kumpuaa järkiperaisestä perusteesta ylläpitää toimintakykyäni ravitsemalla itseäni ja toteutan päätökseni hallitusti. Samaan aikaan tämä tilanne voidaan nähdä tavanomaisena, rutiininomaisena toimintana, jossa tietoisella harkinnalla ei ole kovinkaan suurta osaa, eikä rutiini ole välttämättä täysin toimijan hallinnassa, ainakaan siinä määrin, että hän pystyisi sitä helposti ja aina tarvittaessa muuttamaan. Tavat täyttävät arkipäiväisen elämän ja tuntuvat keskeiseltä inhimillisen toiminnan kuvauksessa, mutta muodostavatko ne samalla uhan yksilön vapaalle toiminnalle?

Nykyisessä teonfilosofisessa keskustelussa tavat ja rutiininomainen toiminta eivät tyypillisesti nouse esiin (vrt. esim. Mele 1990, Sandis & O'Connor 2013, Dancy & Sandis 2015). Jossain määrin tavat ja rutiininomainen toiminta kuitenkin huomioidaan ja ne nähdään olennaisena osana arkipäiväistä toimintaa. Tavanomainen toiminta ei automaattisuutensa vuoksi hukkaa toimijan resursseja turhaan harkintaan tai sellaiseen keskittymiseen, joka ei ole tarpeen esimerkiksi pyörällä ajamisessa: toimijalle jää enemmän aikaa huomioida liikenteessä tapahtuvat asiat, jos hänen ei tarvitse keskittyä siihen, että jokainen polkaisu vie häntä kohti määränpäättä. Tietyissä mielessä tavanomainen toiminta siis säästää toimijan resursseja ja tällä tavoin mahdollistaa myös sellaisen toimijuuden, joka vaatii harkintaa ja keskittymistä. Tällaisessa yhteydessä tavat tuntuvat siis harmittomilta, eivätkä ole uhkana toimijan vapaudelle toimia harkintansa mukaan.

Filosofisessa kirjallisuudessa tavat kuitenkin mainitaan myös toisenlaisen toiminnan yhteydessä, esimerkiksi addiktioiden yhteydessä. Tällöinkään habituaaliseen toimintaan ei silti ole tyypillisesti kiinnitetty paljoa huomiota, vaan se on sivuutettu muutamalla lauseella ja keskitytty pakottavan halun vaikutuksiin addiktin toiminassa.¹ Pakottavien halujen on nähty olevan uhka toimijan vapaudelle toimia toisin. Samaan aikaan on kuitenkin todettu, että osittainen syy siihen, miksi addiktoituneilla yksilöillä on vaikeuksia lopettaa addiktiivista toimintaansa, on juuri siinä, että addiktiosta on muodostunut tapa (esim. Pickard 2012).

On totta, että tapojen muuttaminen tai lopettaminen voi toisinaan olla hyvinkin hankalaa. Tämä koskee muitakin toimijoita kuin addikteja ja heidän toimintaansa. Tällaisessa yhteydessä tavanomaisesta toiminnasta muodostuukin uhka toimijalle, erityisesti silloin, kun hän on päättänyt toimia toisin. Kirjoituksessani tarkastelen näiden kahden tapoja koskevan näkemyksen välistä jännitettä tavanomaisen, habituaalisen, toiminnan selittämisessä. Kysyn, ovatko tavat ja niissä ilmentyvä habituaalinen toiminta jotain sellaista, joka kyseenalaistaisi toimijan hallinnan omasta toiminnastaan. Katoaako toimijan vapaus, kun hän rutinoituu johonkin toimintaan?

1. Vapaasta toiminnasta

Jotta analyysini etenisi, avaan muutamia keskeisiä käsitteitä, joita kirjoituksessani käytän.

Vapaan toimijan nähdään tyypillisesti hallitsevan omia tekojaan, olipa kyse sitten metafysisestä kyvystä toimia toisin tai oikeanlaisesta teon syntyhistoriasta. Yksi näkökulma vapaaseen toimijuuteen on sanoa, että vapaus on perusteherkkyyttä: toimijalla on mahdollisuus ja kyky reagoida saatavilla oleviin perusteisiin ja toimia niiden mukaisesti (Fischer & Ravizza 1996). Sitoudun kirjoituksessani tähän määritelmään, koska se korostaa sitä jännitettä, mitä habituaalisen ja vapaan toiminnan välillä voi olla, varsinkin jos pidetään kiinni ajatuksesta, ettei habituaalista toimintaa tehdä perusteista (Pollard 2003).

Toiminnan analyysini lähtökohtana käytän Donald Davidsonin kausaalista teonteoriaa. Davidsonin mallissa halu tehdä jotain ja uskomus sen

¹ Ainakin yksi poikkeus on kuitenkin Timothy Schroeder, joka käsittelee tapoja ja addiktioita useampaankin otteeseen. Ks. Schroeder 2010 ja Arpaly & Schroeder 2013.

toteuttamisesta muodostavat intention, joka saa aikaiseksi teon. Davidson olettaa, että toimija toimii aina vahvimman halunsa mukaisesti ja jos toimija arvioi, että kyseisessä tilanteessa olisi parempi tehdä teko a kuin tehdä teko b, hän myös haluaa tehdä teon a enemmän kuin tehdä teon b. (Davidson 1980.) Arvot liittyvät siis motivaatioon. Halu-uskomus-pari (tai intentio) muodostaa perusteen, joka rationalisoi ainakin yhden kuvauksen teosta (Davidson 1963).

Kun käsittelen vapautta eli perusteherkkyyttä toimijuuden ja tekojen suhteen tarkoitan, että teko on vapaa pakosta, uhkailusta ja kiristyksestä tai muusta ongelmallisesta manipulaatiosta ja toiseksi, että toimijalla on mahdollisuus tehdä kyseinen teko. Davidsonin mallissa premissi vahvimmasta halusta käsittelee osittain näitä aspekteja:

Jos toimija haluaa tehdä teon x enemmän kuin teon y ja hän uskoo olevansa vapaa tekemään joko x:n tai y:n, hän tulee intentionaalisesti tekemään x:n, jos tekee intentionaalisesti joko x:n tai y:n. (Davidson 1980, 23.)²

Davidson siis ottaa huomioon, että tilanteessa voi olla mukana kontrolloivia vaikutteita, jotka voivat estää teon, mutta jos toimija on vapaa tuollaisista vaikutteista ja hän päättää ryhtyä tekoon, hän toimii mielensä mukaan, eli sen mukaan, mitä haluaa eniten. Toimijan uskomus siitä, että on vapaa viittaa siihen, että toimija on motivoitunut tekemään juuri sen kyseisen teon, jos hän ylipäätään tekee jommankumman teoista. Tämäkin viittaa siihen, että toimija on perusteherkkä. Jos jotain muita perusteita ilmaantuu, hän kykenee toimimaan niiden perusteella.

Vapaa toiminta tässä kohtaa ei tietenkään ole äärimmäistä vapautta, vaan sillä on rajoitteita: toimija ei voi toimia toisin siinä mielessä, että hän toimisikin esimerkiksi toiseksi vahvimman halunsa tai minkään muun halunsa mukaisesti kuin sen, mikä on vahvin. Davidson ei kuitenkaan anna substantiaalisia ehtoja sille, mitä vahvin halu on toisin kuin esimerkiksi rationaalisen valinnan teoria, jonka mukaan yksilö toimii aina maksimoidakseen hyötyä. Davidson (1963) sallii teon motivaatioksi kaikenlaisia haluja, periaatteita, taloudellisia ennakkoluuloja, sosiaalisia

² "If an agent wants to do x more than he wants to do y and he believes himself free to do either x or y, then he will intentionally do x if he does either x or y intentionally." (Davidson 1980, 23.)

käytäntöjä, moraalisia näkemyksiä, yksityisiä ja julkisia tavoitteita ja arvoja, eikä niiden täydy olla sidottuna utiliteettiin tai palkkioihin. Tällaisessa viitekehyksessä motivaatio habituaaliseen toimintaan ei pitäisi olla ongelmallinen ainakaan sen suhteen, onko habituaalisessa toiminnassa esiintyvä motivaatio uhka tekoa ohjaavalle periaatteelle (kuten esimerkiksi hyödyn maksimoimisen periaatteelle). Toimija voi toimia vapaasti erilaisista perusteista. Jos kuitenkin korostetaan sitä, että toimija toimii perusteesta joka rationalisoi teon, kokonaisuus ei välttämättä olekaan ongelmaton, kun oletetaan, ettei habituaalisessa teossa ei ole toimintaan johtavia perusteita läsnä (vrt. Schroeder & Arpaly 2013). Seuraavaksi tarkennan mitä tarkoitan habituaalisella toiminnalla, jotta voin esittää ensimmäisen “haasteen” tavanomaisen toiminnan vapaudelle.

2. Tavoista

Angloamerikkalainen analyttinen teonfilosofia ei ole paljolti kiinnittänyt huomiota tapoihin tai tapakäyttäytymiseen, vaikka analyysejä automatisoidusta toiminnasta löytyykin (esim. Annas 2011, Browstein 2014, Di Nucci 2011, Railton 2009, Velleman 2008). Jossain mielessä tavanomainen toiminta tuntuu itsestään selvältä: toimija tietoisesti tai tiedostamattaan toistaa jotain tekoa tietyissä tilanteissa ja tällaisesta toiminnasta tulee hänelle tapa – jotain, jonka hän todennäköisesti suorittaa ollessaan samankaltaisissa tilanteissa ilman sen suurempaa pohdintaa miten tai miksi näin pitäisi tehdä. Habituaalinen toiminta on kuitenkin jotain, jonka toimija tekee eikä se vain tapahdu hänelle. (Pollard 2003.)³

Näyttäisi siltä, että tällaisen määritelmän ollessa kyseessä habituaalinen toiminta kattaa laajan joukon erityyppisiä toimintoja kynsien pureskelusta aamukahvin juomiseen tai teatterissa käymiseen. Analyysini kannalta on kuitenkin keskeistä valita paradigmaattinen toiminta, jota voidaan kutsua tavaksi ja jonka yhteensopivuutta voidaan analysoida vapaaseen toimintaan.

³ Pragmatismissa tapa ymmärretään täysin eri tavoin. Siinä missä perinteinen analyttinen filosofia näkee tavat ensisijaisiksi osiksi, joista tapa muodostuu, pragmatismissa tapa nähdään olevan ensisijainen tekoon nähden: teko on inhimillisen toimijan tavan ilmentymä (Kilpinen 2015). En kuitenkaan tässä kirjoituksessani lähde siitä, että tavat ovat yksilöiden itsensä toteuttamisen taipumuksia, jotka realisoituvat tekoihin, vaan lähden analyttisen filosofian perinteisestä näkökulmasta: teoista muodostuu tapoja ja habituaalista toimintaa.

Sellaista voisi olla habituaalinen toiminta, joka on pitkäjänteisen intentionaalisen opettelun tulosta. Erilaiset taidot opetellaan usein näin. Toisaalta habituaalista toimintaa ovat kynsien pureskelu, hiusten sukiminen, jalan heiluttelu, tai kynän naputtelu. Tällainen toiminta on kyllä toimijan aikaansaamaa liikettä, mutta tällaista toimintaa tyypillisesti tehdään tiedostamatta, eikä sitä tyypillisesti myöskään opetella tai pyritä jalostamaan paremmaksi. Pollardin (2003) mukaan habituaalinen toiminta eroaa kompulsioista ja reflekseistä siinä, että se on kuitenkin jotain, jonka toimija tekee eikä vain jotain, joka tapahtuu hänelle. Tämä viittaa siihen, että kompulsioista ja reflekseistä poiketen toimija hallitsee toimintaansa jossain määrin, mutta koska toiminta ei sisällä syvällistä harkintaa toimimisesta tai toimimatta jättämisestä, Pollard ehdottaa, että tällaista hallintaa kutsuttaisiin interventiohallinnaksi.

Olen kirjoituksessani kiinnostunut habituaalisesta toiminnasta, joka on jotain näiden kahden – taitojen opettelun ja kynsien pureskelun – väliltä: olen kiinnostunut arkipäiväisten toimien automatisoitumisesta. Näitä tekoja ei siis välttämättä suoriteta siksi, jotta kehityttäisiin tällaisessa toiminnassa, vaan ne automatisoituvat suurimmalta osaksi siksi, että toimija toistaa niitä riittävän usein.

3. Tekojen hallinta

Nyt voimme kysyä, onko habituaalinen toiminta uhka vapaalle toimijuudelle, koska se ei näytä olevan toimijan hallinnassa samalla tavalla kuin esimerkiksi silloin kun toimija harkitsee pitkään johonkin asiaan ryhtymistä ja suorittaa teon keskittyneesti. Onko habituaalinen toiminta kuitenkin riittävästi toimijan hallinnassa? Kuten edellä totesin, habituaalinen toiminta näyttää eroavan tällaisista täysiverisistä teoista siinä, ettei toimija samalla tavoin paneudu siihen kuin johonkin sellaiseen tekoon, jota edeltää ensinnäkin pohdinta siitä, tehdäänkö teko ylipäätään ja toisaalta siitä, kuinka paljon huomiota teon itse tekeminen toimijalta vaatii.

Habituaalisen toiminnan hallinta tulee olennaiseksi erityisesti silloin, kun kyseinen tapa ei ole enää toimijan intresseissä, ja hän haluaa joko muuttaa sitä tai lopettaa sen kokonaan. Addiktioita voidaan pitää esimerkkeinä pinttyneistä tavoista, sillä ne havainnollistavat hyvin tällaista vaikeutta. Tupakoitsija voi hyvinkin päätyä näkemykseen siitä, että hänen tapansa tupakoida on huono ja

hän haluaa lopettaa sen. Toimija on siis perusteherkkä siinä mielessä, että uusien perusteiden ilmaantuessa hän päättää muuttaa toimintaansa. Hän reagoi perusteisiin. Osaksi ongelmaa muodostuu kuitenkin se, ettei toimija ole habituaalisessa toiminnassa välttämättä “läsnä” ohjaamassa toimintaa senkaltaiseksi minkä hän on arvioinut olevan siinä kohtaa parasta. Toiminta on automatisoitunutta, jolloin tällaista toimintaa vastaan muodostetut harkitut päätökset eivät ole mukana toiminnan synnyssä. Habituaalinen toiminta näyttäisi tällöin uhkaavan toimijan parhaaksi näkemää toimintaa ja näin toimijan vapautta.

Intention muodostaminen ja siinä pitäytyminen vaativat toimijan aktiivista vaivannäköä ja yritystä, mutta koska tavat puolestaan tuntuvat sijoittuvan tietoisesta intentionaalisen toimijuuden syrjäseuduille tai reunoille, itsehallinnan yritykset voivat olla hankalia toteuttaa. Tällainen on helposti ymmärrettävissä ainakin sellaisista tavoista, kuten kynsien pureskelusta, joissa toimintaan ryhdytään melkein salakavalasti, huomaamatta. Tällaisesta toiminnasta puuttuu selkeästi toimijan sellainen hallinta, joka edellyttää teon alullepanemista. Pollard käyttää tällaisesta hallinnasta nimitystä “suora” hallinta ja toteaa, että habituaalisessa toiminnassa tyypillisesti tällaista hallintaa ei välttämättä ole, vaan toimijalla on hallinnan välineenä sen sijaan kapasiteettia puuttua toimintaansa (esimerkiksi lopettaa kynsien syöminen heti kun sitä huomaa tekevänsä).

Vaikka habituaalinen toiminta näyttäisi välttävän toimijan “suoran hallinnan”, se ei merkitse automaattisesti sitä, että se kyseenalaistaa vapaan toimijuuden. Otetaan addiktio taas esimerkiksi. R. Jay Wallace (2003) tarkastelee addiktiiviseen käyttäytymiseen liittyvää automaattisuutta. Hän toteaa, ettei siitä, että addiktio sisältää habituaalista toimintaa, voida vielä vetää johtopäätöstä, että addiktiossa esiintyvä habituaalinen toiminta viihjaa ongelmiin toimijan tahdossa, itsehallinnan ongelmiin. Siitä tosiasiasta, että toimija ei hallitse toimintaansa reflektiivisen pohdinnan kautta, ei voida päätellä suoraan, että hän ei siihen pystyisi. Habituaalinen toiminta ei kyseenalaista toimijan tahtoa siten, että se automaattisesti uhkasi toimijan hallintaa omista teoistaan. Jos tarkastelun kohteeksi ottaa jonkinlaisen muun toimijan kuin esimerkiksi addiktoituneen tai muun vastaavan toimijan, joka mielletään usein vähemmän vapaaksi, tämä ajatus näyttäisi tuntuvan luonteelta. Esimerkiksi toimija, joka käy aamuisin lenkillä, ei vaikuta

holtittomalta juoksunsa suhteen. Jokapäiväinen juoksu voi hyvinkin muodostua tavaksi, jonka toteuttamista juoksija ei lähde aamuisin pohtimaan, vaan hän suorittaa toiminnan rutiinilla. Silti on helppo kuvitella, että toiminta on toimijan suorittamaa alusta loppuun, eikä hänen tahtonsa lähtökohtaisesti kyseenalaistu tällaisessa toiminnassa.

Habituaalisen toiminnan hallinta voidaan nähdä analogisena vapaan tahdon ja moraalisen vastuun keskusteluun. Vapaalle toimijuudelle pitäisi olla riittävää, että hallinta on sellaista, joka mahdollistaa interventiot (esim. Fischer & Ravizza 1998). Tällainen interventiohallinta on riittävää siihen, että toimintaa voidaan pitää toimijan omina tekoina. Tavat voivat olla vaikeita muuttaa, mutta niiden muuttaminen ei ole mahdotonta. Tavan muuttaminen tai poisoppiminen voi vain vaatia (hyvinkin suuren määrän) vaivaa ja päättäväisyyttä. Tämän pohdinnan valossa habituaalinen toiminta ei luokitella sellaiseksi kontrolloivaksi toiminnaksi, jossa vapaa toimijuus olisi automaattisesti uhattuna.

Kuten edellä totesin, habituaalisen toiminnan määritelmään kuuluu, ettei tämä ole reflektion ja tällaisen rationaalisen toimijuuden tulosta. Tuntuu kuitenkin ilmeiseltä, että aamukahvin keittäminen on toimijan hallinnassa, vaikka hän ei toiminnan alkamiseksi ensin pohdi ja tee päätöstä kahvinkeittämisestä. Vaatimus tällaisesta tuntuisi antavan aivan liian intellektuaalisen kuvan inhimillisestä toiminnasta, ja se olisi kuva, joka ei tunnu vastaavan todellisuutta.

Palatakseni Davidsoniin (1980), hän toteaa, että toimiakseen intentionaalisesti toimijan täytyy haluta tekoa a enemmän kuin tekoa b ja uskoa olevansa vapaa toimimaan halunsa mukaisesti. Myös Hanna Pickard (2012) on todennut, että muodostaakseen intention, yksilön pitää uskoa, että hän pystyy toimimaan intentionsa mukaisesti tilanteen huomioiden.

Toinen haaste vapaalle toimijuudella onkin, ettei habituaalinen toiminta ole toimijan vapaata toimintaa, koska sitä ei saada aikaan järkiperustein. Tällainen keskustelu pohjautuu kysymyksiin erilaisista toimintaan motivoivista lähteistä. Toimijuus ei ole yksinkertaisesti sellainen kenttä, jossa kaikki toimintaa motivoivat seikat ovat tavoitettavissa samassa määrin. Habituaalisen toiminnan ollessa kyseessä vaikeudet sen hallitsemiseksi eivät ole palautettavissa mihinkään addiktiossa tai muussa oireyhtymässä tapahtuvaan aivojen virhetoimintaan, vaan sen haasteellisuus piilee tapojen

olemuksessa yleisemmin. Ihmiset kamppailevat tapojensa kanssa muutenkin kuin vain addiktioissa ja muissa häiriöissä, olipa kyseessä sitten uuden tavan opettelu, vallitsevan tavan muokkaaminen tai yksinkertaisesti jonkin tavan lopettaminen.

Habituaalinen toiminta selittää myös addiktioiden ja muiden häiriöiden haasteellisuutta: mitä kauemmin toimija esimerkiksi polttaa, sitä vaikeammaksi tavan muuttaminen tulee, ei pelkästään nikotiiniriippuvuuden vuoksi, vaan myös toisto vahvistaa tapaa entisestään.

Habituaalisen toiminnan voidaan nähdä myös olevan intentionaalista toimintaa, missä toimijalla ei näytä aina olevan tietoista intentiota toiminnan suorittamiseen.⁴ Jossain mielessä habituaalinen toiminta voidaan nähdä samankaltaisena kuin toiminta, joka toteutetaan välinpitämättömästi. Toimija toimijana ei ole täysillä mukana. Pahoissa tavoissa tämä näkyy hyvin, koska silloin toimijan aktiivisuus voi itseasiassa asettua habituaalista toimintaa vastaan jossain mielessä. Vastentahtoinen tupakoitsija voi kamppailla muokatakseen tapaansa, mutta se voi olla haasteellista, koska koko toiminta tuntuu tapahtuvan jossain muualla kuin toimijuuden keskiössä. Tällainen taistelu ei kuitenkaan ole taistelua kahden voimakkaan halun välillä, joista voimakkain sitten voittaa, vaan pikemminkin taistelua siitä, millä kentällä/tasolla toiminta tapahtuu, eikä kysymys silloin palaudu mihinkään toimijan valintaan tai halun voimakkuuteen. Toimijan aktiivisuutta habituaalisissa teoissaan ei siis tulisi mitata pelkästään reflektion kautta.

Asian ongelmallisuus näyttää kuitenkin erilaiselta kun tarkastellaan toimijan aktiivisuutta tai passiivisuutta enemmän positiivisten tapojen valossa. Jo Aristoteles ja Hume, esimerkiksi, näkivät tavat täysiverisen moraalisen toiminnan kulmakivinä. Tällaisen katsantokannan mukaan tavat ovat jotain, mikä toteuttaa ihmisluonnon periaatetta. (Pollard 2013, 76.) Tällaisessa näkökulmassa sen, ettei toimija tiedosta ja ole valppaana omista perusteistaan ja intentioistaan toiminnan hetkellä, ei tarvitse olla merkki passiivisuudesta. Sen sijaan se voi olla merkki todella hyveellisestä ihmisestä, joka yksinkertaisesti tietää pohtimatta mitä tehdä ja miten toimia hyveellisesti. Usein ymmärrämme opitut taidot tämänkaltaisesti. Niinkin arkiset taidot kuin

⁴ Joissain tapauksissa voidaan sanoa, ettei toimijalla ole toimijana välttämättä ollenkaan intentiota toimia tavalla, jolla toimii. Henkilö joka syö kynsiään, ei välttämättä ole koskaan tahtonut pureskella kynsiään, hänellä ei ole ollut siihen perustetta, silti hänen toimintansa voidaan nähdä intentionaalisena.

pyörällä ajaminen tai käveleminen eivät tyypillisesti vaadi toimijan täydellistä huomiota kyseisen suorituksen aikana tai sen alkuunpanossa. Itse asiassa, tällainen automatisoituminen vapauttaa toimijan mielen muunlaiseen aktiviteettiin (Pollard 2005) ja tällä tavalla omalta osaltaan mahdollistaa vapaan toiminnan, koska näin resursseja jää reflektoinnille ja vaivannäölle.

Jossain mielessä tavat ovat siis jotain hyvin rakentavaa toimijuudelle. Ne ilmaisevat autenttisuutta: toimijan olemista ja toimimista maailmassa. Kiinnostava kysymys kuitenkin on, miten autenttisuus suhteutuu autonomiaan, jos autonomia ymmärretään itsehallintana.

Nähdäkseni ihmiset eivät tyypillisesti kiinnitä huomiota arkipäiväisiin rutiininomaisiin tekoihinsa paitsi jos olosuhteet muuttuvat poikkeukselliseksi. Esimerkiksi kävelyreitti onkin aivan jäässä ja jokaisen askeleen ottaminen vaatii keskittymistä. Tällöin he näyttäisivät kuitenkin olevan perusteherkkiä toimintansa suhteen ja näin hallitsevansa toimintaansa. En näe järkevänä näkemystä, missä ollakseen inhimillistä toimintaa teko edellyttää tietoisia perusteita tai intentioita. Se olisi jälleen liiallista inhimillisen toiminnan intellektualisoimista. Silti tuntuu huolestuttavalta, että habituaalinen toiminta ei ole täysin saatavilla ja muokattavissa itsekurillisilla toimenpiteillä.

Tässä huolella tuntuu silti olevan piilotettu premissi. Huoli pätee vain sellaiseen toimintaan, joka on jollain tavalla ongelmallista. Ja tämä ongelmallisuus tuntuu liittyvän rationaalisuuteen. Hyveellisen toiminnan kanssa huoli ei näytä realisoituvan. Kuitenkin näyttää siltä, että habituaalinen toiminta näyttää täyttävän jonkinlaisen minimaalisen rationaalisuuden ehdon, myös pahoissa tavoissa. Otetaan addiktio taas esimerkiksi. Wallace (2003, 428) kuvailee addiktion habituaalista luonnetta seuraavasti:

Sellaiset automatisoituneet rutiinit mitä addikti toteuttaa oletettavasti kehittyvät habituaation kautta, sopeutumisenä riippuvuuden tilaan, sopeutumisenä tapoihin, joilla välttää jatkuvasta aineen puutostilasta johtuvia epämiellyttäviä vaikutuksia.⁵

⁵ "Such automatic routines as the addict exhibits presumably develop through habituation, as adjustments to one's state of dependency, ways of staving off the unpleasant effects of continued substance deprivation" (Wallace 2003, 428).

Tämä kuvaus havainnollistaa, että voi olla yksinkertaisesti rationaalista kehittää tapa, joka täyttää addiktiiviset halut, jos toimija näin välttyy epämiellyttäviltä vieroitusoireilta. Toiminta vaikuttaa siis jossain mielessä järkevältä, eikä itsessään kerro vielä mitään hallinnallisista ongelmista. Kuten totesin, se, että toimija aloittaa päivänsä tupakalla, kiinnittämättä toimintaansa paljoakaan huomiota, ei vielä tarkoita, että kyseisen toimijan tahto olisi kyseenalaistunut, vaikka toimijan tahto ei sillä hetkellä olisikaan kovin aktivoitunut. Vaivattomuus tuntuu pikemminkin mahdollistavan kuin rajoittavan toimijan toimintaa.

Jos rationaalisuus pitää sisällään muutakin kuin instrumentaalisen rationaalisuuden vaatiman koherenssin keinojen ja päämäärien välillä, kysymys habituaalisen toiminnan ongelmallisuudesta nousee uudelleen esille. Tämä kuitenkin vaatii substantiaaalisten rationaalisuuden ehtojen muodostamista, ja se on jotain mitä en tässä aio lähteä käymään läpi.⁶

4. Lopuksi

Olen tässä kirjoituksessani tarkastellut onko habituaalisessa toiminnassa – sen määritelmällisessä luonteessa – jotain sellaista joka kyseenalaistaa vapaan toimijuuden. Tarkasteluni valossa habituaalinen toiminta ei itsessään toteuta sellaista (toimijasta riippumatonta) hallintaa, joka suoraan uhkaisi toimijan vapautta. Habituaalinen toiminta näyttäisi myös olevan riittävästi toimijan hallinnassa, jotta sen voidaan nähdä olevan toimijan toimintaa eikä esimerkiksi refleksin- tai pakonomaista liikehdintää. Tietoisien reflektoinnin puuttuminen ei näytä automaattisesti uhkaavan vapaata toimintaa. Järkiperusteet kuitenkin tuntuvat olevan keskeisiä vapaalle toimijuudelle, ja kysymys habituaalisen toiminnan vapaudesta näyttää riippuvan näistä perusteista. Jotta voitaisiin hyväksyä, että habituaalinen toiminta on uhka vapaalle toimijuudelle, pitää olla enemmän oletuksia siitä, minkälaisesta rationaalisuudesta habituaalisessa toiminnassa kulloinkin on oikeastaan kyse.

⁶ Pollard (2005) ei pyri määrittelemään rationaalisuutta antamalla sille ehtoja, vaan määrittelemällä sen mikä on irrationaalista ja argumentoimalla, että kaikki mikä ei ole irrationaalista, on rationaalista. En näe kuitenkaan tämän menetelmän väistävän niitä haasteita, mitä substantiaaalisen rationaalisuuden määrittelemisessä nousee esiin.

Totean loppuun, että jos habituaalinen toiminta kyseenalaistaa vapaan toiminnan, muita oletuksia tulee olla läsnä kuin pelkästään ne, jotka koskevat puhtaasti tapoja tai habituaalisen toiminnan luonnetta.

Helsingin yliopisto
Turun yliopisto

Kirjallisuus

- Annas, Julia (2011). *Intelligent Virtue*. New York: Oxford University Press.
- Arpaly, Nomy & Schroeder, Timothy (2014). *In Praise of Desire*. Oxford: Oxford University Press.
- Brownstein, Michael (2014). "Rationalizing Flow: Agency in Skilled Unreflective Action", *Philosophical Studies* 168, 545–568.
- Dancy, Jonathan & Sandis, Constantine (toim.) (2015). *Philosophy of Action: An Anthology*. Chichester: Wiley-Blackwell.
- Davidson, Donald (1963). "Actions, Reasons, and Causes", *Journal of Philosophy* 60 (23), 685–700.
- Davidson, Donald (1980 [1970]). "How Is Weakness of Will Possible?", teoksessa Donald Davidson, *Essays on Actions and Events*. Oxford: Clarendon Press, 21–42.
- Di Nucci, Ezio (2011). "Automatic Actions: Challenging Causalism", *RRM* 2, 179–200.
- Fischer, John Martin & Ravizza, Mark, S. J. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge Studies in Philosophy and Law. Cambridge: Cambridge University Press.
- Kilpinen, Erkki (2015). "Habit, Action, and Knowledge from the Pragmatist Perspective", teoksessa Ulf Zackariasson (toim.), *Action, Belief and Inquiry—Pragmatist Perspectives on Science, Society and Religion*. Nordic Studies in Pragmatism 3. Helsinki: Nordic Pragmatist Network, 157–173.
- Mele, Alfred. R. (toim.) (1990). *The Philosophy of Action*. Oxford Readings in Philosophy. Oxford: Oxford University Press.
- Pickard, Hanna (2012). "The Purpose in Chronic Addiction", *AJOB Neuroscience* 3 (2), 40–49.
- Pollard, Bill (2003). "Can Virtuous Actions Be both Habitual and Rational?", *Ethical Theory and Moral Practice* 6, 411–425.
- Pollard, Bill (2005). "The Rationality of Habitual Actions", *Proceedings of the Durham-Bergen Postgraduate Seminar* 1, 39–50.
- Pollard, Bill (2013). "Habitual Actions", teoksessa Constantine Sandis & Timothy O'Connor (toim.), *Blackwell's Companion to Philosophy: A Companion to Philosophy of Action*. Chichester: Wiley-Blackwell, 74–81.
- Railton, Peter (2009). "Practical Competence and Fluent Agency", teoksessa D. Sobel & S. Wall (toim.), *Reasons for Action*. Cambridge: Cambridge University Press, 81–115.
- Sandis, Constantine & O'Connor, Timothy (toim.) (2013). *Blackwell's Companion to Philosophy: A Companion to Philosophy of Action*. Chichester: Wiley-Blackwell.

