

Rakentaako Chen evaluaatiotutkimukselle sen tuolista puuttuvat kaksi jalkaa - vai vahvistaako hän jo olemassa olevaa? Entä selkänoja? Tarvitseeko evaluaatiotutkimus laajemmasta näkökulmasta oman irtotuolin sijasta penkin vai torin, jolla ei ole edes penkkejä?

Timo Oksanen

ABSTRACT

Does evaluation need a three-legged stool? Should there be even more legs in the stool for evaluation or should evaluation have a bench? Or would it be best or more practical for evaluation if it was a square or marketplace without any kind of stools or bench? What is Huey-Tsyh Chen doing for evaluation in his new "Practical Evaluation"?

This article has two main aims. Firstly, short description will be presented about Huey-Tsyh Chen's latest book "Practical Program Evaluation. Assessing and Improving Planning, Implementation, and Effectiveness, Sage, 2004, 279 s." (PPE). I will focus on the two main contributions of the book for the evaluation research: the conceptual framework of program theory and a practical taxonomy of means and ends of program evaluation. With these conceptual and practical advices Chen is trying to combine two challenges which cannot be easily combined and have not yet been successfully met: (1) theoretically strict and justifiable way to understand the target and nature of evaluation, which makes possible rational development and accountability of the

evaluand (the target of the evaluation), and (2) practical advice to help the evaluator and stakeholders (the subject of the evaluation) to face their practical problems and situations. Has Chen been lucky, clever or wise enough to make this complex combination work (well)?

Secondly, the study of the PPE will awake more general problems and questions of evaluation research. Evaluation research will encounter many challenges in future. Thomas Cook has described them in a metaphor where evaluation research is sitting on a one-legged stool. The methods of evaluation have been successfully cultivated and developed. However, the self-understanding and identity of evaluation research as a discipline has developed slowly. Evaluation research needs more and better evaluation theory. Expect this, Cook also defines another gap: theories about successful and unsuccessful programs have not been developed and elaborated. The utilization of the substantive knowledge base of substantive disciplines in evaluation is restricted. Besides that researchers will not make enough syntheses or reviews about of their own main results.

Because Chen is one of the few cultivators of theory-oriented evaluation he has to take

seriously the challenges mentioned above even though the main purposes of PPE are practical. It seems that Chen's aims are ambitious but they are ambiguous, too. If we think about the perspectives of Cook's writings mentioned above (21st century) maybe it is not realistic to achieve remarkable progress in Chen's aims in near future. However, the practical aims of the PPE can be seen realistic and successful if we consider them in a narrow, modern meaning. Chen presents the goals and means of evaluation analytically and usefully in his practical taxonomy of program evaluation. In the taxonomy he presents the logic of evaluation in which strategies and approaches of evaluation depend on the needs and purposes beyond evaluation and on the stage of the program, which will be evaluated. This means that the ultimate quality of the first and only leg of the evaluation research stool (evaluation methods) will depend on how the needs, purposes, strategies and approaches of evaluation are articulated and refined. This starting point has important implications on the evaluation enterprise and especially for theory-oriented approaches and for Chen's taxonomy, too. Though the logic of taxonomy is basically quite simple and fruitful on the first sight, taxonomy should not be proved mechanically without more pluralistic, comprehensive and complex situation logic.

Although PPE's practical contribution to evaluation research and to other evaluation actors can be clearly seen, Chen is not able to open new (truly classical) ways to see "practice" and new ways to meet new "practical" problems. Chen's conceptual starting point to "practice" is not open enough to meet systemic nature of the world, societies and our health, wealth and survival. Nor has PPE much to give to those who see the main practical challenges of evaluation research in those systemic questions. But are they right? However, Chen has been successful to move evaluation research by and by towards some kind of equilibrium between practicality and reflexivity. Chen is not shortsighted but the distances are formidable. Chen has done a great amount of work (and got a good deal of trouble) and still, a lot remains to be done.

1. TEOKSEN TAUSTA JA ARTIKKELIN LÄHTÖKOHDAT

Niin kansainvälisesti kuin Suomessakin evaluaatiotutkijat ovat ajoittain kiinnittäneet huomiota evaluaatiotutkimuksen peruslähtökohtien selkeyttämiseen. Saadakseni asianmukaisen lähtökohdan artikkelin otsikkokysymyksiin vastaamiselle esitän seuraavassa joukon evaluaatiotutkijoiden esiin ottamia tutkimuksensa kehittämishaasteita. Tutkittuaan evaluaatiotutkimuksen historiaa 25-vuoden ajalta Thomas Cook katsoi evaluaatiotutkimuksen istuneen yksijalkaisella tuolilla. Menetelmäkiistoihin keskittyminen on vienyt liiaksi tilaa perushaasteilta. Evaluaatiotutkimuksen kehittämishaasteiden jäsenyksessä Cookin esiin ottamat evaluaatiotutkimuksen lisäjalat sisältyvät kohtiin 2-3. Cookin artikkelissaan (Cook 1997) käsittelemien kolmen jalan lisäksi hahmotan jäsenyksessä kirjoituksen viitekehikseksi myös neljättä jalkaa sekä tuolin selkänöjää (kohdat 4-6). Kehittämishaasteet voidaan hyvin karkeassa muodossa ilmaista esim. seuraavasti:

- (1) Tutkimusmenetelmien jatkuva ajan tasalla pitäminen,
- (2) evaluaatiotutkimuksen itseymmärryksen kehittäminen tutkimusalueena (tutkimuksen teoria¹),
- (3) yksittäisten evaluaatioiden käyttöarvon ja tieteellisen merkittävyyden lisääminen mm. syntetisoimalla ne tavalla tai toisella laajempiin (perus)tieteiden tietorakenteisiin²,
- (4) tutkijan käytännön valintatilanteissa kohtaamien ongelmien hallinta,
- (5) tieteen- ja tutkimusalojen arvojen, eetosten ja esimerkiksi tutkimusalojen yhteistoiminnan ym. erityistieteiden itsensä vastuulla olevien kysymysten sekä tutkimuksen käytännön haasteiden ja valintatilanteiden näkeminen yhtenä kokonaisuutena,³
- (6) Erityistieteistä ja tutkimuskäytännöistä (yllä kohdat 1-5) filosofian alueelle johtavan kuilun kaventaminen.⁴

Jo erityistieteiden sisälle on muodostunut edellä kohdissa 1-5 karkeasti haarukoituja hyvin erilaisia kehittämistarpeita ja tutkijan profileja, joissa myös portaiden portaiden 1-5 rajoja ylitetään. Niitä yhdistävän langan näkeminen näyt-

tää kuitenkin hankalalta. Tuolin ainokaisen jalan käytännöllistäjien (porras 4), eri suuntiin levottomasti tähyilevien (porras 5) sekä kaukana (filosofi)ympäristöissä tutkimusjalkoja tai niiden perusteita yleisemmin pohtivien, tuolin selkänöjaa rakentavien tai rakentamista valvovien (porras 6) välinen etäisyys on merkittävä.⁵ Evaluaatiotutkimuksen käytännön ongelmien etäisyys esim. tieteenfilosofien, - sosiologien ja metodologiien tyypillisesti analysoimista ongelmista lienee itse kullekin tuttua.⁶

Syvällisesti epäselvää on, miten ja millä ehdoilla tutkimusyhteisön viitekehysten mahdollisen lähentymisen ryhmien 1-6 välillä tulisi tapahtua. Evaluaatiotutkijoiden, erityistieteiden, filosofian vai kenties tieteen ulkoisten ns. eksternaliteettien määrittelemillä ehdoilla? Tulisiko etäisyyksien kenties pikemminkin kasvaa, jotta identiteetin kasvamiselle ja profiloitumiselle jää tilaa? Miksi tieteiden välisen sekä tieteen ja käytännön välisen yhteislinssin rakentamiseen niiden omien alalinssien selkeyttämisen ja kirkastamisen sijasta tulisi ylipäänsä uhrata huomiota? "Istuuko" Cookin tuolimetafora itse vahvalla pohjalla vai yksinkertaistaako se esimerkiksi eri tieteen- ja tutkimusalojen suhteiden näkökulmaa kehittämishaasteisiin liikaa? Tulisiko siihen liittää erikseen vaatimus tuolin jalkojen samanlaatuisuudesta, niiden kyvystä muodostaa teoreettisesti ja metodologisesti yhtenäinen kokonaisuus? Vai tulisiko jaloiltaan tasapainotetun ja vakautetun tuolin sijasta edetä kohti penkkiä - tai peräti toria, jolta jo penkitkin on poistettu?

Smithin, Kirkhardin ja Shadishin ohella Thomas Cook nimesi yllä mainitussa katsauksessaan alabamalaisprofessori Huey-Tsyh Chenin erääksi harvoista tutkijoista, jotka suhtautuvat vakavasti evaluaatiotutkimuksen menetelmäkysymyksiä laajempiinkin haasteisiin. 1980-luvun alussa Taiwanilta Yhdysvaltoihin muuttanut Chen, joka vieraili Suomessakin terveystieteiden lokakuussa tutkijakouluseminaarissa 2004, avasi ongelmakenttää Peter Rossin kanssa tehdyn alkutyön jälkeen varsinaisesti teoksellaan "Theory-Driven Evaluation" v. 1990 (jäljempänä TDE). Yllä esittämässäni karkeassa tyypittelyssä teos haki etäisyyttä evaluaatiotuolin ykkösosalasta ja kehittämishaasteiden ykkösaskelmalta painotuen kolmannelle portaalle, kurottautuen samalla ehkä jo hieman myös viidennelle portaalle. 1990-luvun jälkipuoliskon hieman rauhalliseman vaiheen jälkeen Chen on konkretisoinut ja

ohjelmallistanut TDE:n lähtökohtia teoksessaan "Practical Program Evaluation. Assessing and Improving Planning, Implementation, and Effectiveness. Sage, 2004" (jäljempänä PPE).

Oman kuvauksensa mukaan Chen rikastuttaa PPE:ssä evaluaatiotutkimuksen välineistöä seuraavasti: (seuraavassa kehittämiskohde ja Chenin oma arvio hyödystä)

- 1) kehittämällä evaluaatiotutkimuksen praktisen taksonomian auttamaan evaluoijia ja ohjelmasuunnittelijoita (ml. rahoittajat ja muut sidosryhmät, engl. stakeholders, jäljempänä SH) käsitteellistämään evaluaatiotarpeet ja - olosuhteet sekä evaluaatiostrategiat ja lähestymistavat. Näin muodostuva kartasto auttaa Chenin mukaan evaluaattoreita systemaattisesti ja proaktiivisesti suunnittelemaan nykyiseen ohjelmavaiheeseen liittyviä evaluaatioita sekä myöhempiin ohjelmavaiheisiin liittyviä evaluaatioita,
- 2) identifioimalla SH:n evaluaatiotarpeet kohtavia evaluaatiomahdollisuuksia. Samalla pohditaan mahdollisuuksien ylimenoja, jotta SH:t voisivat tehdä tietoisia päätöksiä valitessaan ohjelmansa evaluointiin soveliaita lähestymistapoja,
- 3) identifioimalla ja kuvaamalla konkreettisin käytännöllisin esimerkein kunkin evaluaatiolähestymistavan toteuttamisessa tarvittavia menettelyjä ja metodeja,
- 4) kytkemällä kussakin ohjelman päävaiheessa (suunnittelu, alkuvaiheen toimeenpano, vakiintuneen ohjelman toimeenpano, lopputulokset) sovellettavat evaluaatiostrategiat ja - lähestymistavat systemaattisesti ohjelmateorian käsitekehukseen,
- 5) suhteuttamalla ohjelmateorian käsitekehys systemaattisesti sen lähikäsitteisiin. Samalla kuvataan niiden merkitystä evaluaattorille, joka työskentelee SH:n kanssa ohjelman käsitteellistämiseksi ja ratkaisujen löytämiseksi. (s. 267-268).

Jäljempänä kirjoituksessa

- a) esitellään PPE:n eräitä ydinkohtia ja arvioidaan sen asemoitumista yllä esitettyyn evaluaatiotutkimuksen kehittämishaasteiden

portaikkoon 1-6. Vahvistaako Chen evaluaatiotutkimuksen ainokaista jalkaa vai rakentaako se sille uusia jalkoja? Entä tuolin selkänöja?

- b) valaistaan Chenin teoksen esittelyn pohjalta myös jossakin määrin laajempaan kokonaisuuteen sisältyviä evaluaatiotutkimuksen ajankohtaisia haasteita, käyttäen apuna yllä esitettyä tutkimuksen kehittämishaasteiden 6-jakoa sekä ponnahduslautana Cookin tuolimetaforaa ja Chenin PPE:ta.

PPE on laaja (279 s.) ja välillä jopa uuvuttavan nyansoidusti evaluaatiotutkimuksen eri strategia- ja lähestymistapayhdistelmiä tyypittelevä. Tämän vuoksi rajoitun teoksen esittelyssä ja arvioinnissa eräisiin peruskysymyksiin, joista käsin työ voidaan asemoida kehittämishaasteisiin 1-6. Jaksoissa 2 ja 3 käsitellään kohtaan a) liittyvät kysymykset. Jaksossa 4 otetaan esiin eräitä laajempia kysymyksiä.

Jaksoissa 2 ja 3 ilmenee, että Chen jatkaa PPE:ssä TDE:ssä avaamaansa black-box -evaluaation kritiikkiä, ja saattaa sen monta askelta käytännöllisemmälle tasolle. Hän kiipeää artikkelini alun kehittämishaasteiden porrastuksen kolmannelta askelmalta neljännelle. Tarkastelun kuluessa ilmenee, että Chenin käytäntökäsite on moderni. Sen heikkoutena on käytäntötermin kapea-alaisuus ja täsmällisemmin lausuen sen klassisen, toiminnan arvoihin ja päämääriin suuntautuvan, ulottuvuuden aliehkittyneisyys tai jopa surkastuminen. Chen jättää samalla jälkeensä black-box'in, jonka olemassa olosta ja sisällöstä ei kuitenkaan vällinne evaluaatiotutkijoiden parissa juurikaan yhteisymmärrystä. Tämä mahdollinen musta laatikko ylittää myös disiplinaariset tai tiedonintressien rajat - erityisesti ns. teknokraattisen ja hermeneuttisen tutkimuksen rajan kriittisen yhteiskuntatutkimuksen suuntaan - mikä vaikeuttaa sen järkiperaistä tunnistamista ja tunnustamista evaluaatiotutkimuksessa. Laatikko näyttääkin paikallistuvan eräänlaiseen "kuolleeseen kulmaan", jonka ainakin autoilijat lienevät tunnustaneet. Autoilija tunnustaa "kuolleen kulman" (näkökentän katkoksen takavasemmalla) kuitenkin ongelmaksi ainoastaan ohitustilanteessa - ja usein vasta jälkikäteen kun onnettomuus on jo tapahtunut. Neljännessä jaksossa äsken mainituista seikoista muodostuvaa monimutkaista tilannetta avataan hieman evaluaatiotutkimuksen disiplinaaristen ja kulttuuristen suhteiden

suuntaan, ryhtymättä suoranaisesti kuitenkaan ohittamaan muiden alojen "autoja". Sen sijaan peräänkuulutan evaluaatiotutkimuksen katsomista refleksiiviseen peiliin: sen kykyä toimia joustavasti muiden suuntaan, eräänlaista uudenlaista systeemisyyttä tai kokonaisuutta, joka mahdollistaa myös mustien laatikoiden ja "kuolleen kulmien" paremman paikallistamisen.⁷

2. OHJELMATEORIA CHENIN YDINLÄHTÖKOHTANA

Ylimalkaisen, mutta käytännössä hyödyllisen viisauden mukaan voimme tehdä oikeita tai vääriä asioita oikein tai väärin. Tämä jäsentely viittaa lähtökohtiimme (arvot, päämäärät, tavoitteet yms., jotka liittyvät "oikeisiin asioihin" tai "oikeiden asioiden parissa toimimiseen"). Lisäksi jäsentely viittaa toimintaamme, keinoihimme ja niiden vaikutuksiin: arvojen, päämäärien, tavoitteiden tai muiden asiointilojen toteutumiseen ja tämän sivu-, kerrannais- tai kokonaisvaikutuksiin, jotka liittyvät "oikein tai tarkoituksenmukaisesti toimimiseen". Tätä jäsentelyä lähtökohtana pitäen jo TDE:n eräksi vahvuudeksi voidaan luonnehtia teoksen tietynlainen laaja-alaisuus. Perinteisen "black-box" evaluaation peruspiirteenä on keskittyminen toiminnan aikaansaantoon ja vaikutuksiin, sen selvittämiseen, onko tehty "oikeita asioita" (onnistuneesti). Vaikka Chen ei suinkaan ole ainoa ns. summatiivisen lopputulosarvioinnin lähtökohtien kapea-alaisuuden evaluaatiotutkimuksessa osoittaneista ja lähtökohtaa avartamaan pyrkineistä, hänen ansionsa tässä suhteessa ovat jo v. 1990 teoksen perusteella noteerattavia. Cookin edellä viitatuut huomiot osoittavat, että ne on kansainvälisen evaluaatiotutkijayhteisön piirissä myös noteerattu.

