

Vapaa tahto taloustieteessä: vastaus Juha Tarkalle

Hannu Vartiainen

Juha Tarkka esittää tässä numerossa joitakin kriittisiä kommentteja edellisessä *Kansantaloudellisen Aikakauskirjan* numerossa hahmottelemistani ajatuksista. Teesini oli, että talousteorian tärkeä olettama on taloudenpitäjän vapaa tahto, mikä näyttäytyy hänen optimointikäyttäytymisessään (Vartiainen 2021). Tarkka pitää tulkintaa jossain määrin banaalina. Hän antaa esimerkin termostaatista, joka optimoi huoneenlämpöä annetun tappiofunktion mukaisesti. Tarkan mielestä ei kuitenkaan ole mielekäs ajatella, että termostaatilla olisi vapaa tahto.

1. Vapaa tahto taloustieteessä

Ennen kuin kommentoin Tarkan erinomaista puheenvuoroa tarkemmin, haluan korostaa, että antamani merkitys vapaalla tahdolle on

konkreettinen ja kirjaimellinen. Esimerkiksi *Stanford Encyclopedia of Philosophy* määrittelee vapaan tahdon *yksilön kyvyksi kontrolloida tekojaan*. Valinnat eivät siis ole sattumanvaraisia tai yksilöstä riippumattomia vaan erityisesti hänen haluamiaan. Tämä vastaa täsmälleen, mitä tarkoitin vapaalla tahdolla. Taloustieteen kannalta on oleellista ainoastaan, että valintaa vastaava lopputulema on haluttu ja ei-valintaa vastaava on ei-haluttu. Kuten Tarkka kirjoittaa, kyse on David Humeen liitetystä näkemyksestä, jonka mukaan yksilön halut ovat ulkoa annettuja ja rationaaliset teot tyydyttävät niitä.

Tällä on tärkeitä seurausvaikutuksia. Ensinnäkin yksilön valintakäyttäytyminen paljastaa nyt hänen halunsa eli preferenssinsä. Taloustiede empiirisenä tieteenä nojaa tähän periaatteeseen, sillä käsityksemme yksilöiden kannustimista ja hyvinvoinnista voidaan nyt

palauttaa heidän käyttäytymiseensä. Toiseksi taloustieteilijän ei tarvitse ottaa kantaa siihen, millaisten prosessien seurauksena yksilö päätyi pitämään juuri tätä valintaa parhaana. Hänen ei siis tarvitse tuntea niitä psykologisia taustamekanismeja, jotka edeltävät valintaa. Tämä suoraviivaistaa analyysiä ja palauttaa sen havaittaviin suureisiin. Juuri tähän Gul ja Pesendorfer (2008) viittaavat näkemyksellään taloustieteestä ”mielettömänä tieteenalana”.

Yhteenvetona: vapaa tahto -olettaman vaikutin on päätöksenteon saattaminen läpinäkyväksi ja analyysin piiriin. Tässä mielessä vapaalla tahdolla on ennen kaikkea instrumentaalinen rooli taloustieteessä.

Yllä sanottu ei ota kantaa vapaan tahdon ontologiseen asemaan, joka on konseptiin liittyvän filosofisen keskustelun pääkohde. Miten mieli voi syntyä materiasta, mihin vapaata tahtoa tarvitaan ja miten se on yhteensovittavissa naturalistisen, välttämättömyyksistä muodostuvan maailmankuvan kanssa? Vaikka kysymykset ovat tärkeitä, niillä ei nähdäkseni ole suurta merkitystä taloustieteen kannalta. Taloustieteilijä voi hyvin ajatella, että vapaa tahto on yksilön mielensisäinen kokemus eikä sillä ole itsenäistä, fysikaalista sisältöä. Vapaa tahto voi myös olla illuusio. Relevanttia on ainoastaan, että yksilö *mieltää* tekevänsä valintoja. Taloustieteen ensisijainen kiinnostuksen kohde on ihminen, ei fysiikka.