- Schroeder, Timothy (2010). "Irrational Action and Addiction", teoksessa Don Ross, Harold Kincaid & David Spurrett (toim.), *What Is Addiction?* Cambridge, MA: MIT Press. (DOI:10.7551/mitpress/9780262513111.003.0016)
- Snow, Nancy E. (2006). "Habitual Virtuous Actions and Automaticity", *Ethical Theory and Moral Practice* 9, 545–561.
- Velleman, J. David (2008). "The Way of the Wanton", teoksessa Catriona Mackenzie & Kim Atkins (toim.), *Practical Identity and Narrative Agency*. New York: Routledge, 169–192.
- Wallace, R. Jay. (2003[1999]). "Addiction as a Defect of the Will: Some Philosophical Reflections", teoksessa Gary Watson (toim.), *Free Will*. Second Edition. Oxford: Oxford University Press, 424–452.

J. V. Snellmanin käsitys oman edun roolista valtiossa¹

Juho Ritola

1. Johdanto

Tämän artikkelin tarkoituksena on tarkastella kansallisfilosofimme J. V. Snellmanin näkemyksiä oman edun ajamisen ja yhteisen hyvän tavoittelun suhteesta poliittisessa keskustelussa. Lähtökohtani on demokraattisen deliberaation teoria, jonka mukaan demokratian legitimitetti perustuu siihen, että sen päätökset voisivat olla kansalaisten, eli päätöksen alaan kuuluvien moraalisten ja rationaalisten agenttien, harkintaan perustuvan vapaan hyväksynnän kohde. Mikä tahansa yhteinen keskustelu ei legitimoisi poliittista päätöstä vaan ideaaliselle keskustelulle annetaan useita ehtoja. Yksi keskeinen vaatimus koko traditiolle on kansalaisten kyky tämän keskustelun aikana asettaa oma etunsa syrjään ja harkita millaiset päätökset parhaiten edistäisivät yhteistä hyvää.²

Snellmanille vaatimus yhteisen hyvän ensisijaisuudesta demokraattisessa päätöksenteossa on keskeinen. Henkilö on valtion jäsen silloin, ja vain silloin, kun hän kohdistaa toimensa yhteiseen etuun:

[...] on helpointa erottaa kansalaisyhteiskunta valtiosta sanomalla: yksilö on edellisen jäsen sikäli kuin hän tavoittelee yksityisiä päämääriään, omaa etuaan loukkaamatta yhteiskunnan muiden jäsenten vapautta. Sen sijaan valtion jäsenenä hänen toimintansa kohteena on yleinen, itse valtion etu [...] (*Kootut Teokset* (KT) V, 76–77).

¹ Tämän artikkelin kirjoittamisen mahdollistivat Voitto Aaltosen testamentissa J. V. Snellmanin tutkimukseen osoitetut rahat. Kiitän myös Juha Räikkää ja Lauri Kalliota Snellmaniin liittyvistä keskusteluista ja kommunikaatioista.

² Bohman ja Rehg (1999, x): ”Ongelma, Kantin termejä käyttäksemme, on saada aikaan ’järjen julkinen käyttö’.” Näitä ehtoja on käsitelty laajasti kirjallisuudessa. Katso esim. Dahl 2006 [1979] ja Cohen 1999 [1989]. Ritola 2015 sisältää lyhyen katsauksen deliberaatiolle asetettuihin ehtoihin.

Vuonna 2010 ryhmä huomattavia deliberaation tutkijoita julkaisi artikkelin, jossa he argumentoivat, että kansalaisen oman edun tulisi olla osa demokraattista päätöksen tekoa:

[...] esitämme, että oman edun, asianmukaisesti rajoitettuna, tulisi olla osa deliberaatiota joka päättyy demokraattiseen päätökseen. Todellakin, jotkut neuvottelun muodot, joissa oma etu esiintyy, täyttävät kaikki ideaalisen deliberaation kriteerit [...] (Mansbridge 2010, 64).³

Mansbridge pyrkii osoittamaan, ettei oman edun ajaminen välttämättä ole ristiriidassa yhteisen edun kanssa, ja joissain tapauksissa se, että jokin ratkaisu tukisi esittäjänsä etua myös edistäisi yhteistä hyvää. Esitän artikkelissani miten Snellmanin näkemys oman edun roolista päätöksenteossa seuraa hänen moraalisisista ja poliittisista näkemyksistään. Esitän snellmanilaisen⁴ rekonstruktion oikeuttamisesta valtiollisessa kontekstissa, joka sopii hyvin yhteen Mansbridgen argumentoinnin kanssa.

Milloin oman edun ajaminen sitten on oikeutettua? Snellmanin voidaan katsoa esittävän kysymykseen perustellun ratkaisun: kansalaisen on voitava ajaa omaa etuaan silloin, kun hänen oma etunsa käy yksiin oikean kanssa. Yhteiskuntafilosofisen keskustelun tarkastelemien periaatteiden perusteella ei kuitenkaan ole mahdollista antaa tarkempaa vastausta siihen milloin näin on. Jokaisen ajan on löydettävä vastaukset omasta, kyseiseen tapaukseen kohdistuvasta vapaasta ja järkeen perustuvasta harkinnastaan. Tämä harkinta on riippuvainen oman aikansa tarjoamasta sivistyksestä ja se on jatkuvassa kehitysprosessissa: kunkin ajan harkinta parantaa edelleen sitä sivistyspohjaa, jonka perusteella tulevat polvet arvioivat eteensä tulevia päätöksiä. Harkintansa perusteella kansalaisten tulee käydä keskustelua, jossa heidän johtopäätöksiä ja perusteluja punnitaan toisten kansalaisten toimesta.

³ Artikkelin kirjoittajiksi ilmoitetaan Jane Mansbridge yhdessä (*with*) James Bohman, Simone Chambers, David Estlund, Andreas Føllesdal, Archeon Fung, Cristina Lafont, Bernard Manin ja José Luis Martí. Alaviitteen * mukaan artikkelin varsinainen kirjoittaja on Mansbridge, mutta artikkeli perustuu yhteisiin ideoihin ja huolenaiheisiin, ja se muotoutui otsikon nimeämien henkilöiden välisessä pitkällisessä vuorovaikutuksessa. Viittaan artikkeliin tästedes Mansbridgen kirjoittamana.

⁴ Esitän yhden rekonstruktion Snellmanin ajatuksista. Kirjoitukseni tavoite ei ole eksegeettinen vaan analyttinen. Rekonstruktion ratkaisee huonojen lakien ongelman fallibilismiin vedoten, toisin kuin Tenkku (1984), jonka mukaan Snellman oli yhteiskunnallisella tasolla relativisti.

2. Oma etu ja deliberaation teoria

Kuten todettua, deliberatiivisen demokratian kantava ajatus on ollut vaatimus yhteisen hyvän⁵ tavoittelusta päätösten teossa. Deliberaation klassisessa mallissa kansalaiset, joita tuleva päätös sitoo, tulevat yhteiselle areenalle erilaisin mielipitein siitä mikä olisi parasta kaikille. He esittävät ja kuuntelevat toistensa perusteluja ratkaisuilleen, ja tämän oikeudenmukaisen ja järkevän prosessin perusteella päätyvät yhteen parhaaksi katsomaansa ja yhteistä etua parhaiten toteuttavaan ratkaisuun yhteisistä perusteista käsin (Mansbridge 2010, 66.)⁶ Klassisen näkemyksen mukaan tällaisen ja vain tällaisen prosessin perusteella syntynyt ratkaisu on legitiimi poliittinen päätös.

Monet kirjoittajat⁷ ovat kuitenkin halunneet laajentaa tätä klassista kuvaa, jonka on koettu viittaavaan valistuksen tavoittamattomaan ja yksiselitteiseen ”järjen” käsitteeseen.⁸ Yksiselitteisen järjen sijaan tulisi riittää, että keskustelijat esittävät toisilleen näkökohtia, jotka kuulijat voivat hyväksyä: keskustelijoiden tulisi pyrkiä vastavuoroiseen oikeuttamiseen. Käytettävissä olevat perusteet eivät aina riitä yksiselitteisen ratkaisun löytämiseen, joten meidän tulisi hyväksyä pluralistisempi käsitys yhteisestä hyvästä. Katsotaan myös, että toisinaan arvot ja intressit voivat olla sovittamattomassa ristiriidassa. Tällöin deliberaation rooli on selventää erimielisyyden luonnetta ja rakennetta ja tarjota siten perusta ei-deliberatiivisille päätöksentekomenetelmille, kuten äänestämislle. Mansbridge haluaa kuitenkin meidän menevän vielä pidemmälle. Hän argumentoi, että kaikkien legitiimien demokratia-käsitysten, myös deliberatiivisen demokratian, tulee sisällyttää ideaaliinsa oma etu ja intressien konflikti, jotta tunnustaisimme vapaiden ja tasa-arvoisten ihmisten erilaisuuden. (Mansbridge 2010, 67–69.)

Väite valistuksen järki-käsitteen mahdottomuudesta tai tavoittamattomuudesta on monimutkainen, emmekä voi siihen tässä kunnolla pureutua.⁹ Mielestäni väitettä ei kuitenkaan tarvitse hyväksyä voidakseen hyväksyä Mansbridgen näkemyksen, jonka mukaan keskustelijan omaa etua ei

⁵ Mansbridge 2010 (alaviite 14) sisältää viitteitä keskusteluun tästä käsitteestä.

⁶ Mansbridgen (2010, alaviite 3) mukaan tämä on regulatiivinen ideaali, joka ei ole kokonaisuudessaan saavutettavissa mutta jota ratkaisun tulisi tavoitella, ja johon sitä verrataan (ks. Kant 1998 [1781/1787], 552, A569/B597; myös A570/B598).

⁷ Ks. Mansbridge 2010, alaviite 12.

⁸ Palaamme tähän myöhemmin.

⁹ Palaamme tähän myöhemmin.

tule sulkea keskustelun ulkopuolelle. Hän (2010, 70–72) esittää neljä erilaista deliberatiivisen neuvottelun¹⁰ tyyppiä, joissa oman edun esilletuominen on intuitiivisesti hyväksyttävää: konvergenssi, epätäydelliset yksimielisyydet, integroivat neuvottelut ja täysin kooperatiiviset distributiiviset neuvottelut.

Konvergenssi: Neuvottelijat aloittavat erilaisista mielipiteistä ja intresseistä mutta päätyvät yhteen yhteiseen lopputulokseen. Voidakseen tehdä niin, heidän on saatava tietää toistensa mielipiteet; se informaatio ja ne intressit, mitä muilla asiaan liittyen on. Jaettuaan tiedot ja tarpeet keskenään, he päätyvät ratkaisuun, joka on paras kaikille.

Epätäydellinen yksimielisyys: Neuvottelijat päätyvät samaan lopputulokseen mutta eri syistä. Voimme olettaa, että heillä on riittävästi yhteisiä välipremissejä, joiden perusteella lopputulema voidaan oikeuttaa, mutta he perustelevat nämä väliperusteet eri tavoin.¹¹

Integroiva neuvottelu: Neuvottelijoilla on eri intressit, mutta he löytävät ratkaisun, joka ratkaisee molempien ongelmat.¹²

Täysin kooperatiivinen distributiivinen neuvottelu: Eroaa edellisistä siinä, että tapahtuu ”toisen hyöty on toisen voitto” eli niin sanottu nollasummapeli, mutta molemmat osapuolet haluavat löytää yhteisen ratkaisun oikeudenmukaisesti, ilman strategista suunnittelua ja avoimesti.

Oma etu voi siis olla mukana keskusteluissa intuitiivisesti hyväksyttävällä tavalla. Osapuolet itse tietävät parhaiten, mikä heidän etunsa on, ja toisinaan heidän täytyy puolustaa itseään. Edelleen, vaikka keskustelun tavoite olisikin konsensus, keskustelijoiden etujen kartoitus on tärkeä osa prosessia. Mansbridge (2010, 74–75) jakaa oman edun roolin kahteen päätyyppiin. Ensinnäkin, se voi olla osa informaatiota, jota keskustelu tarvitsee löytääkseen ratkaisun, joka parhaiten palvelee yhteistä etua. Se, että jokin ratkaisu

¹⁰ Mansbridge (2010, 69) jakaa pakottavat (*coercive*) ja ei-pakottavat interaktiot eri tavoin kuin jako tavallisesti tehdään. Yleensä deliberaatio luokitellaan ei-pakottavaksi ja neuvottelu pakottavaksi. Hänen mukaansa neuvottelut voivat kuitenkin olla myös ei-pakottavia ja deliberatiiviset neuvottelut ovat juuri tällaisia.

¹¹ Tätä näkemystä voidaan tukea eri tavoin. Ernest Sosa on todennut erimielisyyden epistemologiaa käsittelevässä kirjoituksessaan: ”Ajatus, että pystymme aina, tai edes usein, huomaamaan oman operatiivisen ’evidenssimme’, on myytti.” (Sosa 2010, 291.)

¹² Toisin kuin edellisessä tyypissä, tämän tyyppin neuvottelut on alun perin tarkoitettu ratkaisemaan intressikonflikteja. Mary Parker Folletin (1942 [1925]) alkuperäisessä esimerkissä yksi ihminen haluaa ikkunan kiinni, koska sieltä vetää ja toinen haluaa sen auki, koska sieltä tulee raikasta ilmaa. Molempia tyydyttävä ratkaisu on avata ikkuna toisesta huoneesta; niin sanottu *win-win* -ratkaisu.

esimerkiksi tuhoaisi jonkin ryhmän elinkeinon, on painava peruste välttää tätä ratkaisua, edellyttäen, että kyseinen elinkeino ei muutoin ole tuomittava. Toiseksi, joissain tapauksissa yhteinen etu voi olla muodostunut yksittäisten keskustelijoiden omasta edusta. Tällaisessa aggregatiivisessa tapauksessa oma etu olisi sinällään oikeuttava peruste.

Mansbridge kuitenkin korostaa, ettei tämä tarkoita, että mikä tahansa intressi olisi automaattisesti legitiimi. Sen roolia voidaan rajoittaa sekä universaalein moraalisuuteen liittyvin perustein että deliberaatioon liittyvin vastavuoroisuuden ja kunnioituksen periaattein. Esimerkiksi tavoitetta ”olla rikas keinoista riippumatta” ei tarvitse pitää oikeuttavana tavoitteena, mutta tavoite ”olla niin rikas kuin mahdollista, tasolla, jota muut voivat pitää oikeudenmukaisena ja hyväksyttävänä”, voi olla hyväksyttävä. Mansbridgen mukaan kansalaisilla on oikeus vaatia yhteiskunnan instituutioilta, että ne edistävät heidän etujaan, olettaen että nämä edut ovat hyväksyttäviä yleisen moraalin ja deliberaation näkökulmasta.¹³ Tällaisilla deliberatiivisesti rajoitetuilla vaateilla on oma painonsa, ilman että niitä tarvitsee johtaa laajemmasta konsekventialisesta tai deontologisesta teoriasta. (Mansbridge 2010, 77.) Käännymme nyt Snellmanin käsitysten puoleen.

3. Snellman

3.1. Snellman ja oikea

Snellmanin käsitys oikeasta teosta perustuu hänen käsitykseensä legitiimistä yhteiskunnasta ja sen oikeutuksesta. Hänen mukaansa ”[...] valtion ulkopuolella ei näet ole mitään oikean ja väärän korkeinta normia [...]” (KT V, 54).¹⁴ Esitän vaihtoehdoisen tulkinnan kuin Jussi Tenkku (1984),¹⁵ jonka mukaan Snellman on eettinen relativisti: katson, että Snellman vain rajaa

¹³ Mansbridge siteeraa Rawlsia, jonka mukaan ”[kansalaiset ovat] oikeutettuja esittämään vaateita instituutioilleen edistääkseen omaa käsitystään yhteisestä hyvästä, olettaen että nämä käsitykset mahtuvat oikeuden käsitteen sallimaan alaan. Näillä vaateilla [on] oma painonsa ilman, että niitä johdetaan velvollisuuksista, joita poliittinen oikeuden käsitys vaatii, esimerkiksi velvollisuuksista valtiota kohtaan.” (Rawls 1993, 32.)

¹⁴ Myös: ”Kansakunnan tapojen ja lakien noudattaminen on itse kunkin toimien korkein normi ja niin täytyy ollakin.” (KT V, 52.)

¹⁵ Tässä yhteydessä ei ole mahdollista käsitellä Tenkun argumentointia tarkemmin, mutta annan mielestäni koherentin esityksen Snellmanista, joka osoittaa sinällään miksi relativismi ei seuraa hänen ajattelustaan.

moraalin alan ihmisiin ja heidän väliseen toimintaan.¹⁶ Lakien oikeudenmukaisuutta voidaan arvioida monin kriteerein, mutta yksi periaate on ehdoton: ”[...] jokainen laki, joka kieltää persoonallisuuden ja tarkastelee ihmistä vain välineenä, on hylättävä.” (KT V, 113–114.) Snellman ei viittaa *Valtio-opissa* kertaakaan Immanuel Kantiin, jolta tämä periaate on peräisin. Tämä periaate ei sinällään vielä implikoi, että moraali rajautuu vain ihmisiin. Sitaatin perusteella vaikuttaisi kuitenkin luontevalta olettaa, että Snellman hyväksyy, että teko on moraalisesti arvioitavissa jos ja vain jos se vaikuttaa toisiin ihmisiin, jotka ovat rationaalisia ja moraalisia olentoja.

Rationaalisen ihmisen vapaus on Snellmanin yhteiskuntafilosofian lähtökohta. Keskeinen kysymys, johon hän *Valtio-opin* alussa pyrkii vastaamaan on, miten sovittaa yhteen tämä ihmisen perustava vapaus sekä valtio, joka olemuksellaan vaikuttaisi rajoittavan tätä vapautta. Ratkaisu on, että yksilö oman tahtonsa perusteella sitoutuu valtioon:

Onkin myönnettävä, että yhteiskunnan olemassaolo kokonaisuudessaan olisi mahdottomuus ellei yksityinen tahto, samalla kun se säilyttää vapautensa, voisi alistua yleiseen tahtoon. – Yksityisen tahdon mielivaltainen herruus näet purkaisi kaiken ihmisten välisen vapaan yhteyden ja saattaisi voimaan vahvemman oikeuden, kun taas yksilö luopuisi vapaan tahtonsa ohella ihmisyydestään, kieltäisi henkensä olemuksen. Mutta silloin kun monet solmivat vapaaehtoisen sopimuksen, jonka mukaan jokainen sitoutuu alistamaan tahtonsa yleiseen tahtoon, kaikki säilyttävät tästä huolimatta vapautensa, koska jokainen tahto sisältyy yleiseen tahtoon, ja niin ollen yleisen tahdon esittämät lait ovat jokaisen tahdon työtä. (KT V, 34.)

Snellman katsoo, Rousseauhun viitaten, valtion perustuvan sopimukseen, mutta tätä sopimusta ei tule pitää empiirisenä väitteenä (KT V, 34). Se on ”valtion yleisen luonteen ja olemuksen ilmaus.” (KT V, 35.) Edelleen voimme kuitenkin kysyä, miksi tahtoa voidaan pitää vapaana tässä sopimuksessa. Snellmanin mukaan ”[t]ahdon vapautta ei puolestaan saa määrätä

¹⁶ Voimme olla tästä rajauksesta perustellusti eri mieltä ja ajatella, että esimerkiksi eläimiin, ja laajemmin luontoon, kohdistuvat teot kuuluvat myös moraalisen arvioinnin piiriin. Mutta tämä erimielisyys ei oikeuta väitettä relativismista.

mielivaltaisesti, koska ihmisestä tulisi siten satunnaisten nautintojen ja halujen orja ja hän olisi vailla vapautta [...]” (KT V, 36). Pelkkä halujen tyydyttäminen ei ole sitä mitä tarkoitetaan rationaalisen moraalisen ihmisen vapaudella:

Ollakseen vapaa tahdon on siten oltava järjellinen, toisin sanoen sen on haluttava sitä mikä on itsessään ja sinänsä totta ja oikeaa. Itsessään ja sinänsä tosi on puolestaan sitä, mikä jokaisen järjellisen ihmisen täytyy tunnustaa todeksi. Jos jokainen yksityinen ihminen nyt tahtoo vain tätä yleispätevää, järjellistä ja totta, silloin yksityinen tahto käy välittömästi yhteen yleisen tahdon kanssa. Jokainen tahto on silloin yhtäläisesti vapaa, yleinen tahto ei sido sitä enemmistön päätöksen muodossa, vaan se seuraa toiminnassaan sitä, mikä jokaisen järjellisen ihmisen täytyy tunnustaa todeksi ja oikeaksi. (Ibid.)

Tämä kuva herättää kuitenkin kriittisen kysymyksen. Jos on niin, että yhteiskunnan perustana oleva sopimus, ja tämän sopimuksen erilaiset tarkennukset, eli kansalliset lait, perustuvat järkeen, jonka jokainen rationaalinen ihminen voi itsessään käsittää, miten on mahdollista, että niin monissa yhteiskunnissa on voimassa lakeja, jotka vaikuttavat moraalisen harkinnan valossa selvästi epäoikeudenmukaisilta? Edelleen, jos ”valtion ulkopuolella ei näet ole mitään oikean ja väärän korkeinta normia”, onko meidän nyt hyväksyttävä, että moraalit on relatiivista, että meidän moraalimme näkökulmasta ne ovat väärin, mutta kyseisen valtion sisällä ne ovat oikein? Tämä ei ole Snellmanin kirjoitusten mukaan perusteltua:

Erehtyisimme pahasti, jos pitäisimme oikeaa niin sanottuna tyhjänä käsitteenä, jolla ei ole vastaavaa objektia. Silloin oikea olisi mielivaltainen mielikuvituksen tuote, joka saisi uuden muodon kunkin yksilön mielihalun mukaan. Jos sitä vastoin tunnistamme, että oikea itsessään on olemukseltaan määrätty, eikä käsitys siitä saa siis riippua yksilön mielivallasta, vaan oikean luonteen on se määrättävä, niin samalla myönnämme, että oikea on yksilön tiedon kohde, vaikka sen olemassaolon onkin oltava tästä tiedosta riippumaton. (KT V, 43.)

Snellman sitoutuu siis moraaliseen realismiin yksilötasolla. Tämä ei sinällään vielä rajaa pois kollektiivisen tason relativismia, jota Tenkku (1984, 81) katsoo Snellmanin edustavan. Mutta jos kollektiivisen tason päätös, yleinen tahto, on yhteneväinen yksilötahdon kanssa, kuten Snellman oikealta päätökseltä edellyttää (ks. yllä), on vaikea nähdä miten kollektiivinen relativismi olisi Snellmanin mukaan mahdollista.

3.2. Väärän ongelma: fallibilismi ja dialektinen kriteeri

Tämä ei kuitenkaan vastaa ongelmaan menneistä, vallitsevista ja tulevista vääryyksistä. Tähän voidaan Snellmanin kirjoitusten perusteella antaa kaksi vastausta. Toisaalta kyse on kehityksestä ja toisaalta vallitsevista olosuhteista. Snellmanin (KT V, 36) mukaan valtion säilyminen vaatii, että kaikki sen piiriin kuuluvat ajattelevat ja toimivat järkevästi ja siveellisesti; käsitteellinen yhteiskuntasopimus ei riitä ylläpitämään valtiota:

Vain kansakunnan siveellisenä sivistyksenä, kansallishenkenä on siis ajateltavissa se yksityisen ja yleisen tahdon käyminen yksiin, jota jo yhteiskuntasopimus edellyttää – on ajateltavissa näin, jotta se voidaan tietää ja ilmaista ja ilmaistuksi tultuaan pitää voimassa. Tällainen kansallishenki on vuosituhansien työtä, koko ihmiskunnan sivistyksen tuote. (KT V, 38.)

Käsitteet ”kansallishenki”, ”siveellisyys” ja ”isänmaallisuus” ovat Snellmanille tärkeitä termejä, mutta niiden suhde on monimutkainen.¹⁷ Tulkitsen tässä niiden luonnehtivan hieman eri tavoin ihmisen kokonaisvaltaista pyrkimystä toteuttaa omaa moraalista ja tiedollista olemustaan siinä ympäristössä, jossa hän on kasvanut ja elää, sisältäen kielen, maan, tavat, lait ja laitokset. Edelleen:

Kaikki nämä yksilön sivistykseen vaikuttavat ja sen edellytyksenä olevat voimat ovat kuitenkin luonteensa mukaisesti itse riippuvaisia aikakauden

¹⁷ Kansallishenki on siveellistä sivistystä (KT V, 38), ”isänmaallisuus on kansallishengen perusta” (39), mutta ”isänmaallisuus, jonka pitäisi olla kansallishengen perustana, edellyttää jo kehittynyttä kansallishenkeä, kansakunnalle ominaista yleisinhimillisen sivistyksen muotoa.” (40). Toisaalta, ”[k]ansallishengeksi kutsumme puolestaan isänmaallisuutta ja kansallisuutta yhdessä” (40).

ja ihmiskunnan yleisen sivistyksen kehityksestä. Kulttuuri muuttaa nimittäin kaikkein materiaalisimpia voimiakin, maan ja ilmaston ominaisuuksia. (KT V, 40.)

Jokainen yksilö on siis sidottu aikaan ja paikkaan, mutta hänen pyrkimyksensä siveellisyyteen, jalostumiseen, myös vaikuttaa ympäröivään yhteiskuntaan ja antaa kansakunnalle herätteitä edistymiseen. Näin yhteiskunta kehittyy sisältä päin, mutta myös ulkoapäin, kosketuksista muihin kansoihin. Nämä kaksi voimaa muodostavat kehitysprosessin, joka ilmenee kansakunnan lainsäädännössä. (KT V, 42.) Yksilö, joka kasvaa osana tätä prosessia, voi ymmärtää oikean vain omantuntonsa äänen kautta.

On kuitenkin helppo käsittää, että jos kaikki mitä ihminen on pitävinään omantuntonsa äänenä, saisi käydä oikeasta, ei oikealle ja väärälle olisi mitään normia [...] Kokemus opettaa myös, että omatunnon käsky on erilainen eri ihmisillä, eri kansakunnilla ja eri aikoina, että omatunto sanalla sanoen on riippuvainen ihmisen omasta sivistyksestä. Tämä sisältää siis sen, että hyvän ja oikean normi on siinä kasvatuksessa järkeen ja siveellisyyteen, josta ihminen tulee osalliseksi kuulumalla valtioon; ja valtion oletetaan olevan oikean todellisuutta, niin että ihminen voi seurata omatunnon käskyä puhtain mielin vain silloin kun valtiossa ilmenevä siveellinen substanssi täyttää kokonaan hänen omantuntonsa, hänen moraalisen tietoisuutensa. (KT V, 45–46.)

Tulkitsen tätä niin, että Snellman katsoo meidän olevan kykeneviä tavoittamaan sen, mikä on oikein, *a priori*, mutta kykymme tehdä niin on sidottu niihin resursseihin, joita me ympäristömme saamme. Nykyajan filosofian näkökulmasta vaikuttaisi luontevalta sovittaa tämä ristiriita muistuttamalla, että vaikka jokin väite on *a priori* -totuus, kuten oikea Snellmanille viime kädessä on, sen ei tarvitse olla itsestään selvä tai helposti ymmärrettävissä.

Valtio-opin luvuissa 19–21 Snellman tarkastelee sitä, miten jotkin lait voivat olla meidän mielestämme epäoikeudenmukaisia. Hän toteaa: ”Koska lain todellisuuden täytyy viime kädessä olla riippuvainen subjektiivisesta ratkaisusta, siitä syystä laki voi olla oikeudenmukainen tai

epäoikeudenmukainen” (KT V, 92). Mielestäni on selvää, että tämä ei ole askel relativismin suuntaan, vaan vain sen myöntämistä, että laki on ihmisen erehtyväisen harkinnan tulosta. Yksilö voi olla tietämätön, häneltä voi puuttua yleiset edellytykset (johtuen kasvatuksesta ja ympäristöstä)¹⁸ harkita mitä oikea laki edellyttää, tai hän voi korvata oikean omalla mielihalullaan. Näin satunnaisuus ja mielivalta voivat korvata oikean: ”Siinä on siis paljon tehtävää ihmisen viisaudelle ja harkinnalle, jotta lainkäyttö saadaan vastaamaan käsitystä muuttumattomasta oikeudenmukaisuudesta, joka sisältyy lain käsitteeseen.” (KT V, 93.) Snellmanin käsittely vaikuttaa itse asiassa siltä, että hän hyväksyy fallibilismin moraalisen tiedon suhteen:

[...] laissa voi olla vain suhteellista epäoikeudenmukaisuutta, niin että sama laki ei julista oikeaa kaikiksi ajoiksi. Lain ylittävät oikeudenmukaisuudessa myöhemmän ajan lait, aivan kuten kyseinen ajankohta ylittää menneet ajat sivistyksessä. Kuitenkin laki oli omana aikanaan yhtä varma ilmaus oikealle kuin myöhemmän ajan laki on sille ajankohdalle. Kummatkin ovat korkeinta oikeutta. (KT V, 92.)

On syytä huomauttaa, että tämän sitaatin konteksti on Snellmanin luonnonoikeuden kritiikki. Tämä ei kuitenkaan poista sitä tosiseikkaa, että fallibilismin yhdistäminen hänen deontologiseen näkemykseen oikeasta tarjoaisi ratkaisun epäoikeudenmukaisten lakien ongelmaan. Toisaalta Snellman kuitenkin sitoo lain oikeudenmukaisuuden vallitseviin oloihin ja hänen voi tulkita viittaavan ainoastaan tähän fallibilismin sijaan.¹⁹ Luvussa 27

¹⁸ Tämä käsitys toistuu *Valtio-opin* eri kohdissa. Esimerkiksi: ”Onhan isänmaallisuus välittömästi tunne, jonka sisältö voi vaihdella, joten harhaan johdettu isänmaallisuus voi synnyttää hylättäviä tekoja” (KT V, 196).

¹⁹ Esimerkiksi keskustellessaan orjuudesta, jota Snellman pitää epäoikeudenmukaisena, hän toteaa: ”Vain sillä edellytyksellä, että kansakunnan omaksuma valtiomuoto ei suo alamaisille minkäänlaista henkilökohtaista vapautta ilman että orjuutta on – vain tällä edellytyksellä sanomme: orjuus on oikeudenmukainen. Mutta siellä missä laki kykenee turvaamaan kaikkien henkilökohtaisen vapauden, siellä orjuus on epäoikeudenmukainen ja epäjärkevä.” (KT V, 89.) Toisaalta, puhuessaan lapsen kehityksestä kohti moraalitotuusien tajuamista Snellman viittaa yksilön oikeaa koskevan tiedon ja itse oikean väliseen eroon (ks. KT V, 78). Tarkkaan ottaen tämä voidaan käsittää yksilön tiedon ja tiedon kohteen eroksi. Toisaalta, fallibilismin perusajatus on, että tiedoksi laskettavat käsitykset eivät välttämättä perustu konklusiivisille perusteille ja siten sallivat erehtymisen mahdollisuuden. Ajatellen Snellmanin muita tässä kirjoituksessa esitettyjä käsityksiä, hän tuntuu nimenomaan sallivan, että jokin yksilön moraalikäsitelmä voi olla *oikea* ja perustua järkeen, vaikka se myöhemmin osoitetaan väääräksi. Esimerkiksi tuomareiden

Snellman kritisoi kovin sanoin luonnonoikeuden käsitystä, että oikeat lait voitaisiin johtaa suoraan yleisistä periaatteista. Valtion tulee suojata jäsentensä vapautta suhteessa toisiinsa ja itseään suhteessa toisiin valtioihin, ja

[k]aikkea tätä ei saavuteta niillä keinoilla, jotka idiootti luulee saaneensa lahjaksi luonnolta, eikä edes niillä, jotka tutkija ja ajattelija esittävät teoreettisesti oikeina. Keinot saadaan niistä viisaista laskelmista, joita tehdään olosuhteiden kunakin hetkenä asettamista vaatimuksista. Jos filosofi hallitsee oman aikansa sivistyksen, hän voi puolustaa teorian kannalta yhteiskuntajärjestyksen yleisiä periaatteita johtamalla yhden instituution toisesta ja ne kaikki valtion olemuksesta, toisin sanoen esittämällä valtion järjestelmänä. Hän voi ehkä myös osoittaa, mikä on periaatteetonta kokonaisuudesta liian räikeästi erottuvissa instituutioissa. Sen sijaan hän ei voi esittää erityiskysymyksissä oikeita sääntöjä pelkästään teoreettisin[a] perusteina, ja vielä vähemmän antaa sellaisia tietylle valtiolle tietynä ajankohtana. [...] Luonnonoikeus tarkastelee ihmistä pelkästään abstraktisena persoonana; ihminen on kuitenkin jotakin muuta sen lisäksi, nimittäin siveellinen persoona. (KT V, 113.)