Eräs tapa esitellä ja lähteä arvioimaan PPE:n lähtökohtia on kysyä, miten Chenin ohjelmateorian käsite evaluaation lähtökohtana suhteutuu yllä mainittuun tyypittelyyn: mahdollisuuden tehdä inhimillisessä toiminnassa oikeita tai vääriä asioita oikein tai väärin. Chen katsoo ohjelmateorian käsitteen epäselvyyden vaikeuttaneen sen täysimääräistä hyödyntämistä evaluaatiotutkimuksessa. Eriteltään käsitteen "ohjelmateoria" käyttötapoja (s. 16-) Chen toteaa käsitteen erään populaarin käyttötapavan liittyvän kausaalisuhteisiin. Hän viittaa tässä Carol Weissin tapaan pukea ohjelmateoria kysymykseksi: "Kuinka ja

miksi ohjelma toimii, tai ei toimi?" Toinen yleinen tapa käyttää ohjelmateorian käsitettä on liittää se loogisen mallin käsitteeseen. Tässä yhteydessä Chen viittaa Wholeyn tapaan määritellä ohjelmateoria seuraavasti: "program theory incorporates program resources, program activities, and intended program outcomes, and specifies a chain of causal assumptions linking program resources, activities, intermediate outcomes and ultimate goals".⁸

Chen laajensi äsken mainittuja ohjelmateorian määritelmiä jo TDE:ssa. Siinä hän esitti ohjelmateorian tehtäväksi "sen spesifioimisen, mitä asetettujen tavoitteiden saavuttamiseksi tulee tehdä, mitä muita tärkeitä vaikutuksia toiminnasta saattaa muodostua sekä miten tavoitteet ja muut vaikutukset voidaan toteuttaa". PPE:ssa näistä ohjelmateorian tehtävistä ovat käytännölliseltä painoarvoltaan tärkeimpiä ensimmäinen ja viimeinen; syistä, joihin palaan osaksi myöhemmin, teoksen 271 sivusta vain vajaat neljä sivua (s. 262-265) käsittelee toiminnan tarkoitattomia seurauksia ("muita tärkeitä vaikutuksia"). TDE:ssa erilaiset systeemiset arviointiaspektit saivat selvästi enemmän huomiota.

TDE:ssä Chen esittää systeemimallin tavoiteperusteisen ja tavoitteista riippumattoman evaluaatiomallin tärkeimpänä vaihtoehtona. Siinä "sosiaalinen ohjelma nähdään joukkona keskinäisriippuvuudessa olevia alajärjestelmiä, joista jokaisen tulee toimia hyvin, jotta kokonaisjärjestelmä toimisi hyvin". Chen kehittää tätä lähtökohtaa TDE:ssa erityisesti ns. "Impact Evaluation II" teeman alla.⁹

PPE:n eräs tavoite on kehittää v. 1990 teoksen ohjelmateorian käsitettä sen "relevanssin lisäämiseksi ohjelmaevaluoinnissa".

Arviotavan toiminnan suunnittelu ja implementointi perustuu Chenin mukaan SH:n implisiittisille ja eksplisiittisille oletuksille siitä, millaista toimintaa sosiaalisen ongelman ratkaiseminen edellyttää ja miksi toiminta ratkaisee ongelman. Näihin oletuksiin kohdistuvaa analyysia Chen kutsuu PPE:ssä ohjelmateoriaksi (s. 16). Analyysin käsite on tässä yhteydessä keskeinen, onhan teoriaperusteinen arviointi toisinaan identifioitu "holistisesta arvioinnista" viittaamalla nimenomaan arvioinnin kohteen jakamiseen erilaisiin elementteihin tai komponentteihin sekä näiden välisiin suhteisiin¹⁰.

Ohjelmateoriaan sisältyvät taustaoletukset Chen jakaa deskriptiivisiin ja preskriptiivisiin. Jaot-

telulla on yhtymäkohta jo TDP:ssä suoritettuun ohjelmateorian elementtien jakoon. Chen kehittää PPE:ssä ohjelmateorian deskriptiivisten ja preskriptiivisten osien jäsentelyä esittämällä muutosmallin (change model) ja toimintamallin (action model) käsitteet.

Deskriptiiviset oletukset liittyvät kausaalisiin prosesseihin, joihin arvioitavan toiminnan on tarkoitus vaikuttaa. Ne käsitteellistyvät Chenillä intervention (Intervention, Treatment), vaikuttavien determinanttien (determinants) sekä tavoitteiden ja lopputuloksen (goals and outcomes) käsitteissä, jotka seuraavat muutosmallissa toisiaan. Muutosmalli sisältää ja ilmaisee kuvatut osiot. Samalla se ilmaisee arviotavan (ohjelma)toiminnan perusrationaliteetin, tarkemmin sanottuna mallin laatijan käsityksen siitä.

Preskriptiiviset oletukset sisältävät (ohjelma)toiminnan komponentit ja aktiviteetit, jotka SH:t pitävät välttämättömänä implementoida, "jotta muutosmalli voidaan aktivoida". Niistä muodostuu toimintamalli. Toimintamallin tärkeimmät osiot ovat "toimeenpanevat organisaatiot", "toimeenpanijat", "toimintaan läheisesti liittyvät organisaatiot", "ekologinen konteksti", "palvelujärjestelmät ja palvelut" sekä "kohdeväestö".

Kun muutosmalli ilmaisee (ohjelma)toiminnan perusrationaliteetin joukkona perustavia kausaaliuskomuksia, toimintamalli on ennen kaikkea (ohjelma)toiminnan suunnittelun väline, jolla muutosmallin ilmaisemat mekanismit uskotaan voitavan panna liikkeelle. Puhe uskomuksista ilmentää jo sitä, että molempien mallien perustana oleva tieto on usein vaillinaista. Mallien sovelluksessa kumpaan malliin sisällytetyt oletukset voivat olla päteviä tai virheellisiä. Viimeksi mainitussa tapauksessa valittu toimintamalli "ei aktivoi muutosmallia liikkeelle" (toimintamallin virhe) tai toiminnan perusrationaliteetti on virheellinen ja vaatii uuden käsitteellistykseen (muutosmallin virhe). Kun äskeisiin lisätään kolmas tilanne, jossa molemmat mallit ovat päteviä, mutta (ohjelma)toiminnan toimeenpanon virheet estävät toimintamallia aktivoimasta muutosmallia, avautuu näköpiiriin Chenin ja TPA:n ns. black-box-evaluaatiolle tarjoaman vastauksen tai haasteen kokonaisuus (vrt. Stame 2004, 59-63).

Tämän haasteen ytimenä on vaatimus jäljittää sekä arviotavan (ohjelma)toiminnan suotuisa että epäsuotuisa lopputulos ohjelmateorian käsitteen ja teoria-arsenaalin avulla siten, että molemmat voidaan mahdollisuuksien mukaan,

kuitenkin mahdollisimman jäsenyneesti, ymmärtää". "Ymmärtäminen" ei Chenin konseptissa rajoitu yksinomaan intellektuaaliseksi tai "akateemiseksi" arvoksi tai metodiksi. Viime kädessä vain ymmärtämisen kautta tarjoutuu mahdollisuus (ohjelma)toiminnan rationaaliseen ja pitkäjänteiseen kehittämiseen, eri toimijoiden jäsenytymiseen ja oikeudenmukaiseen vastuullistamiseen sekä lukuisten muidenkin arvojen toteuttamiseen (ks. s. 36-38). Tässä suhteessa Chen kuten muutkin evaluaatiotutkimuksen mallien ja lähestymistapojen kehittäjät seisovat - tietoisesti tai tiedostamattaan - pikemminkin klassisen filosofian, eri tiedetraditioiden ja mm. hermeneuttisen (kontinentaalisen) filosofian "jättiläisten harteilla" kuin kehittävät omaehtoisesti perustavat lähtökohdansa ja rationaliteettinsa¹². Se, että evaluaatiotutkijat kuten muutkin erityistieteilijät toimivat muiden tekemän työn varassa, on sinänsä itsestään selvää. Evaluaatio- ym. erityistieteellinen tutkimus toimii tavallaan tietopyramidin huipulla. Samalla vaarana on perusteiltaan hämären tai ristiriitaisten synteisien ja rakennelmien kehittäminen tietopyramidin "huipulle". Kuvatunlainen tilanne voi syntyä, elleivät evaluaatiotutkijat tiedosta ja selkeyttä ratkaisujensa sidoksia niiden perustana oleviin lähtökohtiin, oletuksiin yms. Samalla on tarpeen korostaa evaluaatiotutkijoiden omaa vastuuta tähän liittyvistä kysymyksistä. Kirjoituksen alun kehittämishaasteiden portaikossa tämä tarkoittaa pikemminkin viidennen kuin kuudennen portaan tärkeyden korostusta.

Chenin ohjelmateorian lähtökohta muodostaa monimutkaisen sekoituksen, jonka peruslähtökohdat sijoittuvat habermasilaisittain ilmaisten pikemminkin immanentin kuin autonomisen kritiikin alueelle. Tämä tarkoittaa sitä, että ymmärtäminen ja evaluaation perustavat arvolähtökohdat ja mittapuut ("oikeat ja väärät asiat ...") määräytyvät viime kädessä SH:n ja siis toimijoiden itsensä arvojen, päämäärien ja tavoitteiden pohjalta. Evaluaattorin tehtävänä ei ole niiden normatiivinen tai paljastava kyseenalaistaminen (s. 234-235). Chen puolustautuu avoimesti kritiikoihin vastaan epäämällä nimenomaisesti mahdollisuuden, että evaluaattori voisi itse "teoretisoida" evaluoitavan toiminnan normatiiviset mittapuut. Koska tiede sen kummemmin kuin filosofiaan eivät ole Chenin konseptissa legitiimejä arvomitapuiden asettajia, SH:n lähtökohtia täydentävät "teoriat" koskevat pikemminkin "asioiden teke-

mistä väärin tai oikein" kuin sitä, "mitkä asiat ovat tehtäväksi oikeita tai vääriä". Tässä suhteessa Chenin lähtökohta eroaa merkittävästi esim. Frank Fischerin policyevaluaation metodologiasta. Siinä stakeholder -lähtökohta ei "katkaista" evaluaation mittapuiden oikeuttamisen vaadetta ja ketjua. Varsinainen evaluointi kohteen suhteuttamisena tavoite- tms. mittapuuhun (verifikaatio) täydentyy Fischerillä validoinnin, vindikaation ja rationaalisen harkinnan käsitteillä ja evaluaatiomenettelyillä. Niiden kautta (ja niissä) evaluaation mittapuut ja myös niiden lähtökohdat tulee erikseen oikeuttaa (Ks. Fischer 1980).

Chen luonnehtii toimintamalliaan preskriptiiviseksi, mutta äsken mainitusta seuraa, että Chenin toimintamallin ja toimintateorian käsitteet ovat korkeintaan miedosti preskriptiivisiä tai normatiivisia. "Hyvä" tai "oikea", joka niitä kannattelee, ei omaa liityntöjä esim. oikeusfilosofian, etiikan tai yhteiskuntafilosofian inhimillistä tai yhteisöllistä hyvää koskeviin käsityksiin ("oikeat asiat"), vaan rajoittuu välineellisen hyvän alueelle ("tehdääkö toiminnassa oikein" ao. termin instrumentaaliseen rationaalisuuteen tai välineelliseen hyvään liittyvässä merkityksessä). Lähtökohta sopeutuu hyvin yhdysvaltalaiseen ohjelma-evaluoinnin perinteeseen ja myös länsimaiseen weberiläiseen tiedekäsitteeseen. On kuitenkin syytä muistaa, että yhdysvaltalaisosioologi Edvard Schumann argumentoi jo 1967 monipuolisesti arviointitutkimuksen tarvetta harjoittaa myös avointa ja sisällöllistä arvoargumentointia (Schumann 1967, 29). Chenin varovaisuus tässä suhteessa - johon sisältyy myös vaikuttavuuskonseptin sijoittaminen lähes yksinomaan tavoiteperusteisen arvioinnin traditioon, s. 199 - ei sopeudu hyvin myöskään hänen toisaalla esittämiinsä arvioinnin holistisuuden ja tulevaisuussuuntautuneisuuden yleisiin korostuksiin (s. 6-7)¹³. Edellä on jo viitattu siihen, että tavoittelemattomien vaikutusten tarkastelun osuus PPE:ssä rajoittuu muutamien sivun alueelle. Painoarvoltaan kyseinen tarkastelu on PPE:ssä selvästi suppeampi kuin TDE:ssä. Samoin edellä ilmeni, että Chen rakentaa ohjelmateorian käsitteen keskeisesti ongelmateorian ja samalla myös yhteiskunta-teorian käsitteille ("Millaista toimintaa sosiaalisen ongelman ratkaiseminen edellyttää?"). Hän ei kuitenkaan tarkenna normatiivista lähtökohdtaa myöskään filosofiseen tms. ongelmateoriaan tukeutuen. (vrt. Oksanen 2003, 190 ss., 206 ss.).

Jäljempänä ilmenee, että Chen luonnehtii toista keskeistä lähtökohtaansa - evaluaatiotutkimuksen taksonomiaa - maantiekartastoksi. Tätä vertausta ennakoiden ja täydentäen ohjelmateoria voidaan lähtökohtaisesti tulkita käsitykseksi kartan ja matkan kohteena olevista alueista ja paikkakunnista, millaisia ne ovat (deskriptiivinen osio) ja millaisia niiden tulisi olla (preskriptiivinen osio). Näin ymmärrettynä ohjelmateorian edellä esitetty kritiikki voidaan vertauskuvallisesti rinnastaa tilanteeseen, jossa kartan ja matkan kohteena olevat alueet ja paikkakunnat ja näiden mittapuut johdetaan yksinomaan kunkin toimijan omista itsearviointeista ja toimijoiden itsensä tuottamasta informaatiosta. Evaluaattorin sidos Chenin konseptissa arvioitavan toimijan omiin tavoitteisiin voitaneen tulkita näin. Kuvatulainen lähtökohta on monessa suhteessa oikeutettu. Jos vertauskuvaa vielä hieman jatkaa, se käy sitä kapea-alaisemmaksi ja riittämättömämmäksi, mitä enemmän evaluaation kohteet (paikkakunnat) ovat ymmärrettävissä ja arvoitettavissa vain alueellisten ja muiden ympäristötekijöidensä ja sidostensa pohjalta tai muusta syystä vain toimijan kannalta "ulkopuolisen" informaation avulla. Vertauskuvasta ilmenevät myös ongelmat, joita Chenin TDE:ssä ns. systeemiseksi arvioinniksi luonnehtiman tavoittelemattomien vaikutusten arvioinnin riittämättömyydestä PPE:ssä seuraa.

Chenin ohjelmateorian toisentyypinen rajoitus käytännön evaluaation näkökulmasta liittyy sen yleisyyteen. Toimintamalli ohjaa arvioitsijaa kiinnittämään huomiota mm. arvioitavaa toimintaa toteuttavan toimeenpano-organisaation ominaisuuksiin ja esim. sen ympäristösuhteisiin. Chenin esittämässä muodossa toimintamalli ei yksilöi kovinkaan tarkasti, millainen toimeenpano-organisaation ympäristösuhteen tulee olla. Tätä koskeva ymmärrys on hankittava muualta (esim. implementaatiotutkimuksesta). Toimintamalli sisältää näin käsitteellisen teorian, joka rinnastuu pikemminkin etsintää ohjaavaan väljään heuristiikkaan. Karttametaforaan palaten voidaan viitata väljämittakaavaiseen karttaan. Tätä toimintamallin piirrettä ei ole perusteltua nähdä toimintamallin suoranaisena puutteena. Toimintamallia on tällöin kuitenkin täydennettävä yksityiskohtaisemmalla tiedolla. Mikäli tällaista tietoa ei ole käytettävissä, toimintamallin ongelmat rinnastuvat suunnistajan käytössä olevan väljämittakaavaisen kartan käyttöön liittyviin ongelmiin.

