

Hyviä ja huonoja perusteita kokeelliselle sosiologialle

Petri Ylikoski



VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

Abstrakti

Kokeellisen tutkimuksen suosio on nopeasti kasvamassa eri yhteiskuntatieteissä. Merkittävä osa tästä tutkimuksesta on sosiologisesti kiinnostavaa, ja siksi kaikkien sosiologien on hyvä tutustua sekä kokeellisen tutkimuksen vahvuuksiin että sen heikkouksiin. Tämä kirjoitus aloittaa erittelemällä kokeellisen tutkimuksen nousun taustatekijöitä ja jatkaa sitten esittelemällä kokeellisen tutkimuksen moninaisuutta: erilaiset kokeelliset asetelmat eroavat toisistaan suuresti ja samaakin koeasetelmaa on mahdollista käyttää useaan erilaiseen tutkimukselliseen tarkoitukseen. Tästä moninaisuudesta seuraa, että vaikka kokeellisen tutkimusasetelman erityinen ansio on sen suosiollisuus kausaalisille päätelmille, tulee kokeellinen tutkimus ymmärtää laajemmin kuin pelkkänä kausaalisten vaikutusväitteiden testaamisena. Yleisesti ottaen kokeellisen tutkimuksen lisääntyminen on ehdottoman hyvä asia yhteiskuntatieteiden kannalta. Siihen ei kuitenkaan tule yhdistää kahta ongelmallista ajatusta. Ensimmäinen näistä on ajatus ehdottomasta ja tiukasta näyttöhierarkiasta, joka koskee eri tutkimustapojen soveltuvuutta luotettavien kausaalipäätelmien tekemiseen. Tämä ajatus on osoittautunut toimimattomaksi jo lääketieteellisten toimenpiteiden arvioinnin kohdalla, ja hankaluudet vain kasvavat, jos sitä sovelletaan politiikkatoimenpiteiden arviointiin, puhumattakaan sen yleisemmästä soveltamisesta yhteiskuntatieteissä. Toinen ongelmallinen ajatus on toive, että kokeellisten menetelmien myötä yhteiskuntatieteet voisivat viimein siirtyä nopeasti etenevän ja vakaasti kasautuvan tiedon tuottamiseen. Kokeellisesta menetelmästä ei ole tällaisen haavekuvan toteuttajaksi. Luontevinta onkin ajatella kokeellista tutkimusta yksinkertaisesti tervetulleena täydennyksenä sosiologian ja muiden yhteiskuntatieteiden menetelmävalikoimaan.

ASIASANAT: Kausaliteetti, kokeellinen tutkimus, näyttöhierarkia.

Perinteisesti yhteiskuntatieteissä on ajateltu, että kokeellinen tutkimus on kotonaan psykologiassa ja sosiaalipsykologiassa, mutta

muissa yhteiskuntatieteissä kokeelliset menetelmät ovat lähinnä kuriositeetti. Viimeisten kahdenkymmenen vuoden aikana kokeellinen

tutkimus on kuitenkin ollut vahvassa nousussa yhteiskuntatieteissä. Erityisesti tämä näkyy taloustieteessä, jossa kokeellisen tutkimuksen nousu alkoi jo aiemmin (Kagel & Roth 1995), mutta enenevässä määrin myös politiikan-tutkimuksessa (Druckman, Green, Kuklinski & Lupia 2011). Vaikka sosiologian kohdalla ei voida puhua samanlaisesta mullistuksesta kuin taloustieteessä, on myös kokeellinen sosiologinen tutkimus vahvassa nousussa (Jackson & Cox 2013). Vuosituhannen alun jälkeen kokeellisia asetelmia käyttävien tutkimusten määrä on kasvanut jatkuvasti sosiologisissa julkaisuissa, ja sosiologit käyttävät yhä useammin hyväkseen kokeellisen tutkimuksen tuloksia muista yhteiskuntatieteistä.

Esimerkiksi kokeellisen tutkimuksen sosiologisista mahdollisuuksista käy *Matteus-vaikutus*, tai yleisemmin, *kasautuvien etujen teoria*. Vaikka etujen (ja haittojen) kasautumista on tutkittu paljon sekä empiirisesti että teoreettisesti (DiPrete & Eirich 2006), on silti jäänyt avoimeksi, missä määrin menestys voisi olla puhtaasti alkuvaiheessa saadulle ”ansaitsemattomalle” edulle perustuva itsensä toteuttava ennuste. Laadun, kasautumismekanismien sekä muiden prosessiin vaikuttavien tekijöiden erittelemineen onkin vaikeaa pelkästään kuvailevaa aineistoa käyttävissä tutkimuksissa. Tästä syystä Salganikin ja Wattsin (2006) koeasetelmaa voi pitää tärkeänä.

Salganik ja Watts perustivat verkkoon ilmaisen musiikinjakelupalvelun, jossa he sitten muuntelivat käyttäjien tietoa kappaleiden suosiosta. Tämän sosiaalisen informaation vaikutukset osoittautuivat dramaattisiksi. Todellisilta kulttuurimarkkinoilta tuttu menestyksen epätasainen jakautuminen ja sen vaikea ennustettavuus toteutuivat myös kokeellisessa maailmassa, mutta koeasetelma osoitti, että keskeinen näitä prosesseja ohjaava mekanismi on käyttäjien

nojautuminen sosiaaliseen informaation kappaleiden suosiosta, eivät niinkään kappaleiden yksilölliset erot. Vaikka menestys ei yksin perustunut sosiaaliselle informaatiolle, sen saatuus kärjisti eroja kappaleiden menestyksessä, ja mitä selkeämmin tieto muiden käyttäytymisestä oli esillä, sitä suurempi oli sen vaikutus. Osoittautui myös, että sosiaalinen informaatio lisäsi menestyksen ennustamattomuutta: mitä suurempi sosiaalisen informaation vaikutus, sitä vaikeampi oli ennustaa yksittäisten kappaleiden menestystä.

Toinen samaan aiheeseen liittyvä esimerkki on Van de Rijtin ja kumppaneiden (2014) satunnaisesti kenttäkoe, jossa he tutkivat, kuinka paljon alkuvaiheen menestys vaikuttaa hankkeen myöhempään menestykseen. Kokeen tulokset osoittavat, että riippumatta siitä, millaisesta menestyksestä (rahan kerääminen hankkeeseen, maine arvioijana tai kirjoittajana, nimien kerääminen vetoomukseen) on kysymys, varhaisella menestyksellä on merkittävä vaikutus siihen, tapahtuuko jatkossa menestystä – ja alkuvaiheen menestyksen vaikutukset ovat pysyviä.

Nämä kaksi esimerkkiä tuovat hyvin esille kokeellisen tutkimuksen mahdollisuudet: vaikka se ei korvaa muuta empiiristä ja teoreettista tutkimusta, täydentää kokeellinen tutkimus vakuuttavalla tavalla näiden todistusvoimaa. Aikaisemmin oli vielä mahdollista ajatella, että kasautuvien etujen mekanismin vaikutus todellisen elämän menestykseen on pieni, tai sen vaikutuksen häviävän ajan kuluessa. Näiden tutkimusten jälkeen tällaisen kannan puolustaminen on huomattavasti vaikeampaa. Kokeellinen tutkimus on siis mielenkiintoinen ja merkittävä lisä perinteiseen sosiologiseen todistusaineistoon. Sosiologien kannattaakin kiinnittää huomiota kokeelliseen yhteiskuntatieteelliseen tutkimukseen, vaikka he eivät omassa tutkimuksessaan soveltaisikaan kokeel-

lisiä menetelmiä. Kokeellisessa sosiaalipsykologisessa ja taloustieteellisessä tutkimuksessa nousee jatkuvasti esiin sosiologisesti kiinnostavia teemoja. Nämä tutkimukset ovat myös ohjanneet tutkijoita omaksumaan sosiologiaa käsitteitä ja teorioita, mikä luo pohjaa kiinnostavalle yhteiskuntatieteiden väliselle keskustelulle. Hyvä esimerkki tästä on kokeellinen peliteoria, jossa tutkimustulokset ovat pakottaneet tutkijat ottamaan sosiaaliset normit vakavasti (Bicchieri 2005). Tätä taustaa vasten on paikallaan tarkastella kokeellisen tutkimuksen nousun taustoja, kokeellisen tutkimuksen luonnetta sekä eritellä hyviä ja huonoja perusteita kokeellisten menetelmien käytölle.

Kokeellisen tutkimuksen paluun tausta

Kokeellisen tutkimuksen ei voi sanoa olevan uusi asia sosiologiassa tai yleisemmin yhteiskuntatieteissä. Sosiologian varhaisista vaiheista löytyy monia esimerkkejä yrityksistä tutkia yhteiskuntaa kokeellisesti (Jackson & Cox 2013; Oakley 2000). Totta kuitenkin on, että kokeellinen tutkimus katosi pitkäksi aikaa yhteiskuntatieteistä (pois lukien sosiaalipsykologia) lähes kokonaan, kunnes se alkoi uuden nousunsa taloustieteessä 1980-luvulla (Kagel & Roth 1995) ja muissa yhteiskuntatieteissä hieman myöhemmin (Gerber & Green 2012; Webster & Sell 2007, Willer & Walker 2007). Tässä kirjoituksessa en paneudu kokeellisen tutkimuksen varhaiseen historiaan ja sen hetkelliseen häviämiseen, vaan keskityn sen uuden nousun takana oleviin tekijöihin.

Kokeellisen tutkimuksen uuden nousun taustalla on monenlaisia seikkoja. Yksi merkittävä tekijä ovat kasvaneet vaatimukset kausaalipäätelmien oikeutukselle. Kausaaliset kysymyksenasettelut ovat nousseet aikaisempaa vah-

vemmin esiin yhteiskuntatieteissä, mikä näkyy myös tilastotieteessä: tilastollisen kausaalipäätelyn vaatimukset ja taustaehdot ovat nousseet keskeiseksi huomion kohteeksi (Morgan & Winship 2014; Pearl 2009). Suurempi huomio on merkinnyt kausaalipäätelmille asetettujen vaatimusten tiukentumista. Hienostuneiden analyysimenetelmien sijasta luotettavan tiedon perustaksi nähdään nyt kokeellisen tutkimuksen selkeä tutkimusasetelma (Angrist & Pischke 2010) ja ymmärrys havaintojen taustalla toimivista kausaalista mekanismeista (Hedström & Ylikoski 2010). Kokeellinen manipulaatio yhdistettynä kontrolloituihin olosuhteisiin onkin verraton väline aitojen kausaalisten riippuvuussuhteiden löytämiseen.