Vaikka myös tämä kohta on myös osa luonnonoikeuden kritiikkiä²⁰, on kuitenkin merkillepantavaa, että Snellman asettaa yleiset periaatteet ja siveellisen persoonan harkinnan vastakkain. Tämä asettaa etemme mielenkiintoisen kysymyksen: jos yleiset periaatteet, jotka sinänsä voivat olla tosia, eivät määritä millaisia lakien tulisi olla, ja kaikki päätökset ovat viime kädessä subjektiivisen ja erehtyväisen harkinnan varassa, miten me voimme luottaa säätävämme oikeita lakeja? Snellmanin vastaus on dialektinen kriteeri: ”[...] ihmisen järjellisyys koetinkivenä on edelleen se, että hän ajattelee ja toimii yhtäpitävästi muiden ihmisten ja ylipäätään maailmanjärjestyksen kanssa.” (KT V, 84).

Toisaalla hän toteaa:

oikeudenkäytöstä hän toteaa: ”[...] eikä milloinkaan voi katsoa, että ehdoton oikeudenmukaisuus on tapahtunut.” (KT V, 98.)

²⁰ Snellmanin kritiikkiä luonnonoikeutta kohtaan ei tule lukea yksioikoisesti. Hän toteaa myös, että ”luonnonoikeuden ja moraalin pätevyyttä ei siis kielletä valtiossa; valtio päinvastoin toteuttaa niiden käskyt.” (KT V, 46.)

Jos näin ollen jonkin kansakunnan laissa tai tavassa jokin on ristiriidassa luonnonoikeuden määräysten ja moraalien käskyjen kanssa, on kansakunnan jokaisen jäsenen velvollisuutena ja oikeutena yrittää korjata asia. Tosin korjaamisenkin on tapahduttava lainmukaisessa muodossa, koska sen totuutta ja oikeudenmukaisuutta ei voi mitenkään muuten koetella. (KT V, 47.)

Myös uskonnon esittämien moraalisaäntöjen on alistuttava samaan:

[...] toisaalta historiallisesti annetun uskonnon sisältämät käskyt puolestaan voivat olla ristiriidassa valtion kohtaloita ohjaavan taipumattoman pakon kanssa (esimerkiksi kristinuskon käsky yleiseen ihmisrakkauteen on ristiriidassa vihollisten tappamispakon kanssa), niin sellaiset uskonnon käskyt on alistettava kansalaisyhteiskunnassa vähintäänkin usean tulkinnan ja selvityksen alaiseksi. (KT V, 69.)²¹

Tämä dialektinen oikeuttaminen, koettelu, on Snellmanille tärkeä kriteeri, ja hän viittaa siihen usein. Koulutuksen sisällöstä keskustellessaan hän huomauttaa, että ”ennen kaikkea [nykyinen aikakausi] vaatii järjelle koettelu oikeutta” (KT V, 155). Snellman ei kannata rajoittamatonta sananvapautta²² mutta ennakkosensuurista keskustellessaan hän toteaa:

Jos hallitus yrittää näet yrittää sensuurilla välttää sananvapauden väärinkäyttöä, se asettaa kaikesta huolimatta yhteiskunnan alttiiksi kaikille niille vaaroille, joita [ennakkosensuurin] väärinkäyttö voisi aiheuttaa. Samalla yhteiskunnalle ei tarjota yhtään niistä eduista, joita järvevä julkinen keskustelu yleisistä asioista tarjoaa. Valppainkaan sensuuri ei kykene estämään ulkomaisen lehdistön kirjoittelua. Kaikki niiden sisältämä virheellinen ja turmiollinen aines jää vaille oikaisua ja kumoamista, koska sille ei anneta tilaisuutta [...]. (KT V, 119.)

²¹ Toisaalla Snellman toteaa: ”Itse asiassa tapahtuu niin, että ihminen arvioi uskonnon moraalikäskyjen erinomaisuutta yhteiskuntajärjestyksestä saamansa oikean ja väärän käsitteiden mukaan” (KT V, 182).

²² Snellmanin mukaan ”[...] kukaan järvevä ihminen [ei] voi missään tapauksessa vaatia rajoittamatonta painovapautta.” (KT V, 119.)

Lienee paikallaan koota yhteen tähänastinen keskustelu. Oikea on Snellmanin mielestä todellinen tietämisen kohde, jonka ihminen voi tavoittaa huolellisella harkinnalla. Teko, joka perustuu yksilön parhaaseen järjelliseen harkintaan, omantunnon ääneen ja joka tehdään velvollisuudesta tehdä oikein, on vapaa ja toteuttaa ihmisen rationaalista ja moraalista olemusta. Tätä oikeaa ei ole kuitenkaan helppo saavuttaa: se mitä me teemme oikein, teemme oikein vain sen pitkällisen sivistysprosessin ansiosta, mikä kaikkien kansakuntien tulee läpikäydä. Ympäristönsä avustamana²³ yksilön tulee sivistää itseään ja myös tehdä parhaansa vaikuttaakseen ympäristöönsä, jotta se edelleen sivistyisi ja kehittyisi. Tämä kyky nähdä oikea ei ole erehtymätön, ja sitä miten oikea kulloinkin voidaan toteuttaa, ei voida johtaa suoraan moraalista periaatteista: jokaisen ajan on se päätettävä itse omista lähtökohdistaan. Sellaista filosofiaa, joka osaisi joistain yleisistä periaatteista lähtien kertoa kaikille ajoille ja oloille, miten on toimittava, ei ole olemassa; ei ole tahoja, joka ratkaisisi meidän puolestamme mitä meidän tulee tehdä.²⁴ Ainoa tapa löytää oikea on oma huolellinen harkintamme ja saavuttamiemme käsitysten kriittinen tarkastelu toisten kansalaisten toimesta, jotka myös pyrkivät löytämään oikean oman harkintansa perusteella.

Unohtaen oman edun ajamisen vielä hetkeksi, voidaan todeta, että Snellmanin antama kuva sopii hyvin motivoimaan kansalaisten deliberaation: valtiollisista asioista päätettäessä kansalaisten tulee asettaa oma etunsa syrjään ja keskittyä yhteisen tahdon muodostukseen oman parhaan harkintansa pohjalta. On syytä kuitenkin vielä mainita kaksi seikkaa. Ensinnäkin, Snellman perustaa näkemyksensä nimenomaan valistuksen ”järkeen”, joka Mansbridgen mukaan on koettu liian vaativaksi ja kaukaiseksi. Sen sijaan nykyaikaiset kirjoittajat suosivat vastavuoroista oikeuttamista, oikeuttamista

²³ Tämä prosessi alkaa perheessä ja jatkuu yhteiskunnassa. Snellman käsittelee tätä *Valtio-opin* toisessa osassa, luvut 7–14.

²⁴ Yhteiskuntasopimuksen luonnetta käsitellessään Snellman toteaa: ”Sopimuksen säädökset ovat aina tulkittavissa ja ne vaativat tulkintaa pantaessa sopimusta täytäntöön. Sopimuksen voimassa pysymisen kannalta on tästä syystä ylipäättään tarpeen, että on olemassa kolmas mies, jolla on oikeus ratkaista syntyvät erimielisyydet kahden sopimuspuolen välillä. Yhteiskuntasopimuksesta puuttuu tällainen kolmas ratkaiseva ääni yksityisen ja yleisen tahdon väliltä. – Vain yleinen tahto voi tässäkin olla tuomarina omassa asiassaan ja ratkaista kuinka sopimus on tulkittava ja toteuttava. (KT V, 35.) Edelleen: ”[...] meitä ei auta, jos sanomme: maailman ja sen tapahtumat on järjestänyt ihmisjärjen yläpuolella oleva järki. Maailmanhenki näet tekee tekonsa maailmanhistoriassa meidän päätöstemme, meidän vapaiden toimiemme välityksellä.” (KT V, 46.)

sellaisista perusteista, jotka kanssakeskustelijat voivat hyväksyä. On totta, että termit kuten ”oikea”, ”totuus” ja ”yhteinen etu” ovat ylikäytettyjä poliittisessa keskustelussa ja niitä käytetään retorisesti, vaikutuksen aikaansaamiseksi. Mitään *rationaalista* vaikutusta ei kuitenkaan saada aikaan sillä, että todetaan ”järki sanoo, että p”, ”politiikka r on oikea” tai ”totuus on, että s”: meidän on perusteltava miten järki sanoo, että p, miksi r on oikea ja mistä tiedämme, että s. Snellman kuitenkin hyväksyy, että vaikka oikea on todellinen tietämisen kohde, voimme erehtyä siitä mitä järki tosiasiallisesti sanoo, ja siksi jokaista väitettä on koiteltava. Tämä koittelu ei voi tapahtua muutoin kuin pyytämällä keskustelijoita hyväksymään omat perusteensa. Tämän hyväksymisen puolestaan tulee tapahtua rationaalisesti, sen perusteella mitä muut voivat hyväksyä oman järkensä, kokemuksiensa ja omantuntonsa valossa. Tämä oikeuttaminen siis *on* vastavuoroista. Vaikka kokisimme termin ”järki” sisältävän väärää konnotaatioita, emme voi kokonaan hylätä sitä, mihin termi viittaa; myös moniarvoinen ja vastavuoroinen perustelu tarvitsee järkeä. Vastavuoroisen perustelun vaatimus tulee ilmi *Valtio-opissa* monessa paikassa, mutta haluan nostaa vielä yhden pitkän sitaatin, joka samalla tuo esille sen seikan, ettei Snellmania kuitenkaan voi pitää laajan demokraattisen deliberaation kannattajana:

Muistutan vielä kerran, että filosofian opit eivät sellaisenaan voi olla yleisön tuomittavissa, vaan filosofian on otettava tunnuslauseekseen Horatiuksen [...] [en piittaa asiaan perehtymättömästä rahvaasta, vaan torjun sen]. [...] Kun siis filosofia esittää periaatteen valtion olemassaolosta, periaatteen on puhuteltava aikalaisia ja saatava aikaan parannuksia valtion instituutioissa, jotka vastaavat ajankohtaa. Enemmistö ei voi ryhtyä selvittämään periaatteen totuutta, ei voi todistaa sitä; mutta ajan tarpeiden ja sivistyksen on välittömästi tunnustettava periaate. Muutoin se on epäkypsä ja sitä on kehiteltävä uusin ponnistuksin, on rakennettava uusia järjestelmiä kunnes äkkiä osutaan oikeaan kohtaan, kunnes aikalaiset hyväksyvät katkeralla vaivannäöllä aikaan saadun totuuden omaksi tietämykseksi ja enemmistö uskoo itse ratkaisseensa helpon ongelman. (KT V, 212; alaviite 1.)

Snellman ei katso oman aikansa kansalaisten olevan kypsiä käsittelemään kaikkea sitä mitä oikean toteuttaminen vaatii. Hän katsoo, että sivistyneistön on johdettava kansaa.²⁵ (Mutta kuten hän itse sanoo, vaikka hyväksyisimme sen olleen totta yhdelle aikakaudelle, se ei välttämättä ole sitä toiselle.) Dialektinen vaatimus on kuitenkin ehdoton: omasta mielestään oikean periaatteen löytäneen on ponnisteltava kunnes kansalaiset, kuulijat, hyväksyvät sen. Sitä ennen periaate ei ole kelvallinen peruste politiikkaa valittaessa.

Toiseksi, moderni deliberaatio haluaa tunnustaa (ja juhlistaa) vapaiden ja tasa-arvoisten ihmisten erilaisuutta (Mansbridge 2010, 69). Jotkut voivat kokea Snellmanin kannattaman valistuksen ”järjen” hankaloittavan tätä tavoitetta, koska sen voidaan tulkita pakottavan kaikki samaan, jonkun toisen määräämään oikeaan. Snellmanille sivistyksen tarkoitus on kuitenkin mahdollistaa harkinta siitä, mikä on oikeaa ja totta, ja sitä kautta tuottaa yhä edistyneempää sivistystä.²⁶ Ei ole mitenkään selvää, että tämän prosessin tulos olisi yhden vakaumuksen pakkopaita.²⁷ Uskonnollisen tietämyksen roolia pohtiessaan Snellman toteaa: ”Koska sivistys on mielenkiintoa totuutta kohtaan ja kykyä etsiä vakaumusta, on mielipiteiden sivistyksen lisääntyessä väistämättä lisääntyttävä ja tultava yhä vaihtelevammiksi.” (KT V, 180–181.)

3.3. Oma etu ja valtion etu

Palatkaamme nyt oman edun rooliin ja alueeseen valtiossa. Miksi Snellman vaikuttaa rajaavan oman edun kokonaan valtiollisen toiminnan ulkopuolelle? Voimme mielestämme osoittaa, ettei Snellman tahdo sanoa, ettei oman edun ajaminen voisi oikeutetusti kuulua myös valtiollisen keskustelun ja päätöksenteon piiriin. Ensinnäkin:

²⁵ Snellman pitää laajaa deliberaatiota myös käytännössä mahdottomana toteuttaa: ”Kukaan ei vaatine, että 20–30 miljoonan suuruisten ihmisjoukkojen pitäisi välittömästi säätää lakeja. Näin ollen voimme olettaa, että edustajien valinta on ainoa tapa, jolla ison valtion kansa voi ottaa osaa lainsäädäntöön.” (KT V, 249.) Mutta tämä käytännöllinen argumentti ei pakota meitä hylkäämään demokraattista deliberaatiota päätöksentekomenetelmänä, vain antamaan selvityksen sen muodoista ja järjestämisestä.

²⁶ ”Tälle sivistyksen etenemiselle kansallisuutteen kustannuksella ei ole mitään rajaa” (KT V, 49).

²⁷ Sekin on tietysti mahdollista. Tämän kysymyksen tarkempi tarkastelu edellyttäisi keskustelua vallan käsitteestä ja sen roolista deliberaatiossa. Mansbridgen (2010, luku IV) mukaan valta sinänsä ei ole deliberaatiolle välttämättä haitallista, mutta pakottavan vallan (*coercive power*) poissaolo on olennaista. Emme voi mennä tähän sen syvemmälle tässä.

Jokainen yksilö vaatii, että hän saa tyydyttää nämä kolme tarvetta: sivistys, fyysinen toimeentulo ja mahdollisuus nauttia rauhassa siitä onnesta, jonka hän on pystynyt saavuttamaan. Yhteiskunnan on myös luotava mahdollisuudet niiden tyydyttämiseen. Yhteiskunta puolestaan vaatii yksilöltä työtä ja lainkuuliaisuutta, joilla täytetään myös yhteiskunnan tavoite. Yksilö taas johdatetaan näihin kumpaankin perheessä annetulla kasvatuksella ja yhteiskunta edellyttää siis omalta puoleltaan, että perhe on olemassa ja sen tarpeet on tyydytetty. Niin ollen kumpikaan instituutio ei ole pelkkä väline, vaan ne edellyttävät toinen toisiaan. Kuten näemme, se sama toiminta, jolla yksilö toteuttaa toisen instituution tavoitteen, toteuttaa toisenkin instituution tavoitteen. (KT V, 132.)

Yksilöllä on tietyt perustarpeet, ja koska yksilö on moraalinen ja rationaalinen olento, hänellä on tietyt oikeudet²⁸, joita valtion tulee suojella. Valtion ja yksilön tarpeet ovat kuitenkin usein yhtenevät:

Kansalaisyhteiskunnassa on nimittäin kyse siitä, ettei yksikään yksilö saa tehdä vastoin lakia, joka on yleisen tahdon ilmaus, samalla kun yksilö voi muutoin käyttää voimiaan halunsa mukaan. (KT V, 77.)

Lait tulee suunnitella niin, että kansalainen voi ajaa omaa etuaan parhaansa mukaan. Se, mitä tämä täsmälleen ottaen sallii, on kirjattava lakiin. Laki pyrkii turvaamaan jokaisen oikeudet, ja oman edun ajamisen rajat kirjataan lakiin, sillä yksilö ei tässä toiminnassa saa loukata muiden vastaavaa etua. Yksilö ei saa toimissaan olla esteeksi yleiselle edulle (KT V, 106). Yhteiskunta ei kuitenkaan pysty täyttämään tehtäväänsä, ellei se voi puuttua ja rajoittaa yksilön toimintaa:

On myös selvää, että yhteiskunta ei voi täyttää yksilön vaatimusta aineellisesta toimeentulosta [...] ellei laki määrää miten ja millä tavalla

²⁸ ”Henkilön ja omaisuuden vapaus on yksilön absoluuttinen oikeus suhteessa valtion muihin jäseniin, eikä muutoin voisi ajatellakaan mitään siveellistä vapautta valtiossa” (KT V, 60). Edelleen: ”Yhtäältä ihmisen moraalinen vapaus kuitenkin vaatii, että hän voi päättää vain omantuntonsa mukaan asioissa, jotka eivät koske toisen yksilön tai yhteisön oikeutta.” (KT V, 160.)

kukin yksilö voi hankkia osallisuuden olemassa olevaan kansallisvarallisuuteen, ja miten sitä on säilytettävä suhteessa muihin valtioihin ja lisättävä, jotta kasvavan väestön, lisääntyvän sivistyksen kasvattamat tarpeet tulevat tyydytetyiksi. (KT V, 137–138.)

Tämä näkemys toistuu *Valtio-opin* eri puolilla. Esimerkiksi elinkeinovapaus on tärkeä, ja Snellman pitää myös tärkeänä, että saman alan edustajat voivat liittyä yhteen ja luoda korporaatioita ajaakseen omaa etuaan (KT V, luku 35). Tätä elinkeinovapautta on rajoitettava niin vähän kuin mahdollista (KT V, 145) mutta silloin kun se on tarpeen, valtion oikeus tehdä niin on ehdoton (KT V, 144). Snellman myös katsoo, että tällaisen kansalaisyhteiskunnan luonne sisältyy siveellisyyden luonteeseen (KT V, 78). Toisin sanoen yksilön oikeus ajaa omaa etuaan seuraa järjen periaatteista kuten valtiokin.

Nyt olemme valmiita käsittämään mitä Snellman tarkoitti kirjoittamalla, että

yksilö on [kansalaisyhteiskunnan] jäsen sikäli kuin hän tavoittelee yksityisiä päämääriään, omaa etuaan loukkaamatta yhteiskunnan muiden jäsenten vapautta. Sen sijaan valtion jäsenenä hänen toimintansa kohteena on yleinen, itse valtion etu [...]. (KT V, 76–77.)

Snellman ei tarkoita, ettei kansalaisen oma etu kuuluisi valtiollisen päätöksenteon piiriin perusteena jollekin politiikalle, vaan kyse on toiminnan kohteesta. Jos toiminnan kohteena on lain säätö, valtiollinen päätös, joka koskettaa meitä kaikkia, on toiminnan ajateltava valtion etua. Päivittäisen toiminnan kohde, oma toimeentulo, on lain noudattamisen alue. Mutta sikäli kuin lait käsittelevät yksilöiden mahdollisuuksia tulla toimeen, yksilöiden etu voi olla myös valtion etu ja silloin yksilön etu on relevanttia päätöksen suhteen. Tätä ei pidä kuitenkaan tulkita niin, että väittäisin Snellmanin ajatelleen naivistisesti, ettei yksilön ja yhteisön etu voisi olla ristiriidassa, tai että väittäisin Snellmanin ajatelleen yhteisen edun olevan pelkästään yksilön etujen aggregaatti – kanta, jonka Mansbridge myöntää olevan kiistanalainen. Ensinnäkin, keskustellessaan politiikan ja moraalin suhteesta (KT V, luku 66), Snellman toteaa:

Nyt on niin, että moraali ei voi sisältää mitään erityisiä käskyjä kansalaisyhteiskuntaa tai valtiota koskevista velvollisuuksista, vaan kaikki yhteiskunnalliset velvollisuudet kuuluvat yleiseen kategoriaan ”velvollisuudet lähimmäistä kohtaan”. (KT V, 287.)

Toiseksi, Snellman viittaa useassa kohdassa intressiristiriitoihin ja toteaa, ettei yhden yksilön etu saa loukata toisen etua. Lisäksi hän sanoo eri yhteyksissä, että lait ovat useana aikana olleet sellaisia, että ne ovat hyödyttäneet erityisesti niitä, joilla on valtaa ja omaisuutta. Tämä jonkun osapuolen edun korostuminen laeissa näkyy niin perheen, kansalaisyhteiskunnan, valtion kuin valtioiden välisissä suhteissa. Esimerkiksi keskustellessaan aviottomien lasten oikeuksista ja aviorikoksesta Snellman toteaa:

Sanon epäröimättä, että tässä tapauksessa, samoin kuin perheomaisuuden hallinnassakin, miehet ovat säätäneet lakeja sellaisen oikeudenmukaisuuden perusteella, jota voimakkaammat tavallisesti osoittavat heikommille [...]. (KT V, 64.)

Kuten Mansbridge, Snellman sallii oman edun ajamisen valtiollisessa kontekstissa mutta rajaa pois pakottavan vallan käytön.

4. Lopuksi

Kirjoitukseni tarkoituksena on ollut osoittaa, miten snellmanilaisesta lähtökohdasta käsin voidaan osoittaa yksilön oman edun, tietyn ehdoin, kuuluvan yhteiskunnallisen päätöksenteon piiriin yhtenä perusteena oikeaa ratkaisua haettaessa. Tämä herättää vaikean kysymyksen sen painavuudesta päätöksenteossa. Tarkastelumme perusteella Snellmanin vastaus vaikuttaisi olevan: yksilön etu on otettava huomioon siinä määrin kuin parhaan harkintamme mukaan kussakin tapauksessa katsomme oikeaksi. Yleistä vastausta ei ole, ”säädösten oikeudenmukaisuus voidaan johtaa vain olevista oloista” (KT V, 114). Mutta mikä tahansa oikea vastaus kulloinkin on, sen on kestettävä aikalaistensa kriittinen keskustelu.

Kirjallisuus

- Bohman, James & William Rehg (toim.) (1999). *Deliberative Democracy: Essays on Reason and Politics*. Cambridge, MA: MIT Press.
- Cohen, Joshua (1999 [1989]). ”Deliberation and Democratic Legitimacy”, teoksessa James Bohman & William Rehg (toim.) *Deliberative Democracy: Essays on Reason and Politics*. Cambridge, MA: MIT Press, 67–92.
- Dahl, Robert A. (2006 [1979]) ”Procedural Democracy,” teoksessa Robert E. Goodin & Philip Pettit (toim.), *Contemporary Political Philosophy: An Anthology*. Malden, MA: Blackwell, 107–125.
- Follet, Mary Parker (1942 [1925]). ”Constructive Conflict”, teoksessa H. C. Metcalf & L. Urwick (toim.), *Dynamic Administration: The Collected Papers of Mary Parker Follett*. New York: Harper.
- Kant, Immanuel (1998 [1781/1787]). *Critique of Pure Reason*. Toim. P. Guyer & A. Wood. New York: Cambridge University Press. [A/B]
- Mansbridge, Jane, sekä James Bohman, Simone Chambers, David Estlund, Andreas Føllesdal, Archeon Fung, Cristina Lafont, Bernard Manin & José Luis Martí (2010). ”The Place of Self-Interest and the Role of Power in Deliberative Democracy”, *Journal of Political Philosophy*, 18 (1), 64–100.
- Rawls, John (1993). *Political Liberalism*. New York: Columbia University Press.
- Snellman, Johan Vilhelm (2006 [1842]). *Valtio-oppi. J. V. Snellmanin Kootut Teokset, osa V*. URL = http://www.snellman200.fi/kootut_teokset/fi.jsp
- Ritola, Juho (2015). ”Deliberative Democracy, the Deliberating Agent, and Critical Thinking: An ideal picture and some empirical challenges”, *Studier i Pædagogisk Filosofi*, 4 (1), 29–54.
- Tenkku, Jussi (1984). ”Snellmanin etiikasta”, teoksessa Ilkka Patoluoto (toim.), *J. V. Snellmanin filosofia ja sen hegeliläinen tausta*. Suomen Filosofisen Yhdistyksen järjestämä kansainvälinen kollokvio, Helsinki 4.–5.12.1981. Helsinki: Helsingin yliopiston filosofian laitos.

J. V. Snellman ja Lorenz Lindelöf äänestysteoreetikoina

Eerik Lagerspetz

1. Aluksi

Vuonna 1948 skottilainen taloustieteilijä Duncan Black julkaisi artikkelinsa ”On the Rationale of Group Decision-Making”. Kolme vuotta myöhemmin ilmestyi tulevan Nobel-taloustieteilijän Kenneth J. Arrowin pieni kirja *Social Choice and Individual Values*. Nämä tutkimukset asettivat peruskivet uudelle tutkimusalueelle, sosiaalisen valinnan teorialle. Vaikka sosiaalisen valinnan teoria miellettiin aluksi (ja mielletään osin edelleenkin) taloustieteen osa-alueeksi, sillä on yhteyksiä lukuisiin muihin tieteenaloihin. Kaikkein abstrakteimmalla tasolla sosiaalisen valinnan teorian tutkimusaluetta voidaan luonnehtia seuraavasti: On olemassa joukko yksittäisiä elementtejä ja niiden välisiä mahdollisia järjestysrelaatioita. Sosiaalisen valinnan tutkimusaiheena ovat erilaiset tavat joilla useista samojen elementtien välisistä järjestysrelaatioista voidaan muodostaa uusi järjestysrelaatio. Esimerkiksi erilaisten urheilu- ja taitokilpailujen voittajien määrääminen suoritusten perusteella, opiskelijoiden valinta pääsykoesuoritusten perusteella, tieteellisten aikakauslehtien tai yliopistojen asettaminen paremmuusjärjestykseen ja monia ulottuvuuksia sisältävät päätöksentekoongelmat voidaan hahmottaa sosiaalisen valinnan potentiaalisiksi tutkimuskohteiksi.

Politiikan tutkimuksessa sosiaalisen valinnan teorian keskeinen sovellutus on erilaisten *äänestysmenettelyjen* tarkastelu. Äänestysmenettelyjen teoria on sosiaalisen valinnan teorian vanhin ja tunnetuin sovellutus. Äänestysmenettelyjen teorian tunnetuin osa-alue on puolestaan erilaisten *äänestysparadoksien* tutkimus. Tämä tutkimus on vanhempaa perua. Sen panivat alulle ranskalaisten Jean-Charles de Bordan ja Marie Jean Antoine Nicolas de Caritat’n (eli Condorcet’n markiisin) 1700-luvulla tekemät havainnot (Borda 1995 [1770]; Condorcet 1995 [1785]).

Kirjassaan *The Theory of Committees and Elections* (1958, 156–213) Duncan Black tarkasteli myös äänestysmenettelyjen tutkimuksen varhaishistoriaa. Vaikka

Blackin esityksen oli määrä olla ensimmäinen pikemminkin kuin viimeinen sana aiheestaan, siinä rakennettu yleiskuva on säilynyt melko muuttumattomana myöhemmässä kirjallisuudessa. Iain McLean – joka yhtenä harvoista on Blackin jälkeen ollut vakavasti kiinnostunut äänestysmenettelyjen teorian historiasta – tiivisti tämän yleiskuvan seuraavasti (McLean 1990, 99):

The theory of voting is known to have been discovered three times and lost twice. The work of Condorcet, Borda, and Laplace was entirely ignored from about 1820 until 1952, with the sole exception of E. J. Nanson's paper "Methods of Election", which was read to the Royal Society of Victoria in 1882, published in a British Government Blue Book of 1907, and languished there undiscussed until 1958.

C. L. Dodgson ("Lewis Carroll") discussed Condorcet and Borda methods, and procedures for breaking cycles, in three pamphlets printed in the 1870s; he worked in ignorance of his predecessors, and again was not understood until 1958.

Tässä kirjoituksessa tarkastellaan 1800-luvulla käytyä kahden suomalaisen yliopistomiehen välille puhjennutta kiistaa äänestysmenettelyistä. Kiista on kahdessakin mielessä kiinnostava. Yhtäältä se haastaa kirjallisuudessa lähes yksimielisesti hyväksytyyn käsitykseen äänestysteorian (ja laajemmin sosiaalisen valinnan teorian) varhaishistoriasta. Tulemme näkemään, että vastoin McLeansitaatissa esitettyä käsitystä, Condorcet'n, Bordan ja Laplace'in klassiset tulokset eivät suinkaan vaipuneet täydelliseen unohdukseen 1800-luvulla, eivät ainakaan Suomessa. Toisaalta kiista on lähemmän tarkastelun arvoinen toisestakin syystä. Kiistelevien osapuolten välinen erimielisyys koski koko sosiaalisen valinnan teorian tulkintaa ja teorian relevanssia. Vielä yleisemmin voidaan sanoa, että kysymys oli formaalisten teorioiden merkityksestä ja hyödyllisyydestä yhteiskunnallisten ilmiöiden tarkastelussa. Nämä kysymykset eivät ole menettäneet ajankohtaisuuttaan.

2. Virkaesitykset ja äänestysmenettelyt

Tässä esityksessä tarkastellun kiistan osapuolet olivat molemmat Keisarillisen Aleksanterin yliopiston professoreita ja autonomisen Suomen merkkihenkilöitä. Lorenz Lindelöf (1827–1908) toimi matematiikan professorina. Pariisissa

opiskellut ja Preussin tiedeakatemia palkitsema Lindelöf oli 1800-luvun suomalaisen tiede-elämän kansainvälisimpiä hahmoja. Myöhemmin hänestä tuli yliopiston rehtori ja valtiopäivämies. (Itse asiassa hän tuli eri aikoina valituksi valtiopäiville kolmesta eri säädystä: professorina hänet valittiin pappissäädystä, Helsingin kaupungin edustajana hän kuului porvarissäätyyn, ja aateloituna lopulta aatelissäätyyn.) Lindelöfin kiistakumppani oli kansallisesti vielä suurempi kuuluisuus; itse Johan Vilhelm Snellman (1806–1881), filosofian professori, tuleva senaattori ja armoitettu poleemikko.

Herrojen välit eivät alun perin olleet huonot. Snellman oli toimittamassaan *Litteraturbladissa* esitellyt myönteisesti – ja Lindelöfin elämäkerran kirjoittajan Olli Lehdon mukaan kohtuullisen asiantuntevasti – Lindelöfin ranskankielisen teoksen *Leçons de calcul des variations* (1861). Konsistorissa Snellman oli tukenut Lindelöfin valintaa matematiikan professorin virkaan (Lehto 2008, 59). Snellmanin ja Lindelöfin välinen kiista näyttää saaneen alkusysäyksen yliopiston konsistorissa keväällä 1862 hyväksytystä virkaesityksestä joka koski yliopiston kirjastonhoitajan viran täyttämistä. Tässä asiassa Snellman ja Lindelöf päätyivät eri puolille. Nuorta professori Lindelöfiä ei ärsyttänyt niinkään esityksen sisältö kuin tapa jolla asiasta äänestettiin. Hän julkaisi käytettyä vaalitapaa arvostelevan kirjoituksen *Juridisk Albumissa* (Lindelöf 1862a). Snellman vastasi toimittamassaan *Litteraturbladissa* (Snellman 1862a). Lindelöf jatkoi keskustelua *Juridisk Albumissa* ja sen jälkeen vielä *Litteraturbladiin* lähettämässään vastineessa (Lindelöf 1862b; Lindelöf 1862c); Snellman vastasi kumpaankin (Snellman 1862b; Snellman 1862c). Keskustelun edetessä herrojen käyttämät äänensävyt muuttuivat yhä ärtyneemmiksi. Heidän välinsä eivät kuitenkaan pysyvästi katkenneet. Kun vuonna 1864 asialistalle tuli metrijärjestelmän käyttöönotto, senaattori Snellman huolehti siitä, että käytännön uudistuksista kiinnostunut matemaatikko Lindelöf valittiin asiaa valmistelemaan komiteaan (Lehto 2008, 63).