3. EVALUAATIOTAKSONOMIA PPE:N ERITYISANTINA

Jo TDE:ssa Chen identifioitui ennen kaikkea menetelmälähtöisen evaluaatioajattelun kriitikona. PPE:n ansiot ohjelmateorian käsitteen kehittämisessä voidaan tulkita uudeksi askeleeksi kirjoituksen alussa kuvatulla evaluaatiotutkimuksen kehittämishaasteiden porrastuksella edettäessä menetelmien kehittämisen askelmalta kohti ylempiä portaita. Vai voidaanko? - edellä on jo kuvattu ohjelmateorian käsitteen heikkouksia edettäessä arvioinnille tarpeelliseen muutokseen evaluaation välineistä kohden evaluaation sisältöjä.

Chen kuvaa taksonomiansa maantiekartaksi ja havainnollistaa taksonomian yleistä taustaa kalastusanalogialla. Evaluaatiotutkijan menetelmät ja tekniikat (esim. aineistonkeruussa) rinnastetaan kalastusvälineisiin, joita tutkijan tulee kyetä valistuneesti hyödyntämään. Tähän hän liittyy evaluaatiotutkimuksen tieteellisen ulottuvuuden. Kalastusvälineisiin liittyvistä teknisistä perustiedoista ja -taidoista kalamieheksi kurotettava Chen erottaa kysymykset siitä, missä, miten, milloin ja mitä mikäkin kala syö, mitä välineitä missäkin ympäristössä ja tarkoituksessa kalastettaessa tulee käyttää jne. Taksonomian ajatuksellisenä ytimenä on vertauskuvallisen välineen luominen kalastustaidolle ("the art of fishing") ("evaluaatiotaito"). Chenin ajattelutavan mukaan evaluaatiotutkimuksen lähestymistapojen soveltamista tulee edeltää niiden käyttöyhteyksien ja soveltamisen laajempien horisonttien ymmärtäminen. Kun ohjelmateorian käsitteen anti evaluaatiolle kiteytyy etäisyyden hakemisessa ns. black-box-evaluaatiosta, jossa evaluaation kohteen toimintaprosessit ja vaikuttavuuden edellytykset sivuutetaan, kiteytyy evaluaatiotaksonomian anti äskeisessä metaforassa kalastusvälineiden ja niiden teknisen hallinnan täydentymisenä välineiden käyttöä koskevalla taidolla ja ymmärryksellä. Näin ymmärrettynä Chenin "evaluaatiotaidon" ja edellä siteeratun Ahosen evaluaatiometodologian määritelmien raja jää hyvin ohueksi. Evaluaatiotaitoonkin on Chenin mukaan liittynyt eräänlainen black-box, jota taksonomialla pyritään täyttämään. Ohjelmateorian ja taksonomian avulla Chen hakee etäisyyttä evaluaatiotutkimuksen yksijalkaiseen tuoliin.

Taksonomia kohdistuu ensisijaisesti evaluaa-

<i>Program Stages and Evaluation Purposes</i>	<i>Evaluation Strategies</i>	<i>Evaluation Approaches</i>
Program Planning Stage Providing pertinent information and assistance to help stakeholders in developing program rationale and plan	Background information provision Development facilitation Troubleshooting Development partnership	Needs assessment Formative Research Conceptualization facilitation Concept mapping Relevancy testing Pilot testing Commentary or advisory Bilateral empowerment evaluation
Initial Implementation Stage Providing timely information on implementation problems and the sources for assisting stakeholders to fix the problem and stabilize the program	Troubleshooting Development partnership	Formative evaluation Program review / development meeting Bilateral empowerment evaluation
Mature Implementation Stage Providing information on implementation problems and the sources for improving the implementation process	Troubleshooting Development facilitation	Formative evaluation Program review/ development meeting Conceptualization facilitation Concept mapping
Assessing the quality of implementation for serving accountability needs	Performance assessment	Fidelity evaluation
Monitoring progress of implementation	Performance monitoring	Process monitoring
Holistic assessment of implementation process	Enlightenment assessment	Theory-driven process evaluation
Outcome Stage		
Assessing whether program is ready for outcome evaluation	Development facilitation	Evaluability assessment
Monitoring client's progress of outcomes	Performance monitoring	Outcome monitoring
Assessing the effect on outcomes	Performance assessment	Efficacy evaluation Effectiveness evaluation
Holistic assessment of the program effect for serving accountability and program improvement needs	Enlightenment assessment	Theory-driven outcome evaluation

Kuvio 1. Praktinen taksonomia ohjelmaevaluaation keinoista ja tavoitteista (s. 48)

tiotutkijoiden ja muiden evaluaatiotoimijoiden yhteisöön. Samalla se voidaan ymmärtää myös tutkimuksen teoriaksi tai ainakin evaluaatiomalliksi. Kun ohjelmateoriaa voidaan luonnehtia "arvioinnissa hyödynnetyksi teoriaksi (kohteesta)", taksonomian ytimenä on luoda luokittelun ja systematiikan muodossa oleva "teoria arvioinnista". Tiivistetysti todettuna taksonomian ytimenä on osoittaa ja konkretisoida evaluaatiotavoitteiden (menetelmät, lähestymistavat) riippuvuus arviointitoiminnan sisällöistä (evaluaatiotavoitteet ja

evaluaation tarkoitukset) sekä näiden riippuvuus evaluoitavan toiminnan kulloisestakin vaiheesta (s. 47). Ohjelmavaiheina (taksonomian ilmaisevan taulukon rivit) Chen erottaa ohjelman suunnitteluvaiheen, varhaisen toimeenpanovaiheen, kypsän toimeenpanovaiheen ja lopputulosvaiheen. Taksonomian rakenne ilmenee seuraavassa kuviossa.

Taksonomiataulukon ensimmäisessä sarakkeessa ilmaistaan evaluoitavan toiminnan (ohjelman) etenemisvaiheiden mukaan määritellyt

evaluaation tarkoitukset.

Toisessa sarakkeessa kullekin ohjelmavaiheelle määritellään relevantit evaluaatiostrategiat. Evaluaatiostrategiat määräytyvät kulloisistakin tarpeista johdetuista evaluaation tarkoituksista. Evaluoitavan toiminnan kehitysvaiheesta riippuen strategioiden taustalla olevissa tarkoituksissa korostuvat vaihtelevalla painotuksella kehittäminen, valistaminen ja vastuulistaminen - evaluointina tai monitorointina - päänäkökulman vaihdellessa tilanteesta riippuen ongelmakeskeisestä holistiseen. Evaluaatiostrategiat muodostuvat ohjelman suunnitteluun ja käynnistämiseen soveliaista "taustainformaation tuottamisesta", "kehitysvälineiden rakentamisesta" ja "vianetsinnästä", joilla tuotetaan nopeasti ja kohdennetusti toiminnan suunnittelu- ja käynnistysvaiheessa tarpeellista tietoa. Ohjelmatoiminnan vakiinnuttua lisääntyy tarve ottaa käyttöön myös toiminnan aikaansaannoksiin suuntautuvia sekä ongelmiin jäsentyneemmin etsittyjä evaluaatiostrategioita ("toiminnan arviointi", "valistusarviointi"). (Ohjelma)toiminnan edettyä "kypsään vaiheeseen" on lopputuloksien vaikutuksiin suuntautuvan evaluaation vuoro (tuloksellisuuden evaluointi), johon voi liittyä myös tulosten ymmärtämiseen liittyvä aspekti¹⁴. Kolmannessa eli viimeisessä sarakkeessa erotetaan suuri joukko kullekin ohjelmavaiheelle ja strategialle relevantteja lähestymistapoja.

Myöskin taksonomiaan sisältyvän "evaluaation teorian" tarkastelun lähtökohdaksi voidaan ottaa kirjoituksen alussa esitetty käytännöllinen viisaus, jonka mukaan voimme tehdä oikeita tai vääriä asioita oikein tai väärin. Kyseinen maksimi paljastaa Chenin eräänlaiseksi "puolitien kulkijaksi", jolla on sanottavaa kumpaankin "oikein toimimisen ulottuvuuteen".

Jo taksonomian perusluonne näyttäisi johtavan tulkitsemaan taksonomian siten, että sillä ohjataan lähinnä asioiden tekemiseen "oikein tai väärin". Taksonomiassa on kysymys evaluaatiotutkimuksen uudesta välineestä, jonka avulla evaluaatiohankkeisiin sisältyviä lukuisia haasteita ja valintatilanteita voidaan hallita aiempaa paremmin. Tämän perusteella Chen voitaisiin luonnehtia "evaluaatiotutkimuksen jo olemassa olevan metodijalan vahvistajaksi" - esim. niin, että todettaisiin Chenin opettavan tutkijoille "yksijalkaisella tuolilla paremmin istumisen taitoa" (art). Jo Chenin kalastusanalogia ilmentää kuitenkin osaltaan sitä, ettei taksonomian kuvatulunlainen sijoit-

taminen kirjoituksen alun evaluaatiotutkimuksen kehittämishaasteiden porrastukseen tekisi oikeutta Chenille ja PPE:ssa tehdyille työlle. Sama koskee Chenin sijoittamista taksonomiensa perusteella yksinomaan portaalle neljä (tutkijoiden auttaminen näiden käytännöllisissä valintaongelmissa), vaikka taksonomia tätäkin epäilemättä edistää. Jo taksonomia-aulukonsa otsikossa Chen katsoo taksonomian ilmaisevan sekä evaluaation tavoitteet että sen keinot.

Chen taustoittaa taksonomiaansa tavalla, jolla on jonkinlaisia yhtymäkohtia Cookin odotuksiin evaluaation "suuresta (arviointi)teoriasta" (alaviite 1) samoin kuin eräisiin Kuhnin paradigman määritelmiin. Chenin mukaan taksonomia tarjoaa mille tahansa tieteen- tai tutkimusalalle tärkeän välineen ilmiöiden luokitteluun, se ohjaa tutkimusstrategian valintaa, mahdollistaa kommunikaation sekä kehittää tietoa (s. 47). Kuhnin paradigmatieteoriasta saatavan yleisluotoisen vertailukohtaan lisäksi taksonomian epistemologisen luonteen tulkitseminen näyttäisi edellyttävän tukeutumista Kuhnin ja Hanssonin esittämään havaintojen teoriapitoisuuden teesiin; lisäksi viittaus Habermasin tiedonintressiteoriaan selvittää, joskin myös problematisoi taksonomiaa.

Taksonomian avainsana on (evaluaation) kontekstuaalisuus. Taksonomiassa voidaan erottaa kahdensuuntaista kontekstuaalisuutta: sarakkeiden ja rivin tasoista.

Näistä periaatteellisesti tärkeämpi kontekstuaalisuuslogiikka avautuu sarakkeiden tasolla. Havaintojen teoriapitoisuusteesi ilmenee siinä, että metodit ja tekniikat eivät ohjaa lähestymistapoja, evaluaatiostrategioita, evaluaation tarkoituksia tai evaluaatiotarpeita - yhtä vähän kuin luotettavinkaan peräsin voi kertoa, minne (kalastus)laivan on määrä edetä - vaan rationaalisuussidos esiintyy ensisijaisesti päinvastaisessa suunnassa. Tämä ainakin vertauskuvakielellä yksinkertainen ja intuitiivisesti oikea "periaate" ei ole tieteenfilosofiassa osoittautunut itsestään selväksi, minkä havaintojen teoriapitoisuudesta käyty sitkeä kiista osoittaa. Tuota teesiä on jäsennelly hiljattain jakamalla se viiteen eri tulkintaan. Niistä neljäs näyttäisi tavoittavan evaluaatioteoriaksi tulkittavan taksonomian keskeiset piirteet: "teoreettiset käsitykset ohjaavat tutkimuskohteen valintaa, havaintoaineiston valintaa ja tutkimuksen kysymyksenasettelua" (Kiikeri & Ylikoski 2004, 30-33). Chenin taksonomian avulla tulkittuna evaluaation kannalta erityisen keskeisiä

ovat evaluaation taustalla olevat tarpeet ja niistä johdettavat evaluaation kulloisetkin tarkoitukset. Niitä voidaan kuvata "teorioiksi", jotka ohjaavat kulloisenkin evaluaation konkreettisempia ratkaisuja (lähestymistavat, menetelmät).

Käytännöllisemmin kuvattuna taksonomia auttaa evaluaattoria käsitteellistämään evaluaatiotarpeita, fokuoimaan evaluaation kulloinkin olennaisimpiin kysymyksiin sekä toteuttamaan evaluaatiotarpeita vastaavan evaluaation (s. 47). Tämä toteutuu yhtäältä evaluaattorin parantuneena kykyä tiedostaa taksonomian avulla kulloinkin tärkeät seikat, toisaalta tästä seuraavana SH:n parantuneena kykyä ottaa kantaa olennaisiin kysymyksiin.

Kolmanneksi fokuoituminen ja evaluaatiotarpeita vastaavan evaluaation toteuttaminen varmistuu taksonomiataulukon riveillä, joilla kohtaamme taksonomian toisen kontekstuaalisuuslogiikan. Menetelmät ovat kontekstuaalisia, mutta myöskään temppeleinharjalle nostetut evaluaation sisällölliset tarpeet ja tarkoitukset eivät ole universaaleja, vaan määräytyvät kunkin (ohjelma)toiminnan kehitysvaiheen mukaan. Niitä kuvataan taksonomiataulukon riveillä.

Chenin taksonomiataulukon sarake- ja rivilogiikkaa voitaneen kokoavasti kuvata seuraavalla vertauskuvalla. Peräsimen tai ratin sijasta kalastus- tai automatkaa ohjaa viime kädessä kartta, matkan tarkoitusta ja pääreittejä selkeyttään ennen matkaa ja ehkä vielä alkumatkasta, matkan edistymisestä varmistutaan matkan kuluessa ja vasta matkan loppuvaiheilla kiinnostutaan suoranaisesti sitä, ollaanko maalia lähestymässä ja millä etäisyydellä siitä ollaan. Kussakin vaiheessa tarvitaan erilaisia toimintastrategioita, lähestymistapoja ja menetelmiä. Näin ymmärrettyinä taksonomian viime kätinen rationaliteetti ilmenee ratkaisujen kontekstuaalisuutena. Evaluaation havaintojen teoria(päämäärä, tarkoitus, strategia-)pitoisuus ei kuitenkaan johda siihen, että kaikki suhteellistuisi. Chen ei seuraa Feyerabendia tämän epistemologiseen anarkismiin (kaikki suunnat ja suuntaimet yhtä hyviä), vaan pidättäytyy havaintojen teoriapitoisuusteesein radikaaleimmista tulkinnoista. Peräsimä, suuntaimia ja muita konkreettisia matka- ja kalastusvälineitä tarvitaan, eikä niiden laatu ole suinkaan yhdentekevä (vaikkakin tarkoitussidonnainen), mutta yksinomaan niiden avulla matkaa ei voida tehdä.

Ohjelmateorian tavoin taksonomian normatiiv-

vista voimaa rajoittaa olennaisesti se, että se kantaa kyllä erilaisia filosofiakkin oletuksia matkassaan, mutta taksonomia on kuitenkin viime kädessä tarkoitettu SH:n rationaalista päätöksentekokykyä edistäväksi työvälineeksi. Se ei lausu sisällöllistä teoriaa siitä, mihin evaluaation tulee pyrkiä tai edetä. Ainakin pidemmällä tähtäimellä nimenomaan siihen, että matkan tarkoitus ja suunta määräytyy yksinomaan SH:n kulloistenkin tavoitteiden ja tarkoitusten pohjalta, sisältyy samalla ehkä Chenin suurin rajoitus ajatellen hänen taksonomiensa kykyä parantaa evaluaatiotutkimuksen itseymmänystä tutkimusalueena (kirjoituksen alun evaluaatiotutkimuksen kehittämistarpeiden toinen porras). Tähän liittyvät kysymykset nousevat esiin viitattaessa esimerkiksi Habermasin tunnettuun tiedonintressiteoriaan; Chenin apuvälineistö osoittautuu teknokraattisen suunnittelu- tai hyötytieteen apuvälineistöksi viimeistään pyrittäessä suhteuttamaan se eri tiedonintresseihin, erityisesti kriittis-emansipatoriseen.