Toinen kiinnostava kokeellisen käänteän taustatekijä on perinteisten tieteenalarajojen ja -määritelmien muokkautuminen. Vanhat tieteenalaidentiteetit ovat alkaneet tuntua yhä keinotekoisemmilta, ja yhteiskuntatieteilijät ovat ryhtyneet entistä laajempaan yhteistyöhön kokeellisia menetelmiä soveltavien psykologian ja neurotieteiden kanssa. Onnistuneet esimerkit tällaisesta yhteistyöstä, esimerkiksi behavioraalinen taloustiede, ovat omalta osaltaan lisänneet kokeellisen tutkimuksen näkyvyyttä ja yhteiskuntatieteellistä kiinnostavuutta. Tulevaisuudessa tällainen perinteisten tieteenalarajojen ylittäminen tulee olemaan yhä tavallisempaa, ja toivottavasti myös sosiologit osallistuvat siihen aktiivisesti.

Yhteiskuntatieteilijät ovat heränneet myös kokeellisen tutkimuksen retoriseen voimaan, erityisesti pyrittäessä vakuuttamaan oman erityisalan ulkopuolisia yleisöjä. Asetelmaltaan yksinkertaiset kokeet ja niiden mahdolliset dramaattiset tulokset nousevat helposti keskeiseen osaan popularisoivissa esityksissä, vaikka tutkija ei itse sitä edes itse tarkoittaisi. Tilanteissa, joissa yleisö ei ole täysin vakuuttunut (tai perillä)

tutkimuksen teoreettisista taustaoletuksista, kokeellinen näyttö on usein vakuuttavaa. Ajoittain kokeelliset tulokset ovat jopa liian vahvoja: vaikka systemaattiselta kannalta yksittäinen koetulos on usein lähinnä anekdootti, laajempaa tutkimuksellista asiayhteyttä tuntemattomalle yksittäisen kokeen sujuva ja mielenjäävä kuvaus antaa perustan varsin suoraviivaiselle yleistämislle ja soveltamiselle. Kokeellisen todistusaineiston vakuuttavuus näkyy myös politiikassa. Lääketieteen malliin perustuva ajatus näyttöön perustuvasta politiikasta on nostanut esiin satunnaistamiseen perustuvien kokeellisten menetelmien käytön myös politiikkatoimenpiteiden vaikuttavuuden arvioinnissa. Palaan tähän ajatukseen sekä sen taustalla olevaan ajatukseen näyttöhierarkiasta myöhemmin tässä kirjoituksessa.

Kokeellisella käännteellä on myös teknologiset taustaehtonsa. On esimerkiksi vaikea kuvitella monia behavioraalisen taloustieteen kokeita ilman tietokoneistettua laboratoriotia. Tietokoneet ovat usein osa koeasetelman toteutusta, aineiston keruuta sekä analyysiä. Se, mikä oli aikaisemmin ehkä erittäin vaikeaa tai jopa mahdotonta toteuttaa, on nyt rutiininaimaista. Myös internet on luonut uusia tutkimusmahdollisuuksia: suuret internet-yhtiöt eivät vain tuota massiivisia määriä dataa käyttäjistään, ne voivat myös toteuttaa erilaisia kokeellisia asetelmia, joista osa on aidosti yhteiskuntatieteellisesti kiinnostavia. Internet mahdollistaa myös koehenkilöiden laajemman ja helpomman rekrytoinnin. On vielä hieman kiistanalaista kuinka edustavia esimerkiksi Amazonin *Mechanical Turk*-palvelun (Crump, McDonnell & Gureckis 2013, Shank 2015) kautta löydettävät koehenkilöt ovat, mutta on selvää, että kokeellisen tutkimuksen koehenkilöpooli on laajentunut perinteisistä yliopisto-opiskelijoista. Lisäksi internetin myötä perinteisiin kyselytutkimuksiin voidaan liittää aikaisempaa helpommin kokeellisia ulottuvuuksia, näin laajentaen politiikan tutkijoiden

kehittämien *survey-kokeiden* mahdollisuuksia (Sniderman 2011).

Kokeellisten tutkimusasetelmien kirjo

Pohdittaessa kokeellisen tutkimuksen sosiologista relevanssia on hyvä luopua kaavamaisesta ajatuksesta kokeellisen tutkimuksen yhdenmukaisuudesta. Kokeellista tutkimusta tehdään monessa muodossa, ja kokeiden tiedollinen tarkoitus vaihtelee suuresti. Liian yksinkertainen kuva kokeellisesta tutkimuksesta voi estää näkemästä sen kaikkia tiedollisia mahdollisuuksia sekä johtaa kokeellisen tutkimuksen tulosten merkityksen ja arvion vääryyden arvioitiin.

Ensimmäinen kokeellisen tutkimuksen kirjontulottuvuus ovat tutkimusasetelmat, joita esittelen seuraavassa hyvin yksinkertaistetussa muodossa. Kokeellisen tutkimuksen yksinkertaisin muoto on *kontrolloitu koe*. Sen lähtökohtana on kaksi samoissa olosuhteissa olevaa samanlaista kohdetta, joista toiseen kohdistetaan kokeellinen toimenpide. Toimenpiteen jälkeen kohteita vertaillaan ja päätellään, mikä oli kokeellisen toimenpiteen vaikutus. Olennaista tässä asetelmassa on siis kontrollin luoma samankaltaisuus, kokeentekijän suorittama manipulaatio sekä tapausten vertailu.

Vaikka monet klassiset tieteelliset kokeet ovat tämän mallin mukaisia, ei malli ole erityisen käyttökelpoinen esimerkiksi tutkittaessa biologiaa, psykologisia tai yhteiskunnallisia prosesseja. Tästä syystä *satunnaistettu kontrolloitu koe* on saavuttanut kokeellisen tutkimuksen prototyyppin aseman lääketieteessä, psykologiassa ja yhteiskuntatieteissä. Siinä täydellisen kokeellisen kontrollin sijasta – joka siis takaisi vertailtavien kohteiden mahdollisimman suuren samankaltaisuuden – sallitaan jonkinasteinen muuntelu tutkimuskohteissa. Tähän on kaksi motiivia. Ensinnäkin kokeen pohjalta halutaan tehdä päätelmiä

laajemmasta kohteiden joukosta, jossa esiintyy muuntelua. Toiseksi kaikkia kohteiden eroja ei tunnetta tai niistä ei voida tehdä havaintoja. Tästä syystä satunnaistetussa kontrolloidussa kokeessa vertailu ei tapahdu yksittäisten kohteiden välillä, vaan kohteiden ryhmien välillä. Keskeinen vaatimus on, että kohteiden päätyminen toimenpite- tai verrokkiryhmään on satunnaista. Tämän satunnaistamisen tarkoituksena on rikkoa tuntemattomien kohteiden välisten erojen ja kokeellisen toimenpiteen vaikutuksen välinen tilastollinen yhteys. Satunnaistamisen ansiosta voidaan koe- ja verrokkiryhmien välisestä erosta päätellä kokeellisen toimenpiteen keskimääräinen vaikutus tutkitussa kohteiden joukossa. Kontrolloitu koe ei kuitenkaan perustu yksinomaan satunnaistamiselle, vaan kohteiden kontrollilla on myös keskeinen rooli. Esimerkiksi lääketieteellisissä kokeissa lumetoimenpiteen tarkoituksena on tapauksen samankaltaistaminen vakioimalla mahdollisimman paljon hoitotoimenpiteestä: tavoitteenahan on tutkia juuri lääkeaineen, ei pelkän pillerin syömisen, vaikutusta potilaan terveyteen.

Sosiologinen esimerkki satunnaistetusta kontrolloidusta kokeesta on Robb Willerin ja kumppaneiden tutkimuksen (2013) osio, jossa he testasivat kokeellisesti nk. maskuliinisen ylikompensaation teoriaa. Aluksi molempia sukupuolia edustavat koehenkilöt vastasivat sukupuoli-identiteettikyselyyn. Tämän jälkeen koehenkilöt näkivät, kuinka koeavustajat ”arvioivat” heidän vastauksensa ja palauttivat ne sitten kullekin vastaajalle suljetussa kirjekuoressa. Kirjekuori sisälsi arvion siitä, kuinka maskuliinisia tai feminiinisiä koehenkilön vastaukset olivat olleet. Tosiasiallisesti kirjeiden sisältämät tulosarvot oli tehty jo ennakkoon, eikä vastaajien vastauksilla ollut niihin mitään vaikutusta. Koeavustajat eivät edes tienneet, mitä kuskakin kirjeessä raportoitiin. Koehenkilöt saivat siis tosiasiallisesti satunnaisesti määräytyneitä palautetta sukupuoli-identiteetistään: puolet miehistä kuuli saaneensa maskuliinisuus-feminiinisyys-

skaalalla naisille tyypillisen tuloksen ja puolet naisista kuuli saaneensa miehille tyypillisen tuloksen.¹ Loput kuulivat saaneensa sukupuolelleen tyypillisen tuloksen. Tämän kokeellisen toimenpiteen jälkeen koehenkilöt vastasivat toiseen kyselyyn, jonka kysymykset koskivat katumaastureita, homoseksuaalisuutta ja Irakin sota. Tämän kyselyn vastausten analyysi osoitti, että miehet, jotka olivat saaneet palautetta vastaustensa feminiinisydestä – ja joiden maskuliinisuus oli siis asetettu uhanalaiseksi – suhtautuivat negatiivisemmin homoseksuaalisuuteen, positiivisemmin Irakin sotaan sekä pitivät suuria katumaastureita haluttavampina. Vastaavia vaikutuksia ei havaittu naiskoehenkilöiden vastauksissa.