Keskustelun ymmärtämiseksi on syytä tutustua hieman virkaesityksissä käytettyjen menettelyjen taustaan. Ruotsin valtakunnassa eri intressiryhmät kävivät 1600-luvulla ja 1700-luvun alussa virkamiesten nimitysvallasta pitkällistä kamppailua, joka vapauden ajan alkupuolella päättyi sarjaan kompromisseja. Ne noudattivat samaa perusmallia: itsehallinnollisten korporatioiden kuten kaupunkien ja pitäjien äänivaltaisille jäsenille taattiin oikeus asettaa vapaissa vaaleissa tarjolla olevista ehdokkaista kolme parhaina pitämäänsä

virkaesityksessä *paremmuusjärjestykseen*. Hallitsija tai hänen edustajansa – vapauden ajalla käytännössä valtaneuvosto, paikallisella tasolla maaherra – sitten valitsi näistä kolmesta parhaaksi katsomansa. Hallitsijan aikanaan lähes rajoittamatonta nimitysvaltaa rajattiin huomattavasti, koska hänen oli tehtävä valintansa nimitysesitykseen sisällytettyjen kolmen ehdokkaan joukosta. Kyseessä oli siis eräänlainen ”sinä leikkaat kakun, minä valitsen palan” -periaatteen sovellutus. Tämä yleisperiaate kirjattiin lakeihin. Vuoden 1720 hallitusmuodon § 40 edellytti, että kuningas sai valita kolmen ehdokkaan välillä. Hallitsijan vuonna 1723 antama vastaus kaupunkien yleisiin valituksiin täsmensi käytäntöä kaupunkien hallinnon osalta. Pormestarin, raatimiehen ja kaupunginsihteerin tehtäviä täyttäessä oli järjestettävä vaali, jonka tuloksen perusteella ehdolle asetettiin kolme hakijaa. Näistä hallitsija, tai hänen sijaisenaan maaherra, nimitti yhden. Vuoden 1731 valtiopäivillä tehdyn päätöksen mukaan maaseudun papinvaaleissa siirryttiin samantyyppiseen käytäntöön. Konsistorien (tuomiokapitulien) oli asetettava avoimiin papinvirkoihin kolme ehdokasta, joista seurakuntalaiset valitsivat; kuninkaalliseen määräysvaltaan kuluviissa seurakunnissa seurakuntalaisilla oli oikeus anoa lisäksi neljäs ehdokas, ja kolmen eniten ääniä saaneen lista oli toimitettava kuninkaalle nimitystä varten. Samantyyppistä menettelyä ryhdyttiin soveltamaan myös piispanvaaleissa ja yliopistojen virkaesityksiä laadittaessa.¹

Vaikka hallitsija pääsääntöisesti nimitti ensimmäiselle sijalle asetetut ehdokkaat virkoihin, hän käytti myös harkintavaltaansa. Esimerkiksi 20 Suomen puoleisen kaupungin pormestarin vaaleja tutkinut Mäntylä (1981, 88) on laskenut, että vuosina 1720–1808 nimitetyistä pormestareista 69,1 prosenttia nimitettiin ensimmäiseltä ja 18,2 prosenttia toiselta ehdokassijalta. Alemmalta ehdokassijalta nimittäminen oli helpointa niissä tapauksissa, joissa ehdokassijoista päättäneet enemmistöt olivat kapeita tai joissa menettelyyn liittyi jotakin moitittavaa. Paljon riitaa aiheuttaneessa Turun pormestarinvaalissa vuonna 1745 maaherra pudotti eniten ääniä saaneen raatimies Anders Roosin kolmannelle tilalle. Hallitsija nimitti silti Roosin, koska tämän äänen enemmistö oli niin suuri (Mäntylä 1981, 50).

Äänestysmenettelyjen suhteen päätökset jättivät paljon liikkumavaraa. Virallinen päätöksentekomenettely ennen 1700-lukua oli useimmissa yhteyksissä

¹ Äänestyksiä koskevasta lainsäädännöstä ja äänestyskäytännöistä Ruotsin valtakunnassa ks. Thomson 1951, Matinoli 1954, Mäntylä 1977 ja Mäntylä 1981.

yksimielisyyperiaate. Jos yksimielisyyttä ei saavutettu, ylempi viranomainen saattoi menetellä oman mielensä mukaan. Vapauden ajan alku merkitsi enemmistöperiaatteen läpimurtoa. Vuoden 1719 valtiopäiväjärjestyksessä määrättiin, että valtakunnan säädyt tekivät päätöksensä enemmistöperiaatteen mukaisesti, ja § 47 määräsi, että talonpoikais- ja porvarissäädyn valtiopäivämiehet oli niin ikään valittava enemmistövaalitapaa käyttäen. Virkaesityksissä asetelma oli kuitenkin mutkikkaampi, koska ehdokkaat oli asetettava paremmuusjärjestykseen. Miten tämä järjestys tuotettiin? Vuonna 1746 annettu kuninkaallinen käskykirje pyrki selkeyttämään käytäntöjä jossakin määrin:

Förslaget, hvilka till lediga tjensters besättjande hos Oss ingifvas, endas böra inrättas efter pluraliteten af rösterna, så att de, som fått de mesta *vota* uppföres 1:o *loco*, den som fått mindre 2:o *loco* och den som fått de minsta 3:o *loco*. (*Circulaire till samtelige Collegier*. Kongliga Brefet af den 6. Oktober 1746.)

Tämä täsmennys jätti edelleen tulkinnanvaraa. Mitä tarkoitti kirjeessä mainittu ”äänten pluraliteetti”? Sillä voitiin tarkoittaa joko *absoluuttista* enemmistöä (yli 50 % äänistä) tai *suhteellista* enemmistöä eli suurinta äänimäärää. Ilmaus ”saada eniten ääniä” (”mesta *vota*”) oli samaan tapaan monitulkintainen. Mahdollisia äänestysmenettelyjä on loputtomasti, mutta ainakin seuraavia menettelyjä ilmeisesti sovellettiin myös käytännössä (Mäntylä 1981, 37–38, 46–47, 80–81; Autio 1981, 66–67):

(1) Yksinkertaisin tulkinta oli suoraviivainen *pluraliteettiäänestys*. Äänestys suoritettiin kaikkien hakijoiden välillä ja jokainen äänestäjä äänesti vain yhtä ehdokasta (Mäntylä 1981, 37). Se ehdokas, joka sai eniten ääniä, asetettiin ensimmäiselle sijalle, toiseksi eniten ääniä saanut toiselle sijalle ja kolmanneksi eniten ääniä saanut kolmannelle sijalle. Yksi tämän menettelyn ilmeinen epäkohta oli, että kaikki äänet saattoivat jakautua kahden pääehdokkaan välille, jolloin kolmas sija jäi täyttämättä (Snellman 1862c, 474). Oli myös mahdollista, että äänten keskittyessä yhdelle ehdokkaalle kolmas ja kenties toinenkin sija määräytyivät yksittäisten hajaäänten perusteella (Mäntylä 1981, 81). Äänestäjien suuri enemmistö olisi ehkä sijoittanut myös näille sijoille jonkun muun. Koska nimitysvallan käyttäjä saattoi sivuuttaa selvän enemmistönkin mielipiteen, ei ollut samantekevää, miten toinen ja kolmas sija täytettiin (Autio 1981, 66).

(2) Mainitut pluraliteettäänestyksen ongelmat johtivat seuraavaan menettelyyn. Aluksi äänestettiin ensimmäisestä sijasta kuten edellä, kaikkien hakijoiden kesken. Voittajaksi selviytyi se joka sai eniten ääniä. Seuraavaksi äänestettiin muiden ehdokkaiden paitsi ensimmäisen kierroksen voittajan välillä toisesta sijasta. Voittaja määräytyi jälleen pluraliteetikriteerin mukaisesti. Lopuksi samaa menettelyä sovellettiin jäljelle jääviin ehdokkaisiin kolmannen sijan osalta. Jokainen sija täytettiin ikään kuin ne olisivat erillisiä tehtäviä joihin pyrkivät (osin) samat hakijat. Tätä kutsutaan jatkossa *usean kierroksen menetelmäksi*.

(3) Edellä kuvattu menettely oli sikäli aikaa vievä, että se edellytti kolmea erillistä äänestystä. Ajan säästämiseksi kutakin äänestäjää voitiin pyytää asettamaan kerralla kolme ehdokasta paremmuusjärjestykseen. Virkaehdotuksessa ensimmäiselle sijalle asetettiin se ehdokas, joka sai äänestäjiltä eniten ensimmäisiä sijoja. Toiselle sijalle valittiin jäljelle jäävistä se, jolla oli eniten ensimmäisiä ja toisia sijoja. Kolmannelle sijalle päätyi se, jolla jäljelle jäävistä oli eniten ensimmäisiä, toisia ja kolmansia sijoja. Tämä menettely näyttää vakiintuneen ainakin kirkon tehtäviä täytettäessä (Berg & Nurmi 1988, 96). Kutsun tätä menettelyä jatkossa, Lindelöfiä seuraten, *eliminointimenettelyksi*.

(4) Käskykirje antoi mahdollisuuden myös *hyväksymisäänestyksen* käytölle. Tätä menettelyä käytettäessä jokainen äänestäjä antaa yhdessä äänestyksessä äänensä (enintään) kolmelle ehdokkaalle, ilmaisematta paremmuusjärjestystä. Äännet lasketaan yhteen ja sijat täytetään käskykirjeen ohjeen mukaisesti siten, että eniten hyväksymisiä saanut on ensimmäisellä sijalla jne. (Mäntylä 1981, 37). Vaikka hyväksymisäänestys nykyään usein liitetään sen aktiivisten puolestapuhujien (esimerkiksi äänestysteoreetikko Steven J. Bramsin) nimiin, kyseistä menettelyä on käytetty eri yhteyksissä hyvin kauan.

Kun Suomi siirtyi osaksi Venäjän imperiumia, Aleksanteri I säilytti Porvoossa antamansa lupauksen mukaisesti korporaatioiden erioikeudet ennallaan. Näihin kuuluivat myös niiden sisäiset vaalitavat. Niinpä Keisarillinen Aleksanterin yliopisto, esimerkiksi äänestäessään mainitusta kirjastonhoitajan virkaesityksestä, pyrki noudattamaan Ruotsin valtakunnassa 1700-luvulla säädettyjä lakeja. Lakien mukana periytyivät myös tulkintaongelmat (Autio 1981, 65–67). Vuoden 1862 kirjastonhoitajan viran täytössä konsistori sovelsi tulkintaa (3) eli käytti eliminointimenettelyä. Koska hakijoita oli vain kolme, virkaesityksen kolmas sija määräytyi tässä tapauksessa automaattisesti, kun ensimmäiset ja toiset sijat oli täytetty.

3. Äänestysmenettelyjen matemaattinen tarkastelu

Konsistorin nuorempi jäsen – joka lisäksi oli matemaatikko – pystyi helposti havaitsemaan vallitsevaan käytäntöön liittyvät epäjohdonmukaisuudet. *Juridisk Albumin* lähettämässään artikkelissa Lindelöf lähti liikkeelle tarkastelemalla seuraavaa hypoteettista esimerkkiä (Lindelöf 1862a, 51). Oletetaan 24 äänestäjän päätöksentekoelin, jonka on asetettava kolme hakijaa (A, B, ja C) paremmuusjärjestykseen. Äänestäjät arvioivat hakijoita seuraavasti:

Esimerkki 1:

3	6	4	4	7	äänestäjää
A	A	B	B	C	
B	C	A	C	B	
C	B	C	A	A	

Usean kierroksen menetelmää käytettäessä hakija A asetetaan virkaesityksessä ensimmäiselle sijalle, koska hänen takanaan on pluraliteetti (9 äänestäjää 24:stä). Seuraavaksi järjestetään erillinen äänestys toisesta sijasta. Koska äänestäjien enemmistö (13 äänestäjää 24:stä) on sitä mieltä, että C on parempi kuin B, järjestykseksi tulee $A > C > B$. Lindelöfin mukaan tämä järjestys on kahdessakin suhteessa virheellinen. Ensiksi, peräti 11 äänestäjää pitää A:ta *huonoimpana* ehdokkaana; A:lla on vähemmän ensimmäisiä ja toisia sijoja kuin kummallakaan kilpahakijoistaan. Toiseksi, koska B on ensimmäisenä 8 äänestäjän preferensseissä ja toisena 10 äänestäjän preferensseissä, kun C:llä on vain 7 ensimmäistä ja 10 toista sijaa, B olisi asetettava C:n edelle. Preferenssijärjestyksen tulisi siis olla $B > C > A$. (Ibid., 52–53.) Jos taas käytetään yhteen äänestykseen rajoittuvaa eliminointimenettelyä, A saa silloinkin ensimmäisen sijan. Toinen ja kolmas sija määrätään laskemalla yhteen jäljelle jäävien ehdokkaiden saamat ensimmäiset ja toiset preferenssisijat. Tällöin lopputulos on $A > B > C$. Huomaamme, että yhteen äänestykseen rajoittuva eliminointimenettely *ei* simuloi usean kierroksen menettelyä, koska nämä voivat joissakin tilanteissa johtaa erilaisiin tuloksiin. Tämä oli varmaan yllättävä tulos, koska eliminointimenettely luultavasti miellettiin vain nopeutetuksi tavaksi suoriutua usean kierroksen vaalista.

Eliminointimenettelyyn puutteet liittyivät Lindelöfin mukaan kahteen periaatteelliseen ongelmaan. Virkaesitysten ensimmäiset sijat määräytyivät pluraliteettiperiaatteen mukaisesti pelkästään äänestäjien ykköspreferenssien perusteella; muu informaatio jätettiin huomiotta. Tämä johtaa ensimmäiseen Lindelöfin mainitsemaan ongelmaan. Se voi itse asiassa saada vielä vakavamman muodon kuin esimerkissä 1. Lindelöfin myöhemmin käyttämä esimerkki (Lindelöf 1862c, 465) on seuraava:

Esimerkki 2:

9	8	7	äänestäjää
A	B	C	
B	C	B	
C	A	A	

Tässä esimerkissä absoluuttinen enemmistö (15 äänestäjää 24:stä) pitää pluraliteettivoittaja A:ta huonoimpana vaihtoehtona – tilanne, jota myöhemmässä kirjallisuudessa kutsutaan *Borda-paradoksin* vahvaksi versioksi. Lindelöf huomautti, että äänten jakautuessa monen ehdokkaan kesken pluraliteettiperiaatteen ongelmat kävivät yhä ilmeisemmiksi (Lindelöf 1862c, 468). Lindelöf huomautti edelleen, että esimerkin 2 kaltaisessa asetelmassa pluraliteettivaalin lopputulos itse asiassa riippuu esityslistasta. Jos A:n kilpahakijana olisi pelkästään B tai pelkästään C, A häviäisi kummassakin tapauksessa. Kun molemmat hakevat virkaa, voittajaksi selviytyykin A (Lindelöf 1862c, 468). Lindelöf tuli näin intuitiivisella tasolla muotoilleeksi rationaalisuuschdon, jota myöhemmässä sosiaalisen valinnan teoriassa on kutsuttu nimellä *α -ehto* (englanniksi myös *contraction consistency*). Sen mukaan, jos menettely valitsee jostakin annetusta vaihtoehtojen joukosta yhden, esimerkiksi A:n joukosta {A,B,C}, sen on valittava A myös jokaisesta osajoukosta jossa A on jäsenenä.

Sekä eliminointimenettelyä että usean kierroksen menettelyä käytettäessä virkaesityksen ensimmäinen sija täytetään pluraliteettimenettelyä käyttäen. Toisista ja kolmansista sijoista päätettäessä eliminointimenettely ottaa huomioon myös muita äänestäjien ilmaisemia preferenssejä ja johtaa esimerkin 1 tapauksessa Lindelöfin mukaan oikeudenmukaisempaan tulokseen kuin usean kierroksen menettely. Tässä on kuitenkin eliminointimenettelyn toinen

periaatteellinen ongelma. Se on epäjohdonmukainen yhdistelmä kahdesta eri periaatteelle perustuvasta valintamenettelystä: ensimmäinen sija täytetään eri kriteeriä käyttäen kuin kaksi seuraavaa sijaa (Lindelöf 1862c, 464).

Usean kierroksen menetelmässä toisen ja kolmannen sijan täyttäminen perustuu (kolmen ehdokkaan tapauksessa) jäljellä olevien kahden ehdokkaan väliseen enemmistöparivertailuun. Yksi ajateltavissa oleva mahdollisuus olisi, että *kaikkien* ehdokkaiden välillä suoritettaisiin tällaiset parivertailut. Esimerkin 1 kuvaamassa asetelmassa tämä kuitenkin johtaisi paradoksaaliseen tulokseen. Parittaisissa äänestyksissä B voittaisi A:n äänin 15-9, A voittaisi C:n äänin 13-11 ja C puolestaan voittaisi B:n äänin 13-11. Parittaisiin enemmistövertailuihin perustuva kollektiivinen preferenssi-järjestys olisi siis kehämäinen (Lindelöf 1862a, 56). Nykyisessä kirjallisuudessa esimerkin 1 kuvaamaa asetelmaa kutsutaan *Condorcet-paradoksiksi*. Condorcet-paradoksin tunnistaminen ja sen tarkastelut ovat näytelleet keskeistä roolia sosiaalisen valinnan teorian kehityksessä – esimerkiksi Arrowin kuuluisan teoreeman todistus edellyttää Condorcet-paradoksin mahdollisuutta. Tämän artikkelin alussa mainittu käsitys, jonka mukaan keskeiset äänestysteoreettiset tulokset vaipuivat välillä unohduksiin, perustuu olennaisesti sille olettamukselle että tätä Condorcet'n (1995 [1785]) osoittamaa enemmistövertailun paradoksaalista piirrettä ei tunnettu. Kuten huomaamme, Lindelöfin kohdalla oletamus ei pidä paikkaansa. Lindelöf ei vain muistuttanut paradoksin mahdollisuudesta, vaan myös analysoi yhtä paradoksin keskeistä seurausta. Hän osoitti, että paradoksaalisessa tilanteessa enemmistöpäätöksistä tulee (jälleen nykytermiä käyttäkseni) *polkurippuvaisia*. Kuvatussa tilanteessa kuka tahansa kolmesta hakijasta (A, B, tai C) voidaan saada ykköstilalle riippuen siitä millaista äänestysjärjestystä käytetään.

Konsistorin käyttämän menettelyn tilalle Lindelöf tarjosi omaa vaihtoehtoaan. Se oli Jean-Charles de Bordan vuonna 1770 esittämä niin sanottu Borda-menettely (Lindelöf ei käytä tätä termiä). Yleistetyssä muodossa menettelyä voidaan kuvata seuraavasti: jokainen äänestäjä antaa huonoimmaksi katsomalleen ehdokkaalle a pistettä, seuraavaksi huonoimmalle $a+b$ pistettä, seuraavalle $a+2b$ pistettä jne. Tavallisimmassa versiossa $a = 0$ ja $b = 1$; kolmen ehdokkaan vaalissa painokertoimet ovat siis 2, 1, ja 0. Lindelöfin ehdotuksessa $a = 1$ ja $b = 1$, joten painokertoimet ovat 3, 2, 1. Kuten hän itse huomauttaa, lukujen valinnalla ei ole merkitystä niin kauan kuin niiden väliset suhteet säilyvät

muuttumattomina. Kunkin ehdokkaan saamat pisteet lasketaan yhteen, ja kollektiivinen järjestys määräytyy näiden summien suurusjärjestyksen perusteella. Sovellettuna esimerkkiin 1 saamme seuraavan tuloksen. A: $(9*3)+(4*2)+(11*1) = 46$ pistettä; B: $(8*3)+(10*2)+(6*1) = 50$ pistettä; C: $(7*3)+(10*2)+(7*1) = 48$ pistettä. Kollektiiviseksi paremmuusjärjestykseksi tulee siis $B>C>A$, Lindelöf oli sitä mieltä, että Borda-menettely oli sekä matemaattisesti perustellumpi että tuotti lopputuloksia jotka olivat paremmin sopusoinnussa päätöksentekijöiden tahdon kanssa kuin eliminointimenettely tai monen kierroksen menettely. Olennaista oli, että *kaikkia kolmea sijaa määrättäessä kaikkien äänestäjien kaikki preferenssit otettiin huomioon.* (Lindelöf 1862a, 55–56; Lindelöf 1862b, 126–129.) Lindelöf ei esittänyt suosittamaansa menettelyä omana keksintönään. Vaikka hän saattoi tuntea Bordan alkuperäisen artikkelin, hänen lähteinään olivat maineikkaan ranskalaisen matemaatikon Simon-Pierre de Laplace’in (1749–1827) kirjoitukset.

4. Filosofi vastaa matemaatikolle

Toimittamassaan *Litteraturbladissa* Snellman esitti Lindelöfin tuloksia koskevat epäilynsä. Vaikka Snellman ei ollut matemaattisesti täysin sivistymätön, tällä alueella hän ei pystynyt tarjoamaan Lindelöfille kunnan vastusta. Hänen kommentistaan käy selvästi ilmi, että hän ei kyennyt ymmärtämään kaikkia Lindelöfin argumentteja. Snellman esimerkiksi kirjoittaa toistuvasti Lindelöfin esimerkistä (esimerkki 1 yllä) löytämästään ”laskuvirheestä”; kyse on ilmeisesti vain Snellmanin omasta väärintulkinnasta (Snellman 1862a, 111; Snellman 1862b, 318). Hän näyttää pitävän kiinni käsityksestä, että usean kierroksen menetelmä ja eliminointimenettely tuottavat aina saman tuloksen (Snellman 1862a, 111; Snellman 1862b, 317; Snellman 1862c, 469), vaikka Lindelöf oli esimerkissä 1 osoittanut, että näin ei ole asian laita. Snellman ei kommentoi sen enempää pluraliteettiperiaatteeseen kohdistettua kritiikkiä kuin Condorcet-paradoksiakaan; sen sijaan hän kohdistaa koko kritiikkinsä Borda-menettelyyn ja Lindelöfin sen tueksi esittämiin perusteluihin. Tässä hänellä on myös jotakin huomionarvoista sanottavaa.

Snellman huomauttaa Borda-menettelyn perustuvan olettamukseen, jonka mukaan äänestäjät katsovat ehdokkaiden väliset etäisyydet keskenään yhtä suuriksi: A on välttämättä yhtä ”kaukana” B:stä kuin B C:stä. Ajatellaanpa esimerkkiä jossa hakijat olisivat Laplace (Lindelöfin auktoriteetti) tai Hegel

(Snellmanin auktoriteetti) ja kaksi keskivertoa maisteria. Lindelöfin suosittama menettely johtaa siihen, että äänestäjän on äänestettävä ikään kuin Laplace'in ansioiden ja toisen virkaa hakeneen maisterin ansioiden välinen suhde olisi sama kuin kahden maisterin ansioiden välinen suhde. Äänestäjä saattaa todellisuudessa olla sitä mieltä, että jos maistereiden ansioita arvioidaan luvuilla 1 ja 2, Laplace'in (tai Hegelin) ansioita on merkittävä esimerkiksi luvulla 100. Borda-menettely ei kuitenkaan mahdollista tällaisen mielipiteen ilmaisemista. Menettelyn tarjoama pisteasteikko on mielivaltainen, eivätkä tulokset välttämättä heijasta muuta kuin menettelyn omia ominaisuuksia. (Snellman 1862a, 112.) Tämä Snellmanin esittämä kritiikki on esitetty monesti, jo ennen Snellmania ja myös hänen jälkeensä (esim. Black 1958, 180–183). Lindelöfillä on tähän kritiikkiin kaksi vastausta, jotka nekin ovat aiheeseen perehtyneille tuttuja. Ensiksi hän huomauttaa, että pluraliteettisääntö on vielä mielivaltaisempi ja tarjoaa äänestäjille vielä vähemmän mahdollisuuksia ilmaista todellista mielipidettään. Pluraliteettisäännön tarjoama pisteasteikko on kolmen ehdokkaan tapauksessa 1, 0, 0. Sen mukaan äänestäjän on ilmaistava käsitys, jonka mukaan hänen parhaana pitämänsä ehdokas on *äärettömän* paljon parempi kuin muut ehdokkaat, eikä muiden hakijoiden keskinäisillä eroilla ole mitään merkitystä. Jos hän esimerkiksi pitää Hegeliä parhaana ehdokkaana, hän joutuu samalla allekirjoittamaan mielipiteen, jonka mukaan Laplace'in ja keskinkertaisen maisterin välillä ei ole mitään eroa. Lindelöf katsoo, että *jokainen* käytetty sääntö joutuu nojautumaan joillekin äänestäjien preferenssijärjestyksiä koskeville implisiittisille oletamuksille. Kysymys on vain siitä, miten näiden kilpailevien olettamusten välillä valitaan. (Lindelöf 1862a, 55; Lindelöf 1862b, 126; Lindelöf 1862c, 466–467.)

Tässä yhteydessä Lindelöf tarjoaa positiivisen argumentin Borda-asteikon tueksi. Se on sama kuin Laplace'in esittämä (Laplace 1812, 273–274; Laplace 1814, 87–92). Intuitiivisesti muotoiltuna argumentti on seuraava. Oletetaan, että äänestäjien arviot kolmen kilpailevan ehdokkaan ansioista voidaan muuttaa numeeriseen muotoon ja normalisoida jollekin lukuvälille. Oletetaan, että arviot vaihtelevat satunnaisesti toisistaan riippumattomalla tavalla. Tällöin voidaan osoittaa, että arviointilanteiden (tai arvioitsijoiden) lukumäärän lähestyessä ääretöntä, arvioita ilmaisevien keskimääräisten lukujen suhteet lähestyvät suhdetta 3:2:1. Tämä on sama kuin Lindelöfin suosittama Borda-asteikko.

Laplace’in tulos voidaan yleistää n :än ehdokkaan arviointitilanteeseen. (Lindelöf 1862b, 126–128.)

Snellmanin vastaus on kiinnostava. Lindelöfin lähtökohta on, että jokaisen äänestäjän arvio kahden kilpailevan ehdokkaan keskinäisistä ansioista voidaan *aina* liittää johonkin lukuarvoon. Äänestys on itse asiassa eräänlainen mittausoperaatio. Snellman kiistää tämän tavan tulkita äänestyksiä. Hakijoiden ansioiden keskinäinen vertailu ei ole samanlainen toimitus kuin fyysisen voiman tai tenttisuoritusten vertailu. On täysin mahdollista, että kahden tai useamman hakijan ansiot ovat keskenään vertailemattomia. Äänestäjät voivat esimerkiksi olla perustellusti sitä mieltä, että sekä A että B ovat selkeästi parempia kuin C, mutta A:n ja B:n ansiot ovat keskenään niin erilaisia että vertailu on mahdotonta. Modernein termein ilmaistuna, äänestäjien arviot hakijoista eivät muodosta järjestystä joka olisi *täydellinen* ja *tiukka* (engl. *strict*). Tilanne saattaa kuitenkin olla sellainen, että jokin päätös on tehtävä. Tällöin äänestystulos on väistämättä tietyssä määrin sattumanvarainen. Lindelöfin suosittama menettely vain lisää tätä sattumanvaraisuutta, koska se pakottaa äänestäjät ilmaisemaan preferenssejä joita heillä ei todellisuudessa ole. Sen käyttö perustuu sille kyseenalaiselle *a priori* -olettamukselle, että äänestäjät pystyvät tuottamaan tiukan ja täydellisen järjestyksen missä tahansa hakijoiden joukossa. (Snellman 1862b, 315–316; 321–322.)

Ensimmäisessä kommentissaan Snellman oli syyttänyt Lindelöfiä ja tämä kunnioittamaa auktoriteettia Laplace’ia yrityksestä ulottaa matemaattinen argumentaatio ”sen oman alueen ulkopuolelle” (Snellman 1862a, 110). Tällainen syytös sai matemaatikon raivostumaan (Lindelöf 1862b, 119). Eikö filosofi ymmärtänyt, että jokainen matematiikan menestyksellinen sovellutus, myös luonnontieteissä, edellytti aina joitakin ei-matemaattisia premissejä? Tässä toteamuksessaan Lindelöf oli tietysti oikeassa. Samalla hän tuli kuitenkin paljastaneeksi aukon puolustuksessaan. Sillä toisaalta hänen perusargumenttinsa oli, että ”järki” ja ”tiede”, erityisesti todennäköisyyslaskenta, riittivät perustelemaan Borda-menetelyn ylivoimaisuuden (Lindelöf 1862b, 120). Hän näytti puolustavan omaa kantaansa pelkästään *a priori* -argumenteilla ja syytti Snellmania matematiikan vastaisuudesta. Kieltämättä Snellmanin väite, jonka mukaan matematiikkaa ei voi soveltaa ”hengen maailmaan” (Snellman 1862b, 314), oli tulkittavissa tähän tapaan. Lindelöfin oma huomio ei-matemaattisten premissien välttämättömyydestä mahdollistaa kuitenkin myös toisen tulkinnan.

Hegeliläisen ”hengen maailman” sijasta voitaisiin myös puhua äänestysmenettelyjen valintaan välttämättä liittyvästä *normatiivisesta* aspektista. Toisin kuin niissä tilastotieteellisten metodien sovellutuksissa joista Lindelöf oli kiinnostunut, äänestysmenettelyjen valinnassa ei ollut kyse empiiristen ilmiöiden analysoimisesta. Jos argumentin tarkoituksena oli osoittaa, että Borda-menettely oli, paitsi loogisin, myös *oikeudenmukaisin* (”hwilar på fullt rättwis och logisk grund”; Lindelöf 1862a, 54), oli nojaututtava myös joihinkin normatiivisiin premisseihin. Näitä premissejä hän ei artikuloinut, vaan tukeutui vain esimerkkeihinsä liittyviin intuitioihin jotka hän oletti lukijoittensa jakavan.²

5. Matematiikkaa, filosofiaa, juridiikkaa – vai sittenkin politiikkaa?

Äänestysmenettelyjen oikeudenmukaisuutta koskevat intuitiot eivät kuitenkaan olleet niin yksiselitteisiä kuin Lindelöf oletti. Yleisen filosofisen kritiikin ohella Snellman pystyi, matematiikkaan kohdistuvista epäluuloistaan huolimatta, esittämään myös teknisempiä vastaväitteitä. Hän huomautti, että Lindelöfin esittämä Borda-menettely saattoi johtaa lopputulokseen, jossa enemmistön parhaana pitämä ehdokas jäi valitsematta. Näin käy seuraavassa esimerkissä (Snellman 1862c, 471):

Esimerkki 3:

13	1	9	1	äänestäjää
A	A	B	C	
B	C	C	B	
C	B	A	A	

² Lindelöfin hyväksymä Laplace’in argumentti *voidaan* kyllä muotoilla uudelleen tavalla, joka sopii paremmin normatiiviseen tarkasteluun. Sen sijaan että puhuttaisiin äärettömien havaintosarjojen raja-arvoista, voidaan esittää, että äänestysmenettelyn painokertoimien valinnassa on kysymys Rawlsin ja Harsanyin käyttämän ”tietämättömyyden verhon” takana tehdystä päätöksestä. Yksinkertaistaen: hypoteettisten valitsijoiden tehtävänä on päästä sopimukseen sellaisista painokertoimista jotka maksimoivat vaalituloksiin liittyvät utiliteettien odotusarvot. Koska he eivät tiedä, miten he tulevaisuuden äänestystilanteissa tulevat eri vaihtoehtoja painottamaan, he olettavat Laplace’in nimiin pannun periaatteen mukaisesti kaikki mahdolliset tilanteet yhtä todennäköisiksi. Tiettyjen lisäoletusten vallitessa he päätyvät valitsemaan äärettömän monen mahdollisen painotustavan joukosta Bordan ja Lindelöfin suosittaman aritmeettisen asteikon. Tämä argumentti nojautuu Lindelöfin toiveen mukaisesti pelkästään ”järkeen ja todennäköisyyslaskentaan”. Siihen pätevät kuitenkin samat kritiikit kuin muihinkin hypoteettisiin sopimusargumentteihin.