Chen on onnistunut osaltaan rajoittamaan kahta keskeistä evaluaatiotutkimuksen blackboxia: evaluaation kohteen toimintaprosesseihin sisältyvää sekä menetelmien ja tekniikoiden kaikkivoipaisuus(tai universaalisuus)oletuksesta nousevaa. Evaluaation tuottama tieto on sanan tärkeässä, joskaan ei rajattomassa, mielessä teoriapitoista. Chen tekee tästä PPE:ssä implisiittiseksi jäävästä lähtökohdasta lukuisia valistuneita johtopäätöksiä, mutta ei ilmaise niiden teoreettisia perusteita selkeästi. Edellä kuvatuista syistä PPE:n kriittinen tai refleksiivinen potentiaali on sekä ohjelmateorian että taksonomian osalta tärkeällä tavalla rajoittunut. PPE:tä ei voitaneen luonnehtia pitkälle meneväksi askeleeksi kirjoituksen alussa kuvatun evaluaatiotutkimuksen toisen kehittämisportaan suuntaan (evaluaatiotutkimuksen itseymmärryksen lisääminen). PPE:ssä hahmotetun konseptin nousemista käytännölliseksi teoriaksi ja apuvälineistöksi sanan laajassa merkityksessä (porras 5) - sen samanaikaista merkitystä evaluoijien käytännöllisenä apuvälineistönä ja tutkijoiden refleksiivisenä välineenä - on puolestaan vielä tässä vaiheessa mahdotonta tyhjentävästi arvioida¹⁵. Tämä ei estä luonnehtimasta PPE:tä siten, että siinä edetään epäilemättä neljännen portaan suuntaan (tutkijoiden käytännöllisten valintatilanteiden helpottaminen), vaikka työn täsmällinen merkitys tässäkin suhteessa jää vasta tulevaisuudessa nähtäväksi.

Tähänastisessa tarkastelussani olen tietoisesti jättänyt taka-alalle neljännen portaan; tarkemmin sanoen olen pyrkinyt orientoitumaan vallitsevaa käytäntökäsitettä laajempaan, klassiseen käytäntökäsitteeseen, jossa korostuu käytännöllisen viisauden toiminnan perusteisiin (arvoihin ja päämääriin) liittyvä aspekti. Jotta PPE:n valitsevaan käytäntökäsitteeseen liittyvä aspekti ei jää kohtuuttomasti sivuun, esitän seuraavassa joitakin lisänäkökohtia erityisesti PPE:n kyvystä antaa evaluaatioiden toteuttajille (mm. tutkijoille) konkreettisia keinoja ja apuvälineitä evaluaatioiden toteuttamiseen. Tällöin korostuu käytännöllisen viisauden päämääräaspektin sijaan sen keinoaspekti, teoksen kyky ohjata konkreettisesti tutkijan kykyä toteuttaa oikeita tai vääriä asioita oikein.

PPE:ssä ja varsinkin evaluaatiotaksonomiassa omaksuttu tapa jäsentää evaluaatiostrategiat ja lähestymistavat lähtien evaluaatiotarpeista ja evaluaatiotoiminnan tarkoituksesta ja eteneminen vasta näiden pohjalta konkreettisiin menetelmiin ja aineistoihin vastannee useimman käytännön evaluoijien intuitiota, ja monessa suhteessa myös (tavoiteltavaa) käytäntöä. Edellä esitetty vertaus, jossa matkaa todetaan käytännössä suunnattavan (tai ainakin suuntaa etsittävän) viime kädessä kartan eikä peräsimen avulla, ilmaisee tämän verrattain konkreettisesti. Kuten peräsin tai suuntimet ovat konkreettinen ja käytännöllinen tapa "vastata" kalastuslaivan matkan määränpäättä, tarkoitusta ja etenemisväyliä koskeviin jo suoritettuihin harkintoihin, arvioinnin lähestymistavat (esim. 1. formatiivinen evaluaatio, 2. käsitekarttojen laatiminen, 3. vaikuttavuuden arviointi) ovat keino vastata arviointistrategioiden (esim. 1. vianetsintä ja 2. kehitysvarustuksen luominen osana toiminnan kehittämisstrategiaa sekä 3. toiminnan lopputulosten evaluaatio osana toiminnan vastuullistavan evaluaation strategiaa) asettamiin haasteisiin. Lähestymistavat puolestaan määrittelevät soveliaita menetelmiä - tai menetelmät ovat tapa vastata lähestymistapoja koskeviin jo suoritettuihin harkintoihin ja ratkaisuihin.

Toisaalta rationalististen toimintamallien kriitikki on kuitenkin vakuuttavasti osoittanut, että keinojen saatavilla oloon liittyvät mahdollisuudet ja ongelmat on otettava huomioon silloinkin, kun tämä näyttää johtavan tarpeiden ja niistä johdettujen tarkoitusten/tavoitteiden kannalta alioptimaaliseen tulokseen. Evaluaatiotarpeiden ja

mahdollisuuksien sovittelu toisiinsa monimutkaisissa käytännön tilanteissa saattaa edellyttää esim. aineiston riittävyyden tai valikoitumisen huomioimista tavalla, jota ei ole helppoa sovittaa taksonomian yleisiin periaatteisiin, niin loogisia kuin ne lähtökohtaisesti ovatkin.¹⁶

Oman tärkeän kysymyksensä muodostavat eräät institutionaaliset ja kontekstuaaliset seikat, joihin Chenin taksonomia voi olla sen joustavuudesta huolimatta hankalaa sovittaa. Johdettaessa evaluaation konkreettiset muodot evaluaation tarkoituksesta ja tarpeista, spesifien arviointi-instituutioiden "paikan" määrittelyssä kohdataan vaikeuksia. Esimerkkinä voidaan mainita valvonta- ja tarkastusorganisaatiot, vaikkapa USA:n GAO tai Suomen VTV. Arviointiorganisaatioksi ymmärrettynä esim. VTV:lla on omat valtiosääntöiset tarkastus- ja evaluaatiovastuunsa. Näitä vastuita toteuttavista tarkastustoiminnoista ainakin ns. toiminnantarkastuksen sijoittaminen Chenin ohjelmaprosessin taksonomian loppupäähän - evaluoitavien toimintaprosessien loppupäähän, jolloin toiminta on jo kypsää vaikuttavuusarviointiin - näyttäisi luonteelta. Taksonomian erään peruseriaatteen mukaanhan toimijoita vastuullistavaa lopputulosevaluaatiota tulisi harjoittaa vasta kypsän toimeenpanon vaiheessa, tai sen "jälkeen". Lisäksi evaluaatiomenetelmien tulisi tällöin olla hyvin vakaita, ennakoitavia (rigor) läpinäkyviä ja luotettavia, koska toimijoita tullaan vastuullistamaan evaluaation tuloksista (s. 165).

Jos evaluaation menettelyt ja vaatimukset tulkitaan kovin yksioikoisesti evaluaation institutionaalisesta tarkoituksesta käsin, voidaan hyvin perustein kysyä, miten ja missä määrin tarkastusorganisaation sitominen sen institutionaalisen ja valtiosääntöisen tehtävän perusteella yksinomaan kuvatulaiseen vastuullistavaan lopputulosevaluaatioon /tarkastukseen on läpivietävissä silmälläpitäen esimerkiksi valvonta- tai tarkastusorganisaatioiden käytössä olevia resursseja, toimintamalleja ja muita institutionaalisia puitteita. Chenin ja lukuisien muiden tutkijoiden vaikuttavuuden arvioinnin nykytilasta esittämät huomiot (joihin palataan 4. jaksossa) rekisteröiden voi kysyä, mitä ylipäänsä ovat ne "hyvin vakaat, ennakoitavat, luotettavat ja läpinäkyvät menetelmät ja lähestymistavat", joilla valvonta- tai tarkastusorganisaatio (tai mikä tahansa muukaan organisaatio!) voi legitimitä harjoittaa vastuullistavassa tarkoituksessa vaikuttavuusarviointia.¹⁷

Chen peräänkuuluttaa teoksensa lopussa "teorian" ja "käytännön" uutta yhteyttä ja katsoo PPE:ssä tarjoamansa välineistön edistävän tuon yhteyden rakentamista (s. 268-271). Kirjoitukseni alun evaluaatiotutkimuksen kehittämishaasteiden porrastuksessa tämä merkitsi etenemistä viidennelle portaalle. Taksonomian ja näin myös PPE:n kokonaisuudessaan kyky edistää tätä siirtymää ei ole kuitenkaan selvä. Mikäli taksonomiaa ja sen kontekstuaalisuutta tulkitaan liian yksioikoisesti - vaakasuunnassa vain vasemmalta oikealle ja pystysuunnassa hyvin "matalasti" - taksonomian kontekstuaalisuus myös rajoittaa eri toimijoiden kykyä kokonaisuuden hahmottamiseen sekä kykyä ymmärtää muiden evaluaatiotutkijoiden tai -toimijoiden "konteksteja" ja ylittää näitä. Tällaisia "konteksteja" ovat evaluaation tarkoitukset, muodot, menettelyt, instituutiot, puitteet jne. Rajojen tai kuulujen ylittäminen voi olla tarpeen eri syistä, esim. evaluaation yli- tai aliresponsiivisuuden välttämiseksi tai rajoittamiseksi.

Chenin moitteettomasti perusteleman ja jalostaman evaluaatiotaksonomian analyttisyydellä ja kontekstuaalisuudella on siis hintansa. Se on syytä tiedostaa, jotta ylilyönneiltä vältytään. Ryhtymättä analyysoimaan tähän liittyviä kysymyksiä tässä seikkaperäisemmin, rajoitun toteamaan, että esimerkiksi GAO:n Chelimsbyn johdolla kehittelemä evaluaatiosynteesin menetelmä näyttää tässä yhteydessä tarjoavan hedelmällisiä etenemissuuntia. Siitä tarjoutuviin mahdollisuuksiin mm. Thomas Cook on kiinnittänyt huomiota.¹⁶ Näin lausuttaessa joudutaan tavallaan etenemään Chenin taksonomian perus(sarake)logiikasta ("laivan suuntiminen määräytyy kartasta eikä peräsimestä käsin") poikkeavaan ajattelumalliin ("peräsin ja menetelmä/ tekniikka määrittää osaltaan mahdollisia suuntia ja tarkoituksia"). Yllä kuvattu institutionaalinen esimerkki kuvaa yleisemminkin havainnollisesti sitä, että eteneminen suoraviivaisesti taksonomian "deduktiivisella" perus(sarake)logiikalla arvioinnin tarpeesta ja tarkoituksesta lähestymistapoihin ja menetelmiin - kiinnittämättä huomiota päinvastaisesta suunnasta nouseviin sidoksiin ja rajoihin - johtaa käytännössä ongelmiin. Käytännössä taksonomiaa onkin tulkittava molemmilla yllä vertauskuvallisesti kuvatuilla sarakelogiikoilla. Samalla taksonomian heuristinen ohjauskyky ja läpinäkyvyys luonnollisesti kuitenkin heikkenevät.

Jonkinasteisia käytännön ongelmia muodostaa myös Chenin taksonomian rivilogiikka: evaluaation tarpeiden, tarkoitusten, strategioiden jne. sitominen viime kädessä amerikkalaisen ohjelmaevaluaation lähtökohtiin ja ohjelmien elinkaareen. Tästä nousevaa Chenin taksonomian "dekontekstuaalisuusongelmaa" voidaan valottaa kysymällä, miten Chenin apuvälineistö sopeutuu esimerkiksi suomalaisen oikeussääntelyn evaluointiin. Jonkinasteisia käytännöllisiä ongelmia epäilemättä kohdataan. "Sovitteluratkaisuja" esim. suomalaisen sääntelyn ja poliittis-hallinnollisen rakenteen huomioimiseksi joudutaan yleisemminkin tekemään, vaikka taksonomian perusajatus evaluaation sitomisesta sen kohteena olevan toiminnan "kehitys- tai toteutusvaiheeseen" on syytä periaatteessa hyväksyä. Esimerkiksi Sipposen huomiot tarpeesta etsiä oikeussääntelyn vaikutusten arvioinnille uusia kiinnekohtia "lakien elinkaariajattelusta" sopivat luontevasti yhteen Chenin taksonomian ohjelmien elinkaareen sidotun rivilogiikan kanssa (Sipponen 2000, 88). Taksonomiassa ei siis nähdäkseen ole kysymys Petri Virtasen ilmaisua lainaten "alisosiaalitetusta lähestymistavallisesta tms. mallista". Sipposen huomiot osoittavat, että PPE tarjoaa suomalaisille evaluaatiokäytännöille myös käytännöllisiä kehittämismallitteita. Näin on siitä huolimatta, ettei esimerkiksi säädösten elinkaariajattelu itsessään muodosta vielä riittävää lähtökohtaa monimutkaisen sääntelyjärjestelmän sisäisten suhteiden hahmottamiseen esim. ongelmanratkaisujattelun näkökulmasta (ks. Oksanen 2003, 346-359). Toisaalta PPE:n ja esim. evaluaatiotaksonomian "käytännöllisyyttä" ei ylipäänsä tule määritellä ainakaan yksinomaan sen pohjalta, miten suoraviivaisen apuvälineen ne tarjoavat suomalaisen sääntely-, hallinto- ja evaluaatiokäytäntöjen jatkamiseen mahdollisimman vähäisin muutoksin.

PPE:n kykyä tarjota tutkijoille ja muille evaluaattoreille konkreettisia apuvälineitä voidaan arvioida myös teoksen esitysasun ja etenemistavan perusteella. Teoksen esitysasun on selkeä. Chen on varustanut PPE:n lukuisilla käytännöllisillä esimerkeillä, joissa mm. arvioinnin lähestymistapoja havainnollistetaan ja avataan. Teoksen käytännöllistä hyödyntämistä helpottaa myös ohjelmateoriasta ja taksonomiasta versova ja niiden mahdollistama ratkaisu kirjoittaa pääajaksot itsenäisiksi kokonaisuudeksi, joihin voi perehtyä myös erikseen. Tästä seuraa myös eräänlainen

toisto, jonka teokseen kokonaisuudessaan perehtyvä saattaa kokea raskaaksi. Kun lukija kohtaa kymmenien ja lähes satojen sivujen ajan uusia ja toisistaan ehkä vain vivahteikkaasti poikkeavia strategioita ja lähestymistapoja menetelmineen, luetteloita luettelon perään, hän saattaa myös väsyä. Tämä johtuu olennaisesti juuri Chenin tavasta systematisoida esityksensä keskeinen osio ohjelman elinkaaren perusteella (suunnitteluvaihe: strategiat..., lähestymistavat...; varhainen toimeenpanovaihe: strategiat...jne.). PPE on perusuunteeltaan evaluaatiotutkimuksen käsikirja ja epäilemättä myös yliopistojen perusoppikirja. Ensin mainittu seikka tulee ottaa huomioon myös sen raskautta arvioitaessa. Jakson lopuksi on syytä palata kysymykseen, miten äsken mainitut seikat ovat vaikuttaneet Chenin kykyyn edetä kirjoituksen alussa muodostetussa evaluaatiotutkimuksen kehittämishaasteiden portaikossa.

Kirjoituksen alun evaluaatiotutkimuksen kehittämishaasteiden portaita 4 (tutkijoiden käytännöllisten valintatilanteiden helpottaminen) ja 2 (evaluaatiotutkimuksen itseymmärryksen kehittäminen) on erittäin vaikea lähestyä samalla kertaa, koska ne edellyttävät erilaista tutkimus- ja kirjoitusotetta. Tämä voi merkitä sitä, että porras 5, jossa erityistieteilijät kantavat huolta sekä refleksiivisistä että käytännöllisistä kysymyksistä, muodostaa eräänlaisen "pandoran lippaan", jolle tulisi kehittää oma ainutkertainen "taitonsa" (art). Itse asiassa 4 + 2 onkin 6 eikä 5. Vastaavasti on mahdollista, että vasta porras 6 (tutkimuksen perusteita koskeva filosofinen analyysi ja synteesi, jota harjoittavat muut kuin erityistieteilijät eli ammattifilosofit) voi lopulta auttaa evaluaatiotutkimusta paikallistamaan sille mahdollisen refleksiivisen ja praktisen yhdistelmän (portaan 5). Ennen tähän vaihtoehtoon - Cookin kirjoituksen alun metaforaa laventaen selkänöjättömaan tuoliin tai muiden valmistamaan selkänöjaan - alitumista seuraavassa jaksossa pohditaan vielä lyhyesti kirjoituksen alkuporrastuksen kolmannelle portaalle sijoittuvaa tuolin kolmatta jalkaa (evaluaation ja sen tulosten käyttöarvon lisääminen liittämällä ne laajempiin tietorakenteisiin). Sen kautta näyttää löytyvän avain myös tuolin selkänöjan ongelmaan.

