

KENAS BINMORE'AS: SOCIALINIS KONTRAKTAS LOŠIMŲ TEORIJOS KONTEKSTE

Nida Vasiliauskaitė

Vilniaus universiteto
Filosofijos katedra
Didlaukio g. 47, LT-2057 Vilnius
Tel. (370-2) 77 07 06; faks. (370-2) 76 18 90
El. paštas: nvasil@takas.lt

Straipsnyje analizuojamas Keno Binmore'o sumanymas pritaikyti tarpdisciplininės racionalaus pasirinkimo teorijos laimėjimus politinėje filosofijoje, performuluojant lošimų teorijos terminais garsųjį Johno Rawlso „pirminės pozicijos“ konstruktą, ir socialinę dinamiką suprasti kaip sprendimo paiešką totaliniame Gyvenimo Lošime pasitelkus naujai interpretuotą teisingumo kriterijų. Pabrėžiama, kad griežta asmeninių ir empatinių preferencijų perskyra yra būtina Rawlso modelio adaptavimo sąlyga. Šią, taip pat ekonominio ir socialinio laiko perskyrą mėginama suprobleminti ir taip atkreipti dėmesį į nepakankamą Binmore'o koncepcijos nuoseklumą.

Prasminiai žodžiai: socialinis kontraktas, teisingumas, racionalumas, lošimų teorija.

Įvadas

Pastaraisiais dešimtmečiais neoklasikinės ekonomikos pagrindas – racionalaus pasirinkimo teorija (RPT; ang. *Rational Choice Theory*) „vis plačiau naudojama ir disciplinose, nagrinėjančiose reiškinius už tradicinių ekonomikos mokslo ribų: politologijoje ir sociologijoje“ (Norkus 1998: 44). Jos šalininkai siekia kuo daugiau socialinių reiškinų paaiškinti racionalių elgesiu, kurį supranta kaip veikėjo naudos maksimizavimą, t. y. pasirinkimą veiksmų, duodančių kuo daugiau naudos santykinai kuo mažesnėmis išlaidomis. Socialinė-politinė filosofija bei etika mėgina pritaikyti RPT šakos, vadinamos lošimų teorija (ang. *Game Theory*), išvalgas (ši domisi racionalių pasirinkimu strateginės sąveikos situacijose, kai kiekvieno veikėjo pasirinki-

mai priklauso nuo kitų, tokių pat racionalių veikėjų pasirinkimų). Londono universiteto kolekdžo ekonomikos profesoriaus Keno Binmore'o manymu, politinei filosofijai „trūksta visa apimančio sistemiško karkaso“, kuris „suteiktų tvirtą pagrindą etinių problemų analizei“; lošimų teorijos technika kaip tik leistų „mąstyti apie moralę ir socialines reformas nuosekliu ir mokslinė prasme gerbtinu būdu“ (Binmore 1998: VIII). Negana to, ateityje ji padėtų ne tik efektyviai aiškinti socialines bei politines realijas, bet ir projektuoti pageidaujamas bendrabūvio formas, kadangi RPT turi ne vien deskriptyvinių, bet ir normatyvinių ambicijų.

Konkretų pavidalą šiai programai Binmore'as suteikia sintetindamas Johno Rawlso „pirminės pozicijos“ (*original position*) idėją ir RPT teoretiko Johno Harsany metodologiją. Tam

tikru mastu sintezę reikėtų laikyti paties Rawlso intencijų realizavimu, nes jo koncepcija sukonstruota siekiant įgyvendinti egalitarinį „moralinės geometrijos“ idealą „visu griežtumu, kurį numato šis pavadinimas“ (Rawls 1972: 121), tačiau, Binmore'o įsitikinimu, Rawlsui sutrukdė kelios svarbios išankstinės prielaidos, kurias leistų pašalinti lošimų teorija¹. Modifikuotas ir papildytas natūralistiniais postulatais apie žmonių biosocialinę evoliuciją, šis naujas „visuomeninės sutarties“ teorijų variantas kai kurių autorių laikomas „reikšmingiausiu socialinės filosofijos darbu nuo Johno Rawlso *Teisingumo teorijos* pasirodymo“ (Skrms 1998: 1).