Snellmanin esimerkissä 14 äänestäjää 24:stä pitää A:ta parhaana ehdokkaana. Lindelöfin kannattama äänestysmenettely johtaa kuitenkin tulokseen, jossa A saa $(14*3)+(10*1) = 52$ pistettä, kun B saa $(9*3)+(14*2)+(1*1) = 56$ pistettä. Tämä tulos on puolestaan Snellmanin intuitioiden vastainen. Hän mielestään oli käsittämätöntä, että Lindelöf saattoi puolustaa menettelyään parhaana tapana tulkita *enemmistön* mielipidettä (”pluralitetens mening”; *ibid.*).

Keskustelun edetessä siihen liittyi myös kolmas osallistuja. Hänen henkilöllisyytensä ei ole tiedossa, joskin Lindelöf ja Snellman luultavasti tiesivät kuka oli kyseessä. *Finlands Allmänna Tidningissä* julkaistiin lähetetty kirjoitus, jossa lyhyesti kommentoitiin käytyä keskustelua (Anon. 1862). Koska kysymys oli myös lain soveltamisesta, nimetön kirjoittaja – todennäköisesti juristi – tarkasteli yksityiskohtaisemmin vuoden 1746 kuninkaallisen kirjeen sanamuotoja. Tämän perusteella hän torjui Snellmanin esittämän käsityksen, jonka mukaan laki olisi edellyttänyt erillisiä äänestyksiä eri ehdokassijoista ja eliminointimenettely olisi tulkittavissa usean kierroksen menettelyn nopeutetuksi versioksi. Kirjoittajan mukaan käskykirje edellytti nimenomaan *yhtä* äänestystä, jonka perusteella ”mesta *vota*”, ”mindre *vota*” ja ”minsta *vota*” laskettaisiin. Tällä perusteella kirjoittaja katsoi, että laki edellytti *hyväksymisäänestystä* jossa jokainen äänestäjä antaisi kolme – tai kolmen ehdokkaan tapauksessa kaksi – keskenään samanarvoista hyväksymisääntä. Äänimäärät ratkaisisivat ehdokassijat. Koska erityisesti pienissä päätöksentekoeleimissä tasatulokset olivat todennäköisiä, äänestäjien oli samalla ilmaistava preferenssinsä. Preferenssit kuitenkin otettiin huomioon vain siinä tapauksessa että äänet menivät tasan. Jos siis esimerkiksi kolmen ehdokkaan vaalissa kaksi ehdokasta sai saman määrän hyväksymisääniä, ykkössijalle asetettiin se, jolla oli enemmän ääniä ensimmäiselle sijalle.

Lindelöf tulkitsi *Finlands Allmänna Tidningin* kirjoituksen omaa kantaansa tukevaksi, koska kirjoittaja hylkäsi pluraliteettiperiaatteen ja sen myötä Snellmanin tulkinnan, jonka mukaan usean kierroksen menettely perustui lakiin ja eliminointimenettely oli ymmärrettävä sen sovellutuksena. Kummassakin Lindelöfin edellä tarkastellussa esimerkissä (esimerkit 1 ja 2) nimettömän kirjoittajan esittämä menettely tosiaan johtaa samaan järjestykseen kuin Lindelöf-Borda -sääntö (Lindelöf 1862c, 463–464). Snellman ei kuitenkaan jäänyt vastausta velkaa. Hän kiisti sen, että käskykirjeen kieltämättä ylimalkaiset sanamuodot automaattisesti edellyttäisivät hyväksymisäänestystä. Lindelöfille hän huomautti, että vaikka nimettömän kirjoittajan suosittama menettely sattui

tarkastelluissa esimerkkitaapauksissa johtamaan Lindelöfin kannalta ”oikeaan” lopputulokseen, näin ei suinkaan aina ollut asian laita. Snellmanin esimerkissä on neljä ehdokasta, jotka saavat hyväksymisääniä eri ehdokassijoille seuraavasti (Snellman 1862c, 470):

Esimerkki 4:

	I	II	III	ehdokassija
A	9	-	-	
B	-	9	-	
C	-	-	10	
D	1	1	-	

Hyväksymismenettelyn mukaan ensimmäiselle sijalle on asetettava C, joka saa 10 hyväksymisääntä – tosin vain kolmannelle sijalle, mutta *Finlands Allmänna Tidningissä* esitettyä menettelyä käytettäessä sillä ei ole merkitystä. Seuraavia sijoja määrättäessä taas sovelletaan tasatulosten ratkaisuperiaatetta. Näin kolmen ehdokkaan järjestykseksi tulee $C > A > B$. Lindelöfin menettelyä käytettäessä järjestys olisi $A > B > C$. Menettelyt eivät siis johda samaan tulokseen kuin satunnaisesti.

Snellman teki toisenkin Lindelöfin menettelyä koskevan teknisluontoisen huomion. Hän huomautti, että sitä käytettäessä vähemmistöllä on mahdollisuus estää enemmistön tahdon toteutuminen äänestämällä *strategisesti*. Esimerkki tästä (mukailtu esimerkistä Snellman 1862b, 112) on seuraava:

Esimerkki 5:

15	9	äänestäjää
A	B	
B	A	
C	C	

Jos 9 äänestäjää haluaa ajaa B:n ykkössijalle vastoin 15 äänestäjän kantaa, he voivat tehdä sen sijoittamalla A:n äänestyslipuissaan kolmannelle sijalle, vaikka tämä sijoittelu ei edustaisikaan heidän todellista mielipidettään hakijoiden ansioista. Tässä Snellman esittää perinteisen kritiikin, joka aikanaan esitettiin jo Bordalle itselleen (ja johon tämän kerrotaan vastanneen: ”Minun menettelyni

on tarkoitettu rehellisille ihmisille”). Itse asiassa, vedotessaan Laplace’in auktoriteettiin Lindelöf joko syyllistyi lievään huolimattomuuteen tai sitten ei ollut täysin rehellinen lukijoitaan kohtaan. Laplace esitti alun perin argumenttinsa jo vuonna 1795. Vaikka vuonna 1814 ilmestyneessä teoksessaan *Essai philosophique sur les probabilités* Laplace yhä pitää Borda-menettelyä periaatteessa parhaana, käytännössä hän hylkää menettelyn *nimenomaan* esimerkin 5 kuvaaman ongelman takia (Laplace 1814, 90).

Lindelöfiä kaksikymmentä vuotta vanhempi Snellman oli kiistan aikaan jo kokenut poliitikko. Oli mahdollista, että hänen kantaansa vaikuttivat myös seikat joita hän ei halunnut avoimesti julkituoda. Kuten edellä todettiin, nimittävän viranomaisen mahdollisuus käyttää harkintavaltaansa virkaesityksen puitteissa ei jäänyt kuolleeksi kirjaimeksi. Snellman oli itse joutunut poliittisista syistä syrjäytetyksi filosofian professorin valinnassa vuonna 1849, vaikka konsistori oli asettanut hänet ensimmäiselle sijalle. Argumentit jotka heikensivät käytettyjen äänestysmenettelyjen legitimitettä, saattoivat osaltaan tarjota venäläisille vallankäyttäjille lisää perusteita puuttua itsehallinnollisten suomalaisten instituutioiden toimintaan. Oli turvallisempaa puolustaa pääsääntöisesti toimiviksi koettuja perittyjä muotoja kuin pohtia niihin liittyviä abstrakteja ja hypoteettisia ongelmia. Nämä pohdinnat vilahtavat esiin kohdassa jossa Snellman, tuomittuaan ensin Borda–Lindelöf -tyyppisten menettelyjen tuottamat ehdokkaiden väliset äänierot mielivaltaisiksi, jatkaa:

Korkeampi auktoriteetti joka täyttää virat, voi tällöin esim. poliittisista syistä valita sen joka on toisena järjestyksessä, koskapa kerran kyvykkyudessa olevat erot on todistettu vähäisiksi. (Snellman 1862b, 316; käänös EL.)

6. Lopuksi

Lindelöfin elämäkerran kirjoittaja Lehto, tietääkseni ainoa joka on kiistaa kommentoinut, toteaa, että väittelyssä ei ollut voittajaa ”koska kysymys ei ole puhtaasti matemaattinen eikä sillä ole yksiselitteistä ratkaisua” (Lehto 2008, 62). Snellman olisi luultavasti pitänyt tätä tuomiota oman kantansa voittona. Päinvastaisista puheistaan huolimatta Lindelöf näytti käsittelevän kysymystä ikään kuin se olisi ollut puhtaasti matemaattinen. Juuri tähän kohdistui Snellmanin vakavin vastaväite. Vaikka Snellmanin filosofinen kritiikki oli osin

sattuvaa, Lindelöfin äänestysmenettelyihin kohdistuvat huomiot olivat nekin kiistämättömiä. Lopulta kysymys oli siitä, mikä rooli formaalisella teorialla oli käytännön päätöksenteossa. Keskustelu tästä kysymyksestä jatkuu yhä.

Tieteen historiaan professorien kiista lisää vähintään yhden alaviitteen. Suomalainen äänestysmenettelyjä koskenut väittely todistaa osaltaan virheelliseksi sen käsityksen, että ranskalaiset suurmiehet olisivat vieneet äänestysmenettelyjä koskevan teorian mukanaan hautaan. Vaikka on totta, että Condorcet'n ja Arrowin tutkimusten välisenä lähes puolentoista vuosisadan aikana ei sosiaalisen valinnan teoriassa tapahtunut kumulatiivista edistystä, jotakin aiheesta toki tiedettiin, myös täällä Pohjolan perillä.

Turun yliopisto

Kirjallisuus

- Anon. (1862). ”Några ord om förslags upprättande till tjänster”, *Finlands Allmänna Tidning* nr. 213, 867.
- Autio, Veli-Matti (1981). *Yliopiston virkanimitykset 1809–1852*. Historiallisia tutkimuksia 115. Helsinki: Suomen historiallinen seura.
- Berg, Sven & Nurmi, Hannu (1988). ”Making Choices in the Old-Fashioned Way”, *Economia delle scelte pubbliche* 2, 95–113.
- Black, Duncan (1958). *The Theory of Committees and Elections*. Cambridge: Cambridge University Press.
- Borda, Jean-Charles de (1995 [1770]) ”On Elections by Ballot.” Käänt. I. McLean & F. Hewitt. Teoksessa Iain McLean & Arnold B. Urken (toim.), *Classics of Social Choice*. Ann Arbor: The University of Michigan Press, 83–89.
- Condorcet, Marie Jean Antoine Nicolas de (1995 [1785]). ”An Essay on the Application of Analysis to the Probability of Decisions Rendered by a Plurality of Votes.” Käänt. A. B. Urken, R. Pinkham & J. E. McClellan. Teoksessa Iain McLean, Iain & Arnold B. Urken (toim.), *Classics of Social Choice*. Ann Arbor: The University of Michigan Press, 91–112.
- Laplace, Pierre-Simon de (1812). *Théorie analytique des probabilités*. Paris: Courcier.
- Laplace, Pierre-Simon de (1814). *Essai philosophique sur les probabilités*. Paris: Courcier.
- Lehto, Olli (2008). *Tieteen aatelia: Lorenz Lindelöf ja Ernst Lindelöf*. Helsinki: Otava.
- Lindelöf, Lorenz (1862a). ”Om röstberäkning wid voteringar”, *Juridisk Album* 1:3, 50–57.
- Lindelöf, Lorenz (1862b). ”Ytterligare om röstberäkning”, *Juridisk Album* 2:1, 119–130.
- Lindelöf, Lorenz (1862c). ”Slutord i frågan om röstberäkning”, *Litteraturblad* 10/1862, 463–469.
- Matinelli, Eero (1954). ”Enemmistöperiaate ja papinvaali-instituutio”, *Historiallinen Aikakauskirja* 52, 16–24.
- McLean, Iain (1990). ”The Borda and Condorcet principles: three medieval applications”, *Social Choice and Welfare* 7, 99–108

- Mäntylä, Ilkka (1977). *Yksimielisyydestä kauppiaiden valtaan. Raatimiesten vaalit 12 kaupungissa 1722–1808*. Historiallisia tutkimuksia 104. Helsinki: Suomen historiallinen seura.
- Mäntylä, Ilkka (1981). *Valitut, ehdollepannut ja nimitetyt. Pormestarin vaalit 20 kaupungissa 1720–1808*. Historiallisia tutkimuksia 114. Helsinki: Suomen historiallinen seura.
- Snellman, Johan Vilhelm (1862a). ”Om röstberäkning vid förslags upprättande”, *Litteraturblad* 3/1862, 110–113.
- Snellman, Johan Vilhelm (1862b). ”Ytterligare pröfning af Hr Professor Lindelöfs yrkanden angående röstberäkning vid förslags upprättande”, *Litteraturblad* 7/1862, 312–322.
- Snellman, Johan Vilhelm (1862c). ”Pröfning af bevisningen i Hr Lindelöfs Slutord”, *Litteraturblad* 10/1862, 469–475.
- Thomson, Arthur (1951). *Studier i frihetstidens prästvalsagstiftning*. Skrifter utgivna av vetenskaps-societeten i Lund 39. Lund: C. K. W. Gleerup.

Mitä vikaa Granger-kausalliteetissa?

Hannu Nurmi

Kausaliteetin määrittelyssä välttämätön tai lainomainen yhteys syyn ja seurauksen välillä tuntuu vaikeasti yhteen sovitettavalta vapaan tahdon tai aidon toimijuuden käsitteiden kanssa. Kausaalinen terminologia on kuitenkin edelleen käytössä ihmistieteissä. Tämä onkin välttämätöntä muun muassa tehokkaiden toimintastrategioiden erottamiseksi tehottomista. Tässä kirjoituksessa käsitellään yhtä varsin vaikutusvaltaista näkemystä, Granger-kausalliteettia, jonka mukaan kausaliteetti paljolti palautuu ennustettavuuteen. Osoittautuu, että näin voi olla ainoastaan varsin epärealististen oletusten vallitessa.

1. Onko kausaliteetti tarpeen?

Käsitys kausaliteetista on edellytys kaikenlaiselle toiminnalle ja siten myös yhteiskuntapolitiikalle. Poliitiikalla tähdätään tiettyihin tavoitetiloihin, esimerkiksi taloudellisen kasvun kiihdyttämiseen tai työttömyyden pienentämiseen. Yhteiskuntapolitiikan mekanismit ovat kuitenkin harvoin puhtaasti kausaalisia, vaan niihin liittyy tyypillisesti ”teleologinen kehä”, von Wrightin (1971, 85) terminologiaa käyttäkseni. Esimerkiksi korkotason lasku mahdollistaa investoinnit aiempaa pienemmin kustannuksin, joten yritysten voi odottaa investoivan aiempaa enemmän, koska se on entistä halvempaa. Näin syntyy kasvua ja työllisyys kasvaa. Narulla ei kuitenkaan voi työntää; jos yritysten odotukset tuotannon menekin kasvusta ovat negatiiviset, voi korkotason manipulointi olla tehon väline yllä mainittuihin tavoitteisiin pyrittäessä. Tässä suhteessa yhteiskunnalliset selitykset ja ennustukset poikkeavat luonnontieteellisistä. Sopiiko kausaliteetti siis lainkaan ihmistieteisiin?

Yleisemmin: mikä rooli kausaliteetilla on tieteissä yleensä? Vaikutusvaltaisillakin tieteenfilosoifeilla on ja on jo pitkään ollut tästä eriäviä mielipiteitä. Seuraavassa kaksi:

Russell (1913): ”kausaalilaki on – kuten moni muukin filosofian kelpuuttama aihepiiri – menneiden aikojen jäännös, joka on jäänyt elämään vain siksi, että virheellisesti oletetaan sen olevan monarkian lailla harmiton”.

Cartwright (1979): ”ilman kausaalilakeja ei voi tulla toimeen, koska niitä tarvitaan perustelemaan eroa tehokkaiden ja teottomien strategioiden välillä”.

Russell odotti formaalikielisten kuvaustapojen pikkuhiljaa korvaavan arkaaiset ilmaisut ja niiden joukossa syy–vaikutus-terminologian. Näin ei näytä käyneen, vaikka erilaisten formalismien rooli kaikilla tieteenaloilla on voimistunut. Tämä ei silti tarkoita kausaliteetin merkityksen vähenemistä määräytymisen kategoriana. Tätä voidaan havainnollistaa Boylen ja Gay-Lussacin lain soveltamisella (esimerkki on peräisin Woldin (1964, 271) artikkelista). Laki kytkee yhteen ideaalikaasujen paineen (P), absoluuttisen lämpötilan (T) ja tilavuuden (V) seuraavasti:

$PV = cT$, missä c on reaalilukuvakio.

Kaavaa voidaan käyttää seuraaviin päättelyihin:

- (i) $c(T/V) \text{ ----} \rightarrow P$
- (ii) $c(T/P) \text{ ----} \rightarrow V$
- (iii) $PV(1/c) \text{ ----} \rightarrow T$

Näistä (i) ja (ii) ilmaisevat myös kausaalisen mekanismin, ts. niitä voidaan käyttää paineen ja tilavuuden manipulointiin. Sen sijaan (iii) ei tällaiseen sovellu; emme voi lisätä ideaalikaasun lämpötilaa painetta ja tilavuutta manipuloimalla.

Kausaliteetti ja ennustettavuus näyttävät siis kytkeytyvän yhteen. Woldin johtopäätös on, että kausaliteetti implikoi ennustettavuuden, mutta ennustettavuus ei aina kausaliteettia. Samantyyppistä näkemystä on esitelty myös Olli Koistinen (Koistinen ja Nurmi 2008). Ennustettavuus on keskeistä myös Suppesin probabilistisessä kausaliteettiteoriassa (Suppes 1970).

Palaamme jälkimmäiseen tuonnempana. Sitä ennen tarkastelemme näkemystä, joka korottaa ennustettavuuden kausaliteetin keskeiseksi tunnusmerkiksi.

2. Granger-kausaliteetin periaatteet

C. W. J. Grangerin 1960-luvulla esittämä määritelmä kausaliteetista nauttii laajaa suosiota nykypäivän taloustieteilijäin piirissä (Granger 1969). Sen ytimen muodostaa käsitys, jonka mukaan syyt esiintyvät seurauksiaan aikaisemmin ja sisältävät ennusteinformaatiota seurauksista. Granger siis hylkää käsityksen kausaatiosta deterministisenä, poikkeuksettomana säännönmukaisuutena tiettyjen ilmiötyyppien välillä. Keskeinen piirre kausaalirelaatiossa on ajallinen perättäisyys: syyt edeltävät ajassa seurauksia. Myöhemmin kirjoittamassaan artikkelissa Granger (1980) täsmentää kausaliteetti-käsitettään seuraavasti. Lähtökohtana on universumi, jonka kaikkia muuttujia mitataan tiettyinä ajankohtina $t = 1, 2, \dots$. Kunakin ajankohtana n kaikkea käytettävissä olevaa tietoa universumista merkitään Ω_n :llä. Merkitään $\Omega_n - Y_n$:llä kaikkea tietoa lukuun ottamatta muuttujan Y_t arvoja alkuhetkestä hetkeen n asti. Tässä universumissa Granger määrittelee kausaliteetin seuraavien aksiomien avulla.

Aksiooma A: menneisyys ja nykyisyys voi vaikuttaa kausaalisesti tulevaisuuteen, mutta tulevaisuus ei voi kausaalisesti vaikuttaa menneisyyteen.

Aksiooma B: Ω_n ei sisällä redundanttia informaatiota. Niinpä jos jokin muuttuja Z_n riippuu deterministisesti yhdestä tai useammasta muuttujasta (on niiden funktio), silloin Z_n pitää sulkea Ω_n :n ulkopuolelle.

Grangerin yleinen kausaliteetin määritelmä on nyt seuraava. Olkoon A mikä hyvänsä joukko arvoja, jotka muuttuja X_t voi saada.

Määritelmä: Y_n aiheuttaa X_{n+1} :n jos todennäköisyys sille, että X_{n+1} :n arvo on A :n elementti, kun tietomme universumista on Ω_n , eroaa todennäköisyydestä, että X_{n+1} :n arvo on A :n elementti, kun tietomme universumista on $\Omega_n - Y_n$.

Määritelmä ei sellaisenaan sovellu aineistojen analyysiin, joten Granger muotoilee käyttökelpoisemman version lähtien liikkeelle informaatiojoukon käsitteestä ja tavoitteesta spesifioida se, mitä tarkoitetaan sillä, että vektori-jono X_t on toisen vektori-jonon Y_t syy. Informaatiojoukko on määritelty kullekin ajankohdalle ja ilmaisee kaiken käytettävissä olevan tiedon tuona hetkenä. Merkitään ajankohdan n informaatiojoukkoa J_n :llä. Se koostuu vektori-jonon Z_t elementeistä ajankohtina $0, \dots, n$. Siis

$$J_n: Z_{n-j}, \text{ missä } j \geq 0.$$

Oletetaan, että J_n sisältää myös $Y_{t:n}$, mutta ei $X_{t:n}$:tä edes osittain. Merkitään J'_n :llä $X_{t:n}$ menneillä ja nykyisellä arvolla täydennettyä J_n :tä. Toisin sanoen J_n sisältää kaikki tiedot lukuun ottamatta tietoa tarkasteltavasta syyvektorista X_n . J'_n puolestaan sisältää myös tämän lisätiedon. Tässä operationaalisessa kontekstissa Granger määrittelee ei-syy:n ja syy:n seuraavasti.

Määritelmä: X_n ei ole $Y_{n+1:n}$ syy suhteessa informaatiojoukkoon J'_n , jos $F(Y_{n+1} | J_n) = F(Y_{n+1} | J'_n)$.

Tässä F on vektorin Y_t jakautumafunktio. Määritelmä siis sanoo, että tieto $X_{t:n}$:stä ei muuta tietoa $Y_{t:n}$:stä, toisin sanoen ennustekyky suhteessa $Y_{t:n}$:hen säilyy muuttumattomana, kun tieto $X_{t:n}$:stä lisätään informaatiojoukkoon. Syy puolestaan määritellään tässä kontekstissa seuraavasti:

Määritelmä: Olkoon $J'_n = \Omega_n$, ts. $X_{t:n}$ arvoilla täydennetty informaatiojoukko kattaa kaiken tiedossa olevan. Jos nyt

$$F(Y_{n+1} | \Omega_n) \neq F(Y_{n+1} | \Omega_n - X_n), \text{ niin } X_n \text{ on } Y_{n+1:n} \text{ syy.}$$

Hieman tiivistettynä Granger-kausaliiteetissa tietty vektori-arvoinen muuttuja on toisen vektori-arvoisen muuttujan syy, jos tieto edellisen muuttujan aiemmista arvoista muuttaa jälkimmäisen muuttujan myöhempien arvojen ennusteita. Pelkistäen syyt muuttavat seurauksien esiintymistodennäköisyyttä. Lisäksi syyt edeltävät seurauksia.

3. Disjunkttiivinen multipelikausaatio

Suppesin (1970) probabilistinen kausaliteettiteoria perustuu varsin samanlaisiin periaatteisiin kuin Granger-kausaliteetti. Suppesin mukaan aidot syyt löytyvät prima facie -syiden joukosta näennäissyiden eliminoinnin jälkeen. Mikä hyvänsä tapahtuma C voi olla tapahtuman E prima facie -syy, jos seuraavat ehdot ovat voimassa:

- (i) C edeltää ajassa E:tä,
- (ii) C:n esiintymisen todennäköisyys on nolaa suurempi, ja
- (iii) $P(E|C) > P(E)$ tai $P(E|C) > P(E|e_i-C)$.

Jotta prima facie -syy C olisi E:n aito syy, pitää sulkea pois se mahdollisuus, että C on E:n näennäissy. Näennäissy lisää seurauksen esiintymistodennäköisyyttä, toisin sanoen ehdot (i)–(iii) ovat voimassa, mutta C:tä edeltävän tapahtuman B huomioon ottaminen tekeekin C:stä ennustemielessä irrelevantin. Ehto (iii) Suppesilla vastaa siis Grangerin täyden informaation oletusta kausaliteetin määrittelyssä.

Kausaliteetin kytkeminen ennustettavuuteen onkin tärkein Grangerin ja Suppesin näkemyksiä yhdistävä piirre. Intuitiivisestikin syyllä on seurauksen todennäköisyyttä lisäävä luonne: jos tapahtuma tyyppiä C aiheuttaa tapahtuman tyyppiä E, niin on jotenkin luonnollista ajatella, että C-tyypin tapahtuman esiintyminen lisää E-tyypin tapahtuman todennäköisyyttä. Toisin sanoen kausaatiosta voidaan päätellä korrelaatio, vaikka käänteinen päätelmä yleensä torjutaan. Edellinenkin päätelmä, siis Grangerin ja Suppesin näkemyksiä heijastava päätelmä, on kuitenkin helppo osoittaa ongelmalliseksi jo melko yksinkertaisissa asetelmissa.

Tarkastellaan yhteiskuntatieteissä varsin yleistä tutkimusasetelmaa, jossa ollaan kiinnostuneita tietyyppisten ilmiöiden, esimerkiksi yhteiskunnallisten levottomuuksien, syistä. Merkitään tämäntyyppisiä ilmiöitä E:llä. Yhteiskuntatieteissä on tavallista, että laaditaan malleja tutkimuskohteena olevien ilmiöiden taustamuuttujien välisistä suhteista. Nämä mallit, jotka usein koostuvat kausaalisuhteita koskevista hypoteeseista, laaditaan siten, että niistä voidaan vetää empiirisiä tunnuslukuja tai niiden suhteita koskevia päätelmiä. Näitä sitten verrataan empiirisen aineiston pohjalta laskettuihin tunnuslukuihin tarkoituksena selvittää, miten hyvin aineistosta lasketut

tunnuslukusuhteet vastaavat mallista vedettyjä päätelmiä. Käytetyimpiä tunnuslukuja ovat korrelaatiokertoimet tai suhteelliset ehdolliset frekvenssit sekä niiden keskinäissuhteet.

Usein yhteiskunnalliset ilmiöt ajatellaan johtuviksi useista erilaisista syistä. Niinpä yhteiskunnallisten levottomuuksien voidaan ajatella olevan seurausta hallitusten kansalaisvapauksia rajoittavista repressiivisistä toimista. Toisaalta Gurrin (1970) varsin tunnetun näkemyksen mukaan levottomuudet voivat myös seurata taloudellisen kehityksen ja väestön hyvinvointiodotusten välisen kuilun kasvamisesta. Näin esimerkiksi silloin, kun pitkäkö vaurastumisen kausi keskeytyy esimerkiksi vientiteollisuuden keskeisten tuotteiden kysynnän romahdettua. Näkemys sopii esimerkiksi disjunkttiivisesta multippelikausaatiosta: on olemassa useampia vaihtoehtoisia riittäviä ehtoja samalle seuraukselle.

Jo ennen Grangerin ja Suppesin töitä Nowak (1960) tarkasteli disjunkttiiviseen multippelikausaatioon liittyviä päättelyongelmia. Omakohtaiseen kokemukseen perustuen Nowak tarkasteli asetelmaa, jossa tutkimuspopulaation muodostivat asevelvollisuusikäiset puolalaiset miehet toisen maailmansodan aikana. Selitettävänä seurausmuuttujana E oli kuolemanriski. Osallistuminen sotatoimiin R sisälsi luonnollisesti tietyn kuolemanriskin $P(E|R)$. Toisaalta tuolloin tuberkuloosi oli paljon nykyaikaa vakavampi sairaus ja oli syynä monien asevelvollisuusikäisten vapauttamiseen asepalveluksesta ja siirtämiseen hoitoon asutuskeskusten ulkopuolella sijaitseviin parantoloihin. Tuberkuloosi ei suinkaan tuolloinkaan ollut ehdottoman tappava sairaus, mutta tiettyjen potilastyypien D kohdalla sekin johti kuolemaan. Merkitsemällä tuberkuloosiin sairastumista C :llä, voimme laatia yksinkertaisen mallin selittämään kuolemanriskiä asevelvollisuusikäisten puolalaisten miesten keskuudessa toisen maailmansodan aikana. Mallissa esitetään yllä mainitut kaksi kuolemaan johtavaa mekanismia: yhtäältä tuberkuloosiin sairastumisesta ja toisaalta sotatoimiin osallistumisesta johtuvat. Oletetaan, että malli vastaa todellisuutta. Tämä on tietenkin rohkea oletus, mutta teemme sen vain selvittääksemme, päteekö Grangerin ja Suppesin näkemys tämän tyyppisissä tilanteissa.

Olemme erityisesti kiinnostuneita siitä, lisääkö oletetun syyn C (tuberkuloosiin sairastuminen) läsnäolo aina oletetun seurauksen E (kuoleman) todennäköisyyttä. Haluamme siis selvittää, onko todella niin, että

$$(1) P(E|C) > P(E|ei-C)$$

kuten Granger ja Suppes väittävät.

Tarkastelemme ensinnä epäyhtälön (1) vasenta puolta. Se voidaan jakaa seuraaviin osiin:

$$P(E|C) = P(D|C) + P(R|C) - P(D \& R|C)$$

Oikea puoli taas voidaan kirjoittaa seuraavasti:

$$P(E|ei-C) = P(R|ei-C)$$

koska milloin C ei esiinny, ajatellaan ainoan mekanismi E:lle johtuvan R:stä. Onko siis (1) aina voimassa? Ei suinkaan, sillä D:n ja C:n korrelaatio voi olla hyvinkin matala, jos vain suhteellisen harvat tuberkuloosiin sairastuvat kuolevat tautiin. Lisäksi R:n ja C:n korrelaatio on lähes nolla, koska vain ani harvat tuberkuloosiin sairastuvat osallistuvat sotatoimiin. Toisaalta sotatoimiin osallistuminen oli hyvin yleistä tuberkuloosiin sairastumattomien joukossa, joten (1):n vasen puoli saattoi olla suuri ja erityisesti paljonkin suurempi kuin oikea puoli. Näin ollen syyn esiintyminen ei aina lisää seurauksen ennustettavuutta. Tässä esimerkissä siis voitaisiin päätyä ajattelemaan, että tuberkuloosi ei ole syytekijä kuolemantapauksissa. Päinvastoin voitaisiin väittää tuberkuloosiin sairastumisen vähentävän kuolemanvaaraa tietyissä väestöryhmissä.