Luullakseni Tarkan kiinnostus on kuitenkin juuri konseptin filosofisessa ongelmatiikassa. Hänen katsannossaan vapaalle tahdolle tarvitaan perustelu. Sen Tarkka löytää Immanuel Kantin moraaliteoriasta, jonka mukaan vapaata tahtoa tarvitaan, jotta yksilö toimisi vastuullisesti, moraalisesti hyväksyttävällä tavalla. Ilman moraalista dilemmaa vapaalle tahdollekaan ei ole tarvetta.

2. Moraalisääntö vapaan tahdon selittäjänä

Vaikka olen diletantti, kommentoin mielelläni vapaasta tahdosta käytyä filosofista keskustelua. En ole täysin vakuuttunut Kantin moraalinäkökulman hyödyllisyydestä. Se voidaan tiivistää seuraaviin pykäliin.

- i. *Kategorinen imperatiivi* määrää moraalisesti hyväksyttävän toiminataperiaatteen, joka sitoo kaikkia järkeviä olentoja.
- ii. Ihmisellä on muitakin tarpeita, esim. vaitoja, jotka eivät välttämättä ole yhteensopivia kategorisen imperatiivin kanssa.
- iii. Tahdon vapautta tarvitaan, sillä i:n ja ii:n välillä on ristiriita. Moraalinen valinta on yksilön vastuulla.

Kantin malli on ongelmallinen.¹ Tämä on tullut ilmeiseksi aihetta likeisesti käsittelevässä talousteoreettisessa kirjallisuudessa.

Pehmeä kohta on vaatimuksessa, että moraalisääntö sitoo kaikkia järkeviä olentoja. Jos näin on, sen tulisi olla myös olemassa ja pääteltävissä. Moraali on kuitenkin joustava käsite, jolla ei ole yksiselitteistä sisältöä. Erityisen ongelmallista on pyrkiä määrittelemään sääntö, joka sitoisi yksilöä kaikissa olosuhteissa. Miten yksilön päätös olisi yleistettävissä kaikkia koskevaksi laiksi, jos he ovat lähtökohdiltaan erilaisia? Miten tarpeita tulisi vertailla? Kategorista imperatiivia ei yleisenä sääntönä yksinkertaisesti ole olemassa. Kyse on samasta sosiaalisen valinnan

¹ Binmore (1994) tarkastelee kategorista imperatiivia kriittisesti peliteorian näkökulmasta.

ongelmasta, josta Arrow (1955) päätyi esittämään kuulun mahdottomuustuloksensa.

Voimme toki antaa väljemmän tulkinnan kategoriselle imperatiiville. Kenties kyseessä ei olekaan sääntö sanan kirjaimellisessa mielessä vaan kehotus ottaa muut huomioon silloin kun se on mahdollista. Mutta milloin ja miten? Miten yksilön tulisi arvottaa toisen hyvinvointia omaansa nähden? Esimerkiksi kultainen sääntö pätee luontevasti vain tilanteisiin, joissa ihmisten tarpeet ovat samankaltaisia.

Moraalisäännön problematisointi ei ole saivartelua, vaan se havainnollistaa minkä tahansa moraaliiin perustuvan vapaa tahto -teorian syvällisen ongelman. Jos moralilla ei ole selvää *a priori* tulkintaa, on tyydyttävä ajatukseen, että se on ainakin osittain riippuvainen yksilön subjektiivisesta kokemuksesta. Silloin on myös hyväksyttävä, että joidenkin ihmisten moraalii ei minun mielestäni ole kovin korkeatasoista tai ei kenties moraalii ollenkaan. Voimme jopa tehdä ajatusharjoituksen sosiaalisista merkityssuhteista vapaasta ihmisestä, jonka moraalii pelkistyy omiin aistinautintoihin. Esimerkistä käy Albert Camus'n *Sivullinen* (suomennos 1947). Hänellä ei olisi Kantin kuvaamaa moraalista ristiriitaa eikä moraaliiisella vastuulla olisi roolia hänen päätöksenteossaan. Olisiko oikein ajatella, että tällaisella ihmisellä ei ole vapaata tahtoa? Se olisi nähdäkseni mietelöntä, minkä myös Camus'n teos havainnollistaa.