4. ERÄITÄ YLEISEMPIÄ NÄKÖKOHTIA EVALUAATIOTUTKIMUKSEN TAVOITELTAVASTA TULEVAISUUDESTA: TOIMIMISTA TASAPAINOTETULLA IRTO-TUOLILLA, ETENEMINEN IRTOTUOLISTA KOHDEN YHTEISTÄ PENKKIÄ VAI KOHDEN PENKITÖNTÄ TORIA?

Keskityttäessä edellä Chenin ohjelmateoriaan ja taksonomiaan ja niiden tutkimus(kala)matkan tekijälle matkan kohteesta ja sen reiteistä tarjotun ymmärryksen rajoituksiin ja välineiden laatuun, Chenin ansiot kirjoituksen alun portaikon kolmannella portaalla jäivät paitsioon. Kolmannen portaan sisältönä oli yksittäisten evaluaatioiden käyttöarvon ja tieteellisen merkittävyyden lisääminen syntentisoimalla ne laajempiin tietorakenteisiin (Cook 1997, 48-50). Cookin huomion kohteena on tässä muun muassa ns. interdisciplinaarisuutena (monitieteisyys) tunnettu ongelmakokonaisuus. Asiaa voidaan tarkastella myös laajemmassa kehityksessä, ottaen lähtökohdaksi tieteen (eri tieteenalojen) ja yhteiskunnan yhteinen ongelmanratkaisu, jota on luonnehdittu transdisciplinaarisuudeksi (ks. Thompson Klein ym. (toim.) 2000). Seuraavat hyvin alustavat huomiot tähyilevät tätä laajempaa lähtökohdtaa, vaikka liikkeelle lähdetäänkin tieteidenvälisyyden viitekehityksessä.

Chen silmäili evaluaatiotutkimuksen kehittämishaasteiden kolmatta askelmaa ja samalla tuolin kolmatta jalkaa jo TDE:ssä erityisesti ns. systeemisen mallin piirteitä hahmotellessaan (s. 172 ss.). Tällöin Chenille tyypilliset SH-korostukset väistyivät sivummalle huomion siirtyessä arvioitavan toiminnan uskottavien tavoitteiden määrittelyn ongelmiin ja tarkoittamattomien vaikutusten kysymyksiin. Näihin Chen totesi tarvittavan pikemminkin yhteiskuntatieteellistä kuin SH-lähestymistapaa (s. 173-).

Edellä on jo viitattu siihen, että PPE:ssä systeemisen arvioinnin painotukset ovat niukempia. Tämän taustalla lienee mm. PPE:n TDE:a käytännöllisempi painotus, johon systeemimallia on ollut ilmeisen vaikea sijoittaa (TDE s. 173). PPE:ssä arvioinnin tietoperustan laajentamista koskevat korostukset nousevatkin esiin tavoiteperusteiselle evaluaatioperinteelle rakentuvan lopputulosevaluaation yhteydessä. Evaluaatiotutkimuksen tietoperustan laajentamistarvetta PPE:ssä perustellessaan Chen ei siis enää eriytä toisistaan tavoiteperusteista ja "syste-

mistä" mallia, vaan pyrkii tähän erottamalla toisistaan efficiency ja effectiveness -mallit vaihtoehtoisina ohjelmatoiminnan tavoitevaikutavuuden evaluaation muotoina (s. 199-262). Näistä jälkimmäinen, effectiveness -malli, näyttäisi vastaavan Cookin ja esim. Raw Pawsonin esittämään vaatimukseen laajentaa evaluaatio-tutkimuksen tietoperustaa tutkimuksen tieteellisen ja yhteiskunnallisen hyötyarvon lisäämiseksi. Kun tutkimuksen tietoperusta efficiency -mallissa koostuu kunkin evaluaatiotutkijan itse omin tutkimusmenetelmin mahdollisimman kehittyneellä (kontrolloidulla) tutkimusasetelmalla hankkimasta primaariaineistosta, effectiveness -mallissa arvioinnin tietoperusta avautuu kattamaan myös olemassa olevan tieteellisen teorian ja tietämyksen sekä SH:ien ideat (s. 212).

Tilannetta voidaan tässäkin yhteydessä havainnollistaa ottamalla lähtökohdaksi käytännöllinen viisaus mahdollisuutena tehdä oikeita tai vääriä asioita oikein tai väärin. Analyysia voidaan vahvistaa ottamalla esiin Mario Bungen tapa määrittellä aksiologinen rationaalisuus yhtäältä tavoitteiden arvokkuudesta käsin, toisaalta niiden saavutettavuudesta käsin. Aksiologisesti rationaalista on toiminta, jonka päämäärät ja tavoitteet ovat tavoittelemisen arvoisia ja myös käytännössä saavutettavissa (Bunge 1987, 5-6).

Rakentuessaan tavoiteperusteisen arvioinnin lähtökohdille Chenin effectiveness -mallin mahdollisuus suodattaa "oikeita tai vääriä asioita" koskevaa viisautta tai ymmärrystä - kertyy se sitten esim. sosiaalietiiikan, uskonnonfilosofian tai muiden kulttuurin luomusten "alueelle" - jää käytännössä vähäiseksi. Etsiytyminen eri tieteen- ja tutkimusalojen laajempiin tietokerrostumiin voi tällöin merkitä korkeintaan sen arvioimista, ovatko tavoitteet käytännössä saavutettavissa, ja saavutettu. Mahdollisuus päätöksentekijöiden toiminnan normatiiviseen kritiikkiin jää myös disiplinaarisessa ulottuvuudessa niukaksi. Tästä muodostuu Chenin konseptin kolmas black-box - Chenin ohjelmateorian ja taksonomian normatiivisen voiman heikkouksia on jo edellä tarkasteltu.

Yhteiskunnallisen toiminnan tavoittelemattomien vaikutusten ja tarkoittamattomien seurausten kirjo ja yhteiskunnallinen merkitys on maapalloistumisen, talouden ja teknologian vaikutuskirjon lisääntymisen, alueellisen eriarvoisuuden syvenemisen ym. muutosten myötä pikemminkin lisääntymässä kuin vähenemässä.

Systeemiset riippuvuudet lisääntyvät kaikilla tasoilla: globaalilla, alueellisella, kansallisella, seudullisella ja paikallisella. Popper totesi jo teoksessaan Avoin yhteiskunta ja sen viholliset yhteiskuntatutkimuksen perustehtäväksi yhteiskunnallisen toiminnan tarkoittamattomien seurausten tutkimisen¹⁹. Nykytilanteessa sekä kansalaiset ja kansalaisyhteiskunta että poliittiset vallankäyttäjät tarvitsevat yhteiskunnallisista tavoitteista riippumatonta tietoa pikemminkin aikaisempaa kipeämmin kuin vähemmän.²⁰ Tässä tilanteessa teoriaperusteisen arviointikonseptin normatiivisella vajauksella ei tule ryhtyä "ostamaan" arvioinnin selkeyttä eikä myymään sen legitimiisyyttä edes "käytännöllisyyden nimissä".

Näkisinikin, että evaluaatiotutkijat ovat kypsiä ja vastuullisia määrittelemään toiminnalleen myös laajemmat normatiiviset puitteet kuin mihin PPE suoranaisesti tarjoaa tilaa. Teoriaperusteinen evaluaatiotutkimus, jonka "taksonomiaan" ei sisälly vaatimusta itsenäisestä ja päätöksentekijöistä riippumattomasta normatiivisesta argumentoinnista, on lähtökohdiltaan liian kapea-alainen. Tärkeämpää kuin yritys muotoilla omia itsenäisiä koulukuntia, lähtökohtia ja lähestymistapoja, joka useissa tapauksissa perustuu pikemminkin väärinkäsityksiin kuin aitoon refleksiivisyyteen (Oksanen 2003, 529-532, Virtanen 2005, 63), on hahmottaa riittävän monipuolisesti se, mitä tietoa yhteiskunta tutkimukselta nykytilanteessa ja tulevaisuudessa odottaa. Kuitumassa olevan kriittisen yhteiskuntatutkimuksen ja perinteisen (ohjelma)evaluaation välimaastossa on laajeneva black-box, joka ei voi olla yksinomaan evaluaatiotutkimuksen vastuulla, mutta myös evaluaatiotutkimuksen on kannettava siitä oma vastuunsa.²¹

Chen näkee monien muiden tavoin evaluaatiotutkimuksen tulevaisuuden sen disiplinaarisessa itsenäisyydessä. Itsenäisen tieteen- tai tutkimusalan sijasta olisin taipuvainen etsimään evaluaatiotutkimuksen ja myös teoriaperusteisen arvioinnin tulevaisuutta toiselta suunnalta. Sitä voitaisiin haluttaessa kutsua apu- tai arvo-tieteeksi. Sen tehtävänä on kehittää omaa itsemymmärrystään (porras 2), mutta yhtä hyvin epistemologisesti ja metodologisesti kuin tieteen-sosiologisesti evaluaatiotutkimus tullenee aina olemaan tärkeässä suhteessa riippuvainen varsinaisista substanssieteistä ja myös niille alistainen²². Tähän asemaan tulee liittyä avoimuus niin empiristä kuin normatiivista argumentointia

harjoittavien tutkimus- ja tietotoimijoiden suuntaan sekä kyky suodattaa lähtökohtia kriittisesti ja tasapainoisesti eri suunnilta. Tähän haasteeseen sisältyy olennainen osa kirjoitukseni alun evaluaatiotutkimuksen kolmannen kehittämisportaan ideaa. Näin ääriviivoiltaan ymmärrettyinä²³ evaluaatiotutkimuksen haasteet eivät liity yksinomaan tuolin selkänöjan ja kolmannen jalan kehittelyyn, vaan pyrkimykseen itsenäisestä ja tasapainoisesta tuolista kohden penkkiä, jossa disiplinaariset rajat ja jopa laajemmat kulttuuriset rajat häipyvät osaksi näköpiiristä. Samalla kirjoituksen alun kolmannen portaan haaste tulisi tulkituksi Cookia ja Cheniä avarammin, eksplisiittisemmällä toisen ja kuudennen portaan yhdistelmällä, joka voisi aikanaan kypsyä myös viidenneksi portaaksi.

Evaluaatiotutkimuksen kehittämishaastetta ei tällöin rajoitettaisi empiirisen evaluaatiotutkimuksen kehittämishaasteeseen, ymmärrettyinä tulosten nykyistä paremmaksi integroimiseksi perustieteenaloilla jo saavutettuun empiiriseen tietoon ja vm. tiedon nykyistä paremmaksi hyödyntämiseksi evaluaatiotutkimuksessa. Haaste avartuisi tutkimuksen normatiivisten (episteemisten, eettisten) lähtökohtien systemaattisempaan ja eksplisiittisempään avaamiseen ja esittämiseen, pysytellen kuitenkin pikemminkin tiedon kokoajan ja kriittisen suodattajan roolissa kuin hakeutuen omaan disiplinaariseen yksinäisyyteensä.

Tässäkin tapauksessa evaluaation yhteiskunnallisen ja tieteellisen käyttöarvon lisääminen edellyttää pikemminkin disiplinaaristen siteiden hillittyä vahvistamista kuin evaluaatiotutkimuksen omaehtoisuuden pidäkkeetöntä lisäämistä. Samalla on kuitenkin syytä muistuttaa ns. monitieteisyyden ongelmista versovasta ongelmakentästä, josta ns. transdisiplinaarisuuden ongelmakenttä puolestaan osaltaan versoo. Edellä transdisiplinaarisuutta luonnehdittiin lyhyesti viittaamalla tieteen ja yhteiskunnan uudelleen yhteiseen ongelmanratkaisuun. Edellä esitettyyn käytännölliseen viisauteen palaten transdisiplinaarisuuden haasteet ja ongelmat versovat osaksi sen epäilemisestä, etteivät tutkijat halua ottaa tai - kuten perinteisesti on ajateltu - ole kompetentteja ottamaan kantaa arvokysymyksiin, vaan heidän on etsiydyttävä laajempaan toimijakenttään. Viimeksi mainittua ajattelutapaa vastaan argumentoi jo 1970-luvulla erityisen monipuolisesti Johan Galtung ja Suomessa mm.

Erik Allardt.

Tässä yhteydessä ei ole mahdollista pohtia laaja-alaisesti transdisiplinaarisuuden ongelmia. Viittaus eräisiin puheenvuoroihin riittää. Niiden perusteella evaluaatiotutkimuksen tulevaisuus versoisi pikemminkin eräänlaisessa läpensä yhteiskunnallistuneessa torissa (Agora) kuin kiinteässä, jatkuvasti selväpiirteisemmäksi ja vahvemmaksi käyvässä itsenäisessä instituutionaalisisessa ja/tai disiplinaarisessa asemassa. Frederiksen ym. analysoivat kirjoituksessaan tutkimuksen arvioinnin muuttunutta roolia länsimaisissa yhteiskunnissa. Taustalla nähdään tieteen yhteiskunnallisen aseman ja yhteiskuntasuhteen muutos, tutkimuksen ja "muun yhteiskunnan" keskinäisriippuvuuden lisääntyminen sekä ennen kaikkea uusi tilanne, jossa "the rise of a whole new field of technical research evaluation systems should be understood as part of modernization process taking place in all western societies where the close links between social trust, visibility and accountability in any part of society have also reached science" (Frederiksen ym. 2003, 166).

Muutoksen myötä tutkimuksen evaluaation kriteerit, menetelmät ja lähestymistavat ovat jatkuvasti monipuolistuneet ja monimutkaistuneet, evaluaation instituutionaaliset puitteet eriytyneet ja hajautuneet, toimijat lisääntyneet ja agendat eriytyneet. Koko evaluaatiokulttuuri on käymis-tilassa. Pysyviä agendoja ei löydy. Tahoa tai kontekstia, joka "speaks back to science", etsitään kuumeisesti. Kehityksen tai muutoksen, jota tanskalaistutkijat jäljittävät useille muillekin evaluaation kohdealueille, vaihtoehtoisia lopputuloksia ja tulkintaskeemoja tutkijat esittävät kaksi. Yhtäältä vaihtoehtona on evaluaation suuntautuminen uusiin organisatorisiin kehyksiin sekä kriteerien, metodien ja lähestymistapojen edelleen jatkuva monimutkaistuminen - ilman järjestelmän perusteiden syvällistä laadullista muutosta. Tiede itse, tosin jatkuvasti monimutkaistuneessa toimintaympäristössä, määrittelee edelleen jonkinasteisella itsesäätelyllä, mitä evaluaatiot merkitsevät, "onko tehty oikeita tai vääriä asioita oikein tai väärin" ja mitä tästä seuraa. Toisena, tutkijoiden mukaan todennäköisempänä vaihtoehtona on ajautuminen tai hallittu ohjautuminen tilaan, jota tutkijat luonnehtivat kreikkalaisella termillä Agora, tori. Agorassa evaluaatioon osallistujat, evaluaatioforumit, evaluaatioprosessit, evaluaatioiden lopputulokset ja varsinkin niiden

tulkinnat ovat edellistä skenaariota epäselvempiä.

Agorassa toimijat vaihtelevat. Agorassa "kaikki virtaa". Tutkijoiden mukaan tärkeintä Agorassa on se, ettei kukaan kykene ennakoimaan, kuka asettaa keskeiset agendat ja mitkä tulevat olemaan keskeisiä ongelmia. Kiinnostavaksi puheenvuoron tekee se, että tutkijat eivät päätelmissään nojaudu ensisijaisesti teoreettisiin analyyseihin ja esim. Paul Feyerabendin esityksiin, vaan tieteen tutkijoiden laajaan kirjoon, empiirisiin aineistoihin ja case-kuvauksiin. Jos Frederiksenin ym. skenaarit yhdistetään Chenin PPE:n ja tämän kirjoituksen perusteisiin, näyttää siltä, että jo ensimmäisessä skenaariossa lähinnä yhdysvaltalaiselle perinteelle ja järjestelmälle rakentuvan ohjelmaevaluoinnin rakenteet ja arviointimallit, joita Chen esittelee verrattain yleispätevinä, näyttäisivät ajautuvan syvenevään kriisiin. Agora voidaan puolestaan nähdä Chenin ja muidenkin hallintayrityksistä jo tyystin irtautuvana skenaariona.