4. Simpsonin paradoksi

Nowakin esimerkin tyyppisiä varauksia kausaaliiteetin ja ennustettavuuden yhteyteen oli esitetty jo ennen 1960-lukua. Ehkä parhaiten tunnettu aiempi tulos tunnetaan Simpsonin paradoksin nimellä (Simpson 1951). Sen ytimenä on havainto, jonka mukaan kahtia ositetussa populaatiossa voidaan havaita samanmerkkinen korrelaatio kahden muuttujan välillä molemmissa osapopulaatioissa, mutta kun tarkastellaan populaatiota kokonaisuutena korrelaatio häviää. Tai toisin päin katsottuna: populaatiossa havaittu riippumattomuus kahden muuttujan välillä voi muuttua korrelaatioksi

molemmissa tietyn osituksen mukaisissa osapopulaatioissa. Simpsonin oma esimerkki koskee kuvakorttien (sotilas, kuningatar, kuningas) ja ei-kuvakorttien (numerot 1–10) jakautumaa korttipakassa. Lähtökohtana on korttien jakautuma ”reilussa” pakassa. Luvut on ilmoitettu todennäköisyyksinä, so. jaettu korttien lukumäärällä.

	<i>kuvakortteja</i>	<i>ei-kuvakortteja</i>
<i>punaisia (hertoja, ruutuja)</i>	6/52	20/52
<i>mustia (patoja, ristejä)</i>	6/52	20/52

Taulukko 1: Korttien jakautuma reilussa pakassa

Tarkastellaan nyt toista korttipakkaa, jossa on täsmälleen samat kortit kuin reilussa pakassa, mutta tässä pakassa kortit jakautuvat lisäksi kahteen ryhmään: puhtaisiin ja likaisiin kortteihin. Kyseessä on taaskin ositus, toisin sanoen kukin kortti on joko puhdas tai likainen, mutta ei molempia. Tutkimuksen kohteena olevassa pakassa korttien jakautuma on seuraava:

	<i>tahriintuneita</i>		<i>puhtaita</i>	
	<i>kuvia</i>	<i>ei-kuvia</i>	<i>kuvia</i>	<i>ei-kuvia</i>
<i>punaisia</i>	4/52	8/52	2/52	12/52
<i>mustia</i>	3/52	5/52	3/52	15/52

Taulukko 2: Korttien jakautuma tarkastellussa pakassa

Tahriintuneiden korttien joukossa ei-kuvakorttien suhde kuvakortteihin punaisten joukossa on $8/4 = 2$. Mustien tahraantuneiden joukossa suhde on $5/3 = 1.67$. Puhtaiden korttien joukossa suhteet ovat $12/2 = 6$ punaisten ja $15/3 = 5$ mustien korttien joukossa. Siis molemmissa osajoukoissa (tahraantuneet, puhtaat) on assosiaatio kuvakorttisuuden ja punaisuuden välillä. Punaiset kortit näyttävät olevan todennäköisemmin ei-kuvakortteja kuin mustat kortit.

Tässä esimerkissä meillä on oikea tieto kausaalimekanismista, tai pikemminkin sen puutteesta: kortin punaisuus ei tee kuvakortin esiintymistä sen todennäköisemmäksi kuin sen mustuus. Osapopulaatioissa havaittu korrelaatio on kausaalisesti näennäinen. Itse asiassa kausaliteetilla ei ole mitään tekemistä näiden assosiaatioiden kanssa.

Asia saa kuitenkin uuden käänteen, jos annamme riveille ja sarakkeille uuden tulkinnan. Ajatellaan, että 'tahruntuneet' ovat 'miehiä' ja 'puhtaat' 'naisia' ja että kyse on tietystä potilaspopulaatiosta. Oletetaan edelleen, että kuvakorttien asemesta tarkastelemme niitä populaation jäseniä, jotka eivät ole saaneet sairauteensa tiettyä hoitoa, kun taas ei-kuvakorttien asemesta tarkastelemme niitä, jotka ovat saaneet tuota hoitoa. Taulukon ylemmällä rivillä on niiden henkilöiden lukumäärä, jotka ovat hengissä tietyn ajan kuluttua hoidon alkamisesta, ja alemmalle rivillä mainittuun ajankohtaan mennessä kuolleiden populaation jäsenten lukumäärä. Luvut taulukossa ovat täsmälleen samat kuin edellisessä taulukossa.

	<i>miehet</i>		<i>naiset</i>	
	<i>ei hoitoa</i>	<i>hoitoa</i>	<i>ei hoitoa</i>	<i>hoitoa</i>
<i>hengissä</i>	4/52	8/52	2/52	12/52
<i>kuolleita</i>	3/52	5/52	3/52	15/52

Taulukko 3: Hoidon tehokkuus

Suhteet ovat siis samat kuin korttipakkaesimerkissä, mutta erilainen johtopäätös vaikuttaisi mahdolliselta, jopa uskottavalta. Sekä nais- että miespotilaiden populaatioissa näyttää vallitsevan assosiaatio hoidon ja eloonjäämistodennäköisyyden välillä. Koska luvut ovat kuitenkin samat kuin edellisessä esimerkissä, kausaalipäätelyyn ei ole perusteita.

Miten Simpsonin paradoksi sitten syntyy? Tarkastellaan asiaa poliittiseen kilpailuun liittyvän esimerkin valossa. Usein kuulee arveltavan, että paljon rahaa kampanjointiin käyttävät ehdokkaat menestyvät paremmin vaaleissa kuin säästeliäämmät kilpailijansa. Ajatellaan siis hieman yksinkertaistaen, että runsas rahankäyttö kampanjoinnissa on vaalimenestyksen syy. Määritellään dikotominen muuttuja A omaisuudeksi tulla valituksi tarkasteltavissa vaaleissa, B puolestaan esittää ominaisuutta käyttää vähemmän kuin x euroa vaalikamppailuun. Vastaavasti ominaisuudet ei-A ja ei-B tarkoittavat, ettei ehdokas tule valituksi ja että ehdokas käyttää vähintään x euroa kampanjaansa. Maassa on kaksi vaalipiiriä. Ominaisuus C tarkoittaa, että ehdokas on ehdolla vaalipiirissä 1, ja ominaisuus ei-C sitä, että ehdokas on ehdolla vaalipiirissä 2.

Oletetaan, että

$$P(A|B) < P(A|ei-B)$$

Näin ollen enemmän rahaa käyttävät tulevat suuremmalla todennäköisyydellä valituiksi. Simpsonin paradoksi esiintyy nyt, jos

$$P(A|BC) \geq P(A|(ei-B)C) \text{ ja jos } P(A|B(ei-C)) \geq P(A|(ei-B)(ei-C))$$

Paradoksi koostuu siis havainnosta, jonka mukaan näyttäisi olevan assosiaatio suuren vaalibudjetin ja vaalimenestyksen välillä, mutta samalla osoittautuukin, että molemmissa vaalipiireissä pienen budjetin ehdokkaat omaavat suuremmat valituksen tuleminen mahdollisuudet.

Konstruktion ydin on $P(A|B)$:n ja $P(A|ei-B)$:n dekompositio:

$$P(A|B) = [P(C|B)]P(A|BC) + [P(ei-C|B)]P(A|B(ei-C))$$

$$P(A|ei-B) = [P(C|ei-B)]P(A|(ei-B)C) + [P(ei-C|ei-B)]P(A|(ei-B)(ei-C))$$

$P(A|B)$ on selvästikin painotettu keskiarvo $P(A|BC)$:stä ja $P(A|B(ei-C))$:stä. Samoin $P(A|ei-B)$. Jos painot $P(C|B)$ ja $P(ei-C|B)$ ovat samoja, ei paradoksia esiinny. Nämä painot ovat vastaavasti todennäköisyys sille, että alle x euroa kampanjointiin käytävä on ehdolla vaalipiirissä 1 ja 2.

5. Voiko kausaatio olla probabilistista?

Vuoden 2008 dialogissamme professori Koistinen torjuu mahdollisuuden aidosti probabilistisesta kausaliteetista. Toisin sanoen hän katsoo, että syyn vallitessa tulee seurauksen esiintyä poikkeuksetta. Oma kantani tähän on agnostisempi, mutta sillä ei tässä ole merkitystä. Olennaisempaa on todeta, että Simpsonin ja Nowakin havaintojen sanoma asettaa haasteen deterministiselle kausaliteetikäsitykselle. Jotta voisimme sanoa seurauksen aina seuraavan syystä, pitää meidän tehdä pitkälle meneviä oletuksia maailmasta. Yleensä nämä oletukset puetaan ceteris paribus -varauksien muotoon, jotka usein ”unohdetaan” sovellutuksissa. Äärimmilleen vietyinä nämä varaukset voivat esiintyä kaikkietävyyttä hipovina informaatiojoukko-oletuksina, kuten Granger-kausaliteetissa. Muun muassa näiden varauksien

vuoksi empiirisesti orientoitunut kausaliiteettitutkimus usein tyytyy etsimään evidenssiä kausaalisuhteiden olemassaolosta suhteellisista frekvensseistä ja niiden varioinnista. Tämä etenemistapa on täysin yhteen sovitettavissa deterministisen kausaliiteetikäsityksen kanssa: vaikka empiirisissä tarkasteluissa arvelut syytekijät eivät aina johdakaan arveltuihin seurauksiin, voidaan tämän katsoa johtuvan siitä, ettemme tarkoin tunne niitä olosuhteita, jotka yhdessä syytekijäin kanssa muodostavat riittävän ehdon seurauksille. Näin voimme ajatella deterministisen kausaliiteetin saavan probabilistisen asun siksi, ettemme tiedä tarpeeksi kontekstitekijöistä. Fysiikassa tämä näkemys lienee ollut niin sanotun piilomuuttujateorian taustalla. Empiirisen tutkimuksen kannalta deterministisen ja probabilistisen näkemyksen ero on lähinnä siinä, että jälkimmäiseen perustuva tutkimus voi tyytyä ennustettavuuteen, joka ilmenee litteämpinä todennäköisyysjakautumina seurausten suhteen kuin edellinen.

Turun yliopisto

Kirjallisuus

- Cartwright, N. (1979). "Causal Laws and Effective Strategies", *Noûs* 13, 419–437.
- Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross-spectral Methods", *Econometrica* 37, 424–438.
- Granger, C. W. J. (1980). "Testing for Causality. A Personal Viewpoint", *Journal of Economic Dynamics and Control* 2, 329–352.
- Gurr, T.R. (1970). *Why Men Rebel*. Princeton: Princeton University Press.
- Koistinen, O. ja Nurmi, H. (2008). "Yhteiskunnalliset lait ja kausaliiteetti", *Ajatus* 65, 7–23.
- Nowak, S. (1960). "Some Problems of Causal Interpretation of Statistical Relationships", *Philosophy of Science* 27, 23–38.
- Russell, B. (1913). "On the Notion of Cause", *Proceedings of the Aristotelian Society* 13, 1–26.
- Simpson, E. (1951). "The Interpretation of Interaction in Contingency Tables", *Journal of Royal Statistical Society B*: 13, 238–241.
- Suppes, P. (1970). *A Probabilistic Theory of Causality*. Amsterdam: North-Holland.
- Wold, H. O. (1964). "On the Definition and Meaning of Causal Concepts", teoksessa H. O. Wold (toim.), *Model Building in the Human Sciences*. Monaco: Union Européenne d'Éditions, 265–295.
- von Wright, G. H. (1971). *Explanation and Understanding*. Ithaca: Cornell University Press.

Metafilosofinen naturalismi ja antinaturalismi uuden ajan filosofiassa

Henri Pettersson

Olli Koistinen, Turun yliopiston teoreettisen filosofian professori ja ilokseni myös opettajani jo vuosien ajan, ei tämän niteen sivuilla esittelyjä kaipaa. Mutta kuten tällaisen retorisen aloituksen jälkeen hämmentävän usein nimenomaan tapana on, kehunpa vielä kuitenkin: Ollin asiantuntemus ja tutkimusansiot teon filosofian, uskonnonfilosofian sekä etenkin uuden ajan suurfilosofien parissa ovat sanalla sanoen *kiistattomat* millä tahansa mittarilla tarkasteltuna. Professorin virkaroolissaan Olli on väsymättömän kannustava ohjaaja opiskelijoille opintojen kaikissa vaiheissa. Toisaalta hän on myös innostava kollega työtovereilleen – ei vähiten esimerkkinsä kautta. Suureksi osaksi juuri näiden Ollin työn eri vaikutustapojen kautta Turun yliopiston filosofian oppiaine lyö nyrkkeilytermein ”kokoaan isommin” ja on arvostettu kansainvälisestikin korkealle muun muassa uuden ajan filosofian tutkijoistaan ja tutkimuksistaan. Ollin filosofisessa temperamentissa yhdistyy mielenkiintoisella tavalla kaksi luonteenpiirrettä, jotka vaikuttavat olevan keskenään vastakkaiset. Olli nimittäin suhtautuu kokemuksiini mukaan filosofiaan aina huumorilla poikamainen pilke silmäkulmassa, mutta samanaikaisesti kuitenkin kuin se olisi tärkein asia koko maailmassa. Tämä tasapainoinen suhtautumistapa oli minulle laitoksella työskennellessäni aina tärkeä esikuva, sillä etenkin tutkijanurani alkupuolella otin akateemisen ammatin ja siihen liittyvän lakkaamattoman stressin aivan liian henkilökohtaisesti. Keskustelut Ollin kanssa auttoivat lähestymään filosofin työtä oikealla vakavuudella, mistä olen kiitollinen.

Kuten todettua, uuden ajan filosofian historiallinen tutkimus on Ollin työn ansiosta Turun teoreettisen filosofian oppiaineessa korkealla suoritustasolla. Näistä ympäristötekijöistä huolimatta filosofia kutsui minua Turun yksikköme yleisestä profiilista poikkeavien kysymysten pariin: vaikka opiskelin yliopistossa filosofian rinnalla historiaa sivuaineena ja työskentelen kirjoitushetkellä lukiossa tämän oppiaineen opettajana, en ole tuntenut

mielestäni riittävää sisäistä paloa filosofian historiaan kohdistuvaan tutkimukseen. Eksyinkin jo varhain opinnoissani marginaalisempien filosofisten aiheiden pariin piittaamatta kaiken innostukseni keskellä siitä, löytyikö laitokseltamme niille halukasta ohjaajaa (ja tajuamatta omaa parastani – eli siis sitä, että toisenlaisille aiheille *olisi* tarjolla maailmanluokan ohjaaja). Olli oli kuitenkin aina valmis ennakkoluulottomasti ohjaamaan omalaatuisia tutkimuksiani, vaikkeivät ne ehkä vastanneetkaan hänen vahvuusalueitaan – ja tuottivat jo valmiiksi kuormitetulle professorille tätä kautta ylityötä. Ajoittain Olli hienovaraisesti osoitti minulle keskusteluissamme hedelmällisiä kiinnekohtia käsittelemistäni kysymyksistä uuden ajan filosofien ajatuksiin sekä lainasi virikkeellistä luettavaa kirjajhyllystään, mutta pysyin useimmissa tapauksissa ”historiattomissa” tarkastelutavoissa. Olli on suhtautunut tähän jukuripäisyyteen aina niin ymmärtävästi kuin vain on mahdollista, eikä hän asettanut nuoren tutkijanalan filosofiselle uteliaisuudelle rajoja tai ehtoja.

Tämä *Festschrift* tarjoaa oivallisen tilaisuuden ilmaista mittaamatonta kiitollisuuttani Ollin jo vuosien ajan antamaa opetusta ja ohjausta kohtaan kirjoittamalla hänelle läheisistä teemoista, kuten uuden ajan filosofiasta. Tosin tälläkin kertaa lähestyn historiaa omien filosofisten intressieni kautta. Tarkemmin sanoen haluan palata erääseen kysymykseen, jonka Olli on monesti ottanut esiin käymissämme keskusteluissa. Väitöstutkimuksessani suhtaudun varauksellisen myönteisesti näkemykseen, joka tunnetaan *metafilosofisena naturalismina*. Vaikken välttämättä ole vielä valmis tulemaan ulos kaapista naturalismin kannattajana, tuntuu se minusta olevan aihepiirissään kaikista ongelmallisista näkökannoista *vähiten* ongelmallisin (kuten filosofisen teorianvalinnan kohdalla niin usein vaikuttaisi käyvän). Ollin sympatiat tuntuivat puolestaan olevan tässä asetelmassa enemmän jossakin vastakkaisen *antinaturalismin* puolella, joten saimme Ollin kanssa keskusteluissa hedelmällistä jännitettä aikaiseksi Ollin naturalismiin kohdistuneen yleisen kummastelun ja tarkkanäköisen kritiikin kautta. Ollia arvelutti etenkin se, mikä tulisi olemaan filosofian työnkuva naturalisessa kokonaiskuvassa ja tähän liittyen myös se, millaiseksi muodostuu filosofian akateeminen kohtalo. Pysin antamaan kirjoituksessani tähän huoleen ainakin jonkinlaista selvitystä katsomalla, ettei se, mitä naturalistit tarkoittavat ihanteellisella filosofoinnilla, välttämättä ole erityisen vallankumouksellinen kanta länsimaisen filosofian

historian isossa kuvassa. Tarkastelun aluksi on kuitenkin paikallaan lyhyesti kerrata, mitä metafilosofisella naturalismilla ja antinaturalismilla oikein tarkoitetaan.

1. Naturalismi ja antinaturalismi metafilosofiassa

Nimeä ”naturalismi” käytetään filosofiassa lukemattomissa yhteyksissä, mutta metafilosofisessa¹ asussaan naturalismi on *filosofiakäsitys*, joten se esittää naturalistisilla painotuksilla sävytetyn kokonaiskuvan koskien filosofoinnin kohdetta, menetelmiä sekä päämääriä. Naturalismi ei kuitenkaan ole mikään tarkkarajainen tai selkeiksi teeseiksi niputettavissa oleva doktriini, vaan paremminkin yleistasonen ”skeema” tai viitekehys erilaisille mahdollisille muunnoksille tästä perusajatuksesta. Kenties pääpiirteisoin mahdollinen naturalismille annettavissa oleva kiteytys onkin se, että kyseessä on filosofinen asennoitumistapa, joka painottaa tavalla tai toisella filosofisen tutkimuksen saumatonta yhteenkuuluvuutta kaiken muun tieteellisen tutkimuksen kanssa.

Kiistan vastakkaisella laidalla ”antinaturalismi” ei poimi alleen mitään yksittäistä positiota, sillä kyseinen nimilappu lassoaa yhteen eräänlaisen ”kaatoluokan”, jonka jäsenten välisenä pienimpänä yhdistävänä nimittäjänä on oikeastaan vain periaatteellinen metafilosofisen naturalismin hylkääminen ja filosofian omalaatuisten luonteen painottaminen.² Eri antinaturalistisilla filosofiakäsityksillä onkin suuresti vaihtelevia filosofis-metafilosofisia perusteluja ja motiiveja tekemälleen tieteiden ja filosofian erottamiselle. Tämän lisäksi antinaturalistit eroavat toisistaan siinä, millaiseksi he lopulta ajattelevat antinaturalistisen filosofian kohteen, menetelmän ja tavoitteet. Tästä syystä he eivät myöskään näe filosofian autonomis-autoritääristä roolia samalla tavalla. Eräiden radikaalimpien antinaturalistien visioissa filosofian laadullisella erityisyydellä on myös jyrkempiä implikaatioita filosofian ja erityistieteiden marssijärjestyksen suhteen. Heille antinaturalistinen filosofia ei ole poikkeuksellinen tutkimusalojen joukossa yksin autonomisuudessaan, vaan se voi nousta autonomiansa nojalla arvovallaltaan jopa jonkinlaiseksi

¹ Metafilosofia on filosofian osa-alue, joka pohtii filosofoinnin kohteisiin, menetelmiin ja tavoitteisiin liittyviä aiheita käyttäen tässä apuna filosofian omia näkökulmia. Vastaukset näihin metafilosofisiin kysymyksiin tuottavat *filosofiakäsityksiä*, eli filosofisesti perusteltuja kokonaiskuvia siitä, mitä, miten ja miksi filosofiaa tehdään.

² ’Antinaturalismi’ on jo nimenä selvästi naturalismille muodostettu antonyymi.

ensimmäiseksi filosofiaksi (*prima philosophia*) tai tieteiden kuningattareksi (*regina scientiarum*), jolloin filosofialla on myös ainakin jossain määrin auktoriteettia tieteisiin nähden. Antinaturalistinen filosofia voi hyödyntää tätä määräysvaltaansa kommentoidessaan ja kritisoidessaan erityistieteiden tutkimuksia niiden ulkopuolisesta ja myös jossain tietoteoreettisessa mielessä niitä vakaammasta ”Arkhimedeen pisteestään”. Eräät antinaturalistit nostavatkin filosofian *foundationalistiseksi* oppialaksi, joka tarjoaa erityistieteiden ulkopuolelta perusteet niiden keskeisille käsitteille, argumentaatiomuodoille ja evidenssilajeille.

Naturalistit suhtautuvat nuivasti tähän antinaturalistien metafilosofiseen ajatusmalliin, jossa filosofia asettuu jollain poikkeuksellisella mandaatilla tieteiden katraassa itsenäisenä yksikkönä erilleen muista todellisuutta tutkivista aloista – ehkäpä jopa jonkinlaiselle empiirisiä tieteitä edeltävälle tai niitä perustavammalle tasolle, jolla se setvii omilla ehdoillaan kyseenalaistamattomana auktoriteettina ikiomia filosofisia ongelmiaan ja esittää niitä koskevia empiirisestä evidenssistä riippumattomia väitteitään. Naturalistien näkemyksessä filosofinen tutkimus toimii läheisessä ja vuorovaikutteisessa yhteistyössä samojen tutkimuskysymysten kanssa omilla tahoillaan toimivien erityistieteiden – fysiikka, kemia, biologia, kokeellinen psykologia ja niin edelleen – kanssa, joiden tutkimustulokset motivoivat, ohjaavat ja korjaavat filosofien kannanottoja.

Naturalistien ja antinaturalistien tulkintaeroilla filosofian luonteesta on myös luonteivia institutionaalisia seurauksia koskien filosofian akateemista asemaa. Antinaturalistien visio kohottaa filosofian usein itsenäiseksi ja riippumattomaksi toimijaksi, joka takaa filosofialle korkea-arvoisen tehtävän tieteiden perheessä. Sen sijaan naturalistien runollisissa vertauskuvissa filosofian ja empiiristen erityistieteiden esittämien väittämien välillä ei vallitse mitään jyrkkää ”laadullista kontrastia”, joten filosofien ja tieteilijöiden omilla tahoillaan saavuttamat tulokset kutovat yhdessä otettuna eräänlaisen *holistisen* ”verkon”, jossa eri lähteistä tulevat tuotokset punoutuvat saumattomasti osaksi yhtä ja samaa tiedollista kokonaisrakennelmaa. Tähän tapaan ajattelussa kokonaisuudessa onkin monesti hankalaa rajata täsmällisesti missä yksi tutkimusala loppuu ja toinen alkaa. Niinpä eri tutkimushankkeiden – myös filosofia mukaan luettuna – välillä tehdyt rajanvedot eivät ole hengeltään mitään syvälle kiveen hakkautettuja periaatteellisia erotteluja, vaan suurelta

osin erilaisten pragmaattisten näkökohtien sanelemia työnjakoja sekä näille työnjaoille yliopistoissa ja muissa vastaavissa tutkimusyhteisöissä perustuvia institutionaalisia järjestelyjä. Naturalistien mukaan onkin pitkälti merkityksetöntä tehdä jokin mielenkiintoista tutkimusta ”filosofiana” filosofian laitoksella vai jonkin muun laitoksen alaisuudessa. Niinpä esimerkiksi Quinen (1974, 228) näkemyksen mukaan ”filosofian” ja ”matematiikan” kaltaiset otsakkeet ovat lähinnä käteviä apuvälineitä hallinnollisten askareiden kanssa painiville yliopistobyrokrateille ja taksonomisia erotteluja tekeville kirjastonhoitajille.

Aiheellinen kysymys, jonka myös Olli on keskusteluisamme usein ottanut puheeksi, on se, mikä on filosofian lopullinen kohtalo naturalistien maalaamassa kuvassa. Jos filosofia ja tieteet tutkivat samoja kysymyksiä, mutta tieteilijät tekevät kaiken ”oikean” työn, niin mihin filosofeja enää tarvitaan? Uhkaako filosofia jopa hävitä akateemisista instituutioista filosofian arvon vesittyessä? Entä mitä tapahtuu klassisille filosofisille kysymyksille, joita on pohdittu tämän ammattikunnan piirissä jo vuosituhansien ajan?

Pyrin seuraavaksi osoittamaan, ettei naturalistinen filosofiakäsitys ole merkittävä irtiotto länsimaisen filosofian historiasta, jossa vastaavia metafilosofisia ajatuksia on edustettu jo ennen oman aikamme nykynaturalisteja. Aihepiiriin liittyviä näkökohtia on nähdäkseni paikallaan puntaroida siltä varalta, ettei jonkinlaista ”traditionaalisuus-korttia” ainakaan käytettäisi sellaisena ennakkoluuloisena moitteena naturalistiselle filosofiakäsitykselle, jonka mukaan kyseessä olisi jotenkin jo lähtökohtaisesti ”epäortodoksinen” – ja siitä syystä jollain tapaa epäilyttävä – lähestymistapa filosofointiin. Sellaista vallankumouksellista metafilosofista positiota, joka määrittelee filosofoinnin luonteen täysin uudestaan, voidaan nimittäin pitää jo puheenaiheen vaihtamisena: millainen yhteys tällä toiminnalla enää on filosofiaan? Tarkoittaako tällaisen filosofiakäsityksen hyväksyminen todellisuudessa jo filosofian – sellaisena kuin se on totuttu tuntemaan – tappamista ja korvaamista jonkinlaisella toisella toiminnalla? Naturalistit kuitenkin näkevät filosofiakäsityksensä sellaisessa valossa, että se vastaa sitä kuvaa ihanteellisesta filosofoinnista, joka on vaikuttanut filosofian taustalla jo kauan ennen kuin mitään naturalistisen filosofiakäsityksen metafilosofisia rakenneteesejä on muotoiltu filosofisessa kirjallisuudessa julkilausutusti tai tätä nimeä käyttäen. Naturalistit haluavat siten edustaa jatkuvuutta näiden

länsimaisen filosofian varhaisempien metafilosofisten ihanteiden kanssa, ei radikaalia vallankumousta tai uudistusliikettä, jossa filosofian toiminta ajateltaisiin kokonaisvaltaisesti aiemmista filosofeista poikkeavalla tavalla.

2. Naturalismi ja uuden ajan filosofit

Pinnallisten *prima facie* mielikuvien tasolla antinaturalistinen ajatus filosofoinnista empiirisistä tieteistä poikkeavana toimintana on erittäin luonteva – tuntuisihan se sopivan mainiosti yhteen sen kanssa, miten filosofiaa päällisin puolin tyypillisesti harjoitetaan, eli aistihavaintoihin nojaavien koeasetelmien sijaan perushengeltään enemmänkin esimerkiksi matemaattisen tutkimuksen tapaisena ajatustyönä: filosofian laitosten seinien sisäpuolelta ei löydy tehokkaita teleskooppeja tai aikansa viimeisintä huipputekniikkaa edustavia hiukkaskiihdyttimiä, eivätkä filosofit myöskään tee antropologisia ekskursioita alkuasukasheimojen keskuuteen (filosofian historiassa edistysaskeleet eivät ole myöskään olleet samalla tavalla läheisessä yhteydessä teknologisiin edistysaskeleisiin ja uusien tutkimusvälineiden – kaukoputki, mikroskooppi, röntgen-säteet jne. – keksimiseen kuin luonnontieteissä). Tämän sijaan filosofian laitokset koostuvat kirja- ja paperikasojen täyttämistä tutkijankammioista, joissa ahkeroivien ammattifilosofien arki muodostuu tavallisesti sellaisesta nojatuolista käsin suoritettavasta pohdinnasta, jota voi toteuttaa sisäsiististi ”likaamatta käsiään”. Tämä kuvaus vastaa myös sitä, miltä filosofointi vaikuttaa sen historiasta esitetyissä teoksissa. Metodologinen ”ei-empiirisyys” nostetaan monesti yhdeksi filosofian tunnuspiirteeksi, josta poikkeaminen tarkoittaa jo astumista filosofian ulkopuolelle ja jonkin toisen tutkimusalan harjoittamista (Glock 2008, 96). Näiden metodologisten mielikuvien lisäksi kun filosofiaa luonnehditaan usein juuri kaikkein perimmäisimmäksi ja abstraktiotasoltaan yleisimmäksi oppialaksi, on sille luontevaa ajatella kuuluvan poikkeuksellinen autonomia ja auktoriteetti kyseisen erityistehtävänsä menestyksekkääseen hoitamiseen: jotta filosofia voi olla uskottavasti tieteiden kuningatar, on sen oltava jotenkin niiden yläpuolella ja niistä riippumaton.

Tällaisten mielikuvien yllyttämänä on kenties houkuttelevaa ajatella edelleen niin, että filosofia ja empiiriset erityistieteet eroavat myös jossain syvemmässä mielessä laadullisesti toisistaan, jolloin ne tulee tästä seikasta seuraten nähdä erillisinä hankkeina, joilla on kontollaan omat

tutkimuskohteensa ja tavoitteensa. Tätä taustaa vasten metafilosofinen antinaturalismi – muodossa tai toisessa – saattaa nyt näyttäytyä jopa jonkinlaiselta oletusarvoiselta filosofiakäsitykseltä, joka ilmentää filosofialle ominta henkeä ja näkökulmaa. Silloin tällöin kuuleekin puhuttavan traditionaalisesta filosofiakäsityksestä, jonka vallankumouksellista muutoshenkeä edustaneet ”antitraditionalistiset” tai ”jälkifilosofiset” naturalistit ovat haastaneet oikeastaan vasta viime vuosisadan puolivälistä eteenpäin. Samassa naturalistien vaikuttimet voidaan asettaa kyseenalaiseen valoon, sillä naturalismin nousukausi näyttää kulkevan käsi kädessä luonnontieteiden nousun kanssa. Naturalistit näyttäytyvät täten vain luonnontieteiden menestyskulun ihailijoina, jotka haluavat yhdistää filosofointinsa tähän positiivisen auran omaavaan ilmiöön.