Kenttäkoe (Gerber & Green 2012; Teele 2014) noudattaa satunnaistetun kontrolloidun kokeen mallia, mutta siinä kontrollin mahdollisuus on pienempi – kaikki koehenkilöt eivät kohtaa samanlaisia olosuhteita – ja siksi satunnaistamisen

1 Taloustieteellisen ja sosiaalipsykologisen kokeellisen tutkimuksen välillä on mielenkiintoinen ero suhtautumisessa koehenkilöiden harhaanjohtamiseen. Taloustieteilijät eivät tavallisesti hyväksy koehenkilöiden huijaamista, eivätkä kokeellisen taloustieteen julkaisut suostu julkaisemaan käsikirjoituksia, joiden aineistona on huijausta sisältäviä kokeita. Kiellon motivaatio ei liity tutkimusetiikkaan. Taustalla on pelko, että huijaaminen heikentää koehenkilöiden luottamusta kokeentekijöihin, minkä seurauksena he eivät enää ottaisi vakavissaan koeasetelmia, mikä vuorostaan vaarantaisi tulevien tutkimusten luotettavuuden. Sosiaalipsykologit ja muut yhteiskuntatieteilijät pitävät tällaista purismia huonosti perusteltuna ja tutkimusmahdollisuuksia tarpeettomasti rajoittavana tekijänä. Heidän kannaltaan empiiriset tutkimukset eivät tue taloustieteilijöiden olettamusta luottamuksen rapistumisesta (Barrera & Simpson 2012; Bonetti 1998), ja taloustieteilijöiden tekemä hiuksenhieno ero valehtelun ja kertomatta jättämisen välillä on pikemminkin saivartelua kuin merkittävä metodologinen vedenjakaja. Joka tapauksessa tästä metodologisesta kysymyksestä on syntynyt merkittävä kokeellisia yhteiskuntatieteilijöitä jakava kiista (ks. Cook & Yamagishi 2008).

onnistumisella on suurempi merkitys luotettavien kausaalipäätelmien tekemiselle. Kenttäkokeen ansiona on se, että sikäli kuin satunnaistaminen onnistuu, on kokeen ulkoinen validiteetti suurempi. Toisin sanoen, kokeen perusteella voidaan tehdä luotettavampia päätelmiä kokeellisen toimenpiteen vaikutuksista luonnollisissa olosuhteissa kuin usein keinotekoisista laboratorioskokeista.

Esimerkiksi sosiologisesta kenttäkokeesta käy Keizerin, Lindenbergin ja Stegin (2008) tunnettu tutkimus epäjärjestyksen leviämisestä. Tutkimuksen ajatuksena oli testata empiirisesti rikkinäisten ikkunoiden teoriaa kenttäkokeiden avulla (ks. myös Keuschnigg & Wolbring 2015). Tutkijat manipuloivat graffitien määrää ja ympäristön roskaisuutta erilaisissa tilanteissa sekä havainnoivat, kuinka nämä seikat vaikuttivat roskaamista, varastamista ja yleistä järjestystä koskevien normien seuraamiseen. He havaitsivat, että ympäristön epäjärjestys tosiaankin lisäsi myös muiden normien rikkomista. Keizerin ja kumppaneiden tutkimus on tärkeä panos rikkinäisten ikkunoiden teoriasta käytyyn laajaan keskusteluun, sillä se ensimmäistä kertaa konkreettisella tavalla testaa teorian yhtä keskeisistä oletuksista. Koetuloksia tulkittaessa kannattaa kuitenkin olla tarkkana: koasetelma ei testaa rikkinäisten ikkunoiden teorian innoittamaa ajatusta nollatoleranssista keinona vähentää rikollisuutta. Tämän hypoteesin testaaminen edellyttäisi hyvin erilaista koasetelmaa.

Metodologit ovat erimielisiä siitä, onko *luonnollinen koe* (Dunning 2012) aidosti kokeellinen asetelma. Luonnollisessa kokeessa kokeentekijä ei suunnittele koetilannetta eikä toteuta kokeellista toimenpidettä. Tällaisella asetelmalla on kuitenkin selkeä samankaltaisuus kenttäkokeeseen: siinä satunnaistamisesta huolehtii jokin luonnollinen tai hallinnollinen prosessi, minkä vuoksi kausaalisia päätelmiä voidaan tehdä pitkälti samalla tavoin kuin kenttäkokeessa. Olennaista luonnollisessa kokeessa on tietenkin mahdollisuus osoittaa, että kyseinen prosessi todella on

riittävän satunnainen, sillä ilman tätä kyseessä olisi aivan tavallinen havaintotutkimus. Luonnollisen kokeen asetelman tuottavien olosuhteiden löytäminen on pitkälti kiinni sattumasta (ja tutkijan päättävyydestä niiden etsimisessä), joten laajemmassa tarkastelussa ne ovat pikemminkin osoitus siitä, että käytännössä kokeelliset ja kuvailevat tutkimukset muodostavat jatkumon, kuin oma itsenäinen tutkimusmenetelmänsä.

Dalton Conleyn ja kumppaneiden (2013) tutkimus on kiinnostava esimerkki luonnollisesta kokeesta. Siinä tarkasteltiin virheellisesti luokiteltuja saman- ja erimunaisia kaksosia. Tavoitteena oli arvioida, pitääkö käyttäytymisgeneettisissä kaksostutkimuksissa keskeinen mutta kiistanalainen oletus identtisten ja ei-identtisten kaksosten samanlaisesta kasvuympäristöstä paikkansa. Keskeisenä ajatuksena oli, että kaksosten tunnistamisvirhe on satunnainen, ja vertailemalla oikein ja väärin luokiteltuja kaksosia voidaan tutkia, vaikuttaako uskomus kaksosten identtisyydestä heidän kehitykseensä. Tutkimus osoittikin, että kaksosaineistoja hyödyntävät käyttäytymisgeneettiset tutkimukset eivät kyseisen oletuksen seurauksena merkittävästi liioittele piirteiden periytyvyyttä (*heritability*). Näitä alkuperäisiäkin kaksoskokeita voisi kuvata luonnollisiksi kokeiksi. Niissä tutkijan tutkimuksen kannalta satunnainen tekijä hedelmöitymisessä tuottaa sama- ja erimunaisia kaksosia, mikä mahdollistaa päätelmien tekemisen perimän ja ympäristön vaikutuksista tutkittavana olevan ilmiön muunteluun.

Kirjallisuudessa käytetään myös käsitettä *kvasi-koe* (Shadish, Cook & Campbell 2002). Tällaisessa asetelmassa toteutuu ajatus kokeentekijän toteuttamasta toimenpiteestä, mutta toimenpiteen vaikutuksen suoraviivaisen päättelemisen esteenä on vertailutapausten erilaisuus. Kvasikokeellisessa asetelmassa on toki mahdollista oppia kokeellisen toimenpiteen seurauksista, mutta päättelyn tueksi tarvitaan vahvempia teoreettisia taustaoletuksia,

sillä ajatusta tapausten suorasta vertailusta ei voida toteuttaa. Kvasikokeen käsite on kuitenkin tärkeä, sillä suurin osa erilaisista yhteiskunnallisista kokeiluista on luonteeltaan kvasikokeita, ainakin jos niihin liittyy systemaattinen havainnointi ennen kokeellista toimenpidettä, sen aikana ja sen jälkeen.

Edellä kuvatut koeasetelmat ovat ideaalityyppisiä kuvauksia. Usein todelliset koeasetelmat ovat vaikeasti luokiteltavia ja sijoittuvat näiden ideaalityyppien välimaastoon. Millaisia kokeita ovat esimerkiksi politiikan tutkijoiden survey-kokeet (Sniderman 2011)? Niissä lähtökohdalla on surveyn tilastollisesti edustava otos väestöstä, mutta kysymysten muotoilua tai järjestystä muunnellaan systemaattisesti, jotta saataisiin selville, millä tavalla nämä vaikuttavat ihmisten ilmaisemiin asenteisiin. (Sosiologinen esimerkki survey-kokeesta on Phelan, Link & Feldman 2002). Asetelma vaikuttaa tieteellisesti mielekkäältä, mutta ei oikein istu edellä kuvattuihin vaihtoehtoihin. Vastaava tilanne nousee esiin *kokeellisen filosofian* kohdalla (Knobe & Nichols 2008). Kokeellisten filosofien kokeet ovat – ainakin yhteiskuntatieteilijän silmissä – kyselytutkimusta, jossa ihmiset reagoivat kuviteltujen tilanteiden kuvauksiin, mutta jossa ei (ainakaan vielä) kiinnitetä huomiota koehenkilöiden muodostaman otoksen edustavuuteen. Ehkä puhe kokeellisuudesta on näiden tutkimustapojen kohdalla liioittelua, mutta selvästi ne tuottavat kiinnostavaa empiristä aineistoa, jonka avulla voidaan vaikkapa ymmärtää, kuinka kysymyksenasettelu vaikuttaa kyselytutkimuksen vastauksiin, tai kyseenalaistaa analyttisten filosofien perinteinen tapa käyttää ajatuskokeiden tuottamia käsitteellisiä intuitioita todistusaineistona.

Kokeiden tiedollinen tehtävä

Toinen kiinnostava kokeellisen tutkimuksen ulottuvuus koskee kokeiden tiedollista roolia. Tieteen filosofian oppikirjoissa esitetään usein kokeellisen tutkimuksesta erittäin teoriakeskeinen käsitys,

jonka mukaan kokeellisen tutkimuksen lähtökohdalla on teoria ja siitä johdettu hypoteesi. Tässä näkemyksessä koeasetelma suunnitellaan vastaamaan teorian kuvaamaa tilannetta, ja kokeen tarkoitus on testata teoriasta johdetun hypoteesin paikkansapitävyyttä. Jos hypoteesin mukainen ennuste toteutuu, saa teoria tukea. Jos ennuste ei toteudu, on teoria ongelmissa, tai ainakin kokeen suunnittelussa tai toteuttamisessa on tapahtunut jokin virhe. Tässä oppikirjanäkemyksessä kokeellinen havainnointi on alisteista teorian kehittämiseksi ja testaamiseksi. Tieteellisen tutkimuksen todellisuus on huomattavasti monimutkaisempi.

Tarkastellaan esimerkkinä yhtä behavioraalisen taloustieteen kaikkein suosituimmista koeasetelmista: uhkavaatimuspeleä (*ultimatum game*). Tämän pelin perusmuodossa on kaksi toisilleen tuntematonta pelaajaa ja summa rahaa. Kukin pelaaja toimii yhdessä kahdesta mahdollisesta roolista. Tarjoaja tekee ehdotuksen summan – oletetaan sen olevan 10 euroa – jakamisesta pelaajien kesken. Vastaajan tehtävänä on joko hyväksyä tai hylätä tämä tarjous. Jos hän hylkää tarjouksen, kumpikaan ei saa mitään. Jos vastaaja hyväksyy tarjouksen, jaetaan rahat pelaajien kesken tarjoajan ehdotuksen mukaisesti. Pelitilanne on täysin anonymi: kumpikaan pelaaja ei tiedä kuka on vastapuolena, eivätkä pelaajat kohta toisiaan enää toistamiseen. Pelitilanne on siis erittäin yksinkertainen ja molemmat pelaajat tuntevat pelin säännöt. Rationaalisen valinnan teorian kannalta on selvää, millaista jakoa tarjoajan tulisi ehdottaa: hänen tulee tarjota vastaajalle pienintä mahdollista osuutta. Yhtä lailla on selvää, mitä vastaajan tulisi tehdä: hänen tulee hyväksyä mikä tahansa tarjous, jossa hänelle tarjotaan jotakin. Pelitilanne on myös erittäin yksinkertainen, joten ei ole mitään syytä olettaa, että ihmisillä olisi vaikeuksia päätyä tähän teorian kannalta optimaaliseen ratkaisuun.