Filosofeista modernin yhteiskunnan ja hallinnon asiantuntijuuden legitimiisyysongelmia on mielenkiintoisesti ja puhuttelevasti valaissut Alasdair MacIntyre teoksessaan "Hyveiden jäljillä" (alkuteos "After Virtue"). Vaikka MacIntyre ei erikseen käsittele evaluaattoreita vaan pikemminkin yhteiskunnallisten toimintaohjelmien suunnittelijoita (esim. säädösvalmistelijat), hänen argumentaationsa ydin ulottuu myös evaluaation alueelle. Siteeraan erästä MacIntyren asiaa valaisevaa huomiota: "Hallinnon asiantuntijan itselleen vaatima asema ja palkkio osoittautuvat täysin perusteettomiksi, kun huomataan, ettei hänellä ole käytössään pätevää lainomaisten yleistysten joukkoa eikä siten juurikaan ennustevoimaa. Manageriaalisen tehokkuuden käsite sinänsä on yksi uusi kuvitelma aikamme moraalisten kuvitelmien joukossa, ehkä jopa niistä tärkein. Kulttuurimme manipulatiiviseen asenteeseen ei siis liity kovinkaan paljoa todellista ja onnistunutta manipulaatiota. En väitä, etteikö asiantuntijoina esiintyvien toimilla olisi suuria vaikutuksia ja ettemkö niiden takia kärsisi suuriakin vahinkoja, mutta asiantuntijuuden ajatukseen sisältyvä sosiaalisen kontrollin ajatus on vain naamio. Sosiaalisen maailmamme järjestys on hyvin kontrolloimatonta eikä kukaan voikaan olla sen ohjaksissa".²⁴

Todettakoon, että usuaristoteeliseksi luonnehditun joskin Suomessakin verrattain vilkkaasti

siteeratun MacIntyren johtopäätös perustuu seikkaperäiseen yhteiskuntatutkimuksen ja -tiedon epistemologiseen analyysiin.²⁵ Verrattaessa Frederiksenin ym. Agora -kirjoitusta MacIntyren analyyseihin oman kiinnostavan lisänsä tuo se, että MacIntyren eräs keskeinen väite tai tutkimustulos on se, että vanhan ja uuden ajan eettisissä koo-deissa vallitsee tärkeitä jatkumoa, mutta valistus on ideoineen katkaissut tuon jatkuvuuden; näennäisestä samankaltaisuudesta huolimatta (jälki)modernit moraalikäsitteet (oikeudet, velvollisuudet, oikeudenmukaisuus jne.) ovat menettäneet yhtymäkohtansa ja juurensa alkulähteilleen ja samalla käyneet läpinäkymättömiksi. MacIntyre ei näe yhteyden elvyttämistä enää mahdollisena - pohjimmaltaan siksi, että klassinen aristotelinen lähtökohta maailmankuvana, ontologiana ja hyvekäsitteenä on tullut hylätyksi. Tämän huomioiden kirjoittajien viitekehys on samanlaisesti "samanlainen" ja ikään kuin eri aaltopituudella. Kirjoitukseni valistuneen (ja valistukseen tarttuneen) lukijan mielestä MacIntyren analyysia voi ehkä käyttää apuna myös edellä esittämäni Chenin ohuen normatiivisuuden kritiikin vastakritiikissä: kun käsitys luonnollisesta tai lajihyvästä (esim. tutkimus, tutkija, yhteisö, polis) on ohentunut ja murtunut, evaluaatiolla edistettävät ja mittapuina käytettävät hyvät eivät ole enää funktio-käsitteitä, joiden oikeuttaminen mittapuina voidaan palauttaa välineellisen hyvän arvioiden objektiivisuuteen. Kun niiltä näin puuttuu objektiivisuus, on perusteetonta vaatia myöskään tutkimus- tai kalastusmatkakaritoilta sitoutumista joihinkin ehdottomiksi tulkittuihin lähtökohtiin siitä, "mikä on oikein (tai väärin) tekemistä". On keskeytettävä siihen, onko tuo oikea tai väärä "tehty oikein vai väärin".

Ehkäpä muun muassa rohkeus käydä keskustelua toimintansa perusteista on kuitenkin hyve, joka ylittää aikakausien rajat; tälle lähtökohdallehan myös esim. Habermasin tärkeimmät diskurssietiikan kehitelmät osaltaan perustuvat. Esimerkiksi Kotkavirran mukaan Habermas kehiteleekin mittapuitaan viime kädessä jo antiikissa kehkeytyneistä hyvän elämän konsepteista (Kotkavirta 1990).

Paluusta Agoraan, torille, voidaan puhua ainakin kahdessa merkityksessä. Vastaavasti kysyttäessä miksei Chenin taksonomia voi sitoutua mihinkään perustavaan arvolähtökohtaan, voidaan viitata kahteen suuntaan. Voidaan viitata siihen, että antiikissa politiikan teoria oli teoriaa

praksiksesta. Praksiksessa toiminnan arvo sisältyy toimintaan itseensä, eikä toiminnasta muodostuvaan lopputulokseen ja toiminnan vaikutuksiin. Tämä erottaa sen "tuotannosta", poiesiksesta, jonka arvo perustuu toiminnasta "irtautuvaan" lopputulokseen ja tämän vaikutuksiin. Tässä merkityksessä "oikean/väärän tekeminen" jää praktisen teorian ulkopuolelle. Praktinen teoria voi lausua jotakin vain siitä, millainen on "hyvä toimijuus" tai (prosessiarvioinnin termein) "hyvä toimintaprosessi" (vrt. alaviite 14). Paluu "torille", tai Himasen termiä seuraten "Hautomoon" (Himanan 1997), sanan kreikkalaisessa ydinmerkityksessä edustaisi tällöin paluuta praksiksen pariin. Praksiksessa yksilölliset/ yhteisölliset hyveet (esim. "rohkeus") auttavat käymään keskustelua siitä, mikä on "oikeaa ja oikein tekemistä". Chenin ajatus evaluaatiotaksonomiasta evaluaatiotaidon tai jopa - taiteen alueena sopii hyvin praksista "hyvänä toimijuutena" korostavaan praktisen teorian lähtökohtaan.

Kysyttäessä, miksei Chenin ole syytä sitoa taksonomiaansa mihinkään ulkopuoliseen arvo-lähtökohtaan, voidaan viitata myös moderniin politiikan teoriaan kohdistettuun kritiikkiin. Tämä kritiikki suuntautuu erilaisiä klassisia hyvän yhteiskunnan ja hyvän elämän malleja vastaan, mutta etenee tätä pidemmälle asettaessaan kyseenalaiseksi myös erilaiset modernimmat hyvää yhteiskuntaa koskevat mallit ja konseptit. Kullekin yksilölle ja yhteisölle ainakin osaksi spesifin "hyvän" ja "oikean" käsitteiden vallitessa emme kykene enää saavuttamaan yhteisymmärrystä siitä, mikä on kaikille yhteisesti "oikeaa/väärää" tai "tavoiteltavaa/ei-tavoiteltavaa" toimintaa. Lisäksi saatamme joutua hyvää yhteiskuntaa koskevien mallien kanssa tekemisiin utopismin ja jopa totalitarismin ongelmien kanssa (Lagerpetz 2002, 15-17). Paluu "torille" modernin politiikan teorian kritiikin tarkoitamassa merkityksessä merkitsisi luopumista yrityksistä lausua se, mihin tulisi pyrkiä ("oikeiden tai perusteltujen asioiden tekemiseen...") ja rajoittumista käytännöllisyyteen sanan modernissa, välineellistyneessä, ydinmerkityksessä. Tällöin rajoituttaisiin sen lausumiseen, miten tulee toimia tai on toimittu ("... oikein tai tarkoituksenmukaisesti").

Chenin taksonomia näyttää siis saavan tukea sekä klassisesta ajattelusta että moderniin politiikan teoriaan kohdistetusta kritiikistä. Taksonomia osoittaa tutkijoille "hyvin praksiksessa

toimimisen" perusteet (klassinen ajattelu). Toisaalta taksonomiassa luovutaan modernin poliittikan teorian kunnianhimoisista yrityksistä lausua jotakin normatiivista (ja samalla utopistista tai "totalitaristista"?) siitä, mihin tutkijan tulisi lopulta taksonomian avulla pyrkiä. Näin taksonomiassa omaksuttu peruslähtökohta: keskittyä sen lausumiseen mahdollisimman perustellusti, milloin evaluaatiotutkija toimii (oikeiden tai väärin asioiden kanssa) oikein, olisi riittävä ja perusteltu sekä klassisen praktisen (vs. poiesis)teorian lähtökohdian kannalta arvioituna että modernin universalismia tavoittelevan etiikan ja järjen kritiikin lähtökohdista arvioituna. Niistä käsin ajatellen me emme joko voi tietää arvoista, päämääristä tai tavoitteista tai sitten käytännöt ovat hyvää toimijuutta arvona sinänsä, eivät niiden kannalta ulkoisia arvoja tai päämääriä.

Rajoitun lopuksi yhteen vastahuomioon, jossa käytäntöjen arvoille haetaan kriittisen rationalismin hengessä kontrollipistettä inhimillisten käytäntöjen perustavista ongelmista käsin. Yritettäessä lausua jotakin rationaalisesti perusteltua arvoista - tutkimuksen sisäisistä tai ulkoisista - ja näin esim. Chenin taksonomiaan implisiittisesti sisältyvistä arvoista, kohdataan oikeutuksen, justifikaation, ongelma. Samalla kohdataan kehä. Jokaista yritystä lausua jotakin normatiivista seuraa uusi perusteluvaatimus ja samalla päättymättömältä vaikuttava perustelukierre. Chen pyrkii kiertämään tämän ns. justifikationismin tai foundationalismin ongelmaksi luonnehditun modernin tieteen oikeutuksen ydinongelman stakeholder -lähtökohdalle ominaisella pluralismilla. Chenin konseptissa kannanotto siihen, mitkä ovat evaluoitavan toiminnan (ohjelman tms.) kannalta keskeisiä ongelmia ja tavoitteita sekä kannanotto siihen, miten ongelmat ovat ratkaistavissa ja tavoitteet toteutettavissa (ohjelmateoria eli toimintamalli ja muutosmalli) muotoillaan eri osapuolten (SH:t) yhteisenä ratkaisuna. Lopputulosta voidaan luonnehtia esim. eri osapuolten yhteiseksi näkemykseksi siitä, mikä on tarkoituksenmukainen tiekartta, tai siitä, mihin suuntaan laivaa tulisi suuntia (vrt. Kant 1997, 47).

Varsinkin ihmisen ympäristösuhteen kannalta tilanne jää näin meneteltäessä usein ongelmalliseksi, koska asetelmasta puuttuu perustava edellytystekijä tai "Agora", fysis, ja samalla keskeinen "stakeholder": luonto. Hyvän yleiskuvan tilanteesta tarjoaa tässä suhteessa esim. tuore ympäristöraportti "Maailman tila 2005". Käytän-

nössä on tyypillistä, että juuri ihmisen ympäristösuhte ja luonnon ihmiselle asettamat vaatimukset on ulkoistettu pohdittaessa sitä, "mikä on tarkoituksenmukainen tiekartan mukainen suunta tai laivamatkan reunaehto". Kriittisen rationalismin lähtökohdista kuvattua lähtökohtaa on pyritty vahventamaan ja tavallaan objektivimaan sitomalla ongelmatulkinnat inhimillisessä elämässä evoluution seurauksena kohdattujen ja kohdattavien (elämän)ongelmien asettamiin vaatimuksiin. Esim. John Kekes (1987) on määntellyt elämänongelmat kokonaisuutena, joka muodostuu ihmisen suhteesta itseensä, ihmiseen suhteesta yhteisöön ja ihmisen suhteesta luontoon. Näin määriteltävät elämänongelmat on ymmärrettävä ja kohdattava kokonaisuutena. Larry Briskman (1987) on kehittänyt asetelmaa dialektisen epistemologian suuntaan. Siinä ongelmanratkaisut ja arvot kontrolloivat dialektisesti, vastavuoroisesti, toisiaan.

Evaluaatiotutkimuksen kannalta äskeinen merkitsisi sitä, että jokaista ohjelmateoriaa kontrolloitaisiin kahdelta suunnalta: ensiksikin vaatimuksella käsitteellistä ja määntellä ongelmat kokonaisuutena. Mikä tahansa holistinen ongelmanratkaisu ei vielä riitä toiminnan oikeuttamiseen, vaan käytettyihin keinoihin kohdistuu erilaisia eettisiä rajoitustekijöitä. Näin laajennettunakaan tieteen ulkoinen rationaliteetti- ja legitimitetilähtökohta ei vielä yksinomaan määrittele evaluaatioasetelmaa, vaan lisäksi on huomioitava tieteen sisäisiä arvoja (luotettavuus, objektiivisuus jne.). Kun yllä mainittu tieteen ulkoinen rationaliteetti- ja legitimitetilähtökohta ohjaa evaluaatiota "oikeiden asioiden pariin" ja asettaa jo keinoillekin eräitä perusrajoitteita, jälkimmäinen tieteen sisäinen rationaliteetti ohjaa evaluaatiotutkijaa "toimimaan oikein". Lisäksi "toimimista oikein" on tarpeen täydentää tutkimuksen eetosta koskevilla arvoilla ja lähtökohdilla.

Chenin taksonomian kannalta äskeiset ääniviivat merkitsisivät sitä, että SH:ien arvioinnin tarpeelle ja tarkoitukselle asettamia evaluaation lähtökohtia kontrolloitaisiin vaatimuksella ratkaista tai hallita elämänongelmat kokonaisuutena, kokonaisuutena. Tämä merkitsee evaluaation huomiokentän laajentamista ja johtanee käytännössä usein SH-ongelmateorian modifioimiseen esim. ihmisen ympäristösuhteen näkökulmasta. Kuvatunlainen asetelma veisi Chenin taksonomiaa kohden hänen TDE:ssä kehittelemäänsä systemisen arvioinnin lähtö-

kohtaa. Samalla asetelma lähestyy eräitä uusaristotelismin nimellä kulkevan teoriakonseptin peruslähtökohtia. Jotta evaluaatiotutkimus ei ajaudu esimerkiksi operoimaan ylivoimaisten laajojen ongelmanratkaisujen kanssa ja kykenee tuottamaan perusteltua tietoa, ongelmateoriaa ja ohjelmateoriaa on ohjattava myös mm. tiedon luotettavuutta ja yleistettävyyttä koskevilla vaatimuksilla. Nämä vaatimukset johtanevat usein evaluaation huomiokentän supistamiseen sekä evaluaation kysymyksenasettelujen rajoittamiseen. Olennaista kuitenkin on, että huomiokentän ja arvojen rajoittamista kontrolloitaisiin tavalla, jota esim. Chenin taksonomian SH-lähtökohtiin ei lainkaan sisälly. Edellä esitetty kahdensuuntaisen kontrollin vaatimus, joka edellyttää tasapainoilua huomiokentän laajentamisen ja supistamisen välillä, vastaa itse asiassa lähtökohdiltaan edellä esitettyä vaatimusta Chenin taksonomian yhdensuuntaisen sarakelogiikan muuntamisesta kaksisuuntaiseksi (ks. edellä s. 13-14). Erona on kuitenkin se, että tässä yhteydessä Chenin arvioinnin tarkoituksista ja tarpeista johtamaa taksonomiaa on laajennettu taksonomian kannalta ulkoistaa arvojen suuntaan.