Straipsnis skiriamas Binmore'o pasiūlyto visuomenės modelio analizei ir siekiama atskleisti jo gana prieštaringas implikacijas. Teigiama, kad:

- 1) neaiškiai apibrėžta esminė teorijai asmenių ir empatinių preferencijų dichotomija;
- 2) ekonominio ir socialinio laiko skirtis prieštarinčiai modeliuoja socialinių reformų įgyvendinimo sąlygas;
- 3) „teisingumo“ kriterijaus stabilumo reikalavimas yra dviprasmiškas.

Pirmoje dalyje labai glaustai pristatomos būtinės tolesnei analizei lošimų teorijos idėjos. Antroje aptariami binmoriškos visuomenės mechanikos pagrindai, išryškinkamas svarbiausias koncepcijos uždavinys (rastigalimų reformų atrankos kriterijų) ir kartu paaiškinamas lošimų teorijos ir Rawlso politinės filosofijos aljansas. Trečioje nurodomas tokio lydinio problemiškas, o ketvirtoje kritiškai vertinama Binmore'o rasta jo galimybės sąlyga. Penktoje dalyje nagrinėjama keletas kitų su šia sinteze susijusių keblumų.

¹ Tačiau Binmore'as nėra neriboto, absoliutaus racionalizmo šalininkas; atvirkščiai, jis tęsia skeptiškąją reliatyvistinę Davido Hume'o tradiciją (protas – tik iracionalių veiksmų, kitaip tariant, aistrų vergas).

1. Pagrindinės lošimų teorijos prielaidos

RPT individualiai elgsenai aiškinti pasitelkia tam tikrą schemą, kurią vaizdžiai išreiškia norvegų filosofo ir politologo Jono Elsterio pasiūlytas „dviejų filtrų“ modelis. Pirmiausia logiškai galimų (neprieštarinčių) veiksmų aibė susiaurinama paliekant tik tuos, kurie galimi fiziškai, realiai. Antrąjį „filtrą“ sudaro veikėjo lūkesčiai (*beliefs*; subjektyvus situacijos suvokimas, įsitikinimai, nuomonės) ir preferencijos (*preferences*; santykiniai norai, pirmenybės, rodančios subjektyvų veiksmų pasekmių „gerumo“ įvertinimą). Būdamas racionalus („racionalumas“ čia reiškia tik preferencijų ir elgesio suderinamumą; *consistency*) ir pakankamai informuotas apie aplinkybes, lemiančias jo veiksmų pasekmes, veikėjas *nuosekliai* remiasi savo lūkesčiais, norais ir renkasi tą realiai galimą veiksmą, kurio rezultata (*pay-off*) laiko labiausiai pageidaujamu. Kitaip sakant, jis vadovaujasi Bayeso racionalaus pasirinkimo kriterijumi, pagal kurį būtina maksimizuoti savo naudą (ji nespecifikuojama ir paliekama lošėjo nuožiūrai). Tačiau RPT netvirtina, kad „veiksmas *b* pasirinktas vietoje veiksmo *a* todėl, kad pirmasis garantuoja didesnę naudą“, ar kad racionalūs žmonės privalo siekti naudos. Atvirkščiai, „veiksmo *b* teikiama nauda *apibrėžiama* kaip didesnė už veiksmo *a* naudą, *jeigu b* pasirenkamas tuomet, kai įmanoma pasirinkti ir *a*“ (Binmore 1994: 169). Tiesiog aiškinant veikėjų elgesį patogu tarti, kad jie elgiasi taip, *tarsi* maksimizuotų naudą savo preferencijų, apie kurias sprendžiama iš jų elgesio, atžvilgiu².

² Tuo nebūtinai numatomas egoizmas, ir altruizmas nėra kontrargumentas, kadangi savanoriškai besiaukojantis kitų labai individas elgiasi taip, *tarsi* tokio poelgio pasakmes (jomis gali būti kad ir maloni būseną atlikus „savaimc gerą“ veiksmą) laikytų labiau priimtiniomis nei alternatyvas, t. y. norą aukotis galima įtraukti į individo naudos funkciją.