Ensimmäisissä uhkavaatimuspeleissä (Günth, Schmittberger & Schwarz 1982) testattiinkin tätä

ennustetta, ja todettiin se selkeästi vääräksi. Myöhemmät sadat koesarjat tukevat tätä havaintoa: tyypillisin tarjous on tasajako, ja keskimäärin hyväksyjälle tarjotaan 40–45 prosentin osuutta. Rationaalisen valinnan teoria ei siis ennusta ihmisten käyttäytymistä kovin hyvin, eikä se pysty selittämään, miksi käyttäytyminen poikkeaa ennustetusta. Sivuutan tässä yhteydessä kiistat siitä, missä rationaalisen valinnan teoria menee vikaan, ja keskityn siihen, mitä uhkavaatimuspeiliä soveltavassa kokeellisessa tutkimuksessa on tapahtunut myöhemmin.

Ensimmäisten teorian kyseenalaistavien tulosten jälkeen koeasetelmaa ryhdyttiin innokkaasti toistamaan ja muuntelemaan, jotta saataisiin selville, onko kyseessä vain virheellisestä koejärjestelystä johtuva tulos tai muutoin ei-yleistettävä, jollakin tavoin idiosynkraattisen asetelman tuottama poikkeuksellinen tilanne. Nämä kokeilut osoittivat, että tulos on toistettavissa ja erittäin vakaa. Vakaus tarkoittaa tässä yhteydessä sitä, että muutokset asetelmassa eivät heilauta tuloksia rationaalisen valinnan teorian ennusteiden mukaisiksi. Erilaiset koeasetelman muunnokset eivät kuitenkaan aina tuottaneet samanlaisia tuloksia vaan osoittautui, että on erilaisia systemaattisia tapoja, joilla voidaan vaikuttaa tarjousten reiluuteen ja niiden hyväksymiseen. Tässä joitakin näistä havainnoista (ks. esim. Bicchieri 2005; Camerer 2003):

- jos tarjoaja on ihmisen sijasta tietokone, ovat vastaajat valmiita hyväksymään pienempiä tarjouksia
 - jos tarjoajalla ei ole vaihtoehtoja, ovat vastaajat valmiita hyväksymään pienempiä tarjouksia
 - jos panokset ovat suuria, ei tämä merkittävästi vaikuta tarjouksiin, mutta hyväksyttävän tarjouksen kynnys laskee hieman
 - jos pelitehtävät on jaettu arvan sijasta huutokaupalla tai tietokilpailun palkintoina, ovat tarjoukset pienempiä
 - jos panokset ovat arvokkaampia tarjoajalle kuin vastaajalle (eikä vastaaja tiedä tätä), tarjoukset heijastavat pikemminkin sitä, mikä vastaajan kannalta näyttäisi reilulta, kuin mikä olisi tasajako tarjoajan laajemman tiedon perusteella
 - jos peliasetelma muunnetaan diktaattoripeliksi, jossa vastaaja ei voi hylätä tarjousta, vastaajalle tarjottava summa pienenee; useimmiten hänelle tarjotaan silti jotakin (keskimäärin 20–30%) mutta vähemmän kuin uhkavaatimusasetelmassa
- Tätä listaa voisi jatkaa varsin pitkään, sillä valtaosa uhkavaatimusasetelmaan liittyvästä tutkimuksesta liittyy tällaisten peliasetelman muutosten vaikutusten kartoittamiseen. Näitä kokeita voi toki edelleen kuvata hypoteesien testaamiseksi, mutta olennaista on, että hypoteesien lähtökohtana ei ole enää rationaalisen valinnan teorian kaltainen yleinen teoria, vaan testattavat hypoteesit nousevat arkijärkeilystä ja kokeiden tekemisen yhteydessä tehdyistä havainnoista. Tällaisen tutkimuksen tarkoituksena ei olekaan yleisen teorian testaaminen, vaan pikemminkin
- jos tarjoajalle esitetään vastaajan valokuva, ovat tarjoukset hieman parempia
 - jos tarjoaja ja vastaaja ovat henkilöiden sijasta ryhmiä, ovat tarjoukset matalampia, ja vastaajat hyväksyvät alhaisempia tarjouksia; ryhmän koolla ei ole merkitystä
 - jos pelaajat ovat pieniä lapsia, ovat he valmiita hyväksymään alhaisempia tarjouksia

eräänlainen *ilmiön kartoittaminen*, jolla tarkoitan pyrkimystä löytää sellaisia tekijöitä, joiden avulla voidaan ennakoitavalla tavalla muokata kokeen lopputulosta. Tällaisessa tutkimuksessa kokeen tekijän saavutus onkin teorian testaamisen sijasta kehittää ja jalostaa toistettavia ja vakaita ilmiöitä, joiden kokoelmat – Francesco Guala (2005) puhuu *ilmiöiden kirjastoista* – ovat jotakin sellaista, jota ilmiöjoukkoa koskevan teorian tulisi kyetä selittämään.² Olennaista on, että tällaisen yleisemmän teorian puuttuminen ei ole este kokeellisen tutkimuksen jatkamiselle ja uusien ilmiöiden löytämiselle.

Kokeellinen ilmiöiden kartoitus perustuu siis olemassa olevan koeasetelman muunteluun sekä toistettavien ilmiöiden tuottamiseen ja mittaamiseen. Tutkimuksen jatkuessa riittävän pitkään koeasetelman ja sen muuntelemisen mahdollisuudet tulevat tutkijoille tutuiksi. Tämä mahdollistaa ”hyvin käyttäytyvän” koeasetelman uudenlaisen käytön: *koeasetelman käytön mittaussvälineenä* (Guala 2008). Siinä missä ilmiöiden kartoittamisessa keskeistä on koeasetelman muuntelu, tässä kokeiden käyttötavassa olennaista on samana pidetyn koeasetelman soveltaminen vaihtelevissa olosuhteissa. Tällöin tutkimuksen tarkoituksena on tuoda esiin yksilöiden, ryhmien tai erilaisten tilanteiden välisiä eroja. Esimerkki uhkavaatimuspelin käytöstä mittaussvälineenä on Joseph Henrichin ja kumppaneiden laaja kulttuureja vertaileva tutkimushanke (Henrich et al. 2005). Sen ideana oli tutkia, pätevätkö (pääasiassa) länsimaisilla yliopisto-opiskelijoilla tehdyt havainnot myös pienissä yhteisöissä elävien metsästäjä-keräilijöiden ja paimentolaisten kohdalla.

Tulokset olivat erittäin yllättäviä: siinä missä kaikissa maissa yliopisto-opiskelijoiden tyypilliset tarjoukset olivat hyvin samankaltaisia, löytyi

antropologien uusissa kokeissa erittäin suurta hajontaa sekä tehtävien tarjousten suuruudessa että siinä, millaisia tarjouksia ihmiset olivat valmiita hyväksymään. Olennaista tässä yhteydessä on, että tällaisten havaintojen tuottaminen edellyttää koeasetelman pitämästä mahdollisimman samankaltaisena. Asettamalla eri ryhmiin kuuluvat ihmiset samanlaisen tehtävän eteen voidaan tehdä kiinnostavia vertailevia havaintoja. Tässäkään tutkimusasetelmassa ei olla varsinaisesti testaamassa tiettyä yleistä teoriaa, vaan pikemminkin luomassa havaintoja ilmiöistä, jotka kaipaavat selitystä.

Miksi esimerkiksi jotkin Papua-Uuden-Guinean ryhmät (*Au* ja *Gnau*) tarjosivat vastaajalle yli puolta potista, mutta nämä ehdotukset tavallisesti hylättiin? Kokeen suorittaneiden antropologien mukaan taustalla on paikallinen käytäntö luoda lahjojen avulla kiitollisuudenvelkoja. Tarjoajat tulkitsivat tilanteen mahdollisuudeksi luoda vastapuolella kiitollisuudenvelka, kun taas vastaajat vuorostaan katsoivat parhaaksi välttää tällaisen rasitteen ottamista. Tässä valossa koeasetelma onkin keino valaista eri ryhmien sosiaalisten normien ja tilanteiden tulkinnallisten kehysten eroja.³

Kokeellisen tutkimuksen mittausvälinetulkinta soveltuu myös imitaatiopeliin (Collins & Evans 2014), jota Ilkka Arminen ja Mika Simonen käyt-

3 Yksittäisten havaintojen lisäksi tämän kaltaiset vertailevat tutkimukset toki myös asettavat kyseenalaiseksi suuressa osassa psykologista tutkimusta tehdyn oletuksen siitä, että länsimaisilla yliopisto-opiskelijoilla tehdyt kokeelliset havainnot voidaan vaivattomasti yleistää myös muihin ryhmiin. Henrich, Heine ja Norenzayan (2010) esittävätkin, että yksinomaan WEIRD (*Western, Educated, Industrialized, Rich and Democratic*) -ryhmistä tehdyt tutkimukset johtavat yhteiskunnallisen ja yksilöllisen muuntelun vähättelyyn sekä harhaanjohtaviin päätelmiin kaikille ihmisille tyypillisistä psykologisista piirteistä ja prosesseista. Jos tämä kritiikki pitää paikkansa, täytyy kokeellisen tutkimuksen käytäntöjä muuttaa tavalla, joka korostaa sosiologisen ja antropologisen tiedon merkitystä.

2 Havaintoaineiston ja ilmiön käsitteistä ks. Kiikeri & Ylikoski 2004, luku 2.

tävät aineistona tämän numeron kolmannessa artikkelissa. Imitaatiopelin koeasetelma ei ensisijaisesti testaa tiettyä teoreettista hypoteesia, eikä se jäljitä koeasetelman muutosten vaikutuksia kiinnostuksen kohteena olevaan vasteeseen. Pikemminkin se tuottaa havaintoja ryhmien välisistä eroista, jotka sitten nousevat selittämistä kaipaaviksi ilmiöiksi.