Ollakseen mielekäs konsepti, vapaan tahdon kokemuksen tulee mielestäni päteä ihmiseen yleensä, eikä se saa olla riippuvainen hänen mielisäisten moraaliiisten tai muiden konfliktien ominaisuuksista. Vapaa tahto ei myöskään tule koskea rajattua joukkoa yksilön valintatilanteista vaan kaikkea hänen päätöksentekoaan. Näihin maksiimeihin on vaikea sovittaa moraaliiiseen vastuuseen nojaavaa teoriaa.

3. Vastuu ja moraalii taloustieteessä

Yhdyn kuitenkin Tarkkaan siinä, että inhimilisellem vastuulle voidaan muotoilla myös taloustieteelle hyödyllinen tulkinta. Mutta toisin kuin Tarkka, ajattelen että sitä ei tarvitse kytkeä moraliteettiin. Sen sijaan vastuuta voi ajatella osana *konsekventialismin* periaatetta, jonka mukaan ainoastaan lopputulemalla – ei miten siihen päädytään – on väliä. Vastuu näyttäytyy tällöin yksilön velvollisuutena ymmärtää tekojen seuraukset ja toimia järkevästi suhteessa niihin. Itse asiassa taloustiede nojaa tähän ajatukseen rutiininomaisesti. Esimerkiksi kannustintien teoriaa ei voisi edes olla ilman että ihminen on vastuussa teoistaan. Silloin voimme rangaistusten ja palkintojen avulla ohjata häntä toimimaan suunnittelijan haluamalla tavalla. Taloudellisilla mekanismeilla kuten markkinoilla on juuri tämä tehtävä.

Mutta voisiko kantilainen moraaliiinäkökulma tarjota myös substanssia taloustieteelliseen ihmiskuvaan, kuten Tarkka ehdottaa? Hän viittaa Sudgenin (1992) noteeraamaan ristiriitaan instrumentaalisen, yksilötasaisen rationaalisuuden ja normatiivisen, moraaliiisen rationaalisuuden välillä. Näkökulma, jota itse tässä tekstissä edistän, on ensimmäistä lajia. Nämä kaksi katsantoa eroavat tavoitteiltaan. Normatiivisen rationaalisuuden visio hyvästä teoriasta on deskriptiivinen. Sen avulla halutaan ymmärtää myös psykologisia prosesseja, jotka johtavat valintaan. Teorialta haetaan totuudenkaltaisuutta, mihin perustuu myös suuri osa behavioraalista taloustieteestä. Kuten yllä olen pyrkinyt selittämään, instrumentaalinen rationaalisuus on ”ikään kuin” muotoa. Se tunnustaa, että taustalla olevat mielen prosessit ovat perustavaa laatua olevalla tavalla yksityisiä eikä tutkija voi päästä niihin käsiksi muutoin

kuin spekulatiivisella tasolla. Siksi ainoa seikka, joka voidaan uskottavasti tietää päätöksentekijästä, on hänen paljastetut preferenssinsä. Moraaliset ja muut seikat, jotka vaikuttavat päätöksentekoon, heijastuvat niissä. Esimerkiksi kategorisen imperatiivin noudattaminen on Kantin kuvaaman vastuullisen ihmisen preferenssien mukaista valintakäyttäytymistä. Tämä kuitenkin edellyttää, että valinnat ovat todellisia, siis vapaita.