Tällöin väitteenä on se, että antinaturalistiset linjaukset filosofian autonomiasta ja auktoriteetista suhteessa luonnontieteelliseen tutkimukseen ovat olleet filosofian historiassa valtavirtänäkemymiä. Esimerkiksi Bealer (1996, 1) on kuvaillut autonomia-teesin olleen filosofien keskuudessa ”dominoiva” näkemys ”läpi intellektuaalisen historiamme”. Se onkin hänen mukaansa syrjäytetty tältä jalustaltaan vasta 1900-luvun ”skientistisessä” ilmastossa. Niin ikään naturalisti Papineau (1993, 3) kutsuu samassa äänensävyssä antinaturalismia ”traditionaaliseksi asenteeksi” filosofointia kohtaan, joten tällaisia näkemymiä naturalismin ei-traditionaalisuudesta on esitetty naturalismi-kysymyksen rajalinjan kummaltakin puolelta. Naturalistit korostavat usein retoriikassaan tarpeettomasti filosofiakäsityksensä metodologista ja historiallista vallankumouksellisuutta. Motiivina on tällöin ilmeisesti halu erottautua selvemmin muista filosofisista koulukunnista ja perinteistä.³

Eräät filosofit ovat kuitenkin esittäneet tulkintoja, joiden mukaan jonkinlainen maltillinen naturalismi on ollut käytännössä filosofien keskuudessa suosituimpi lähestymissuunta filosofian ja tieteiden

³ Toisaalta monesti ”tieteelliseltä suunnalta” tulevat kielteiset – jopa pilkkaavat – kommentit filosofian hyödyllisyydestä ovat peräisin tieteilijöiltä, ei naturalistifilosoifeilta. Esimerkiksi fyysikot Hawking ja Mlodinow (2010, 5) aloittavat tuoreen teoksensa esittämällä lukuisia kosmologisia kysymymiä maailmankaikkeuden alkuperästä ja perimmäisestä luonteesta, joiden jälkeen he toteavat: ”Perinteisesti nämä kysymykset ovat kuuluneet filosofialle, mutta filosofia on kuollut. Filosofia ei ole pysynyt tieteen – etenkin fysiikan – nykyaikaisten kehitysten perässä. Tieteilijät ovat ottaneet kantaakseen tiedon etsinnän soihdun.” Kaikki tekstissä esiintyvät suomennokset ovat omiani, ellei lähdeluettelossa toisin mainita.

kanssakäymiseen. Kuten Kitcher (1992) on huomauttanut, olivat useat uuden ajan filosofian ehdottomat kärkinimet, kuten Descartes, Locke, Leibniz, Hume, Kant ja Mill, paitsi erinomaisesti perillä oman aikansa biologiassa, fysiikassa ja psykologiassa tehdystä tieteellisestä tutkimuksesta, niin he myös hyödynsivät ahkerasti näitä tutkimustuloksia liukuvasti osaksi omia filosofisia teorioitaan, ja yrittivät tyypillisesti myös sovittaa filosofiset teoriansa yhteen näiden tieteellisten näkemysten kanssa.⁴ Esimerkiksi Kantin ajatukset synteettisen *a priori* -tiedon sisällöstä ovat läheisessä yhteydessä newtonilaisen luonnonfilosofian oletuksien kanssa (Friedman 2004, 78–79). Tältä pohjalta voidaan myös spekuloida sillä, että jos Kant olisi tiennyt miten nykypäivän fysiikka ja matematiikka eroavat newtonilaisesta mekaniikasta ja euklidisestä geometriasta, olisi hän saattanut tehdä tämän tiedon pohjalta myös merkittäviä muutoksia näitä tiedonaloja sivuaviin näkemyksiinsä

Niin ikään monet niistä filosofista ongelmista, joita pidettiin näiden ajattelijoiden tahoilla ratkaisemisen arvoisena, syntyivät tieteellisen kehityksen sivutuotteena. Rosenbergin (2000, 8) mukaan mikään toinen ilmiö ei ole innoittanut filosofiä samassa määrin kuin tiede sen 1600-luvulla otettujen vallankumouksellisten edistysaskeleiden jälkeen. Esimerkiksi Galileon, Kopernikuksen ja Newtonin luonnonfilosofisten tutkimusten inspiroiva vaikutus uuden ajan filosofiin oli huomattava, samoin kuin myöhemmin 1800-luvulla Darwinin evoluutioteorian ja kehittyvän kokeellisen psykologian tuon ajan ajattelijoihin. Näin jälkikäteen katsottuna näiden filosofien tuotannoista on mahdollista nähdä miten ne ovat vahvasti tieteellisten aikakausiensa tuotteita, eikä niitä siten suinkaan kehitelty minkään ”filosofisen kuplan” sisällä. Esimerkiksi tieteellisen vallankumouksen aikakauden Descartesilla näyttäisi olleen ”konkreettiset” tieteelliset pyrkimykset mielessään filosofoidessaan. Descartes halusi nimittäin filosofisilla argumenteillaan syrjäyttää skolastisessa luonnonfilosofiassa keskeisellä sijalla olleen aristoteelisen teleologisen selitysmallin kausaalis-mekanistisella vaihtoehdolla (jota Descartes itse sovelsi tunnetusti muun muassa eläinten toiminnan kuvailussa) sekä samalla hylätä Aristoteleen tieteenfilosofiaan

⁴ Katso myös Quine 1984, 190–1; Keil 2008, 281–282; Raatikainen 2010, 165. Kornblith (1994, 49) kirjoittaa: ”Ajatus siitä, että filosofian tulee jollain tapaa perustua tieteisiin, ei ole mitenkään uusi, ja se on tarjonnut huomattavan määrän filosofisia ideoita. Descartes, Locke, Leibniz, Kant, Marx, Reichenbach ja lukemattomat muut pyrkivät osoittamaan, että heidän ideansa sopivat hyvin yhteen parhaan heidän käytössään olleen aikalais-tieteen kanssa.”

sisältyvän *hylomorfismin*, jossa kappaleiden käyttäytyminen ymmärretään niiden muodon ja materiaalisen koostumuksen kautta. Mainitut filosofit eivät myöskään olleet missään nimessä kapeasti koulutettuja tai fakkiintuneita yhden alan spesialisteja nykyaikaisessa mielessä, vaan tuottivat ”renessanssineroina” itsekin filosofointinsa ohessa merkittäviä tieteelliseksi luokiteltavia läpimurtoja (Glock 2008, 163). He toisin sanoen harrastivat erityyppisiä kokeellisia tutkimusasetelmia ja muita empiirisen tutkimuksen muotoja sekä esittivät erilaisia spekulatiivisia hypoteeseja. Tällainen tieteiden moniottelijana toimiminen oli tuohon aikaan vielä mahdollista, sillä toisin kuin nykyään, yhden oppineen mielen oli mahdollista sisäistää aikansa keskeiset tieteelliset saavutukset ja osallistua eturintamassa niiden kerryttämiseen.⁵

Kitcherin (1992, 55; ks. myös Friedman 1992, xi–xii) mukaan yksi mahdollinen syy näiden näkökohtien unohtumiselle tulee siitä ilmiöstä, että kun historiaa katsotaan ”jälkiviisaasti” taaksepäin nykynäkökulmasta käsin, sysätään menneiden filosofien (kenties nyt jo vanhentuneet) tieteelliset taustavaikuttimet helposti marginaaliin, samalla kun ”ajattomat” filosofiset osuudet nostetaan narratiiveissa etualalle. Samalla luodaan hieman vinoutunut rekonstruktio näiden filosofien ajatusmaailmasta.⁶ Lukuisat uuden ajan hahmot ovatkin jääneet elämään historiankirjoihin ennen kaikkea filosofeina, vaikka alun perin heidän maineensa saattoivat perustua matematiikan ja luonnontieteen aloilla tehtyihin oivalluksiin. Näiden näkökohtien sivuuttamisen kautta filosofeista on tehty jälkikäteen ”filosofisempia” kuin mitä he todellisuudessa olivatkaan, vahvistaen samalla mielikuvia metafilosofisen antinaturalismin traditionaalisuudesta.⁷

⁵ Vastavuoroisesti sen lisäksi että monet nykyisin ensisijaisesti filosofeina pidetyt hahmot puuhastelivat tieteellisten kysymysten parissa, käsiteltiin monen saman aikakauden luonnontieteilijän tuotannossa filosofisilta vaikuttavia ongelmia, kuten esimerkiksi Newtonin pohdinnat avaruuden metafysisen luonteen yhteydessä osoittavat.

⁶ Samaan tapaan kuin luonnontieteiden nerohahmoiksi kohotettujen tieteilijöiden erilaiset nykyvinkkelistä irrationaalisilta vaikuttavat uskonnolliset ja epätieteelliset – astrologia, alkemia jne. – harrastukset ja taustavaikuttimet ovat kiusallisia tahroja luonnontieteen historiassa, jotka helposti unohdetaan ”rationalisoivissa” historiikeissa.

⁷ Rortyn (1979, 132–133n2) mukaan filosofian historiasta kirjoitetuissa esityksissä näkyy selvä muutos Kantin jälkeen siinä, miten niissä rakennetaan filosofian menneisyydestä antinaturalistinen narratiivi, jonka mukaan filosofia on aina ollut empiirisistä tieteistä erillinen ja autonominen hankeensa.

Kitcherin tulkinnassa antinaturalistinen ajatusmalli ”puhtaan filosofian ideaalista” syntyi tietoisessa mielessä vasta Fregellä 1800-luvun loppupuolella, innoittaen sittemmin varhaisen analyttisen filosofian tärkeiden ydinhahmojen (Wittgenstein, loogiset positivistit, Oxfordin luonnollisen kielenkäytön filosofit ja niin edelleen) parissa ”antinaturalistista vallankumousta”, joka syrjäytti hetkellisesti aiemman naturalistisen tradition. Etenkin aikarajauksen 1930–60 väliin jäävät vuosikymmenet olivat analyttisen filosofian sisällä eräänlaista kulta-aikaa antinaturalismille, vaikka erilaiset antinaturalistiset ajattelijat ja koulukunnat esittivätkin erilaisia perusteluja antinaturalismille ja näkivät antinaturalistisen filosofian tehtävät eri tavoin (Glock 2008, 137). Vastaliikkeenä tälle antinaturalismin kultakaudelle analyttisessä filosofiassa tapahtui 1960-luvulta alkaen amerikkalaisen W. V. Quinen johdolla naturalistinen käänne, joka on palauttanut naturalismin vallitsevaksi metafilosofiseksi kannaksi – ainakin pohjoisamerikkalaisen analyttisen filosofian sisällä, sillä naturalismin suosio ei ole missään vaiheessa kasvanut yhtä suureksi Euroopan puolella (Hacker 2006, 231).⁸ Kitcher (1992, 52) toteaa tähän liittyen hieman provosoivasti, että nykyiset “naturalistit saattavat nähdä Fregen käynnistämän [antinaturalismin] suuntauksen vain ohimenevänä poikkeamana filosofian

⁸ Tosin näin tiiviit kertaukset filosofiakäsitysten asetelmien vaihtelusta yksinkertaistavat ja kärjistävät väkisin naturalistisen käänteen luonnetta. Naturalismin ja antinaturalismin valtakausien vaihtelut eivät tietenkään olleet todellisuudessa näin jyrkkiä, sillä filosofiassa vaikutti jo ennen 1960-luvulla tapahtunutta naturalistista käännettä merkittäviä (proto-)naturalisteja. Ensinnäkin: vaikka Fregen, Wittgensteinin ja Carnapin kaltaiset analyttisen filosofian tärkeät suunnannäyttäjät olivat filosofian luonnetta koskevilta ajatuksiltaan antinaturalisteja, voidaan näiden nimien vastapainoksi mainita naturalistisemmin viritetty Russell. Samaten vaikka Wienin piirin loogiset positivistit ja Oxfordin luonnollisen kielenkäytön koulukunnat olivat jäsenkuntiansa metafilosofisilta ajatuksilta valtaosin antinaturalistisesti painottuneita, oli niidenkin sisällä Neurathin ja Austinin kaltaisia oppositionimiä. Toisaalta jo ennen naturalistisen käänteen tapahtumista amerikkalaisessa filosofiassa oli olemassa naturalististen filosofien joukko (mm. R. W. Sellars, John Dewey, Ernest Nagel, George Santayana, Sidney Hook), joista osa jopa käytti tätä nimitystä itsestään. Osalla näistä filosofeista, kuten Deweylla, naturalismi kietoutui yhteen pragmatismien kanssa. Amerikkalaisessa filosofiassa oli tätä kautta siten jo valmiiksi olemassa naturalistinen pohjavire, joka nousi valtavirta-ajatteluksi samalla kun kansainvälisen filosofian painopiste siirtyi toisen maailmansodan jälkeen Euroopasta uudelle mantereelle. Loogisten positivistien ja muiden keskieuropalaisten filosofien ”aivovuoto” kansallissosialismin seurauksena toi heidän vankan tieteellisen painotuksen mukanaan Amerikkaan, joka yhdistyi nyt naturalismiin ja pragmatismiin. Viime vuosisadan alkupuolen Australiassa oli puolestaan vallalla realistis-materialistinen asennoituminen, joka tarjosi hedelmällisen kasvualustan naturalistiselle filosofiakäsitykselle. (Glock 2008, 38, 84–85, 135–136, 141, 254; Raatikainen 2010, 170–172.)

historiassa, filosofian todellisen tehtävän ja todellisten juurien hetkellisenä hylkäämisenä.”

Kitcherille näkemyksille vastakkaisesta kulmasta esimerkiksi Keil (2008, 281–2) toteaa, että pelkän tiedesensitiivisyyden ja empiirisen valveutuneisuuden nostaminen jonkun filosofin naturalistisuuden tunnuspiirteeksi on liian väljä kriteeri, jolloin ”naturalismi” lakkaa olemasta metafilosofisesti informatiivinen nimilappu poimiessaan alaansa oikeastaan kaikki uuden ajan filosofit. Mikään empiirisen tietämyksen täydellinen laiminlyönti ei voi olla tunnuspiirre antinaturalismille, sillä tällöin antinaturalisteiksi jäävät vain todella radikaalisti empiiriseen tietoon ja tieteisiin suhtautuvat filosofit. Fumerton (1999, 21, alaviite poistettu) kirjoittaa:

Ensivaikutelmalta näyttäisi selvästi siltä, että nykyfilosofia on kääntynyt julkilausutusti ja dramaattisesti poispäin siitä a priori ”nojatuolifilosofiasta”, johon lukuisat meistä analyttisen tradition jäsenistä kasvatettiin. Olisi tietenkin mahdollista väittää, että 1900-luvun painotus filosofiasta a priori oppialana oli itsessään vääristymä ja sivupolku alamme historiallisessa kehityksessä. Selvästikään Aristoteles, Descartes, Locke, Hume, Hobbes ja lukemattomat muut eivät aina erottaneet tiukasti filosofisia tutkimuksiaan fysiikan, biologian, sosiologian ja psykologian tutkimuksista. Filosofian historia olisi kuitenkin mahdollista esittää tarkemmin argumentoimalla, että menneisyyden jättiläiset olivat yksinkertaisesti kiinnostuneita useammasta kuin yhdestä tutkimusalasta. He eivät yrittäneet ratkaista filosofisia kysymyksiä tieteellisten tutkimusten avulla. Sen sijaan he olivat yksinkertaisesti kiinnostuneita lukuisista muistakin kysymyksistä filosofisten kysymysten ohella.

Kysymyksen arvioinnissa onkin otettava huomioon yksittäisten filosofien tieteellistä harrastuneisuutta ”filosofisemmin” heidän metafilosofiset ajatuksensa filosofisen tiedon luonteesta sekä tällaisen filosofisen tiedon suhteesta muihin tiedon lajeihin, ennen kaikkea empiiriseen tietoon. En tahdo kuitenkaan ottaa näihin kiistoihin kantaa näin yleisellä tasolla. Tämä edellyttäisi sekä menneiden filosofien metafilosofisten kannanottojen

perinpohjaista lähilukua että naturalismin kriteerien yksiselitteistä naulaamista. Varsinkin ensin mainitun tehtävän kohdalla loppuu oma asiantuntemukseni filosofian historian parissa, joten jätän pinnallisen tarkasteluni tältä erää tähän.

3. Kant antinaturalismin kantaisänä

Lopuksi tarkastelen sitä, millainen rooli Kantilla on yhtenä antinaturalismin monesti ehdotettuna kantaisänä. Kuten edellä todettiin, määrittyy antinaturalismi metafilosofisena positiona ensisijaisesti naturalismin negaationa ilman mitään yksityiskohtaista julkilausuttua kannanottoa filosofian luonteesta, sillä nämä julkilausutut kannanotot ovat eri antinaturalisteilla erilaiset. Antinaturalisteista ei muodostu mitään yksimielistä naturalismin vastaista rintamalinjaa, eivätkä antinaturalististen filosofiakäsitysten sukupuun oksat myöskään palaudu johonkin kaikille antinaturalisteille yhteiseen varhaisempaan kantamuotoon, jonka yksin pintapuolisissa seikoissaan toisistaan poikkeavia muunnelmia ne olisivat – vaikka eri antinaturalististen sukuhaarojen innoittajiksi on toki mahdollista nimetä filosofian historiasta merkittäviä yhdistäviä vaikuttajahahmoja.

Vaikutusvaltainen historiallinen roolimalli nykyisille antinaturalistisille filosofiakäsityksille on etenkin Kant, jonka tuotantoa hallitsevat metafilosofiset pohdinnat filosofian identiteetistä ja mahdollisuuksista. Kuten Koistinen (2009, 130) kirjoittaa, ”Kant oli huolestunut filosofian mahdollisuudesta, ja tämä sai hänet pohtimaan filosofian luonnetta perusteellisesti.” Voidaan jopa sanoa, ettei yksikään toinen filosofi ollut ennen Kantia keskittynyt ajattelussaan yhtä eksplisiittisesti ja huomattavassa määrin filosofian oman erityisperustan ja -tehtävän reflektointiin (Glock 1997, 287). Arvostettu Kant-tutkija Allison (1983, 25) onkin todennut, että Kantin ”[t]ranssendentaalista idealismia tulee luonnehtia ensisijaisesti metafilosofisena tai metodologisena ’positiona’, eikä suoraviivaisena metafyyssisenä doktriinina koskien ihmiskognition kohteiden luonnetta tai ontologista statusta.” Kuten Koistinen (2010, 119–120) kiteyttää, oli Kantin metafilosofinen visio leimallisesti antinaturalistinen:

Kant erotti filosofian selvästi muista tieteistä. Filosofiasa pyritään ymmärtämään tietävän subjektin panos maailman ja luonnon

luomiseen. Inhimillisellä kokemuksella on ennakkoehtonsa. Luonnon perustavimmat lait ovat ihmisen itsensä maailmaan asettamia. Kant ei näe ihmistä aivojen ohjaaman aluksen hämmentyneenä matkustajana, vaan moraaliseen arvokkuuteen ja autonomiseen toimintaan ainakin periaatteessa kykenevänä toimijana. Ihminen poikkeaa eläimistä ymmärryksensä ja järjensä suhteen; enkeleistä ruumiillisuuteensa perustuvien yllykkeiden suhteen. Ihmisellä on oma erityinen paikkansa, vaikka todellisuuden perimmäinen luonne pysyy hänelle arvoituksena.⁹

Katalyyttinä Kantin metafilosofisille pohdinnoille toimi osaltaan se 1600- ja 1700-lukujen kuluessa tapahtunut ja tieteenharjoittamisen ammatillistumiseen liittynyt ”ulkofilosofinen” kehitys, jossa tieteellinen kenttä jäsenyi yliopistoissa aiempaa selkeämmin erityistieteiksi, joilla oli nyt vastuullaan omat rajatut reviirinsä. Aikaisemmin kaikki yliopistolliset oppialat – kenties teologiaa, oikeustiedettä ja lääketiedettä lukuun ottamatta – kuuluivat (luonnon)filosofian nimen alaisuuteen. Samalla metodologinen keskustelu eri tieteenalojen tutkimuksellisista lähtökohdista lisääntyi, ja näin myös filosofiaan kohdistui painetta määrittää suhteensa tieteiden perheeseen. Luonnontieteiden ja empiirisen psykologian nousu asetti filosofian tässä suhteessa nyt puolustuskannalle, sillä ne uhkasivat ”kypsinä tietinä” omia itselleen ne kysymykset, jotka olivat aiemmin olleet filosofien kontolla. Samassa naturalismi-kysymys tuli ajankohtaiseksi metafilosofiseksi ongelmaksi, joka vaati filosoifeilta jonkinlaista kannanottoa. Tällä aikakaudella vaikuttaneella Kantilla muodostui tavallaan ensimmäistä kertaa selvä ammatillinen itsetietoisuus filosofian akateemisista tehtävistä.¹⁰ Rorty (1979, 131–132) kirjoittaa:

Ajatus siitä, että on olemassa erillinen ja autonominen oppiala nimeltään ”filosofia”, joka yhtäältä eroaa uskonnosta ja tieteestä sekä toisaalta on arvostelmissaan kummankin yläpuolella, on alkuperältään verrattain tuore. [...] Nykyaikainen filosofian ja tieteen välinen erottelu vakiintui vasta Kantin jälkeen.

⁹ Kattavamman kuvan Kantin filosofiakäsityksestä saa teoksesta Koistinen 2011.

¹⁰ Kantin roolista filosofian akateemisen identiteetin linjaamisessa ks. myös Glock 2008, 24–26.

Vaikka aikaisemmat filosofit – kuten Descartes – olivat puolustaneet filosofian itsenäisyyttä ja riippumattomuutta ulkopuolisten toimijoiden tekemisiltä, liittyivät nämä ajatukset Rortyn mukaan enemmän filosofian ja teologian väliseen suhteeseen, eikä niillä ollut vielä pitkälle meneviä seurauksia filosofian itseymmärryksen tai institutionaalisen aseman kannalta. Sen sijaan Kantin myötä filosofialle vakiintui oma paikkansa saksalaisissa yliopistoissa, ja Kantin maalaama filosofian ”omakuva” tuli myös osaksi filosofian professorien käsitystä siitä, mitä he ovat tekemässä.

Rortyn lavealla pensselillä maalaamien näkemysten historiallisesta tarkkuudesta voi olla monta mieltä, mutta kyseiset kiistanalaisuudet eivät kuitenkaan ulotu Kantin metafilosofiseen tärkeyteen. Esimerkiksi Glock (2008, 149) kutsuu koko naturalismi-kysymystä ”kantilaiseksi ongelmaksi”:

On olemassa yleisempi kantilainen ongelma, joka koskee sitä, että voidaanko filosofia ajatella empiirisistä tieteistä erilliseksi autonomiseksi tutkimusalaksi, ja mikäli näin on, niin millä perusteella tämä tapahtuu. Kyseinen ongelma on ollut esillä useiden niiden analyttisten filosofien kirjoituksissa, jotka ovat syventyneet metafilosofiseen reflektointiin.

Kantin nimi samaistetaan monesti myös kokonaisuudessaan antinaturalistisen metafilosofisen position kanssa. Esimerkiksi Glendinning (2002, 214–5) puhuu antinaturalismin yhteydessä ”jälkikantilaisesta traditiosta”, ja myös jaottelun toiselta puolelta naturalisti Papineau (lainattu teoksesta Glock 2008, 258) on maininnut samantyyllisen jaottelun (tosin Papineau puhuu jälkikantilaisten sijaan ”uuskantilaisista”):

[U]usi ja potentiaalisesti hedelmällisempi jako on kehitymässä englantia puhuvan filosofisen maailman sisällä. Vanhan analyttisen-mannermainen-jaon tilalla meillä on nykyään vastakkainasettelu naturalistien ja uuskantilaisten välillä. Naturalistit etsivät lähtökohtaa filosofoinnille tieteestä, siinä missä uuskantilaiset aloittavat tarkastelunsa tietoisuudesta.

Mitä tästä ”jälki-” tai ”uuskantilaisuuden” nimen käyttämisestä antinaturalismin synonyyminä tulisi ajatella? Puhetapa on nähdäkseni

epäonnistunut siitä syystä, että vaikka Kant onkin länsimaisen filosofian kehityshistoriassa kiistatta äärimmäisen vaikutusvaltainen hahmo, eivät kaikki antinaturalistit kuitenkaan hyväksy mitään ilmiselvästi *kantilaiseksi* luokitettavaa visiota filosofian tutkimusmenetelmästä, -kohteesta tai -tehtävistä. Ei, vaikka ”kantilaisuuden” kriteerit rajattaisiinkin mahdollisimman sallivasti ja Kantin vaikutuksen sallittaisiin olevan välittyntä toisten filosofien kautta. Osa antinaturalisteista nimittäin suhtautuu Kantin metafilosofisiin ajatuksiin jopa avoimen kielteisesti etsien metafilosofiset esikuvansa antinaturalismilleen muualta länsimaisen filosofian kaanonista. Edustavaksi esimerkiksi tässä kohdin voidaan poimia vaikkapa viime vuosikymmeninä suosiota saanut uusaristoteelisen metafysiikan suuntaus, jossa metafysiisiä aiheita lähestytään ”esikantilaisella”, mutta silti samalla julkilausutun antinaturalistisella otteella. Tahko (2012, 1) kuvailee tätä lähestymistapaa seuraavasti:

Nimi ”aristoteelinen metafysiikka” osoittaa sitoumusta siihen näkemykseen, että on olemassa tutkimusta, joka on erilaista kuin luonnontiede, ja joka edeltää sitä. Metafysiikka on ”ensimmäinen filosofia”, eli kaiken maailmaa koskevan filosofisen ja rationaalisen tutkimuksen ydin ja alku. Metafysiikan tehtävä ei ole palvella tiedettä tai avata käsitteellisiä sekasotkuja, vaan tutkia olevaa ja todellisuuden kaikkein perustavinta rakennetta kaikkein yleisimmällä tasolla. [...] Aristoteelinen metafysiikka voidaan asettaa [...] vastakkain kantilaisen metafysiikan kanssa: kategoriat ovat keskeisiä molemmille, mutta aristoteelisessa metafysiikassa nämä ovat olevan kategorioita, kun taas kantilaisessa metafysiikassa ne ovat ymmärryksen kategorioita.

Kantiin ja antinaturalismiin liittyen esille voidaan ottaa lisäksi toisesta suunnasta huomio, että jos antinaturalismia ei voi samaistaa ongelmattomasti jälkikantilaisuuden kanssa, ei vastaavasti jälkikantilaisuutta voi samaistaa sen paremmin täysin antinaturalismin kanssa. Näin siksi, että Kantin luonnostelemasta metafilosofisesta mallista esitettiin jo 1800-luvun keskusteluissa sellaisia Kantin omaa filosofiaa naturalistisempia tulkintoja, joissa hänen teorioitaan mielen aktiivisesta toiminnasta kokemuksen jäsentämisessä käytettiin empiirisinä hypoteeseina, joille haettiin sitten tieteellistä vahvistusta tuon ajan fysiologispsykologisen tutkimuksen suunnilta

(Glock 2008, 163). Myös omana aikanamme erilaiset filosofit (ks. esim. McDowell 1994) ovat yrittäneet luoda sopuisaa synteesiä Kantin transsendentaalifilosofiasta ja naturalismista, vaikka näiden hankkeiden ”naturalistisuuden” aitous on toisaalta kyseenalaistettu keskusteluissa.

Oli miten oli, jälkikantilaisuuden sisäpuolelle mahtuu joka tapauksessa selvästi myös monenlaisia naturalistisia alavirtauksia, jolloin jostain naturalismille vastakkaisesta ”jälkikantilaisuudesta” puhuminen uhkaa jättää liiallisena yksinkertaistuksena nämä asianhaarat huomiotta. En ylipäättäänkään näe mitään erityistä syytä sille, miksi antinaturalismista pitäisi väkisin puhua kantilaisuutena, kun huomattavasti yksinkertaisemmassa nimessä ”antinaturalismi” ei ole lähtökohtaisesti mitään vikana.

Näistä pienistä varauksista huolimatta Kantin merkitys metafilosofiselle naturalismikysymykselle ja antinaturalismin muodoille on kiistaton. Kant tunnetusti loi paljon uutta teknistä terminologiaa ja antoi nykyäänkin käytetyt merkitykset vanhoille ilmaisuille. Samassa hän ”sanoitti” myös naturalismin ja antinaturalismin välisen keskustelun tarjoamalla ne termit, joita eri naturalistisissa ja antinaturalistisissa filosofiakäsityksissä käytetään niiden peruspilarien paaluttamiseen (apriorisuus, analyttisyys, aleettiset modaliteetit jne.).

Kantin ja hänen seuraajiensa filosofinen työ heijastui vahvasti institutionaaliselle tasolle vakiinnuttaen filosofian akateemista asemaa. Kantin jälkeen saksaa puhuvan Euroopan vallitsevaksi suuntaukseksi tuli idealismi, johon sisältyi vahvasti antinaturalistisia painotuksia filosofian tehtävästä ja asemasta. Nämä näkemykset heijastuivat myös tiedepoliittisiin keskusteluihin, sillä monet idealistifilosofit ottivat kantaa filosofian asemaan kehittyvässä yliopistolaitoksessa.¹¹ Aihe oli ollut tärkeä jo Kantille, mutta häntä seuranneista filosofeista etenkin Fichte ja Schelling olivat tärkeässä asemassa puolustaessaan filosofian asemaa tieteiden kuningattarena erilaisten utilitarististen yliopistovisioiden sijaan. Preussilaisen Wilhelm von Humboldtin perustaessa vuonna 1810 Berliiniin uuden yliopiston, oli tämän hankkeen yhtenä taustaideologina juuri Fichte. Kun asiaa katsotaan tässä

¹¹ Yliopistojen tiedekuntien perinteisen arvojärjestyksen muuttuminen heijastui myös Suomeen, jossa filosofian arvo ja asema tulivat puheeksi pohdittaessa tieteenalojen hallinnollista jäsentämistä Helsingin Keisarillisen Aleksanterin Yliopiston sisällä. Filosofian professori Thiodolf Rein oli keskeinen hahmo tässä prosessissa, sillä hän jäseni filosofian oppiaineen uuteen yliopistoon ja oli perustamassa Suomen Filosofista Yhdistystä 1873.

valossa, ei olekaan sattumaa, että antinaturalistiset filosofiakäsitykset tulivat julkilausutuiksi juuri näihin aikoihin: samalla kun muut tiedonalat ammatillistuivat yliopistoissa, joutui filosofiakin etsimään omia rajojaan, arvoaan ja tarkoitustaan yliopistojen oppiaineiden joukossa. Antinaturalistisilla filosofiakäsityksillä oli tarjota yhdenlainen vastaus tähän haasteeseen ja osoittaa filosofialle oma reviirinsä akateemisessa maailmassa (Hamlyn 1992, 71–73, 82–83, 91–92). Collins (1998, 4) kirjoittaa: ”Idealistinen suuntaus kehittyi juuri oikeaan aikaan samalla kun kamppailut saksalaisten yliopistojen muokkaamiseksi olivat käytössä. Tämän seurauksena turvattiin filosofian laitosten autonomisuus ja syntyi nykyaikainen tutkimusyliopisto.”

Antinaturalistiset metafilosofiset positiot nähdäänkin monesti antinaturalistien retoriikassa välttämättöminä filosofian institutionaalisen itsenäisyyden turvaamisen kannalta. Pasnaun (2011, 92ff) tulkinnessa Descartesin antinaturalistinen filosofiakäsitys oli pelastanut filosofian myöhemmän saksalaisen idealismin tavoin muutamaa vuosisataa aiemmin kun uuden kokeellisen luonnontieteen vallankumous uhkasi pyyhkäistä filosofian skolastiikan mukana menneisyyteen kuuluvana ilmiönä. Nämä kommentit kuulostavat niiltä nykykeskusteluissakin esiintyviltä antinaturalistisilta huolilta, joiden mukaan naturalistinen filosofiakäsitys johtaa filosofian katoamiseen akateemiselta kartalta. Esimerkiksi Kant-spécialisti Hannan (2008, 192) mukaan filosofian kohtalonkysymyksenä on tällä hetkellä se, pystyykö filosofia kantilaisessa hengessä osoittamaan itsensä autonomiseksi oppialaksi, vai katoaako se pelkäksi erityistieteiden ”alaosastoksi”. Nähdäkseni tällaisiin ennakkoluuloisiin pelkokuviin ei kuitenkaan ole aihetta, sillä naturalistien tavoitteena ei ole tappaa filosofiaa – ei sen paremmin metafilosofisessa kuin institutionaalisessakaan mielessä. Yhtä lailla myös naturalistit voivat suhtautua ”nurkkapatrioottisesti” oman oppialansa akateemiseen edunvalvontaan.

Mitä nämä huomiot kenties osoittavat on se, että naturalistien ja antinaturalistien välisissä keskusteluissa esiintyy puolin ja toisin liikaa ennakkoluuloja ja olkiukkomaisia kärjistyksiä vastapuolen näkemysten sisällöstä ja tarkoituseristä. Niinpä aihepiirissä on tilausta tasapuoliselle selventävälle tutkimukselle ennen kuin tämä kiista pääsee yltymään aikamme filosofiassa aikaisempien suurten jakolinjojen veroiseksi ”kylmäksi sodaksi” –

kuten tapahtui 1900-luvulla analyttisen ja mannermaisen suuntauksen välillä. Kiinnostunut lukija löytää syvällisempää pohdintaa näistä kysymyksistä väitöstutkimuksestani. En kuitenkaan kehtaa päättää kirjoitustani näin häpeilemättömään mainokseen, joten toivotan Ollille vielä kerran onnea merkkipäivän johdosta!