Sama koeasetelma voi siis toimia teorian testinä, selitettävien ilmiöiden kartoittamisen välineenä sekä erojen havaitsemiseen käytettävänä instrumenttina. Näiden tehtävien lisäksi kokeella ja sen tuloksilla saattaa olla vielä yksi tehtävä, joka on helpoiten havaittavissa tutkimustuloksia popularisoivissa julkaisuissa. Siinä kokeet toimivat jonkinlaista moraalista opetusta dramatisoivina esimerkkeinä. Koetulosten väitetään paljastavan jotakin ”syvällistä” ihmisluonnosta. Augustine Brannigan toteaaakin kokeellisen sosiaalipsykologian kritiikissään, että ”suuri osa sosiaalipsykologiasta on kokeelliseksi tutkimukseksi naamioitua moralioppia” (Brannigan 2004, 18, oma käännös).

Branniganilla on erityisesti mielessään Stanley Milgramin 1960-luvun auktoriteettikokeet ja Philip Zimbardon 1970-luvulla toteuttamat vankilakokeet, jotka kaikista tieteellisistä heikkouksistaan huolimatta ovat edelleen jatkuvan huomion kohteena. Tässä käytössä olennaista ei ole kokeen tulosten käyttö todistusaineistona jonkin tietyn väitteen puolesta, vaan kokeen käyttö havainnollistuksena ja mieleenpainuvana opetuksena. Tähän tarkoitukseen soveltuvat parhaiten yksinkertaiset stilisoidut tarinat, jotka sivuuttavat tutkimuksen metodologiset yksityiskohdat ja niiden mahdolliset heikkoudet. Kokeiden käyttö moraalisisina allegorioina ei rajoitu vain menneiden aikojen sosiaalipsykologiaan, vaan on vahvasti läsnä myös monissa kokeellisen peliteorian esityksissä. Tällöin uhkavaatimuspele ei vain kerro, miten ihmisen käyttäytyvät hyvin

erikoislaatuudessa tilanteessa, vaan paljastavat jotakin yleisemmin pätevää inhimillisen toiminnan perimmäisistä motiiveista. Tällainen ihmisluonnon pohdinta on toki lähtemätön osa ihmistieteitä, mutta on syytä varoa, ettei siitä muodostu korviketta koetulosten systemaattiselle teoreettiselle jäsentämiselle.

Laajennettu kokeellisen tutkimuksen luonnehdinta

Perinteinen hypoteettis-deduktiivinen tulkinta kokeellisesta tutkimuksesta ei siis anna oikeaa kuvaa kokeellisen tutkimuksen mahdollisuuksista. Toki joissakin tilanteissa kokeellinen asetelma soveltuu erityisen hyvin teoreettisen hypoteesin koettelemiseen – koeasetelma voidaan rakentaa juuri teoreettisten oletusten mukaiseksi ja sopivan yksinkertaiseksi, mikä tekee ennusteen testaamisesta huomattavan suoraviivaista. Tämä ei kuitenkaan ole ainoa tai edes merkittävin käyttö kokeelliselle asetelmalle yhteiskuntatieteissä. Tarvitaan siis laajempi määritelmä, joka ottaa paremmin huomioon kokeellisen tutkimuksen koko kirjon.

Avain osuvampaan luonnehdintaan on huomion keskittäminen kokeellisen kontrollin ja satunnais-tamisen mahdollistamiin luotettaviin päätelmiin kokeellisen toimenpiteen vaikutuksista. Suosiollisuus kausaalipäätelmille on keskeinen kokeellisen tutkimuksen ansio ja tärkein syy sille, miksi sitä pidetään keskeisenä tieteellisen menetelmän elementtinä. Kokeellisen tutkimuksen anti ei kuitenkaan rajoitu kokeellisen toimenpiteen kohteena olevan muuttujan ja vastemuuttujan välisen kausaalisen yhteyden paljastamiseen. Koe ei ole välttämättä arvoton, vaikka kaikki satunnaistetun kontrolloidun kokeen ideaaliset ehdot eivät toteutuisikaan. Kaikki kokeelliset tutkimusasetelmat eivät edes pyri tällaiseen kausaalisten vaikutusten mittaamiseen. Esimerkiksi käyvät vaikkapa

useimmat tässä numerossa käsitellyt koeasetelmat.

Laajemman luonnehdinnan lähtökohtana on ajatus, että kokeellinen tutkimus tuottaa uutta havaintoaineistoa. Usein uusi havaintoaineisto koskee olosuhteita, joita ei esiinny luodun koeasetelman ulkopuolella, tai jotka ovat muulla tavoin harvinaisia. Tämän vuoksi kokeellinen tutkimus mahdollistaa ajoittain aivan uusien ja yllättävien ilmiöiden löytämisen. Eristetyn kokeellisen asetelman muokkaaminen puolestaan mahdollistaa ilmiöiden ja niiden taustalla olevien mekanismien toiminnan systemaattisen kartoittamisen. Näin voidaan esimerkiksi tutkia, kuinka vakaa tai herkkä tutkittava ilmiö on taustatekijöiden muutoksille. Havaintoaineiston tuottaminen mitä erikoislaatuisimmista olosuhteista ei tietenkään ole itseisarvo. Kokeellisen tutkimuksen merkitykseksi riittää, että sen tuottamat havainnot ajoittain johtavat uusien teoreettisten oivallusten syntymiseen tai mahdollistavat kilpailevien teorioiden vertailun.

Toinen keskeinen kokeellisesti tuotetun havaintoaineiston tärkeä piirre on se, että havaintoaineiston syntyy johtaneet prosessit ovat yleensä paremmin tiedossa ja ymmärrettyjä kuin ei-kokeellisissa havainnoissa. Tämän ymmärryksen perustana on kokeellinen kontrolli ja toistuvat havainnot muunnelluista olosuhteista. Ymmärrys havaintoaineiston tuottaneista prosesseista lisää kyseisen havaintoaineiston luotettavuutta. Asian mahdollinen kääntöpuoli voi olla se, että kokeellisten havaintojen relevanssi on pienempi. Koeasetelma ei välttämättä mittaa kiinnostuksen kohteena olevaa asiaa (heikko sisäinen validiteetti) tai havaintoja on vaikea yleistää koeasetelman ulkopuolelle (heikko ulkoinen validiteetti). Mutta sikäli kuin sisäinen ja ulkoinen validiteetti ovat kohdallaan, parempi ymmärrys havaintoaineiston tuottaneista prosesseista on kiistaton tiedollinen ansio.

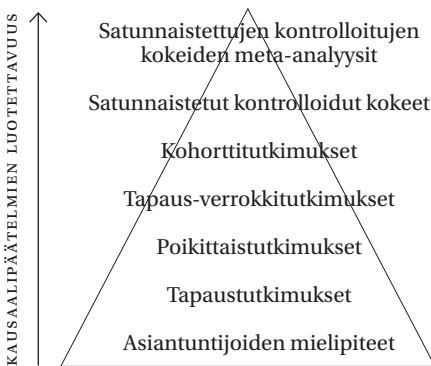
Tämän laajennetun kokeellisen tutkimuksen luonnehdinnan ansio on, että sen kautta kokeellinen tutkimus ja ei-kokeellinen tutkimus muodostavat pikemminkin jatkumon kuin kaksi erillistä luokkaa. Kokeellinen asetelma, joka toteuttaa satunnaistetun kontrolloidun kokeen vaatimukset on tässä näkemyksessä pikemminkin erityistapaus kuin laajempaa joukkoa edustava prototyyppi. Tämä on tärkeää, sillä valtaosa kokeellisesta yhteiskuntatieteellisestä tutkimuksesta sijoittuu jonnekin tiukan kontrolloidun kokeen ja etnografisen havainnon tai rekisteriaineiston analyysin väliin. Satunnaistetun kontrolloidun kokeen ideaalituypille rakentuva ymmärrys kaikesta kokeellisesta tutkimuksesta tuottaa harhaista ymmärrystä tutkimuksellisesta näytöstä, kuten tulemme kohta näkemään. Esittämäni laajennettu luonnehdinta tekee myös paremmin oikeutta tutkimusasetelmille, joissa on joitakin, mutta ei kaikkia kokeellisen tutkimuksen oluttuvuuksia.

Näyttöhierarkia

Laajennettu kokeellisen tutkimuksen luonnehdinta antaa välineitä tarkastella näyttöön perustuvasta lääketieteestä (Howick 2011) peräisin olevaa *näyttöhierarkian* käsitettä, joka on viime aikoina alkanut välittyä myös yhteiskuntapoliittiseen ja -tieteelliseen keskusteluun (Cartwright & Hardie 2012). Näyttöön perustuvan lääketieteen perusidea on suosia sellaisia hoitoja, joiden vaikuttavuudesta on näyttöä. Lähtökohta on mielekäs: kun resurssit ovat rajalliset, niitä ei kannata tuhata hoitoihin, joilla ei ole vaikutusta potilaan terveydentilaan. Toisaalta potilaalla on oikeus saada sellaista hoitoa, jolla on tosiasiallinen mahdollisuus parantaa hänen vointiaan. Mutta mitä on näyttö ja kuinka arvioida sitä johdonmukaisesti?

Näyttöhierarkia (Kuvio 1) on ratkaisuyritys tähän ongelmaan. Koska kiinnostuksen kohteena ovat hoito-toimenpiteiden vaikutukset, nousee satunnaistettu

kontrolloitu koe näyttöhierarkian keskeiseksi elementiksi, sillä se suoraviivaisesti testaa tietyn toimenpiteen kausaalisia vaikutuksia. Näin se asettuu hierarkiassa korkeammalle kuin epidemiologiset ja muut tilastolliset tutkimukset, tapaustutkimukset, asiantuntijamielipiteet tai lääketieteelliset anekdootit. Mutta koska yksittäiset kokeelliset tutkimukset saattavat olla huonosti suunniteltuja, epäonnistuneita tai tilastolliselta voimaltaan liian heikkoja, nousevat näyttöhierarkian huipulle systemaattiset satunnaistettujen kokeiden meta-analyysit. Meta-analyysi mahdollistaa vahvemmat tilastolliset johtopäätökset ja pienentää puhtaan tilastollisen satuman vaikutuksia johtopäätöksiin (on tilastollisesti odotettavissa, että osa tutkimuksista tuottaa tilastollisesti merkittäviä tuloksia, vaikka tutkittavalla toimenpiteellä ei olisi mitään vaikutuksia).