On toki totta, että suuri osa sovelletusta talusteoriasta olettaa tietyt, tyypillisesti yksinkertaista ja egoistista muotoa olevat preferenssit, jotka eivät missään järjellisessä mielessä ole ”paljastettuja”. Tällainen spesifikaatio ei kuitenkaan ole tarkoitettu tosissaan otettavaksi, todellisuutta luonnehtivaksi. Sovelletussa tutkimuksessa mallin yksityiskohdat valitaan tavalla, jotka palvelevat mallinnuksen tavoitteita. Yleensä ne ovat muualla kuin yksilöiden valinnoissa, esimerkiksi markkinoiden toiminnassa tai politiikkateoissa. Tällaisissa oloissa Tarkan ehdottama moraalinen tai sosiaalisten preferenssien huomioiminen voi tarjota uutta ymmärrystä. Itsestään selvää se ei kuitenkaan ole, sillä mallien perusrakenteiden monimutkaistaminen johtaa helposti läpinäkymättömään sotkuun. On siitä huolimatta selvää, että preferenssien realistisuuden lisääminen on tärkeä behavioraalisen taloustieteen anti sovelletulle taloustieteelle. Se on minulle kuitenkin epäselvää, missä määrin tämä johtaa laajempaan ”vapaa tahdon” ilmenemiseen taloustieteessä. Mallintamiemme agenttien valinnat perustuvat edelleen optimointiin.

4. Vapaan tahdon komputaatiotulkinta ja taloustiede

Tässä yhteydessä on hyödyllistä pohdiskella lyhyesti filosofista vapaan tahdon ongelmaa, jonka mukaan determinismi ja vapaa tahto ovat jännitteisiä käsitteitä. Miten vapaa tahto tai tietoisuus voivat syntyä luonnonlakien määräämässä fyysikaalisessa maailmassa? Miten tahto voisi koskaan olla ”vapaa” jos valintamme on jo ennalta määrätty optimoinnin kautta?

Luullakseni ongelmassa on kyse väärinkäsityksestä, joka juontuu siitä, että sekoitamme ihmisen sisäisen kokemuksen ulkopuoliseen havainnointiin. Mielikuvamme valinnasta, siihen liittyvästä epävarmuudesta ja sen vapaudesta perustuu tilanteeseen ennen valintaa, jolloin emme vielä tiedä mitä tulisi valita. Tunnistamme kenties vaihtoehdot ja vasta punnitsemme mikä olisi järkevä valinta. Sen jälkeen alamme prosessoida hyvää valintaa. Voidaan ajatella, että mieleemme toimii kuin algoritmi, joka käy vaihtoehtoja läpi ja valitsee tavoitteitamme parhaiten vastaavan ratkaisun.² Mutta samalla tavoin kuin tietokone ei osaa vastata, minkä tuloksen algoritmi sylkee ulos ennen kuin se on ajettu, mekään emme tiedä mitä tulemme valitsemaan. Olemme ennen valintaa aidosti epävarmoja. Tämä korostuu erityisesti uusissa tilanteissa, jollaisia kaikki ovat ainakin jossain mielessä. Siksi olemme väistämättä ainakin jossain määrin agnostisia siitä, millaisia valintoja tulemme tekemään.³

² *Mielen komputaatiotulkintaa ehdotti Putnam (1967). Moderni johdatus aiheeseen esim. Pinker (1997).*

³ *Gul (1999) esittää vastaavan huomion.*

Voimme toki kysyä, miksemme näe oman (tulevan) ajattelumme läpi ja välittömästi tiedä, mitä tulemme valitsemaan. Voisiko harjaantuminen tuollaiseen korkeamman tason ajatteluun poistaa valintoihimme liittyvän epävarmuuden? Ikävä kyllä vahvojen tietoteoreettisten tosiasioiden takia on syytä uskoa, että algoritmisesti ajatteleva ihminen ei koskaan pääse eroon omaa valintakäyttämistään koskevista epävarmuudesta.⁴

Sen sijaan ulkopuolisen tarkkailijan näkökulmasta, joka havaitsee ainoastaan asetelman ja valinnan mutta ei prosessia joka johtaa valintaan, valintakäyttäytyminen näyttää mekaanisena ja vääjäämättömänä. Taloustieteilijän rooli on tuon ulkopuolisen tarkkailijan.