Skirtingai nuo kitų RPT šakų, lošimų teoretikus domina racionalių veikėjų (lošėjų) tarpusavyo sąveika. Minėta nerealistiška „visos informacijos“ (*situation under certainty*) prielaida čia reiškia *bendro žinojimo (common knowledge)* apie racionalumą principą: lošėjas ne tik pats yra racionalus, bet ir įsitikinęs savo partnerių racionalumu bei tuo, kad jie žino jo fizines galimybes, lūkesčius ir preferencijas. Renkamasi ne veiksmi, bet veikimo strategijos, o „rezultatu“ vadinamas dalyvaujančių lošėjų strategijų derinys (situacija, kuri susikuria, kai kiekvienas lošėjas pasirenka tam tikrą „veiksmų planą“, atsižvelgdamas į kitų lošėjų numanomus pasirinkimus). Lošimas laikomas išspręstu, jeigu galima numatyti jo rezultatą esant nustatytiems lošėjų preferencijų profiliams. Vienas iš sprendimo paieškos būdų – Nasho pusiausvyros (*Nash equilibrium*) nustatymas, t. y. tokios būklės, kuriai esant kiekvieno lošėjo strategija tam tikromis sąlygomis yra geriausias atsakymas į kitų lošėjų strategijų pasirinkimus ir todėl niekas nesuinteresuotas *pirmas* nuo jos nukrypti. „Geriausiu rezultatu“ vadinamas ne didžiausių lošėjo troškimų išsipildymas, bet optimali *realiai pasiekiamą* padėtis esant tam tikram tarpusavyo jėgų santykiui: bandymui ją vienašališkai dar labiau pasigerinti kiti priešinasi taip, kad išlaidos viršija tokio poelgio teikiamą naudą. Deja, taip išsisprendžia tik nedaugelis lošimų; dažniausiai jie turi daugiau nei vieną pusiausvyros tašką. Tokiais atvejais žiūrима, kuris pusiausvyros taškas tenkina italų ekonomisto ir sociologo Vilfredo Pareto pasiūlytą principą – pagerina bent vieno lošėjo padėtį nepablogindamas kitų. Tačiau dažnai visi pusiausvyros taškai būna pagal Pareto efektyvūs (nebeturintys Pareto pagerinimų; *Pareto-efficient*)³ ir tuo pat metu nepalyginami pagal

³ Binmore'as įprastesnę „Paretooptimumo“ sąvoką keičia „Pareto efektyvumu“, norėdamas išvengti klaidingų asociacijų, esą nepagerinama pagal Pareto būklė visada „socialiai optimali“, t. y. labai patraukli visiems sociumo nariams.

šį standartą tarpusavyje (lošėjų preferencijos jų atžvilgiu visiškai priešingos). Tuomet lošimas sprendžiamas *mišrųjų strategijų* metodu, bet jo apžvalga mūsų temai nėra aktuali. Čia apsiribojame minimaliu lošimų teorijos pristatymu, reikalingu Binmore'o koncepcijai aptarti.