KUVIO 1. NÄYTTÖHIERARKIA.

Näyttöhierarkian ajatus on kiistanalainen lääketieteessä ja siitä onkin keskusteltu laajasti. Kritiikin seurauksena alkuperäisiä muotoiluja on jo aika lailla lievennetty. Keskityn tässä ainoastaan siihen, kuinka ajatus näyttöhierarkiasta, ja erityisesti kokeellisen tutkimuksen tiedollisesta erityis asemasta, soveltuu yhteiskuntatieteisiin.

Koko näyttöhierarkian ajatuksen keskeinen ongelma on hierarkian perustana olevan idealisoi-

dun koeasetelman ja tutkimuksissa tosiasiallisesti toteutuvien asetelmien välinen kuilu. Ero on merkittävä jo lääketieteellisessä tutkimuksessa, mutta se kärjistyy siirryttäessä yhteiskuntatieteisiin. Otetaan esimerkiksi lumetoimenpiteisiin (Benedetti 2009) liittyvät ongelmat. Ideaalisessa koeasetelmassa koehenkilöt eivät pysty päättämään, ovatko he saaneet oikeaa vai lumelääkettä. Todellisuudessa koehenkilöt saattavat esimerkiksi oppia tunnistamaan, kumpaa lääkettä he saavat. Tämä voi tapahtua vaikkapa aidon lääkkeen vaikutusten tai sivuvaikutusten kautta. Jos henkilö uskoo saaneensa lumelääkettä, on olettavasti lääkkeen lumevaikutus pienempi, ja henkilö myös todennäköisemmin kokeilee myös muita hoitoja vaivaansa. Molemmat seikat tietenkin heikentävät kokeen pohjalta tehtäviä päätelmiä lääkkeen vaikuttavuudesta. Vastaavat vaikeudet yhteiskuntatieteissä saavat oikean mittakaavan, kun huomataan, että yhteiskuntatieteellisessä tutkimuksessa on usein mahdotonta edes löytää mitään vastinetta lumetoimenpiteelle. Millaista olisi vaikkapa lumesosiaaliturva, tai millainen olisi lumekouluavustaja? Vastaavan lumetoimenpiteen olemassaololla ei tietenkään aina ole merkitystä, mutta se havainnollistaa, kuinka kaukana tyypillinen yhteiskuntatieteellinen koe on satunnaistetun kontrolloidun kokeen ideaalista.

Vastaavia ongelmia liittyy oletukseen kokeellisten toimenpiteiden vaikutusten erillisyydestä (Sampson 2010). Siinä missä lääketieteellisessä kokeessa voidaan useimmiten olettaa, että koe-ryhmän jäsenen saama lääkitys ei vaikuta verrokkiryhmän jäsenen terveyteen, on tämä oletus usein ongelmallinen yhteiskuntatieteellisissä kenttäkokeissa. Jos esimerkiksi koeasetelmassa yhden kunnan asukkaille annetaan terveysvinkkejä, ei liene kaukaa haettua olettaa, että nuo terveysvinkit päätyvät myös ainakin joidenkin verrokkiryhmänä toimivan naapurikunnan asukkaiden tietoon (ks. Warpenius tässä numerossa). Ihmiset ovat keskenään vuorovaikutuksessa, min-

kä vuoksi on usein vaikeaa pitää koe- ja verrokki-ryhmät erillään.

Samanlainen ongelmakimppu muodostuu satunnaistamisesta. Satunnaistaminen ei ole lääketieteellisissäkään tutkimuksissa aina ongelmatonta, mutta yhteiskuntatieteellisissä sovellutuksissa ongelmat ovat jälleen kertaluokkaa suurempia: aito satunnaistaminen ei välttämättä ole käytännössä mahdollista, tai sen perustaksi ei ole riittävä tilastollista tietoa, joka mahdollistaisi yleistämisen laajempaan väestöryhmään. Siinä missä lääketieteessä meta-analyysi (Bohlin 2012) tarjoaa mahdollisuuden korjata yksittäisten kokeiden rajoituksia ja vinoumia, on sen käytettävyys yhteiskuntatieteissä huomattavasti rajoitetumpaa. Ensinnäkin aineistona käytettäviä kokeita on tavallisesti vähemmän johtuen niiden kalleudesta ja toteuttamisen vaikeudesta. Toiseksi yksittäiset kenttäkokeet saattavat erota toisistaan suuresti niin tutkittavien toimenpiteiden kuin tutkimuspaikankin suhteen. Tämä vertailtavuuden puuttuminen vaikeuttaa meta-analyysin pohjalta tehtäviä johtopäätöksiä. Kolmas ongelma on kokeellisten asetelmien epäedustavuus: koeasetelman rakentamisen vaatimukset saattavat johtaa siihen, että kokeellisia tutkimuksia on mahdollista toteuttaa vain olosuhteissa, jotka eivät kovin hyvin vastaa niitä olosuhteita, joista halutaan tehdä päätelmiä.

Näiden huomioiden tarkoituksena ei ole osoittaa, että kokeellinen todistusaineisto on arvotonta yhteiskuntatieteissä. Argumenttini kärki kohdistuu pikemminkin ajatukseen näyttöhierarkiasta ja sen soveltamisesta yhteiskuntatieteelliseen tietoon. Jo lääketieteen kohdalla havaitaan merkittävä erotodellisen tutkimuksen ja hierarkian perustana olevien ideaalimallien välillä, ja yhteiskuntatieteiden kohdalla tämä kuilu on vielä suurempi. Vaikka siis hyväksyisimme ajatuksen, että ihanteellinen koeasetelma on luotettavin tapa hankkia kausaalista tietoa, tulee suhtautua erittäin

varauksellisesti ajatukseen, että satunnaistettu kontrolloitu koe olisi jonkinlainen yhteiskuntatieteellisen näytön ”kultastandardi”. Käytännössä tämä tarkoittaa sitä, ettei meidän tulisi hyväksyä kriittikittömästi esimerkiksi seuraavia väittämiä:

- ainoastaan kokeellinen näyttö tukee kausaalipäätelmiä
- kokeellinen näyttö on aina parempaa kuin muu näyttö
- kokeellinen näyttö riittää yksin (kiinnostaviin) kausaalisiin päätelmiin
- jos kokeellista näyttöä on saatavilla, muu näyttö voidaan sivuuttaa

Ensimmäinen väittämä nojautuu vaaralliseen ”näyttöpurismiin”. Onnistuneesti toteutetun kokeellisen asetelman tuottama näyttö on ehkä vakuuttavampaa, mutta tämä ei tee muusta näytöstä arvotonta. Tällainen johtopäätös edellyttäisi riippumattomia argumentteja. Toisen väittämän uskomus kokeellisen näytön yliveraisuudesta puolestaan perustuu oletukseen ideaalisten kokeiden ominaisuuksien soveltumisesta myös tosiasiallisiin kokeisiin. Tällöin kokeiden toteuttamisen käytännön yksityiskohdat unohtetaan. Kolmas ja neljäs väittämä puolestaan sivuuttavat koeasetelmasta yleistämiseen tarvittavan taustatiedon roolin ja sen, että usein toteutettavissa olevat koeasetelmat ovat varsin kaukana niistä tilanteista, joista varsinaisesti ollaan kiinnostuneita. Kokeellisten puristien (tuttavallisesti *randomistas*) argumenttien taustalla on tieteenfilosofisesti erittäin pulmallinen ajatus, että kokeellinen tutkimus on jollakin ratkaisevalla tavalla teoriavapaata. Kokeellinen tutkimus ei toki ole vapaata taustaoletuksista, vaikka sen taustaoleukset eroavatkin perinteisen havaintoaineistoon perustuvan tilastollisen analyysin oletuksista

(Sampson 2010). Taustaoletusten uskottavuus tulleekin arvioida tapauskohtaisesti ja pitää silmällä sitä, millaisia päätelmiä aineistosta oikeastaan halutaan tehdä. Tästä syystä perinteinen epidemiologinen todistusaineisto saattaa hyvinkin tarjota paremman perustan kausaalille päätelmille kuin poikkeusolosuhteissa hankittu kokeellinen aineisto. Jos hyväksymme tämän, hylkäämme ajatuksen tiukasta näyttöhierarkiasta.

Tätä taustaa vasten näyttöhierarkian ajatusta voi-kin pitää pikemminkin toiveajattelun muotona kuin uskottavana todistusaineiston arvioinnin välineenä. Taustalla on periaatteellinen ongelma: heterogeenisen tieteellisen näytön arvon ja vakuuttavuuden arviointiin ei ole olemassa kiistatonta ja yhtenäistä menetelmää (Cartwright & Hardie 2012). Osittain tämä perustuu jo siihen, että näytön laatua ei voi arvioida ennen kuin tiedetään, *minkä puolesta* sen pitäisi toimia näyttönä. Riippuen tarkoituksesta muutoin laadukas näyttö voi hyvinkin olla puhtaasti irrelevanttia tai vaihtoehtoisesti mahdotonta tuottaa. Tämä ei tarkoita, etteikö näyttöä voisi arvioida järkipärisesti. Näytön arvioinnin järkipärisuus ei vain palaudu kaavamaiseen näyttöhierarkiaan.

Kokeellista näyttöä liiaksi painottava näyttöpurismi on uhka yhteiskuntaa koskevan tiedon kehitykselle. Vaikka se lisää kokeellisen tutkimuksen määrää – mikä sinänsä on hyvä asia – on sillä myös haitallisia vaikutuksia. Ensinnäkin, jos hyväksyttävä näyttö rajataan ainoastaan kokeelliseen tutkimukseen, rajoittaa tämä käytettävissä olevaa näyttöä dramaattisesti ilman erityisen hyvää perustelua. Tämä jos mikä olisi näyttöön perustuvan ajattelun perusidean vastaista. Toiseksi, jos tutkijat keskittyvät vain kokeellisiin asetelmiin, ohjautuu suuri osa tutkimuksesta sen perusteella, mihin kokeelliset asetelmat soveltuvat, eikä sen perusteella, mikä olisi käytännön sovellutusten kannalta hyödyllisesti tai teoreettisesti kiinnostavaa. Tällainen tutkimuksen yksipuolistaminen ei palvele tieteen tiedollisia

päämääriä. Vastaavalla tavalla tutkimusten tuottama tieto voi muuttua vähemmän kiinnostavaksi: esimerkiksi jos ei-kokeellisissa tutkimuksissa kaavamaisesti haetaan kysymyksenasetteluja, joihin voitaisiin soveltaa luonnollisen kokeen asetelmaa (esimerkiksi siksi, että tällaiset tutkimukset saa esille paremmilla julkaisufoorumeilla) voi olla, että huomattavasti kiinnostavampien muuttujien suhteet jäävät selvittämättä.