Talusteoreettiset tulokset perustuvat usein hyvin käyttäytyvän, optimointiin perustuvan valintafunktion olemassaoloon. Tämä on pätevä lähtökohta tyypillisissä taloudellisissa asetelmissa ja varsin heikkojen säännöllisyysominaisuuksien vallitessa. Pelkästään valintafunktion olemassaolon ja eräiden primitiivisten ominaisuuksien turvin voimme todistaa esimerkiksi tasapainojen olemassaolon tai hyvinvointiteoreemoja. On kuitenkin selvää, että valintafunktion laaja olemassaolo edellyttää matemaattisen valinta-aksooman soveltamista, mikä mahdollistaa algoritmisten reunaehtojen ylittämisen. Samalla voimme ulkopuolisina tarkkailijoina sivuuttaa ne ennakoimattomuuden piirteet, jotka valintaan liittyy yksilön kannalta.

On myös syytä huomata, että mallimaailmassamme optimointi on automaattisesti banaali versio todellisesta valintakäyttäytymisestä. Mallit ovat äärimmilleen vietyjä yksinker-

taisuksia. Todellisuudessa kohtaamamme asetelmat eivät ole. Ne sisältävät monimutkaisia syy-seuraussuhteita, joiden ennakoiminen ei ole helppoa. Arkisetkin valintatilanteet koostuvat usein lukemattomista elementeistä. Päätökset ovat myös keskenään dynaamisesti linkittyneitä. Nämä seikat tekevät todellisesta optimoinnista huomattavan vaativaa. On minusta uskottavaa, että preferenssien maksimointi on todellisuudessa vaativa tehtävä eikä sen ennakoiminen ole mahdollista kuin rajatessa joukossa valintatilanteita.

Sovelletuissa asetelmissa valitsemme tyypillisesti myös ympäristön, jossa ratkaisu karakterisoituu käytännöllisessä muodossa. Tämä on tarpeellista esimerkiksi tasapainoanalyysin kannalta. Vaikka karakterisaatio synnyttää usein mielikuvan ilmeisestä ja vääjäämättömästä valinnasta, se ei suinkaan ole itsestään selvä ennen kuin karakterisaatio on johdettu. Tämä on se asetelma, jossa tarkastelemamme päätöksentekijä ryhtyy tekemään valintaansa. Siksi hänellä on myös illuusio vapaasta tahdosta.

Päätöksenteko, joka on sisältäpäin katsoen kaoottista ja epävarmaa, saattaa ulkoapäin näyttää banaalilta optimoinnilta. Tästä nähdäkseni juontuu vapaan tahdon mysteerisyys. Se on subjektiivisesti todellinen kokemus, joka kumpuaa tiedon prosessoinnista haasteista. Sillä ei tarvitse olla tekemistä moraalisen tai muun tahdonvoimaa edellyttävän sisäisen konfliktin kanssa. Esimerkiksi Dennetin (2003) mukaan näin käsitetty vapaa tahto on luonnollinen evoluution tulos.

Näin ymmärrettynä vapaa tahto on myös yhteensovittavissa determinisimin kanssa. Yksilön päätökset ovat todellisuudessa ennakoituja ja vääjäämättömiä, hän ei vain tunnista niitä etukäteen. Vapaan tahdon ei kuitenkaan tarvitse olla ontologisessa mielessä todellinen,

⁴ Riittävän rikkaan (vähintään aritmetiikan sisältävän) formaalin järjestelmän sisältä ei voida ennakoita kaikkia järjestelmän implikaatioita.

kuten ei tietoisuudenkaan, vaan kyseessä voi ongelmatta olla päätösongelmaa ratkovan koneen muodostama illuusio.