2. Visuomenės modelis

Binmore'as *visą* socialinę tikrovę modeliuoja kaip totalinį Lošimą, sudarytą iš dviejų lygmenų: Gyvenimo ir Moralės. „Gyvenimo Lošimas“ (*the Game of Life*) atitinka pirmąjį, objektyvų, RPT „filtrą“: „jo taisyklės nulemtos fizikos ir biologijos dėsnų, geografinių ir demografinių faktų, technologinių ir fiziologinių apribojimų (*constraints*)“ ir leidžia visus veiksmus, kuriuos gebame atlikti (Binmore 1998: 4). „Moralės Lošimas“ (*the Game of Morals*) tapatinamas su antroju „filtru“. Sąvoką „moralė“ Binmore'as vartoja norėdamas ne tik pabrėžti minėtų apribojimų žmogišką kilmę ir taip juos priešinti santykinai natūraliems „paties gyvenimo“ sąlygotumams, bet ir suteikti lošimų teorijos agentams *socialinį-istorinį* parametą, bendrą pagrindą, kuris leistų apie juos kalbėti kaip apie *visuomenės narius*, o ne vien kaip apie idiosinkratiškais lūkesčiais bei preferencijomis besivadovaujančius atomus. Plačiai suprasta Moralė, kaip individus siejanti eksplcitinių ir implicitinių socialinių normų, reguliuojančių jų interakcijas, visuma (teisė, moralė siaurąja prasme, tradicijos, papročiai, įpročiai, praktinės taisyklės, mados, etiketas ir t. t.), ir yra visuomenės egzistavimo būdas, jos „substancija“: „bendri supratimai (*common understandings*) yra tie metmenys ir ataudai, iš kurių nuaustas visuomenės audeklas“ (1994: 3). Taigi socialumas *per se* tenkina *bendro žinojimo* reikalavimą: norma nėra norma, jeigu nėra bendrai žinoma visų tų, kuriuos ji liečia, t. y. jeigu negalima pasakyti, kad ją „visi žino, visi žino,

kad visi tai žino, visi žino, kad visi žino, kad visi tai žino, ir t. t.“ (ten pat: 139).

Moralė, arba pačių lošėjų susikurtos Lošimo taisyklės, padeda nuspėti partnerių situacijos suvokimo būdą, jų vertinimų skalę, vadinasi, ir jų „ėjimus“. Tokiomis konvencijomis pagrįstas hipotetines visuomenės būsenas Binmore'as vadiną „socialiniais kontraktais“⁴, pabrėždamas jų dirbtinį, sukonstruotą, tam tikra prasme savanorišką pobūdį ir siūlo traktuoti kaip Nasho pusiausvyros taškus. Tuomet galima suformuluoti labai svarbią politinę tezę: „Stabilioms visuomenėms nereikia rišamosios medžiagos – jos sukonstruotos remiantis tuo pačiu principu, kaip ir sausų akmenų sienos, kuriomis piemenys atskiria savo ganyklas: (...) kiekvieną akmenį laiko vietoje greta esančių akmenų svoris, o jis pats prisideda prie kitų palaikymo. (...) Gerai sutvarkytoje visuomenėje (*well-ordered society*) kiekvienas pilietis gerbia socialinį kontraktą todėl, kad jis pats suinteresuotas taip daryti, jeigu jo bendrapiliečiai daro tą patį“ (1998: 267, 5). Visuomenė savą konfigūraciją išsaugo, jei ir tik jei ją *savanoriškai* remia visi jos nariai, *turintys galią* ją pažeisti, t. y. jeigu laikytis normų jiems naudingiau negu nesilaikyti (tų, kurie nepajėgūs nepaklusti, nesvarbi nei parama, nei priešiškas). Išorinė kontrolė nereikalinga ir neefektyvi, nes vienintelis jos funkcijų vykdymo garantas (darant prielaidą, kad ji racionali) – suinteresuotumas, o suinteresuotas tvarkos užtikrinimo aparatas nebėra išorinis savo interesų pagrindu operuojančios sistemos atžvilgiu ir veikia taip pat, kaip ir kiti jos nariai. Politinę valdžią reikia modeliuoti kaip vieną iš lošėjų, kurio elgseną nuolatos įdėmiai kontroliuoja kiti lošėjai, o ne kaip Lošimą prižiūrinčią, atsparią

⁴ Be abejo, tai tik metafora. Lošimų teorijai socialinę tikrovę paranku traktuoti *tarsi* racionalių agentų sudarytą sandėrį.