On myös syytä pitää mielessä politiikan ja tieteen päämäärien erillisuus. Vaikka kokeellinen asetelma periaatteessa toimii politiikkatoimenpiteiden vaikuttavuuden arvioinnissa, on yhtä lailla selvää, että suurin osa tällaisesta tutkimuksesta on vailla laajempaa yhteiskuntatieteellistä kiinnostavuutta. Poliittisen puolueen tai järjestön kannalta voi olla kiinnostavaa tutkia, aktivoiko äänestäjiä paremmin ehdokkaan kättely vai puhelinsoitto, mutta politiikan tutkijan tai sosiologin kannalta tällainen kysymys on kovin vähämerkityksinen. Kokeellisen käänteen huumassa onkin hyvä muistaa, että vaikka tieteellinen kiinnostavuus ja käytännön intressit käyvät joissakin tapauksissa yksiin, useimmiten tutkijaa kiinnostava kysymyksenasettelu on perustellusti erilainen kuin poliitikon tai muun toimeksiantajan.

Tuottaako kokeellinen tutkimus normaalitiedettä?

Näyttöhierarkian ohella toinen huono peruste kokeelliselle sosiologialle on uskomus, että kokeelliseen tutkimukseen siirtyminen muuntaa tieteenalan nopeasti kehittyväksi ja vakaasti edistyväksi (Collins 1994) kuhnlaisiksi normaalitieteeksi. Tieteenfilosofisten fantasioiden lisäksi tällaisen haaveen taustalla voi olla havainto tieteen rahoitusjärjestelmän toimintaperiaatteista: rahoittajat näyttävät suosivan lähestymistapoja, joka tuottavat nopeaan ja ennustettavaan tahtiin julkaitavan kokoisia tuloksia. Kun rahoitusjärjestelmä

on suunniteltu enemmän lääketieteellisen kuin antropologisen tutkimuksen mallin mukaiseksi, ehkä yhteiskuntatieteilijöidenkin tulisi muokata tutkimusstrategiaansa?

Tällainen haave on kuitenkin epärealistinen. Ensinnäkin kokeellisia tuloksia jatkuvalla syötöllä tuottavan ”tehtaan” toiminta edellyttää koeasetelman ympärille rakennettua toimivaa kokeellista järjestelmää (Tuunainen 2001). Tällaisen kokeellisen järjestelmän luominen vaatii usein huomattavasti aikaa, eikä sen onnistuminen ole millään tavalla taattu. Uuden kiinnostavan kokeellisen asetelman luominen onkin yllättävän vaikeaa – juuri tästä syystä valtaosa taloustieteellisestä ja sosiaalipsykologisesta kokeellisesta tutkimuksesta perustuu muutamaaan harvaan kokeelliseen paradigmaan. Suunniteltua koeasetelmaa ei ehkä edes pystytä muokkaamaan sellaiseksi, että se tuottaisi vakaita ja ennustettavia ilmiöitä (Kiikeri & Ylikoski 2004), jotka herättäisivät yleisempää mielenkiintoa. Lisäksi kokeellinen tietotaito karttuu pitkälti yrityksen ja erehdyksen kautta, ja siksi on vaikeaa ennustaa, kuinka pitkä aika menee julkaisukelpoisten tulosten saamiseen.

On totta, että onnistuessaan kokeellinen tutkimus tuottaa vakaan tiedollisen edistyksen vaikutelman: tutkimus tuottaa jatkuvan virran uusia ”löytöjä” ja kumottuja hypoteeseja. Tämän lisäksi menetelmät kehittyvät, tulosten arviointikriteerit tiukentuvat ja käsitteet täsmentyvät jatkuvaa tahtia. Tällainen edistys on kuitenkin pitkälti paikallista. Jos tarkastellaan esimerkiksi kokeellisen sosiaalipsykologian kehitystä pidemmällä aikavälillä, havaitaan tällaista paikallista edistystä. Sen sijaan laajamittaista tiedollista edistymistä löytyy vähemmän. Yksi syy tähän on se, että kokeellisen erikoistumisen seurauksena tapahtuu merkittävää horisontin kapenemista: tutkijat kyllä tuntevat erityisen kokeellisen asetelman ominaisuudet ja mah-

dollisuudet yhä paremmin, mutta samalla he usein menettävät näköpiiristään alkuperäistä asetelmaa motivoineet laajemmat kysymykset. Ja kun kokeellinen paradigma on viimein ammennettu tyhjiin – ts. se ei enää tuota kiinnostavia uusia tuloksia – se korvautuu pian uusilla asetelmilla. Hyvin vähän vanhan asetelman ympärille rakentuneesta tietotaidosta välittyä seuraaville tutkijasukupolville. Toisin sanoen erikoistuneen tutkimuslinjan tulokset ja saavutukset häviävät tieteenalan kollektiivisesta muistista. Onkin kiinnostavaa nähdä esimerkiksi, kuinka suurta osaa tällä hetkellä tehtävää kokeellisesta peliteoriasta tullaan pitämään parinkymmenen vuoden kuluttua yksinkertaisesti mielenkiinnottomana tai metodologisesti kyseenalaisena.

Tutustuminen viimeaikaisen kokeellisen psykologian ja sosiaalipsykologian toistettavuuskriisiin⁴ on hyvää vastalääkettä haaveille kokeellisesta normaalitieteestä. Kriisi on asettanut monet näiden alueiden keskeiset kokeelliset löydökset ainakin hetkellisesti kyseenalaisiksi. Kriitikkojen mukaan pitkälti tieteenalojen julkaisukäytäntöjen seurauksena on syntynyt vinoutunut käsitys esimerkiksi tiedostamattomia vaikutuksia koskevista tutkimuksista. Heidän mukaansa monet tutkimukset on toteutettu liian pienellä otoksella, eikä niitä ole asianmukaisesti toistettu. Ajatuksena on, että psykologien suosima käsitteellinen toistaminen (*replication*) ei riitä. Kokeet tulisi toistaa myös täsmällisesti samoina, jotta vältetään riski, että kyseessä on vain sattumalta tilastollisesti merkitseväksi nouseva tulos. Kriisin toinen elementti ovat lehtien julkaisukäytännöt: ne julkaisevat yleensä vain tilastollisesti merkitseviä tuloksia, ja vain harvoin negatiivisia tuloksia, mikä synnyttää julkaisuvinouman: julkaistut tutkimukset näyttävät johdonmukaisesti puhuvan ilmiön puolesta,

4 Ks. *Perspectives on Psychological Science* November 2012, 7:6.

vaikka tarkemmassa tarkastelussa kyseessä on vain tilastollisten analyysien tuottama artefakti.

Kokeellinen psykologia ja sosiaalipsykologia tulevat luultavasti selviämään tästä kriisistä muuttamalla tutkimus- ja julkaisukäytäntöjään. Kriisi kuitenkin osoittaa, kuinka vaikeaa kokeellinen tutkimus todellisuudessa on. Voi hyvinkin olla, että esimerkiksi kokeellinen politiikkatoimenpiteiden tutkimus ei selviäisi vastaavasta toistettavuuskriisistä. Toistettavuuskriisi ei kuitenkaan ole esimerkki kokeellisen tutkimuksen erityisestä heikkoudesta. Vastaava keskustelu analyysi- ja julkaisukäytäntöjen toimivuudesta on paikallaan kaikessa tilastollisia menetelmiä käyttävässä tutkimuksessa. Ja jos toistettavuuden ajatus otetaan vakavasti – kuten se pitäisi ottaa – olisi paikallaan käydä laajempaa keskustelua myös kvalitatiivisten analyysien toistettavuudesta. Näissä keskusteluissa kokeellisen psykologian nykyinen kriisi voi hyvinkin nousta esiin esimerkkinä siitä, kuinka tieteenala selviää menetelmällisiä perusteitaan koskevasta kriisistä.

Lopuksi

Olen tässä kirjoituksessa pyrkinyt laajentamaan perinteistä kuvaa kokeellisesta tutkimuksesta kuvaamalla sen moninaisuutta. Olen myös tarkastellut joitakin hyviä ja huonoja perusteita kokeelliselle sosiologiselle tutkimukselle. Kokonaisuutena kokeellisen tutkimuksen palaaminen yhteiskuntatieteisiin on nähdäkseni erittäin hyvä asia, ja kokeellista yhteiskuntatieteellistä tutkimusta tarvitaan lisää. Kokeellinen tutkimus ei tule korvaamaan muita menetelmiä ja aineistoja, mutta se on tärkeä lisä nykyisten menetelmien kokonaisuuteen. Sosiologien tulisikin pohtia kokeellisen tutkimuksen mahdollisuuksia omissa tutkimuskohteissaan ja seurata kokeellisen tutkimuksen kehitystä ja tuloksia lähitieteissä. Kokeellisella tutkimuksella – oli se sitten sosiologien

tai muiden tieteenalojen edustajien tekemää – on paljon annettavaa sosiologialle.

Kokeellisella tutkimuksella tulee olemaan aina omat rajoituksensa yhteiskuntatieteissä. Tästä syystä on hyödyllistä erottaa toisistaan kokeellinen ajattelutapa ja varsinaisten kokeiden tekeminen. Kokeiden toteuttaminen on usein hankalaa ja joskus mahdotonta, mutta ei tämä ole este kokeellisen ajattelutavan harjoittamiselle. Vaikka sosiologi käyttäisi muita kuin kokeellisia menetelmiä, voi hän aina ajatella tutkimuskohdettaan mahdollisten kokeiden kautta. Näitä kuviteltuja kokeita ei useinkaan voi käytännössä toteuttaa, mutta voi olla erittäin hedelmällistä miettiä, millaisin asetelmin omaa tutkimuskohdetta voisi periaatteessa tutkia. Kuvitteellisen koeasetelman konkreettinen ajattelu on sosiologisen mielikuvituksen käyttöä, joka edesauttaa oman teoreettisen hahmotuksen taustaoletusten artikulointia sekä sen mahdollisten mustien pisteiden paljastamista. Vain omassa mielikuvituksessa suunniteltujen ja toteutettujen kokeiden lopputuloksilla ei tietenkään ole varsinaista empiiristä näyttöarvoa, vaan näiden ajatuskokeiden merkitys ja arvo on teoreettisen ajattelun selkiytymisessä.