Yllä luonnosmaisesti kuvaamaani yritystä hallita evaluaatiotutkimuksen ja esim. Chenin taksonomian justifikaatio-ongelmaa voidaan kritisoida mm. siitä, ettei myöskään siinä esitetty evaluaatioasetelma "vakaudu" tasapainopisteeseen, ristiriidattomaan tilaan, vaan perustelukettu näyttäisi jatkuvan arvojen - tieteen sisäisten ja ulkoisten - ja ongelmanratkaisujen dialektiikkana loputtomiin, ad infinitum. Ehkäpä tällaisessa arvojen ja ongelmanratkaisujen dialektiikassa kuitenkin ilmenee juuri torin, Agoran, ilmikuva tai keskeinen ulottuvuus. Rationaalisuus käsitteestä itsessään ei seuraa kuvatunlaisen ristiriidan, dialektiikan, torjumista (Tranoy 1980, 199-200), tai "torin sulkemista". Niille, jotka kritisoivat kriittisen rationalismin edellä kuvattua lähtökohtaa käytännöllisyyden (yksinkertaisuuden) tai vaikkapa PPE:n perusnäkökulmasta, tyydyn vastaamaan kysymyksellä: eikö kuvatunlainen ongelmanratkaisujen ja arvojen jatkuva tasapainottelu ja vuorottelu, jota on mahdotonta pukea teoreettiseen kaavaan ja yksiselitteiseen metodisääntöön tai metodisääntöjen kokoelmaan, juuri luonnehdi syvällisesti evaluaatiota käytännössä? Olennaista on lisäksi se, että vm. käytäntökäsité pitää sisällään käytännöllisesti viisaan toiminnan molemmat aspektit: päämäärät ja keinot,

se sisältää sekä "oikeat asiat" että "toimimisen niiden parissa oikein".²⁶

Entä tulevaisuus? Edellä ääriviivoiltaan hahmottelemieni ristiriitaisten skenaarioiden perusteella evaluaatiotutkimuksen edessä ja alla voi olla tuoli, jonka jalat vahvistuvat ja monipuolistuvat, penkki, jossa rajoja ennakkoluulottomasti ja rohkeasti (tai nöyrästi, nöyryys ei tosin ollut MacIntyren mukaan klassinen hyve, TO:n huom.) ylitetään ja rakennetaan - tai tori, jolla ei ole lainkaan edes penkkejä saati tuoleja. Kussakin näistä skenaarioista käytännöllinen viisaus mahdollisuutena tehdä oikeita tai vääriä asioita oikein tai väärin saa oman tulkintaskeemansa. Chen rakentaa PPE:ssä evaluaatiotutkimuksen tuolille ensisijaisesti lisäjalkoja, mutta joitakin aineksia hän tuo myös laajemman ja pidemmän penkin rakennusaineiksi. Toria hän ei suunnittele - itse asiassa PPE voidaan luonnehtia ensisijaisesti määrätietoiseksi yritykseksi saada evaluaatiokenttä hallintaan tavalla, joka on pikemminkin vastakohta tanskalaistutkijoiden Agora-skenaarille, vaikkakin SH-korostukset tuovat myös tähän suuntaan ehkäpä viisaasti kurottautuvan perusuunteen (vrt. Ahonen 2001, 103). Ainoa mikä näyttää selvältä on se, että tilanne ei ole kovin selkeä. Jo tämänkin vuoksi keskustelun tulisi jatkua, ellei torilla niin penkillä tai tuolilla. Koska evaluaatiotutkimuksen suunta ei ole ennalta määrätty, voisimme ehkä katsoa jokaisen evaluaatiotutkijan velvollisuudeksi osallistua tulevaisuutta koskevaan keskusteluun ja ainakin pohdintaan (vrt. Lagerspetz 2002, 26-27).

5. LOPUKSI

Ainakin 1930- ja vielä 1940 -luvullakin Suomessa laulettiin laulua "Raja railona aukeaa - takana länttä ja Eurooppaa, edessä Aasia, Itä ... - .. varjelen vartija sitä". Vaikka Chen tavallaan edustaa laulun antiteesiä - ainakin tähänastisella elämäнкаarellaan²⁷ - voi kysyä, onko evaluaatiotutkimuksen kokonaisuudessaan otettava kantaa tähän lauluun sen "henkeä" ennakkoluulottomammin ja kokonaisvaltaisemmin. Ehkäpä myöskään evaluaatiotutkimuksen rajoja ei voi nähdä niitä ylittämättä. Lieneekö tässäkin hankkeessa jälkiviisaus parasta viisautta? Mutta kenties kykymme tehdä oikeita asioita oikein - ilmaisun koko merkityssyvyydessä - ilmeneekin vasta jälkikäteenä (ex post). Tämähän on eräs evaluaatiotutkimuksen ydinajatus, joskin sek

alkaa jo olla hämärtymässä.

VIITTEET

¹ Evaluation alue oli Cookin mukaan vuosituhanteen vaihdetta lähestyttäessä sekavassa tilassa. Selostettuaan useiden tutkijoiden yrityksiä Cook toteaa evaluaatiotutkijoilta edelleen puuttuvan yhtenäisteoria evaluaatiosta, sen luonteesta, organisoinnista ja toteuttamisesta, institutionalisoinnista sekä tulosten käytön stimuloinnista. Samalla hän kiinnittää huomiota siihen, että tämän tasoisten yritysten perustaksi tarvittava tutkijoiden kriittinen massa on vähenemässä tutkijoiden ikäännyttyä, siirryttyä muille alueilla ja/tai nuorempien keskittyessä kapeampiin metodologisiin kysymyksiin. Viitattuaan yrityksen vaativuuteen ja "yksijalkaisenkin tuolin" merkittävään edistymiseen hän toteaa lopuksi eräänlaisena yhteenvetona, että "I want to see a field that is much more than social science methods applied to the study of social interventions" (1997 s. 50). Näin lausueessaan Cook tulee mm. korostaneeksi, että äsken mainittujen interventioiden luonnetta, vaikutuksia jne. koskevat havainnot ovat joka tapauksessa teoriapitaisia, ja että evaluaatiotutkimuksen on tunnistettava ja tunnustettava tämä myös teoreettisella tasolla. Vrt. Virtanen 2005.

² Cook 1997 s. 47-50.

³ Tähän liittyvästä yrityksestä ympäristötutkimuksen alueella ks. Oksanen 2002 s. 113 ss. Ks. myös Virtanen 2005, 64 sekä ns. transdisiplinaarisuuden ongelmista yleisemmin kokoavasti Thompson Klein ym. (toim.) 2001.

⁴ Terho Vuorela kiinnitti artikkelikäsitteilytoimuksesta antamassaan kommentissa huomiota kehittämissaasteiden portaikon mahdollisiin ongelmiin. Portaikko ei jäsennny loogisesti yhden kriteerin perusteella, vaan sitä "rakennettaessa" on esim. epistemologisten seikkojen ohella kiinnitetty huomiota tieteen- ja tutkimusalojen suhteisiin liittyviin kysymyksiin (esim. erityistieteet-filosofia: 5-6 porras; evaluaatiotutkimus omana (apu)tieteenä suhteessa menetelmätieteisiin: 2-1 porras, jne.). Myös 3. ja 2. portaan suhde voi olla ongelmallinen, vaikka Cook itse ei näe näiden portaiden välillä perustavaa ristiriitaa. Lisäksi kolmannella portaalla viitattu synteesi saattaa lisätä evaluaatiotutkimuksen tulosten tieteellistä käyttöarvoa, mutta ei välttämättä niiden yhteiskunnallista vaikuttavuutta. (Tältä osin Cook itse viittaa evaluaation kehittämissuuntana konsultteihin, jotka kykenevät hänen mukaansa hyödyntämään eri disipliinien tietorakenteita paremmin kuin evaluaatiotutkijat).

⁵ Vrt. Kiikeri & Ylikoski 2004, 21. Kiikerin ja Ylikosken huomion kohteena oleva (empiirinen) tieteen tutkimus voidaan tosin nähdä myös kuvatus aukon välimaaston täydentäjänä, sijoittuahan se tavallaan tieteenfilosofian ja perustieteiden (ja niiden mahdollisesti harjoittamien refleksiivisten pohdintojen) välimaas-

toon. Toisaalta henkilökohtaiset kokemukseni tuosta välimatkasta ovat johtaneet täysin päinvastaisinkin pohdintoihin. Arviointitutkimuksen alueelle sijoittuneen väitöskirjani (Oksanen 2003) ensimmäinen vastaväittäjä, oikeusfilosofi ja yleisen filosofiankin puolella väitellyt Juha-Pekka Rentto, totesi väitöstilaisuudessa kuulleensa evaluaatiotutkimuksesta vasta verrattain hiljattain. Silti (tai siksi?) hän kykeni tarjoamaan koko joukon vaikeita antiteesejä, joihin arviointitutkimuksen erityistieteellinen paradigmatonaisuus ei juuri tarjoa suuntaviivoja. Hänet lienee sijoitettava yllä esittämässäni luokituksessa "3" tutkijoihin, vaikka Rentto ei arviointitutkijan luonnehdintaa epäilemättä hyväksyisikään. Vrt. Sihvolan ns. soveltavan etiikan hyödyllisyydestä esittämään (2004 s. 28-31, ks. myös Oksanen 2002 s. 113-116.).

⁶ Kuten Ahonen huomauttaa, ilman evaluaatiometodologiaa evaluaatiometodien tarkastelu jää kuitenkin luetteloinniksi, kuvailuksi ja rämminäksi metodisessa suossa. Ahosen mukaan sen tehtäväksi jää metodien konkreettisten käyttöyhteyksien täsmentäminen, metodien sovelluksien suhteuttaminen kysymykseen tulevien toimijoiden strategisten päämäärien edistämiseen sekä metodien läpikäymistä kysymykseen tulevilla laajassa mielessä kulttuurisilla merkityksillä (1994 s. 7).

⁷ Olen saanut artikkeliin virikkeitä väitöskirjani ohella erityisesti Huey-Tsyh Chenin kanssa marraskuulla 2004 Tampereella, Helsingissä ja Porvoossa käymistäni keskusteluista. Chen kannusti pohtimaan teoriaperusteisen evaluaation kysymyksiä ja antoi lukuisia virikkeitä jatkopohdintoille. Lisäksi haluan erityisesti kiittää Terho Vuorelaa hänen artikkelikäsitteilyluonnoksesta antamistaan arvokkaista kommentteista. Tekijä vastaa luonnollisesti kaikilta osin yksin kirjoituksesta painotuksineen ja huomioineen.

⁸ Chen itse eriyttää toisistaan loogisen mallin ja ohjelmateorian niiden syvyyden ja tarkoituksen perusteella, vaikka toteaaakin loogista mallia voitavan käyttää eräänlaisena ohjelmateorian yksinkertaistettuna versiona: Looginen malli tarjoaa panos-, prosessi- ym. elementteineen "virstanpylväitä" (milestones) ohjelman jatkuvalle monitoroinnille kvantifoimalla seurattavia asioita yksinkertaisiksi tunnusluvuiksi, kun taas ohjelmateoria "strive to ask, what makes a program sound and how this is accomplished" (s. 36). Lisäksi ohjelmateoria tarjoaa mm. arviointistrategioita.

⁹ 1990, 143 ss. Systeemimallin ongelmaksi evaluaatiotutkimuksen näkökulmasta Chen toteaa sen peruskäsitteiden epäselvyyden ja sen, että poliittinen prosessi tarvitsee joka tapauksessa ajan tasalla olevaa tietoa tavoitteiden toteutumisesta (s. 173). TDE:ssä Chen valottaa eräitä tähän liittyviä kysymyksiä. PPE:ssä "systeemimallista" on pääosin vaiettu, vaikka nimenomaan käytännön evaluaation näkökulmasta juuri kyseisen mallin selkiytymättömyys olisi kipeästi perustellut myös laaja-alaisemman arviointiaselman tarkastelua. T & K-toiminnan arvioinnin näkökulmasta ks. esim. Rogers ym. 2001 sekä yleisemmin Stame 2004.

¹⁰ Erään jo vuosia sitten jäljittämäni nettilähteen mukaan, jota en tätä artikkelia kirjoittaessani enää samasta osoitteesta tavoittanut (kirjoituksessa ei ole identifioitu tekijää), "TDE looks at the theory according to which the components cooperate to produce the outcomes". Näin ymmärrettynä TDE kuuluu laajempaan analyttisten arvioiden "perheeseen". Siinä evaluaation kohde identifioidaan joko kohteen komponentit ja/tai niiden suhteet eriyttämällä (esim. kellon osat arvioidaan osina tai niiden suhteina eli järjestelmänä kuten TDE:ssä) tai kohteen laadulliset dimensiot eriyttämällä (esim. tarkkuus, esteettisyys, kompaktisuus...). Ns. holistisessa arvioinnissa arvioinnin kohdetta ei eriytetä komponentteihin tai dimensioihin, vaan sitä arvioidaan kokonaisuutena (esim. kellon "tyydyttävyyttä"). Vrt. Stame 2004, 61-62, joka identifioi "teoriaorientoituneen evaluaation" eri muotoja sen perusteella, miten nämä määrittelevät ja ratkaisevat "metodiperusteisen evaluaation" black-box-ongelman. Se liittyy Stamen ajattelussa ja paljon yleisemminkin komponentteihin, mutta tässä artikkelissa eräänlaista black-box'ia on jäljitetty myös normatiiviselle puolelle, eli dimensioita koskeviin normatiivisuuteen teorioihin.

¹¹ Eräänä lukuisista esimerkeistä Chen ottaa esiin koulutusohjelman evaluoinnin, jolla pyritään parantamaan opiskelijoiden menestystä (s. 36-37). Ohjelmateoriakonseptin mukaan evaluoinnissa tehdään kolmenlaisia tärkeitä asioita: 1) hankitaan tietoa tavoitteiden saavuttamisesta, 2) kysytään kuinka tehokkaasti toimintamalli implementoitiin sekä 3) tutkitaan toiminnan perustana olevien kausaalimekanismien roolia. Näiden pohjalta evaluaattori voi keskittyä dokumentoimaan ohjelman toteutusta, kuinka opiskelijat rekrytoitiin sekä kuinka opettajat opettivat ja miten motivoituneita he olivat ohjelman hyödyntämiseen. Kausaalimekanismeista voidaan tutkia, edistääkö tavoitteiden saavuttaminen todella opintoja. Onko tavoitteet saavutettu opettamalla testit opiskelijoille (manipulointi) tai saatamalla heikosti menestyvät opiskelijat epäoikeudenmukaisten sanktioiden kohteeksi? Tällainen evaluointi mahdollistaa ohjelman toimivien ja toimimattomien osioiden ("kontekstien") erottamisen toisistaan. Korostus yhdistää Chenin ainakin Pawsonin ja Tilley'n realistisen evaluaation eräisiin keskeisiin lähtökohtiin sekä GAO:n Chelimsyn johdolla kehittelemän ns. evaluaatiosynteisin joihinkin ydinajatuksiin.

¹² Ks. siten esim. Niinistö 1984 s. 91 kvantitatiivisen ja kvalitatiivisen "arviointiparadigman" laajemmista liitynnöistä (filosofinen ja teoreettinen perusta, tiedon käsitys jne.) esim. aristoteliseen ja galileiseen tiedetraditioon, joita arviointitutkijat tyyppillisesti varsin niukasti tuovat esiin. Vrt. Virtanen 2005, 63 sekä Chelimsky 1997, 25, joka puolestaan tähdentää evaluaatiotutkimuksen luonnetta "eklektisenä yrityksenä".

¹³ Chen korostaa kyllä "ulkoisten arvojen" merkitystä toiminnan "sisäisten arvojen" rinnalla tähdentäessään holistisen lähestymistavan tärkeyttä (s. 9-10, 234). Hän ei kuitenkaan systematisoi tai artikuloi tähän liittyviä lähtökohtia, eräitä yleisluonteisia viittauksia lukuunot-

tamatta (esim. s 18-19 viittaukset Weberiin ja Durkheimiin). Galtungin trilateraaliseen tiedekonseptiin (1977 s. 60) suhteutettuna Chenin viittaukset Weberiin ja Durkheimiin voi tulkita niin, että selittävän teorian ohella Chenin katsoo tarvittavan konstruktivisia teorioita, jotka ohjaavat arvojen ja tavoitteiden toteutumisen arviointia (ovatko keinot päteviä). Varsinaisissa normatiivisissa teorioissa hänen lähtökohtansa supistuu kuitenkin yksipuolisesti tavoiteperusteisen arvioinnin alueelle (s. 199-200), jääden samalla arvokysymyksissä Durkheimin lähtökohtaa olennaisesti pidättyvämmäksi (ks. esim. Raatikainen 2004, 139). Tätä lähtökohtaa ns. black-box evaluaation yleinen kritiikki, jota Chen harjoittaa toisinaan selvästi eettisinkin perustein (ks. esim. s. 234 Scrivenin kritiikkiä), ei muuta miksikään. Vrt. kuitenkin Chen 1990, 157-164.

¹⁴ Chen käyttää teoksessaan jossakin määrin ominnakeista käsitteistöä, jolle ei ole välttämättä suoraan löydettävissä sopivaa suomenkielistä vastinetta.

¹⁵ Praktisen teorian käsitteestä ks. seikkaperäisesti Sarvimäki 1988 s. 232 ss. Hänen mukaansa praktisella tieteellä on erilaisia tehtäviä eri tiedonintressien ulottuvuuksilla. Praktisen tieteen tuottaman tiedon avulla oo. alueen toimijoiden tulee voida ymmärtää tilanteensa aiempaa paremmin, orientoitua ympäristöönsä aiempaa paremmin, löytää oikeat ja aiempaa paremmat keinot, asettaa tavoitteensa sekä arvioida toimintaansa. Sarvimäen praktisen tieteen määrittelemää voidaan ehkä pitää "poiesiorientoituneena", onhan sillä selvästi klassisen praksiskäsitteen ylittäviäkin "vaikuttavuustavoitteita".