Turun yliopisto

Kirjallisuus

- Allison, Henry E. (1983). *Kant's Transcendental Idealism – an Interpretation and Defense*. New Haven: Yale University Press.
- Aristoteles (1990). *Metafysiikka*. Suom. Tuija Jatakari, Kati Näätäsaari & Petri Pohjanlehto; selitykset laatinut Simo Knuuttila. Helsinki: Gaudeamus.
- Bealer, George (1996). "On the Possibility of Philosophical Knowledge", *Philosophical Perspectives* 10, 1–34.
- Collins, Randall (1998). *Sociology of Philosophies – A Global Theory of Intellectual Change*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Friedman, Michael (1992). *Kant and the Exact Sciences*, Cambridge, MA: Harvard University Press.
- Friedman, Michael (2004). "Philosophy as Dynamic Reason – The Idea of Scientific Philosophy", teoksessa Havi Carel & David Gamez (toim.), *What is Philosophy?* King's Lynn, UK: Continuum Publishing, 78–96.
- Fumerton, Richard A. (1995). *Metaepistemology and Skepticism*. Lanham, Maryland: Rowman & Littlefield.
- Glendinning, Simon (2002). "The Analytic and the Continental", teoksessa Julian Baggini & Jeremy Stangroom (toim.), *New British Philosophy – the Interviews*. Lontoo: Routledge, 201–218.
- Glock, Hans-Johann (1997). "Kant and Wittgenstein: Philosophy, Necessity and Representation", *International Journal of Philosophical Studies* 5 (2), 285–305.
- Glock, Hans-Johann (2008). *What is Analytic Philosophy?* Oxford: Oxford University Press.
- Hacker, P. M. S. (2006). "Passing by the Naturalistic Turn: On Quine's Cul-de-Sac", *Philosophy* 81 (316), 231–53.
- Hamlyn, David W. (1992). *Philosophy Becomes a Profession – The History of a Practice*. Lontoo: Routledge.
- Hanna, Robert (2008). "Kant in the Twentieth Century", teoksessa Dermot Moran (toim.), *The Routledge Companion to Twentieth Century Philosophy*. Lontoo: Routledge, 149–203.
- Hawking, Stephen & Mlodinow, Leonard (2010). *The Grand Design*. Lontoo: Bantam Press.
- Keil, Geert (2008). "Naturalism", teoksessa Dermot Moran (toim.), *The Routledge Companion to Twentieth Century Philosophy*. Lontoo: Routledge, 254–307.
- Kitcher, Phillip (1992). "The Naturalists Return", *Philosophical Review* 101 (1), 53–114.

- Koistinen, Olli (2009). ”Kantin kopernikaaninen kumous”, teoksessa J. Manninen & R. Vilkkonen (toim.) *Ajattelun välineet ja maailmat – kirjoituksia Jaakko Hintikan filosofiasta*. Helsinki: Gaudeamus, 130–155.
- Koistinen, Olli (2010). ”Filosofia analyysinä ja synteessä: Kant filosofiasta ja sen tehtävästä”, teoksessa Henrik Rydenfelt & Heikki A. Kovalainen (toim.), *Mitä on filosofia?* Helsinki: Gaudeamus, 110–121.
- Koistinen, Olli (2011). *Kant ja puhtaan järjen kritiikki*. Turku: Areopagus.
- Kornblith, Hilary (1994). ”Naturalism: Both Metaphysical and Epistemological”, *Midwest Studies in Philosophy* 19 (1), 39–52.
- McDowell, John (1994). *Mind and World*. Cambridge, MA: Harvard University Press.
- Papineau, David (1993). *Philosophical Naturalism*. Oxford: Blackwell.
- Pasnau, Robert (2011). *Metaphysical Themes 1274–1671*. Oxford: Oxford University Press.
- Quine, Willard Van (1974). *Roots of Reference*. La Salle, IL: Open Court.
- Quine, Willard Van (1984). *Theories and Things*. Cambridge, MA: Belknap.
- Raatikainen, Panu (2010). ”Se paha naturalismi”, teoksessa Henrik Rydenfelt & Heikki A. Kovalainen (toim.), *Mitä on filosofia?* Helsinki: Gaudeamus, 160–173.
- Rorty, Richard (1979). *Philosophy and the Mirror of Nature*. Oxford: Blackwell Publishing.
- Rosenberg, Alex (2000). *Philosophy of Science – A Contemporary Introduction*. Lontoo: Routledge.
- Tahko, Tuomas E. (2012). ”Introduction”, teoksessa Tuomas E. Tahko (toim.), *Contemporary Aristotelian Metaphysics*. Cambridge: Cambridge University Press, 1–7.

Graciánin viisaudet

ELI JESUIITTAISÄ BALTASAR GRACIÁNIN LAATIMAT KIRJOITUSOHJEET

VALIKOINUT, TOIMITTANUT JA LISÄSELITYKSIIN VARUSTANUT

Juha Räikkä

1. Johdanto

Baltasar Gracián (1601–1658) oli itsepäisestä luonteestaan tunnettu aragonialainen jesuiitta, jonka kynän terävyys oli hänen veljeskuntansa johdon jatkuvana huolenaiheena. Graciánin pääteoksena on tapana pitää kolmiosaista romaania *El Criticón*, mutta suuren yleisön tietoon Zaragozan filosofi on epäilemättä päätenyt teoksensa *Viisauden käsikirja (Oráculo Manual y Arte de Prudencia, 1647)* ansiosta. Teoksessa Gracián neuvoo miten yksilö löytää tiensä menestykseen maailmassa, jossa ihmiset ovat toisinaan kateellisia, kiittämättömiä ja ymmärtämättömiä. Kolmestasadasta neuvosta ja niiden selityksistä koostuva käsikirja on vedonnut varsin erilaisiin ajattelijoihin. Nietzschen kerrotaan ihailleen Graciánin teosta, ja Schopenhauer jopa laati kirjasta omaa filosofiaansa mukailevan saksannoksen. Graciánin tekstiä voi lukea kyynikon tekstinä. Näin tulkittuna se sopii melko hyvin yhteen sekä Nietzschen katkeruuden että Schopenhauerin synkkämielisyyden kanssa.¹

Gracián on kuitenkin monitulkintainen kirjoittaja eikä yksioikoinen luenta tee oikeutta hänen antamiensa ohjeiden syvällisyydelle. Esimerkiksi melko yleinen käsitys, että *Viisauden käsikirjassa* eettinen ajattelu alistetaan strategiselle ajattelulle, on kyseenalainen (Maurer 1999). Graciánin ohjeissa korostetaan usein sitä, että tietyt asiat tulee salata tai jättää julkisesti

¹ Jos Gracián oli kyyninen, hänen kyynisyytensä oli erilaista kuin Machiavellin, sillä teoksessa *El Héroe* (1637) arvostellaan Machiavellia. – Spinoza ei Graciánin tuotannosta tietävästi pitänyt (Salomaa 1950). Mutta Spinozan kirjahyllyssä olikin ehkä vain *El Criticón* (ks. MacCormack 1991).

sanomatta, muttei tämä tarjoa todistusaineistoa siitä, että etiikka olisi jätetty syrjään. Elämässä on paljon tilanteita, joissa meillä on moraalinen velvollisuus olla hiljaa, ja vielä enemmän tilanteita, joissa meillä on oikeus siihen. Se seikka, että tällainen käytös sattuu usein olemaan myös oman etumme mukaista ja sosiaalisesti toivottavaa, ei tee siitä erityisen ”strategista”. Vuoden 1953 englanninkielinen käännös käsikirjasta – *The Oracle, A Manual of the Art of Discretion* – tuo esiin sen olennaisen seikan, että teoksen aiheena on pitkälti *tahdikkaus*, ei siis oveluus (Gracián 1953).

Graciánin keskeinen ohje menestykseen on se, että meidän on esiinnyttävä sopivalla tavalla. Tämä ohje lienee myös hänen tekstejään koskevien väärinkäsitysten taustalla, sillä ”esiintyminen” on tavallisesti tulkittu teeskentelyksi ja esiintymiseen kuuluva harkittu vaikeneminen taas epärehellisyydeksi. Tuodessamme itsestämme esiin vain tiettyjä piirteitä tietyissä yhteyksissä meidän ei kuitenkaan tarvitse teeskennellä vähääkään, eikä epärehellistä ihmistä tunnista siitä, ettei hän paljasta sisimpäänsä kaikkien ihmeteltäväksi (Räikkä 2007). Sosiaalisen persoonan muotoileminen tilanteiden vaatimusten mukaan ja sisäisen minän häveliäs verhoaminen on hienotunteisuuden ja sivistyksen merkki, jolla ei ole mitään tekemistä sen paremmin tekopyhyiden kuin valehtelunkaan kanssa. Vaikka ihmisen pitää (Graciánin ohjeen 118 mukaan) pyrkiä esimerkiksi kohteliaan maineeseen, ”kaikessa tulee aina olla puolet enemmän sisäpuolta kuin ulkopuolta” (ohje 48).

Saattaa olla, että osa Graciánin ohjeista on turhan suorasukaisia, mutta väite niiden yleisestä epäeettisyydestä on selvästi liioiteltu. Kyseenalaisiltakin vaikuttavat ohjeet voivat olla sellaisia vain näennäisesti. Esimerkiksi ohje (numero 149), jonka mukaan paha pitää usein säilyttää toisten niskaan, voi kuulostaa moraalisesti epäilyttävältä, mutta lienee tosiasiassa ongelmaton. Kollektiivisessa toiminnassa muutaman henkilön kannattaa ottaa epäonnistumiset vastuulleen, jotta ryhmän avainhenkilöiden maine varjeltuu ja yhteinen toiminta onnistuu. Poliittiset ja uskonnolliset ryhmittymät ovat kautta aikain toimineet tuolla tavalla. Niiden toiminta ei ole erityisen moraalitonta, olettaen että yksilöt ovat suurin piirtein vapaaehtoisesti rooleissaan. Sikäli kun ryhmän johdon moraalisenä velvollisuutena on edistää ryhmänsä etuja, heidän tosiaan *pitää* säilyttää pahat asiat syntipukeiksi valittujen niskaan – aivan kuten oppinut jesuiittaisä sanoo.

Jos Graciánin ajatuksia on moitittu väärin perustein, on häntä myös kehuttu virheellisistä lähtökohdista. Joidenkin mielestä *Viisauden käsikirja* on mestariteos, koska elämä nähdään siinä jännittävänä pelinä ja imagonluojat voivat hyödyntää sitä kehitellessään erilaisia temppuja sosiaalisen nousun varmistamiseksi (vrt. Suni 1999). Vaikka tällainen katsanto voi jossakin mielessä pitää paikkansa, Graciánin projekti on täysin vastakkainen kuin nykyajan menestysoppaista tuttu menetelmä, jossa voittoja haetaan psykologisen kikkailun avulla (*Sen teet mihin uskot* 1974, *Turha arkuus pois* 1987, *Menesty myönteisen ajattelun avulla* 2001). Aragonialainen kehottaa meitä katsomaan sekä maailmaa että itseämme täysin realistisesti. Tässä suhteessa *Oráculo* on yhteensopiva vanhan, itsetuntemukseen kannustavan, viisauden kanssa: tunne itsesi (vrt. ohjeet 69 ja 89). Ulkoiseen todellisuuteen tulee taas suhtautua siten kuin vaikeasta masennuksesta kärsivät kai tekevät. ”Älköön siis kukaan kuvitelko turhia, sillä ihmisten ilkeys tulee panemaan merkille kaikki hänen huonot tekonsa eikä yhtäkään hyvää” (ohje 168). Gracián ei väsy varoittamasta meitä toiveajattelusta, perusteettomasta optimismista ja erityisesti itsepetoksesta, joka vaanii yksilöä kaikkialla, odottaen että tämän valppaus hetkeksi herpaantuu. ”Hengen kypsyys ilmenee uskomisen hitautena” (ohje 154).

Tässä kohdassa on kuitenkin paikallaan muistaa, ettei *Viisauden käsikirjaan* tarvitse suhtautua liian vakavasti. Arvuuttelematta kirjoittajan omia päämääriä on sanottava, että teos on paikoin varsin humoristinen eikä sen aforistisella sanonnalla ole muotoiltu päättelyketjuja, joita voitaisiin kutsua argumenteiksi. Barokkityylistä konseptismia edustava kirja on täynnä monimerkityksisiä lauseita, mikä tuskin on vahinko, ja esitettyjen näkemysten perusteita täytyy etsiä kirjoittajan muista teoksista (Tuulio 1973). Jotkut ohjeet näyttävät keskenään ristiriitaisilta. Käsikirjan viiheellinen arvo vaikuttaa selvästi korkeammalta kuin sen filosofinen arvo, eikä tässä ole mitään pahaa. Ajatuksia herättävä viihde on hyvää viihdettä. Vaikka Nietzsche, Schopenhauer ja muutama muu näyttää ottaneen *Oraculón* tosissaan, luultavasti suuri osa ihmisistä on lukenut sitä pilke silmäkulmassa. Tästä kertoo muun muassa se, että teos on poikunut erilaisia mini-oppaita kaikenlaisiin tarkoituksiin – kuten aloittelevien poliitikkojen ohjeistamiseen.²

² Ks. Jaakko Korpi-Anttilan puheenvuoro ”Poliitikon käsikirja” Uuden Suomen keskustelussa: <http://jaakkokorpi-anttila.puheenvuoro.uusisuomi.fi/192543-poliitikon-kasikirja> (1.12.2015).

Poliitikon on epäilemättä hyödyllistä tietää esimerkiksi se, että ”viisaan mielipidettä ei tunne siitä, mitä hän puhuu torilla, koska hän siellä ei puhu omallaan, vaan yleisen typeryyden äänellä, vaikka hänen sisimpänsä kuinka olisikin toista mieltä” (ohje 43).

Tämä kirjoitus kuuluu *Viisauden käsikirjan* innoittamaan mini-opas-genreen. Graciánin viisauksia ei tietävästi ole aiemmin sovellettu akateemisen kirjoittamisen ohjeiksi ainakaan filosofian alalla, mutta nyt on. Kokeneen kirjoittajan ei tarvitse ohjeista välittää, mutta jotkut muut voivat kenties harkita, josko Graciánin ohjeita olisi järkevää seurata, ainakin toisinaan.³ Huomautettakoon, ettei ohjeita ole valittu mukaan sillä perusteella, että niitä kannatettaisiin. Valinnan peruste on ollut yksinkertaisesti se, että ohjeet näyttävät sopivan kirjoitusohjeiksi eivätkä ole *ilmiselvästi* vääriä – tai jos ovat, niissä voi silti piillä totuuden siemen. Joitakin ohjeita on lyhennetty toimituksellisista syistä. Ohjeita on kymmenen sen vuoksi, että ”kymmenen” on mukava luku. Erilaisissa menestysoppaissa on lähes aina ”kymmenen vinkkiä” tai ”kymmenen ohjetta”. *Raamatussa* on kymmenen käskyä. Miksei siis myös alla olevassa luettelossa? Olisi omituista, jos ohjeita olisi esimerkiksi yhdeksän tai yksitoista. Niiden totuudenmukaisuutta alettaisiin luultavasti epäillä heti kättelyssä.

Oráculo Manual y Arte de Prudencia on suomennettu kolme kertaa. Sirkka Salomaan vuonna 1950 ilmestynyt suomennos *Viisauden käsikirja: kolmesataa sääntöä maailmanviisautta* perustuu Schopenhaurin tekstiin ja on siis käännetty taskuoppaan yhdestä saksankielisestä versiosta. Tyyni Tuulion suomennos *Viisauden käsikirja* ilmestyi 1954, ja se on laadittu paitsi alkuperäistekstin myös monien apukäännösten pohjalta. *Maallisen viisauden käsikirja* ilmestyi Annikki Sunin suomentamana 1999 ja sen lähtötekstejä ovat Graciánin teos ja Christopher Maurerin englanninkielinen bestseller-käännös *The Art of Worldly Wisdom: A Pocket Oracle* vuodelta 1992. Seuraavassa hyödynnetään Tuulion (esteettisesti onnistunutta) suomennosta. Vuoden 1999 käännöstä on helppoa lukea, mutta kun Graciánin alkuperäistekstien kerrotaan olevan niin vaikeaselkoisia, että ”itse espanjalaisetkaan eivät niitä ymmärrä” (Tuulio 1973, 5), niin valitkaamme vähän kryptisempi käännös.

³ Esimerkkinä kokeneen kirjoittajan tekstistä, ks. Koistinen 2015.

2. Graciánin ohjeet akateemisille kirjoittajille

Graciánin ohjeet ja niiden selitykset akateemiselle kirjoittajalle filosofian alalla ovat seuraavat. Sitaatteja seuraavat lisäselitykset ovat vain tulkintoja ja heuristisia vihjeitä. Niihin kannattaa epäilemättä suhtautua asianmukaisella varauksella.

1. EI SAA OLLA LIIAN LUJA KANNANOTOISSAAN. Tyhmä on aina varma asiastaan, ja asiastaan varma on aina tyhmä, ja mitä erheellisempi hänen mielipiteensä on, sitä sitkeämmin hän pitää siitä kiinni. Myöntäminen on kunniaksi silloinkin kun asia on päivänselvä, sillä oikeassa oleminen tulee kyllä tiedoksi ja ritarillisuus muistetaan. Itsepäisellä kiistelyllä menettää enemmän kuin voitosta saattaa hyötyä. Sellainen ei ole totuuden puolustamista, vaan karkeutta. On oikeita rautapäitä, joita on äärimmäisen mahdoton saada kääntymään, ja kun itsepintaisuuteen vielä yhtyy haihattelu, on tuloksena purkamaton liitto tyhmyyden kanssa. Sitkeys kuulukoon tahtoon, ei arvostelukykyyn. (Ohjeesta 183.)

Kirjoitusohjeessa 1 muistutetaan kirjoittajaa varmaan siitä, että hänen kannattaa muotoilla väitteensä sopivasti *kvalifoiden* ja jättää erehdyksen varalta takaportti auki silloinkin, kun asia on jokseenkin ongelmaton. Kun kiistoja syntyy, omaa kantaa voi puolustaa muita näkemyksiä kunnioittavin sanakääntein, olivatpa ne järkeviä tai eivät.

2. TYHMYYTTÄ EI PIDÄ JATKAA. Jotkut tekevät epäonnistuneesta yrityksestä veloituksen ja kerran eksytyään katsovat luonteenlujuudeksi jatkaa samaa tietä. Sisäisen tuomioistuimen edessä he syyttävät, mutta ulkonaisen edessä puolustavat erehdystään. Jos heitä alkuvaiheessa pidettiin varomattomina, niin erheen jatkuessa heidät leimataan tyhmiksi. Harkitsematon lupaus ja virheellinen päätös eivät ole sitovia. Mutta muutamat jatkavat ensimmäistä tökeryyttään ja kehittävät kyvyttömyyttään. He haluavat olla vakiotyhmyreitä. (Ohje 261.)

Filosofisen tekstin laatimiseen sovellettuna kirjoitusohje 2 ehkä kertoo, ettei johdonmukaisuudella ole itseisarvoa, vaan sen arvo riippuu siitä, minkä suhteen ollaan johdonmukaisia. Omaa kantaa voi välillä tarkistaa.

3. EI SAA KOSKAAN PUHUA ITSESTÄÄN. Joko kehumme itseämme, mikä on turhamaista, tai moitimme, mikä on raukkamaista. Puhujan viisaudessa on vikaa ja kuulija kiusaantuu. Jos tätä on vältettävä tavallisessakin puheessa, niin vielä enemmän esiinnyttäessä korkeilla jalustoilla, kun puhutaan yleisölle ja pieninkin vivahdus tähän suuntaan tuntuu typerältä. Yhtä hankalaa on viisaan puhua läsnäolevista, koska silloin vaarana on kaksi salakaria, imartelu tai moite. (Ohje 117.)

Yksi mahdollinen tulkinta kirjoitusohjeesta 3 on, että siinä varoitetaan kirjoittajaa puhumasta omaehtoisesti aiemmista teksteistään tai käyttämästä ”kuten olen aiemmin osoittanut” -tyyppisiä ilmaisuja, jotka saattavat aiheuttaa lukijoissa myötähäpeän tunteita. Omaan tekstiin voi tietenkin viitata, ja usein niin pitää tehdäkin.

4. EI SAA KOSKAAN SUUTTUA. Suuri viisauden pykälä on olla koskaan joutumatta suunniltaan. Se todistaa sydämen avaruutta, koska kaikkea sielun suuruutta on vaikea järkyttää. Intohimot ovat sielun nesteitä, ja vähinkin liikamäärä aiheuttaa häiriöitä viisaudessa. Jos paha ehtii suuhun asti, on maine vaarassa. Viisas hallitkoon itsensä siinä määrin ja olkoon niin suuri, ettei parhaankaan myötäkäymisen tai pahimmankaan vastoinkäymisen sattuessa kukaan voi sanoa häntä kiihtyneeksi, vaan hänen ylevyyttään ihailtaan. (Ohje 52.)

Tyylivalinnan neuvona kirjoitusohje 4 varmaankin sanoo, ettei riidanhaluinen teksti ole tyylikästä tekstiä. Pahinta Graciánin mukaan on, jos teksti paljastaa kirjoittajan olevan oikeastikin riidanhaluinen.

5. AHKERA JA ÄLYKÄS. Ahkeruus toteuttaa nopeasti sen minkä äly on laajasti harkinnut. Kiirehtiminen on tyhmien intohimo, ja kun he eivät huomaa kareja, he toimivat ilman varokeinoja. Viisaiden vikana

taas on usein hitaus, koska tarkkaavaisuus johtaa varovaisuuteen, ja toiminnan tehottomuus joskus pilaa onnistuneen suunnitelman. Nopeus on menestyksen äiti. Paljon on tehnyt se, joka ei ole jättänyt mitään huomiseen. (Ohjeesta 53.)

Voi hyvin olla, että kirjoitusohjeen 5 tarkoitus on todeta, että kirjoitukset syntyvät kirjoittamalla ja ettei ahkera kirjoittaja siirrä tehtävänsä tuonemmas ilman erityisen painavia syitä. Toisinaan tuollaisia syitä tietysti on.

6. EI PIDÄ VIIPYÄ NARRIEN SEURASSA. Narri on se joka ei narreja tunne ja vielä enemmän se joka tuntee eikä karta heitä. Narrit ovat vaarallisia pinnallisessa seurustelussa ja tuhoisia tuttavallisessa. Ja vaikka oma arkuus ja toisten huolenpito saisikin heidät joksikin aikaa pysymään aisoissa, he lopulta tekevät tai sanovat tyhmyyden, sitä juhlallisemman, mitä kauemmin sitä sai odottaa. Huonosti edistää toisten mainetta se jolla sitä ei ole itselläänkään. Narrit ovat onnettomia, sillä onnettomuus liittyy narriuteen kuin nahka luuhun ja sopii sen kanssa yhteen. Ainoa vähemmän paha puoli heissä on se, että vaikka viisaat eivät ole heille miksikään hyödyksi, he hyödyttävät viisaita paljonkin, joko havaintoesineinä tai varottavina esimerkkeinä. (Ohje 197.)

Tässä täytyy olla kyse niin kutsutusta huuhaa-filosofiasta. Kirjoitusohje 6 liittyy selvästikin lähteiden käyttöön, aiheiden valintaan ja siihen, kenen kanssa ryhtyy filosofiseen keskusteluun.

7. EI PIDÄ ARKIPÄIVÄISYYDEN PELOSTA OLLA ARVOITUKSELLINEN. Molemmat äärimmäisyydet ovat haitaksi. Kaikki arvokkuudesta poikkeava vivahtaa typeryyteen. Paradoksi on eräänlaista hämäämistä, alussa suosittua uutena ja pirteänä, mutta harhan haihduttua varsin halveksittua. Se on tavallaan huijausta ja valtion asioissa tuhoisaa. Ne jotka eivät hyveen tietä pääse tai uskalla pyrkiä sankaruuteen, turvautuvat paradokseihin ja herättävät ihailua tyhmissä, mutta innoittavat viisaita oikeihin päätelmiin.

Paradoksaalisuus todistaa arvostelukyvyn horjuvuutta ja on kaukana viisaudesta, ja jollei pohja aina ole väärä, on se ainakin epävarma, mikä on vaarallista tärkeissä asioissa. (Ohje 143.)

Tämänkin ohjeen on pakko liittyä filosofiseen temppuiluun. Kirjoitusohjeella 7 Gracián varoittaa kirjoittajaa varmaankin siitä, että keksimällä uusia sanoja ja välttämällä esittämästä asioita ymmärrettävästi ja tutuin käsittein (silloinkin kun se olisi mahdollista) hän voi saavuttaa suosiota ainoastaan niiden parissa, jotka kunnioittavat vain ”sitä mitä eivät tajua” (vrt. ohje 253). Tietenkään kaikki asiat eivät taivu arkipäivän käsitteisiin, ja silloin tulee edetä ongelman ehdoilla.

8. EI SAA OLLA VÄSYTTÄVÄ. Samaa puuhaa ja puhetta jauhava ihminen on tavallisesti kiusallinen. Lyhyys on viehättävämpää ja asiallisempaa. Kohteliaisuus korvaa puheen lyhyden. Hyvä on lyhyenä kaksin verroin hyvää. Eikä pahakaan ole niin pahaa, jos sitä on vähän. Ytimekkyys on tehokkaampaa kuin lavertelu. Ja yleinen totuus on, että monisanainen on harvoin viisas, mikä ei niinkään ilmene puheen sisällössä kuin muodollisessa puolessa. Hyvin sanottu on pian sanottu. (Ohjeesta 105.)

Ehkäpä Gracián haluaa kirjoitusohjeella 8 väittää, että monisanainen ja tilaa vievä esitystapa voi saada lukijan kysymään, eikö kirjoittaja sittenkään viitsinyt tai osannut pohtia asiaa riittävästi ja halveksiiko kirjoittaja lukijoitaan, kun julkaisee tekstinsä, vaikkei valmista koskaan tullutkaan.

9. HENKILÖKOHTAISET OMINAISUUDET YLITTÄKÖÖT VIRKATEHTÄVÄT EIKÄ PÄINVASTOIN. Oli asema miten korkea tahansa, on osoitettava että ihminen on suurempi. Avara henki kasvaa ja näkyy yhä enemmän tehtävien mukana. Ahdassydäminen joutuu pian kiinni ja horjahtaa lopulta tehtävistään ja asemastaan. Suuri Augustus piti kunnianaan olla suurempi ihmisenä kuin ruhtinaana. Tällöin on kysymys mielen ylevyydestä, ja viisas itseluottamus on myös suureksi avuksi. (Ohje 292.)

Kirjoitusohje 9 koskee ilman muuta kirjoittamistyyliä ja siinä Gracián varoittaa *pompöösistä* ja *grandioottisesta* tyylivalinnasta, jottei kirjoittaja joutuisi naurunalaiseksi.

10. EI SAA OLLA LUOKSEPÄÄSEMÄTÖN. Kukaan ei ole niin täydellinen, ettei joskus olisi huomautuksen tarpeessa. Parantumaton tyhmyyttä on olla kuulematta. Vapautuneimmankin tulee suoda sija ystävälliselle neuvolle, eikä edes kuninkaan mahti saa olla taipumaton. Jotkut ihmiset ovat luoksepääsemättömyydessään parantumattomia, he syöksyvät turmioon, koska kukaan ei uskalla estää heitä siitä. Oivallisimmankin tulee pitää ovi auki ystävyydelle, ja siitä tulee avun ovi. Tarvitaan ystävä, jolla on vapaus neuvoa, jopa ojentaa toista. Tämän hänen arvovaltansa tulee perustua omaan tyydytykseen ja ylevään käsitykseen hänen uskollisuudestaan ja viisaudestaan. Kaikille ei tule helpottaa pääsyä tähän arvoon ja luottamukseen, mutta itsevarjelun sisimmässä kammiossa olkoon luotettavana peilinä uskottu, jonka suuriarvoinen ojennus vapauttaa erheestä. (Ohje 147.)

Sovellettuna akateemisen tekstin laatimiseen kirjoitusohjeessa 10 ilmeisesti todetaan, että kirjoitusprosessin osiin kuuluu myös kommenttien hankkiminen ja huomioiminen. Ken saa hyvät kommentit, mutta jättää ne huomiotta, pelaa Graciánin mukaan korttinsa väärin.

3. Lopuksi

Baltasar Gracián y Morales oli merkittävä kirjailija, jonka ajatuksista voimme vielä nykyäänkin iloita – tai pitäisikö sanoa, erityisesti nykyään. Omana aikanaan Gracián herätti monissa hyväksynnän tunteita, mutta myös paljon murhetta ja huolta. Kirjojaan Gracián julkaisi salanimellä, mutta niiden tekijästä tuskin oli epäselvyyttä. Graciánin saarnastuolitemsaus, jossa hän ilmoitti saaneensa postia suoraan Helvetistä ja aikovansa lukea sieltä lähetetyn kirjeen, ei ollut 1600-luvulla erityisen tasapainoinen veto jesuiittapapilta. Herääkin kysymys, välittikö Gracián laatimistaan ohjeista omissa valinnoissaan. ”Selkkauksia kartettava”, sanoo ohje 47, mutta Gracián ajautui jatkuvasti selkkauksiin. Ehkäpä kyse oli siitä, että Gracián hyväksyi kulmikkaan luonteensa välttämättömänä pahana, josta piti vain olla tietoinen.

Ohjeen 161 mukaan yksilön ”pitää tuntea helmasyntinsä”, sillä ”täydellisimmälläkin ihmisellä on sellainen ja hänellä on siihen joko laillistettu tai salainen lemmensuhde”. Ohjeiden rikkominen on syntiä, mutta hyvin ymmärrettävää syntiä, jos henkilön suhde siihen on aivan erityinen. Raadollisen ankarista ohjeista rakentuvan kilven takana näyttää sykkineen suvaitseva sydän.

Turun yliopisto

Kirjallisuus

- Gracián, Baltasar (1950). *Viisauden käsikirja: kolmesataa sääntöä maailmanviisautta*. Suom. S. Salomaa. Porvoo: WSOY.
- Gracián, Baltasar (1953). *The Oracle: A Manual of the Art of Discretion*. Engl. L. B. Walton. Lontoo: Dent.
- Gracián, Baltasar (1973). *Viisauden käsikirja*. Suom. T. Tuulio. Porvoo: WSOY.
- Gracián, Baltasar (1992). *The Art of Worldly Wisdom: A Pocket Oracle*. Engl. C. Maurer. New York: Doubleday.
- Gracián, Baltasar (1999). *Maallisen viisauden käsikirja*. Suom. A. Suni. Juva: WSOY.
- Koistinen, Olli (2015). ”Towards an Adverbial Theory of Spinoza’s Modes”, *Res Cogitans* 10, 101–122.
- MacCormack, Sabine (1991). ”Demons, Imagination, and the Incas”, *Representations* 33, 121–146.
- Maurer, Christopher (1999). ”Saatteeksi”, teoksessa Baltasar Gracián, *Maallisen viisauden käsikirja*. Juva: WSOY, 5–22.
- Räikkä, Juha (2007). *Yksityisyyden filosofia*. Helsinki: WSOY.
- Salomaa, Sirkka (1950). ”Johdanto”, teoksessa Baltasar Gracián, *Viisauden käsikirja: kolmesataa sääntöä maailmanviisautta*. Porvoo: WSOY, 5–10.
- Suni, Annikki (1999). ”Sananen suomennoksesta”, teoksessa Baltasar Gracián, *Maallisen viisauden käsikirja*. Juva: WSOY, 23–24.
- Tuulio, Tyyni (1973). ”Alkusanat”, teoksessa Baltasar Gracián, *Viisauden käsikirja*. Porvoo: WSOY, 5–8.