jo logikai jėgą, todėl tenka iš karto atmesti bet kokius paternalistinius siūlymus pasikliauti visagaliu biurokratinio aparato („nesaugomo sargu“) geranoriškumu⁵. Socialinį kontraktą Binmore'as supranta kaip sudėtingą save palaikantį (*self-policing*) ir abipusiškumu (*reciprocity*) grindžiamą mechanizmą, kurio visi sraigteliai patys patikimai sergsti vieni kitų sukimęsi „pagal taisyklę“. Žinoma, būtina sąlyga – racionalumas. Tai pilietybės ir apskritai priklausymo socialiniam Lošimui cenzas: „Individai, kuriems stinga žinių, intelekto ar vaizduotės galių, kad visuomenėje galėtų sėkmingai derinti savo elgesį su kitais (...), turės būti nušalinti. Tačiau tai nereikia, kad 'teisėta' tokius nepiliečius plėšti, išnaudoti, nesirūpinti jais ar ignoruoti“ (1994: 73). Tiesiog sąveika su jais būtų jau nebe strateginė, o parametrinė (analogiška kontaktams su medžiais, gyvūnais ar psichikos ligoniais).

Darant prielaidą, kad Lošimas yra tęstinis (*repeated*; daugelio etapų), kiekviena nusistovėjusi pusiausvyra savo ruožtu traktuojama kaip atspirties taškas naujo sprendimo paieškoms. Šis privalo būti Pareto pagerinimu (irgi stabilium) ankstesniojo atžvilgiu – kitaip transformuotis nepavyktų. Kadangi tokių taškų esti daug ir konkretus pasirinkimas nėra neišvengiamas, realu tikėtis rasti būdą *pakreipti* įvykių eigą labiau pageidaujama linkme. Beje, pats būdas turi būti įgyvendinamas: stabilumu privalo pasižymėti ne tik pradinė ir galutinė būklė, bet ir perėjimo nuo vienos prie kitos procesas, kiekviena senajo kontrakto virsmo nauju pakopa⁶. Vadinasi,

⁵ Tai artima Friedricho von Hayeko nuomonei, kad socialinės institucijos nėra kieno nors sąmoningų veiksmų rezultatas; jos susikuria savaime, spontaniškai, individams siekiant privačių tikslų (Hayek 1978).

⁶ Binmore'as kažkodėl daro prielaidą, kad socialinė raida – diskretiškas procesas, bet tokiu atveju iškyla senoji Zenono problema: galima kalbėti apie pereinamąsias stadijas *tarp* tarpinių pakopų *ad infinitum*.

teks „įtikinti visus, turinčius galios, kad jie šį tą laimėtų pakeisdami savo požiūrį“ (1998: 270). Reformos prasideda nuo bendro žinojimo, kultūros, Moralės transformacijų. Tačiau Binmore'as neketina siūlyti jokių *turingų* naujų konvencijų dėl bendrojo gėrio – tokie užmojai būtų ne tik nepriimtini, bet ir neįgyvendinami, kadangi naivu tikėtis visus įtikinti *a priori* sugalvotų reformų gerumu. Atvirksčiai, jas reikia sukonstruoti remiantis visuomenėje *jau esančiais* „gėrio“ supratimais (sena retorikos taisyklė nurodo argumentus grįsti ne savo, o adresatų prielaidomis). Be to, norint įtikinti, visai nebūtina sukelti džiaugsmingą pritarimą, pakanka pripažinimo, kad tam tikras sprendimas jį priimančiam asmeniui yra mažiau blogas nei jo *realios* alternatyvos.

Tuomet pagrindinė problema – kokiais *formaliais* kriterijais turėtų vadovautis visuomenė pasirinkdama tam tikrą perspektyvą, kad išvengtų destabilizuojančio blaškymosi?

3. Rawlso „pirminė pozicija“ kaip reformų atrankos mechanizmas

Binmore'o įsitikinimu, pageidaujamo socialinio kontrakto paieškoms puikiai tinka Rawlso „pirminės pozicijos“ idėja⁷, tik ją reikėtų reformuluoti lošimų teorijos sąvokomis. Tai nėra viintelis įmanomas problemos sprendimas, jo pasirinkimo priežastys grynai pragmatinės – dermė su mūsų ikianalitinėmis intuicijomis apie teisingumą, kadangi „praktiškai jau naudojame negrabias šio mechanizmo versijas bendraudam-