Voimme havainnollistaa ilmiötä myös Tarkan esiin nostamassa termostaatti-esimerkissä. Ulkopuolisen tarkkailijan näkökulmasta termostaatin valintakäyttäytyminen vaikuttaa triviaalilta. Tämä ei kuitenkaan mielestäni oikeuta päättelemään, että termostaatilla ei olisi vapaata tahtoa. Oleellista on termostaatin oma kokemus. Luulenkin, että se on hyvinkin epävarma, miten tulee reagoimaan mikrosekunti sitten saapuneeseen energiainformaatioon. Voimme jopa argumentoida, että termostaatilla on myös sisäisiä intentionaalisia tiloja tai tunteita, kuten ihmisellä. Niitä ovat ainakin ”täällä on liian kuuma”, ”täällä on liian kylmä” ja ”onpa sopivan lämmintä”.

Voisiko neurotiede ratkaista vapaan tahdon ongelman, kuten Tarkka spekuloi? Suhtaudun ajatukseen skeptisesti. Neurotieteen ensimmäinen ongelma taloustieteen näkökulmasta on, että neurotiede voi kertoa paljon aivojen toiminnasta, mikä on eri asia kuin mielen toiminta. Täydellinenkin ymmärrys aivoista tuskin riittää selittämään ihmisen ajattelua kaikessa rikkaudessaan. Siksi emme kykene tyhjentävästi selittämään myöskään ihmisen valintoja tarkastelemalla hänen aivojaan. Tämä pätee erityisesti nk. korkeampaan ajatteluun.

Vaikka olen tässä kirjoituksessa korostanut rationaalisuuden ja vapaan tahdon instrumentaalista merkitystä taloustieteelle, en halua sulkea pois tai vähätellä vaihtoehtoisten lähestymistapojen potentiaalia. Esimerkiksi rajoitetun rationaalisuuden mallit ovat tärkeitä useissa ympäristöissä. On tarpeellista ymmärtää säännönmukaisuuksia, joilla ihmiset ovat epärationaalisia. Haasteena kuitenkin silloin on selittää, miksi yksilö on tehnyt virheellisiä valinto-

ja. Samalla menetämme mahdollisuuden paljastettuihin preferensseihin, jotka ovat empiirisen taloustieteen kulmakivi. □

Kirjallisuus

- Arrow, K. (1950), “A difficulty in the concept of social welfare”, *Journal of Political Economy* 58: 328–346.
- Binmore, K. (1994), *Game Theory and the Social Contract, Volume 1: Playing Fair*, MIT Press, Cambridge, MA.
- Camus, A. (1947), *Sivullinen*, (alkuteos *L'Étranger*, 1942), Otava, Helsinki.
- Dennett, D. (2003), *Freedom Evolves*, Penguin Books, London.
- Gilboa, I. (1999), “Can free choice be known”, teoksessa Bicchieri, C., Jeffrey, R. C. ja Skyrms, B. (toim.), *The Logic of Strategy*, Oxford University Press, Oxford.
- Gul, F. ja Pesendorfer W. (2008), “The Case for Mindless Economics”, teoksessa Caplin, A. ja Shoter, A. (toim.), *The Foundations of Positive and Normative Economics*, Oxford University Press, Oxford.
- Pinker, S. (1997), *How the Mind Works*, Norton Books, New York, NY.
- Putnam, H. (1967), “Psychophysical Predicates”, teoksessa Capitan, W. ja Merrill, D. (toim.), *Art, Mind, and Religion*, University of Pittsburgh Press, Pittsburgh, PA.
- Sudgen, R. (1991), “Rational choice: A survey of contributions from economics and philosophy”, *Economic Journal* 101: 751–785.
- Vartiainen, H. (2021), “Mitä ajatella talusteoriaista?”, *Kansantaloudellinen aikakauskirja* 117: 115–169.