Vaikka ajatus kokeellisesta sosiologiasta saattaa tuntua aluksi vieraalta, ei sosiologien ole syytä maalata itseään nurkkaan jättämällä kaiken kokeellisen tutkimuksen muille yhteiskuntatieteilijöille. Jotta kokeellisen tutkimuksen mahdollisuudet voisivat toteutua täydessä mitassaan, sosiologien tulee myös kehittää omia kokeellisen tutkimuksen asetelmiaan, eikä vain esittää sosiologisia tulkintoja muiden tekemistä kokeista. Ei ole olemassa mitään tiedollista syytä määrittellä sosiologiaa vain ei-kokeellisia menetelmiä käyttäväksi yhteiskuntatieteeksi. Sosiologia on yleinen yhteiskuntatiede, joka ei ole sidottu mihinkään tiettyyn tutkimusmenetelmään. Sen historia

on uusien menetelmien ja aineistojen omaksumisen historiaa. Tästä näkökulmasta kokeelliset menetelmät ovat luonteva osa sosiologien laajenevaa menetelmäarsenaalia. Kokeellisten menetelmien omaksuminen on syytä nähdä haasteena sosiologiselle mielikuvitukselle: kuinka luoda sosiologisesti kiinnostavia kokeellisia tutkimusasetelmia?

Kirjallisuus

- ANGRIST, JOSHUA D. & JÖRN-STEFFEN PISCHKE. 2010. "The Credibility Revolution in Empirical Economics. How Better Research Design Is Taking the Con Out of Econometrics." *Journal of Economic Perspectives* 24:2, 3–30.
- BARRERA, DAVIDE & BRENT SIMPSON. 2012. "Much Ado About Deception Consequences of Deceiving Research Participants in the Social Sciences." *Sociological Methods & Research* 41:3, 383–413.
- BENEDETTI, FABRIZIO. 2008. *Placebo Effects. Understanding the Mechanisms in Health and Disease*. Oxford: Oxford University Press.
- BICCHIERI, CRISTINA. 2005. "The Grammar of Society. The Nature and Dynamics of Social Norms." Cambridge: Cambridge University Press.
- BOHLIN, INGEMAR. 2012. "Formalizing Syntheses of Medical Knowledge. The Rise of Meta-Analysis and Systematic Reviews." *Perspectives on Science* 20:3, 273–309.
- BONETTI, SHANE. 1998. "Experimental Economics and Deception." *Journal of Economic Psychology* 19:3, 377–395.
- BRANNIGAN, AUGUSTINE. 2004. *The Rise and Fall of Social Psychology. The Use and Misuse of the Experimental Method*. New Brunswick: Aldine Transaction.
- CAMERER, COLIN. 2003. *Behavioral Game Theory. Experiments in Strategic Interaction*. Princeton: Princeton University Press.
- CARTWRIGHT, NANCY & JEREMY HARDIE. 2012. *Evidence-Based Policy. A Practical Guide to Doing it Better*. Oxford: Oxford University Press.
- COLLINS, HARRY & ROBERT EVANS. 2014. "Quantifying the Tacit. The Imitation Game and Social Fluency." *Sociology* 48:1, 3–19.
- COLLINS, RANDALL. 1994. "Why the Social Sciences Won't Become High-Consensus, Rapid-Discovery Science." *Sociological Forum* 9:2, 155–177.
- CONLEY, DALTON, EMILY RAUSCHER, CRISTOPHER DAWES, PATRIK K. E. MAGNUSSON & MARK L. SIEGAL. 2013. "Heritability and the Equal Environments Assumption. Evidence From Multiple Samples of Misclassified Twins." *Behavior Genetics* 43:5, 415–26.
- COOK, KAREN S. & TOSHIO YAMAGISHI. 2008. "A Defense of Deception on Scientific Grounds." *Social Psychology Quarterly* 71:3, 215–221.
- CRUMP, MATTHEW J. C., JOHN V. MCDONNELL & TODD M. GURECKIS. 2013. "Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research." *PLoS ONE* 8(3):e57410.
- DIPRETE, THOMAS A. & GREGORY M. EIRICH. 2006. "Cumulative Advantage as a Mechanism for Inequality. A Review of Theoretical and Empirical Developments." *Annual Review of Sociology* 32:1, 271–297.
- DRUCKMAN, JAMES N., DONALD P. GREEN, JAMES H. KUKLINSKI & ARTHUR LUPIA. 2011. *Cambridge Handbook of Experimental Political Science*. Cambridge: Cambridge University Press.
- DUNNING, THAD. 2012. *Natural Experiments in the Social Sciences. A Design-Based Approach*. Cambridge: Cambridge University Press.
- GERBER, ALAN S. & DONALD P. GREEN. 2012. *Field Experiments. Design, Analysis, and Interpretation*. New York: W. W. Norton.
- GUALA, FRANCESCO. 2005. *The Methodology of Experimental Economics*. Cambridge: Cambridge University Press.
- GUALA, FRANCESCO. 2008. "Paradigmatic Experiments. The Ultimatum Game From Testing to Measurement Device." *Philosophy of Science* 75:5, 658–69.
- GUALA, FRANCESCO. 2012. "Reciprocity: Weak or Strong? What Punishment Experiments Do (and Do Not) Demonstrate." *Behavioral and Brain Sciences* 35:1, 1–15.
- GÜTH, WERNER, ROLF SCHMITTBERGER & BERND SCHWARZ. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization* 3:4, 367–388.
- HEDSTRÖM, PETER & PETRI YLIKOSKI. 2010. "Causal Mechanisms in the Social Sciences." *Annual Review of Sociology* 36, 49–67.
- HENRICH, JOSEPH ET AL. 2005. "Economic Man' in Cross-Cultural Perspective. Behavioral Experiments in 15 Small-Scale Societies." *Behavioral and Brain Sciences* 28:6, 795–815.
- HENRICH, JOSEPH, S. J. HEINE & A. NORENZAYAN. 2010. "The Weirdest People in the World?" *Behavioral and Brain Sciences* 33:2–3, 61–83.
- HOWICK, JEREMY. 2011. *The Philosophy of Evidence-Based Medicine*. Wiley-Blackwell.
- JACKSON, MICHELLE & D. R. COX. 2013. "The Principles of Experimental Design and Their Application in Sociology." *Annual Review of Sociology* 39:1, 27–49.
- KAGEL, JOHN K. & ALVIN E. ROTH (TOIM.) 1995.

- Handbook of Experimental Economics*. Princeton: Princeton University Press.
- KEIZER, KEES, SIEGWART LINDENBERG & LINDA STEG. 2008. "The Spreading of Disorder." *Science* 322:5908, 1681–85.
- KEUSCHNIGG, MARC & T. WOLBRING. 2015. "Disorder, Social Capital, and Norm Violation. Three Field Experiments on the Broken Windows Thesis." *Rationality and Society* 27:1, 96–126.
- KIIKERI, MIKA & PETRI YLIKOSKI. 2004. *Tiede tutkimuskohteena. Filosofinen johdatus tieteututkimukseen*. Helsinki: Gaudeamus.
- KNOBE, JOSHUA & SHAUN NICHOLS (TOIM.) 2008. *Experimental Philosophy*. Oxford: Oxford University Press.
- MORGAN, STEPHEN L. & CHRISTOPHER WINSHIP. 2014. *Counterfactuals and Causal Inference. Methods and Principles for Social Research*. 2nd ed. Cambridge: Cambridge University Press.
- OAKLEY, ANN. 2000. "A Historical Perspective on the Use of Randomized Trials in Social Science Settings." *Crime Delinquency* 46:3, 315–29.
- PEARL, JUDEA. 2009. *Causality. Models, Reasoning and Inference*. 2nd ed. Cambridge: Cambridge University Press.
- PHELAN, JO C., BRUCE G. LINK & NAUMI M. FELDMAN. 2013. "The Genomic Revolution and Beliefs About Essential Racial Differences. A Backdoor to Eugenics?" *American Sociological Review* 78:2, 167–91.
- SALGANIK, MATTHEW, PETER DODDS & DUNCAN WATTS. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science* 311, 854–856.
- SAMPSON, ROBERT J. 2010. "Gold Standard Myths. Observations on the Experimental Turn in Quantitative Criminology." *Journal of Quantitative Criminology* 26:4, 489–500.
- SHADISH, WILLIAM R., THOMAS D. COOK & DONALD T. CAMPBELL. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- SHANK, DANIEL B. 2015. "Using Crowdsourcing Websites for Sociological Research. The Case of Amazon Mechanical Turk." *The American Sociologist*.
- SNIDERMAN, PAUL M. 2011. "The Logic and Design of the Survey Experiment. An Autobiography of a Methodological Innovation." Teoksessa *Cambridge Handbook of Experimental Political Science*, toim. James Druckman, Donald Green, James Kuklinski & Arthur Lupia. Cambridge: Cambridge University Press, 102–114.
- SUGDEN, ROBERT. 2008. "The Changing Relationship Between Theory and Experiment in Economics." *Philosophy of Science* 75:5, 621–32.
- TEELE, DAWN LANGAN (TOIM.) 2014. *Field Experiments and Their Critics*. New Haven: Yale University Press.
- TUUNAINEN, JUHA. 2001. "Constructing Objects and Transforming Experimental Systems." *Perspectives on Science* 9:1, 78–105.
- VAN DE RIJLT, ARNOUT, S.M. KANG, M. RESTIVO & A. PATIL. 2014. "Field Experiments of Success-Breeds-Success Dynamics." *PNAS* 111:19, 6934–6939.
- WEBSTER, MURRAY JR. & JANE SELL (TOIM.) 2007. *Laboratory Experiments in the Social Sciences*. Amsterdam: Academic Press.
- WILLER, DAVID & HENRY A. WALKER 2007. *Building Experiments. Testing Social Theory*. Stanford: Stanford University Press.
- WILLER, ROBB, C. L. ROGALIN, B. CONLON & M. T. WOJNOWICZ. 2013. "Overdoing Gender. A Test of the Masculine Overcompensation Thesis." *American Journal of Sociology* 118:4, 980–1022.