mi su savo partneriais kasdieniame gyvenime“ (1994: 316)⁸. Be to, jeigu „teisingumo“ (*fairness*) epitetas⁹ ženklina tuos galimus elgesio tipus, kurie socialiai labiau pageidaujami, tai jį nesunku funkciškai sutapatinti su antruoju lošimų teorijos „filtru“ – Morale. Bet tada „teisingumą“ pakanka suprasti kaip formalią, grynai *techninę* sąvoką, savaime neturinčią jokio turinio, išvengiant visuomet keblių specifikacijų. Vietoje jos galima sėkmingai vartoti kurią nors kitą, tik svarbu, kad ji padėtų subalansuoti visuomenės narių veiksmus: „Mane domina tik tai Pareto pagerinimų konstravimas bet kokiomis po ranka pasitaikančiomis priemonėmis. Jeigu socialinė evoliucija nebūtų pirminės pozicijos mechanizmo įdėjusi į mūsų rankas, aš siūlyčiau kokią nors kitą teoriją, pagrįstą kitu koordinuojančiu mechanizmu“ (ten pat: 72). O dabar tereikia pritaikyti tai, ką turime, „platesnio masto koordinavimo problemoms“ ir taip „pamėginti pagerinti piliečių kooperavimo visuomenėje efektyvumą“ (ten pat: 334). Kitaip tariant, butyje įprastassantykių normalizavimo būdas šaukiantis teisingumo praverstų ir svirstant politinę sanklodą: padėtų greičiau susitarti, kuriam konkrečiam esamą padėtį pagerinančiam stabiliam projektui teikti pirmenybę (tam, kuris teisingesnis).

Rawlso „pirminė pozicija“ leidžia „teisingumo“ sąvoką išreikšti techniškai. Ieškomas „*Teisingas (fair) socialinis kontraktas* apibrėžiamas kaip pusiausvyra Gyvenimo Lošime, kuri kvie-

⁸ Tačiau tainesutrukdo anglų ekonomistui radikalčiai atmesti kitas, neparankias jo teorijai intuicijas apie teisingumo turinį, tad kokiu pagrindu vienos iš jų privilegijuojamos kaip patikimesnės?

⁹ Ši Rawlso sąvoka dažnai verčiama kaip „nešališkumas“, bet tada visiškai iškreiptume Binmore'o interpretacijos prasmę, kadangi jam teisingumas neišvengiamai šališkas. Be to, prarastume svarbias asociacijas su ctine problematika ir kasdieniu kalbos vartojimu.

⁷ Hipotetinė situacija, kai laisvi, lygūs ir racionalūs individai tariamai pamiršta visas savo asmenines ir socialines charakteristikas ir anapus šios „nežinojimo uždangos“ tariasi dėl priimtinausios visuomeninės santvarkos principų. Susitarimo rezultata Rawlsas apibrėžia kaip socialinį teisingumą.

čia rinktis tas strategijas, kurios, jeigu būtų pasirinktos Moralės Lošime, niekada jokiam lošėjui nepalikėtų paskatų pasinaudoti savo teise apeliuoti į pirminę poziciją“ (1998: 424). Tai iš esmės tautologija: teisinga ta visuomenė, kurioje niekam nekyla noro ją apkaltinti neteisingumu ir pageidauti reformų, nes tada tik praloštų. Rawlso maksimino kriterijus¹⁰ šiuo atveju sutampa su bejusiškuoju: jeigu kiekvienas racionalus lošėjas siekia optimalaus jo turimų išteklių atžvilgiu rezultato, tai tas pat galioja ir tiems, kuriems labiausiai nepasisėkė, ir „tikimybė, kad pralošusieji, sužinoję savo vaidmenį, panorės atsisakyti sandėrio, yra minimali“ (1994: 317). Dirbtinai sukurta, pranokstanti jų jėgas gerovė ignoruotų realią situaciją ir neišvengiamai žlugtų¹¹. Taip kartu pavyksta išvengti Rawlso problemos: sutartą kontraktą laikant Nasho pusiausvyros tašku, nebereikia jo vykdymo grįsti vien kantiška ištikimybė kartą sutartiems principams ir „savai įpareigojančia“ teisingumo galia (pasak Binmore'o, tai viena iš pagrindinių Rawlso koncepcijos nepakankamo griežtumo priežasčių).