¹⁶ Åberg kiinnittää huomiota arvioinnin perusongelmien (arvioinnin tarkoituksen määrittely = tarkoituserongelma; relevantin ja pätevän tiedon määrittely = tieto-ongelma; relevanttien arviointikriteerien määrittely = arvosteluongelma) keskinäiseen suhteeseen tavalla, jolla on merkitystä myös Chenin taksonomian käytännöllisiä soveltamisongelmia peilattaessa. Chenin taksonomia etenee lähinnä sarakelogiikalla tarkoituserongelma → arvosteluongelma → tieto-ongelma, jossa edellinen määrittelee aina seuraavan puitteet. Käytännössä joudutaan huomioimaan myös toinen Åbergin esittämistä logiikoista: tieto-ongelma → arvosteluongelma → tarkoituserongelma. Tiedon saatavilla olo, laatu ja luonne asettaa osaltaan rajat arvosteluongelman ratkaisemiselle ja määrittelemiselle niin, että arviointi käy ylipäänsä mahdolliseksi. Näin määrittyvä mahdollisuus asettaa sitten osaltaan puitteet sille, mitä tarkoitusta arviointi voi viime kädessä toteuttaa (Ks. Åberg 1997 s. 86). Miten nämä logiikat sovitaan yhteen, siitä taksonomia ei voi kertoa oikeastaan mitään.

¹⁷ Kirjoittajan keskustellessa marraskuulla 2004 Chenin kanssa taksonomian kontekstuaalisuusongelmista esimerkkinä tarkastusorganisaatio, Chen piti "mahdollisena" tarkastusorganisaation suuntautumista arvioimaan vaikuttavuuskysymyksiä. Hän kiinnitti kuitenkin huomiota mm. vaikuttavuusarviointien pitkään aikajänteeseen ja muihin hankaluuksiin samoin kuin

siihen, että parlamentti tarvitsee tarkastusorganisaatioilta ennen kaikkea pikaista palautetietoa. Chen viittasi tässä suhteessa erikseen "vianetsintästrategiaan". Chenin taksonomiassa tätä strategiaa sille ominaisine lähestymistapoineen ja menetelmineen ei kuitenkaan käytetä muussa kuin kehittävässä tarkoituksessa. Chen suorastaan varoittaa käyttämästä "vianetsintästrategiaa" vastuullistavassa tarkoituksessa, koska vianetsinnässä korostuvat menetelmien joustavuus, nopeus ja palautteen tehokas informointi eteenpäin. Nämä vaatimukset eivät ole käytännössä helposti yhteensovittavissa vastuullistavalle arvioinnille välttämättömien "luotettavuuden, läpinäkyvyyden ja ennakoitavuuden" vaatimuksien kanssa.

¹⁸ Cook 1997, 36-37. Myös Chelimsky on korostanut tarvetta nähdä "evaluaatioperspektiivien (tarkoitusten) ylimenoja" ja "evaluaatioperspektiivien komplementaarisuus", joskin kiinnittää huomiota myös niiden eroihin (1997, 14-15, 22).

¹⁹ Popperin annista evaluaatiotutkimukselle ks. yleisesti Rieper 2004 sekä erityisesti teoriaperusteisen arvioinnin näkökulmasta Oksanen 2003, 266 ss., 538, 540-541 ja 544. Kiinnostavan vaihtoehdon Popperin ankaralle falsifikationismille tarjoaa Lakatosin "sofistikoitunut falsifikationismi", jossa tieteen ja tutkimuksen kehittymisen ennakoimattomuus ja epävarmuus saa Popperin falsifikationismia enemmän tilaa (ks. Oksanen 2003, 431-436, 481 ss.).

²⁰ Yhteiskuntatutkimuksen vaikuttavuutta pohtieessaan Jorma Sipilä on perustellut monipuolisesti päätöksentekijöistä riippumattoman itsenäisen yhteiskuntatutkimuksen tärkeyttä (2003).

²¹ Artikkeliluonnosta kommentoidessaan Terho Vuorela kiinnitti monen muun tärkeän seikan ohella huomiota siihen, säilytetäänkö evaluaatiotutkimukselle artikkelissa vastuuta, joita se ei voi eikä sen tule yksinomaan kantaa.

²² Esimerkiksi Pekka Sulkunen kiistää arvioinnilta ja arviointitutkimukselta jopa tutkimuksen statuksen (2004), tähän liittyen ks. myös Virtanen 2005, 64-65.

²³ Ks. seikkaperäisemmin Oksanen 2003, 481-505 ja kokoavasti s. 509-545.

²⁴ MacIntyre 2004, 133. Tanskalaistutkijoita ja tavaltaan ajatuksellisesti MacIntyreäkin lähelle näyttäisi asemoituvan myös Pekka Sulkunen, joka näkee vaikuttavuusarviointien tulokset totuusarvoltaan pääsääntöisesti pikemminkin fiktioina kuin faktoina (2004). Tilanteen monimutkaisuutta ja läpinäkyvättömyyttä kuvaa osaltaan se, että Petri Virtanen näkee puolestaan tuoreessa katsauksessaan valoa pikemminkin arviointitutkimuksen uudessa suuntautumisessa vaikuttavuusarvioinnin kysymyksiin, jota hän luonnehtii "metodologiseksi paluuksi" (2005, 66-67). Myös Chenin effectiveness -mallissa ja sille rakentuvassa vaikuttavuuden arvioinnissa (tai ehkä paremminkin hänen esittämässään bilateraaliossa strategiassa, s. 212) avataan uusia näkökulmia kausaalipäätelmiin myös muissa kuin perinteisissä eksperimentaalisen paradigman mukaisissa koe- ja kvasikoeasetelmissa. Chen

kuitenkin näkee tässä myös vielä tärkeitä uusia tutkimushaasteita (s. 247). Laadullisten menetelmien ja lähestymistapojen mahdollisuuksiin tukea kausaalipäättelmien tekemistä suhtautuu varovaisen optimistisesti mm. Cook (1997, 35).

²⁵ MacIntyre reunustaa epistemologisen analyysinsa itse asiassa ontologiseen analyysiin, jossa nousevat esiin erilaiset ennustamattomuutta yhteiskunnassa tuottavat lähteet (radikaalien käsitteellisten innovaatioiden ennustamattomuus, yksilöiden oman päätösketjun ennakoimattomuus, sosiaalisen elämän peliteoreettinen luonne sekä puhdas sattuma, s. 119-134).

²⁶ Ks. myös Parker 1996, 31-33. Vrt. Stamen Weis- sin (theory-based evaluation) sekä Chenin ja Rossin (theory-driven evaluation) eroista esittämään. Näyttää siltä, että Stamen ns. teoriaorientoituneelle evaluaatiolle esittämät kokoavat piirteet saavat edellä esittämiini PPE:n kehittämistarpeisiin ainakin jonkinlaisia liittymäkohtia (2004, 61, 63).

²⁷ Tilalottuvuudessa Chen liikkuu idästä länteen, mutta ajallisessa ulottuvuudessa hän näyttää välttävän yhtä laajaa liikettä, hylkäähän hän selkeästi paluun Agoraan, torille, sanan kreikkalaisessa perimerkityksessä. Vai hylkääkö - kirjoituksen eräät loppuhuomiot jättävät tilanteen hieman epäselväksi. Holistisuudessa Cheniä odottaa joka tapauksessa haaste, jolla on tärkeitä tilallisia ulottuvuuksia lännestä itään ja ajallisia ulottuvuuksia (post)modernista antiikkiin. Käytäntökäsitteen kokonaisvaltainen analyysi sitä seuraavine synteeseineen sekä edeltää että seuraa tällaista siirtymää. Kirjoitukseni alussa esitetyssä evaluaatiotutkijoiden porrastuksessa filosofit ovat kuudennella portaalla jo ottaneet eräitä askelia tähän suuntaan (ks. yleisesti Pihlström ym. (toim.) 2002). Ehkäpä viides porras saakin tulevaisuudessa huomiota sekä filosofian että erityistieteiden suunnalta. Ellei näin tapahdu ja elleivät käytännön toimijat ole aktiivisesti mukana "hankkeessa", hyviä tuloksia ei liene odotettavissa. Vai voidaanko praksikselta sellaista odottaakaan? Ehkäpä onkin siirryttävä poiesiksen, tuotannon, alueelle?

LÄHTEET

Ahonen, Pertti (1994): Julkisen sektorin uudistaminen ja sen arviointimetodit evaluaatiometodologisesta näkökulmasta, teoksesta Tutkimaton uudistus? Julkisen sektorin uudistukset tutkimushaasteena, Valtionvarainministeriö. Hallinnon kehittämisosasto, Helsinki (Ahonen 1994)

Ahonen, Pertti (2001): Evaluaatio-oppia edistyville. Hallinnon Tutkimus 3/2001 s. 102-117 (Ahonen 2001)

Briskman, Larry (1987): Historicist Relativism and Bootstrap Rationality, teoksesta Rationality: The Critical View, ed. Agassi & Jarvie, Nijhoff International Philosophy Series 23, Dordrecht s. 317-339 (Briskman 1987)

Bunge, Mario (1987): Seven Desiderata for rationality, teoksesta Rationality: The Critical View, ed. Agassi & Jarvie, Nijhoff International Philosophy Series 23, Dordrecht s. 5-15 (Bunge 1987)

Chelimsky, Eleanor (1997): The Coming Transformations in Evaluation, teoksesta Evaluation for the 21st Century. A Handbook, London & New Delhi s. 1-26 (Chelimsky 1997)

Chen, Huey-Tsyh (1990): Theory-Driven Evaluation, Sage, London (Chen 1990)

Chen, Huey-Tsyh (2004): Practical Program Evaluation. Assessing and Improving Planning, Implementation, and Effectiveness, Sage, Thousand Oaks (Chen 2004)

Cook, Thomas (1997): Lessons Learned in Evaluation Over the Past 25 years, teoksesta Evaluation for the 21st Century. A Handbook, London & New Delhi s. 30-52 (Cook 1997)

Diez, Marie-Angeles (2002): Evaluating New Regional Policies. Reviewing the Theory and Practice, Evaluation Vol 8 (3) s. 306-321 (Diez 2002)

Fischer, Frank (1980): Politics, Values and Public policy: The Problem of Methodology, Colorado (Fischer 1980)

Frederiksen, Lars Frode & Finn Hansson & Soren Barlebo Wenneberg (2003): The Agora and the Role of Research Evaluation, Evaluation Vol. 9 (2) s. 149-172 (Frederiksen ym. 2002)

Galtung, Johan (1977): Methodology and Ideology. Essays on Methodology, Kopenhagen (Galtung 1977)

Himanen, Pekka (1997): Hautomo. Verkkojen filosofia, Atena Kustannus Oy, Jyväskylä (Himanen 1997)

Kant, Immanuel (1997): Prolegomena, suom. Oittinen Vesa, Helsinki (Kant 1997)

Kekes, John (1987): Rationality and Problem-Solving, teoksesta Rationality: The Critical View, ed. Agassi & Jarvie, Nijhoff International Philosophy Series 23, Dordrecht s. 264-281 (Kekes 1987)

Kiikeri, Mika & Petri Ylikoski (2004): Tiede tutkimuskohteena. Filosofinen johdatus tieteen tutkimukseen, Gaudeamus, Helsinki (Kiikeri ym. 2004)

Kotkavirta, Jussi (1990): Kriittinen teoria ja diskursiologiikka. Skeptisiä huomioita diskursiologiikasta, Tiede ja Edistys 1/1990 s. 38-50 (Kotkavirta 1990)

Lagerspetz, Erik (2002): Poliittinen teoria ja käytäntö. Teoksesta Käytäntö, toim. Pihlström ym.. Yliopistopaino, Helsinki s. 15-30 (Lagerspetz 2002)

MacIntyre, Alasdair (2004): Hyveiden jäljillä. Moraaliteoreettinen tutkimus, suom. Niko Koponen, Gaudeamus, Helsinki (alkuteos After Virtue. A Study in Moral Theory, Notre Dame, 1981) (MacIntyre 2004)

Niinistö, Kari (1984): Aikuisenkoulutus ja sen evaluointi: uusia sovellutuksia lähinnä tulkinallisesta näkökulmasta. Valtion koulutuskeskus, Julkaisusarja B nro. 23

(Niinistö 1984)

Oksanen, Timo (2002): Ympäristöoikeuden tutkimusintressit erityisesti arvojen näkökulmasta: yksi askel taakse ja kaksi eteen - vai? Teoksesta Oikeudenalojen rajat ja rajattomuus, julkaisu Turun Oikeustieteen päiviltä 23.-24.11.2001, toim. Pekka Viljanen, Turun yliopiston oikeustieteellinen tiedekunta, s. 107-144 (Oksanen 2002)

Oksanen, Timo (2003): Oikeussääntelyn teoriaperusteinen arviointitutkimus. Acta Universitatis Tampereensis 958, Tampere University Press (Oksanen 2003)

Parker, Kelly A. (1996): Pragmatism and Environmental Thought, teoksesta Environmental Pragmatism, toim. Light & Katz, London s. 21-37 (Parker 1996)

Pihlström, Sami & Kristina Rolin & Floora Ruokonen (toim.) (2002): Käytäntö, Yliopistopaino, Helsinki (Pihlström ym. (toim.) 2002)

Raatikainen, Panu (2004): Ihmistieteet ja filosofia, Gaudeamus, Helsinki

(Raatikainen 2004)

Rogers, Juan D & Barry Bozeman & Ivan Chompalov (2001): Obstacles and opportunities in the application of network analysis to the evaluation of R & D, Research Evaluation Vol 10 (3) s. 161-172 (Rogers ym. 2001)

Rieper, Olaf (2004): Exploring Popper's Relevance for the Evaluation Community, Evaluation Vol 10(I) s. 92-100 (Rieper 2004)

Sarvimäki, Anneli (1988): Knowledge in Interactive Practice Disciplines. An Analyses of Knowledge in education and health care, Helsinki (Sarvimäki 1988)

Schumann, Edvard (1967): Evaluative Research. Principles and Practice in Public Service & Social Action Programmes, New York (Schumann 1967)

Sihvola, Juha (2004): Maailmankansalaisen etiikka, Otava, Keuruu (Sihvola 2004)

Sipilä, Jorma (2003): Yhteiskunnan tutkimuksen vaikuttavuus, erillisteema-artikkeli raportista Suomen tieteen tila ja taso 2003. Katsaus tutkimustoimintaan ja tutkimuksen vaikutuksiin, Suomen Akatemian julkaisuja 10/2003, toim. Oksanen Timo & Annamajja Lehvo & Anu Nuutinen, s. 159-167 (Sipilä 2003)

Sipponen, Kauko (2000): Kansalainen - isäntä vai renki? WSOY -lakitieto Oy, Helsinki (Sipponen 2000)

Stame, Nicoletta (2004): Theory-based Evaluation and Types of Complexity, Evaluation Vol 10 (I) s. 58-76 (Stame 2004)

Sulkunen, Pekka (2004): Vaikuttavuus arviointitutkimuksen rivien välissä, teoksesta Tutkimus menetelmien pyörteissä. Sosiaalitutkimuksen lähtökohdat ja valinnat, toim. Räsänen Pekka & Anu-Hanna Anttila & Harri Melin, PS-kustannus, Juva s. 39-52

(Sulkunen 2004)

Thompson Klein J. & W Grossenbacher- Mansuy & R. Häberli & A. Bill & R. W. Scholtz & M. Welti (toim.) (2001): Transdisciplinarity: Joint Problem Solving

among Science, Technology, and Society, Birkhäuser Verlag. Basel - Boston - Berlin (Thompson Klein ym. (toim.) 2001)

Tranoy, Knut-Erik (1980): Norms of Enquiry: Rationality, Consistency Requirements and Normative Conflict, teoksesta Rationality in Science. Studies in The Foundations of Science and Ethics, toim. Risto Hilpinen, Philosophical Studies Series in Philosophy Volume 21, Dordrecht (Tranoy 1980)

Virtanen, Petri (2005): Neljännen sukupolven evaluatian jälkeinen arviointitutkimus? Hallinnon Tutkimus 1/2005 s. 62-76 (Virtanen 2005)

Worldwatch-instituutti (2005): Maailman tila 2005. Globaali turvallisuus, Gaudeamus, Helsinki

Åberg, Jan-Olof (1997): Det Rationelle och det legitima. En Studie av värderingars teori och praktik, Göteborg (Åberg 1997)