Tačiau interpretuodamas Pareto pagerinimų atranką kaip procesą, vykstantį anapus „nežinojimo uždangos“ (*the veil of ignorance*), Binmore'as susiduria su dideliais keblumais. Juk „uždanga“ slepia racionalių derybų dalyvių tapatybes, jie nežino, kokias asmenines ir socialines charakteristikas įgis ją pakėlus, tuo tarpu lošėjai veikia vadovaudamiesi savo preferencijomis, o jas užmiršus būtų sunaikintos būtinos lošimų teorijos ir apskritai RPT prielaidos. Be to, savo teisingumo principus Rawlsas konstruoja kul-

tūriškai sterilioje terpėje *ab ovo*, o Binmore'ui toks abstraktus, universalus, socialiai neutralus teisingumas neišivaizduojamas ir nereikalingas (tik dėl įsitvirtinimo tam tikrame kontekste jis ir gali atlikti koordinavimo funkciją).

4. Asmeninių ir empatinių preferencijų perskyra

Išeitis randama pasitelkiant dviejų tipų preferencijas. Įprastas RPT preferencijas, rodančias individo pomėgių, jo privataus skonio (*taste*) hierarchinę tvarką, anglų ekonomistas pavadina *asmeninėmis* (*personal preferences*); jas ir turėtų laikinai nušluoti Rawlso „uždanga“. Joms priešinos strateginėse situacijose būtinos *empatinės preferencijos* (*empathetic preferences*), kurios padeda lošėjui numatyti partnerių veiksmus ir atitinkamai koreguoti savuosius: „Empatija nurodo procesą, kai įsivaizduojame save kitų padėtyje, kad galėtume pažvelgti į situaciją jų akimis, bet (...) nesutapatiname savo interesų su jų“ (ten pat: 56). Taigi „jei individas sako, kad jis geriau būtų vienu asmeniu esant vienoms aplinkybėms, nei kitu asmeniu esant kitoms aplinkybėms, tai jis išreiškia empatinę preferenciją. Pavyzdžiui, jis gali pastebėti, kad geriau būti Ieva, valgančia obuolį, negu Adomu, dėvinčiu figos lapelį. (...) Aš pats galiu neabejodamas teikti pirmenybę figos lapelio dėvėjimui obuolio valgymo atžvilgiu, jeigu tektų pasirinkti. Bet jei žinau, kad obuoliai labai patinka Ievai ir kad Adomas nemato prasmės slėpti savo nuogumo, tada nieko stebėtina, jei manau, kad Ievos situacija geresnė“ (ten pat: 59)¹². Šis gebėjimas palyginti

¹⁰ Pagal kurį visuomenės teisingumas matuojamas blogiausioje padėtyje atsidūrusių jos narių būkle – ji turi būti kuo mažiau bloga.

¹¹ Kadangi patys nesugebėtų jo apginti, o padėti niekas nebūtų suinteresuotas (priešingu atveju, jeigu suinteresuotų ir pajėgių gynėjų atsirastų, pralošusieji *ex definitio* nelaiikytini tokiais).

¹² Binmore'as apsiriboja minimaliu socialinių veikėjų skaičiumi; juos pavadina Adomu ir Ieva. Dėl šio ir kitų didelių supaprastinimų „progų remtis teorija tiesiogiai pasitaiko tik tais atvejais, kai Adomą ir Ievą galima pagrįstai laikyti monolitinių blokų reprezentantais (pavyzdžiui, tarptautinių santykių analizėje – *N. V.*)“ (1998: 496).

tarpusavyje kitų individų pozicijas, prieš tai jas įvertinus ne savo, o jų pačių asmeninių preferencijų požiūriu, esminis ne tik lošimų teorijai, bet ir „pirminės pozicijos“ idėjai, kadangi be jo nepavyktų įsivaizduoti, ką reiškia būti kitu asmeniu kitomis aplinkybėmis, vadinasi, ir suvokti, ką svarstomas kontraktas žada man pačiam, jau nekalbant apie kitus.

Priešingai, nei mano Rawlsas, „nežinojimo uždanga“ turi būti kuo plonesnė. Adomas ir Ieva pamiršta savo tapatumą (kuris iš jų yra Adomas, o kuris – Ieva), „tačiau kiekvienas lošėjas žino savo empatines preferencijas. Iš čia išeina, kad jis taip pat turi pamiršti, kuri empatinė preferencija priklauso Adomui ir kuri – Ievai. Bet visa kita yra jų bendro žinojimo objektas. Todėl jie žino abiejų Gyvenimo ir Moralės Lošimų taisyklės (t. y. *status quo* ir jo lemiamas konkrečias reformų galimybes; taip išvengiama nerealistinių konstrukcijų – *N. V.*). Taip pat kiekvienas žino Adomo ir Ievos asmenines preferencijas, savo paties ir partnerio empatines preferencijas“ (1998: 429). Bet jei empatinės preferencijos nepriklauso nuo individo interesų ir niekaip neatspindi jo tapatybės, tai kyla klausimas, kodėl svarbu pamiršti jų priklausomybę vienokiam ar kitokiam identitetui? Juk pamiršti tiesiog nėra ką. Adomo ir Ievos empatinės pirmenybės ir taip vienodos, *ex definitio*. Kitaip jos visai nebūtų reikalingos, nes neatliktų „pagrindo prieiti prie kokių nors išvadų“ funkcijos (1994: 58). Jos sumanytos kaip tarpasmeninis gerovės palyginimo standartas, fiksuojantis tam tikrą skirtingų individų naudos vienetų tarpusavio santykį. Pavyzdžiui, jeigu Adomas savo gerovę matuoja figos lapais, o Ieva – obuoliais, tai jų situacijos galima palyginti tik nustačius šių jų vertinamų dalykų vertę vienas kito atžvilgiu išreiškiančią proporciją, bendrą visiems derybų „anapus uždan- gos“ dalyviams. Tad kuo remiasi cituota prielaida, kad Adomo ir Ievos empatinės preferencijos netapačios?

Kita vertus, ar tikrai įmanoma patenkinti šios hipotetinės situacijos sąlygas? Gebėjimas „apsiauti kito batais“ suponuoja žinojimą, kad tai – kito batai, kad ne tik jo situacija, bet ir jo asmeninės preferencijos yra kitokios. Tad jeigu pamiršiu savąsias asmenines preferencijas, kaip galėsiu suvokti, kad „matuoju jo matu“ ir atsiriboti nuo jo interesų? O gal norima pasakyti, kad reikia pamiršti asmeninių pirmenybių turinį, bet ne patį jų buvimo faktą? Bet kaip galiu žinoti, kad turiu savą, kitokią negu partnerio, vertinimų tvarką, jei nežinau jokio konkretaus šių tezę iliustruojančio pavyzdžio? Nematant savo identiteto, toks mintinis eksperimentas virstų visiškai susitapatinimu su kitu asmeniu, kurį Binmore'as vadina simpatijos aktu ir siekia atskirti nuo empatijos. Tai grynai formalus ir trivialus priekaištas, nurodantis, kad negalima išsaugoti vienos dichotomijos pusės, pamirštant kitą (ar, kaip šiuo atveju, simuliuojant pamiršimą).

Dichotomiją Binmore'as pagilina, asmenines preferencijas interpretuodamas kaip biologinės evoliucijos rezultatą, o empatines – kaip socialinės. Pastarosios – ne kas kita kaip tam tikros istoriškai susiformavusios kultūrinės ir socialinės vertybės (*social values*), t. y. Moralės lošimo taisyklės (todėl ir gali atlikti bendro standarto funkciją homogeniškoje visuomenėje). Tada asmeninėmis preferencijomis tenka laikyti visa tai, kas nesocialu, – biologiškai determinuotus polinkius; paradigmiais jų pavyzdžiais anglų ekonomistas laiko potraukius šviežiam maistui, švariam vandeniui ar „sveikam seksui“ (1998: 227). Atrodo, juos reikėtų suprasti kaip bendražmogiškas universalijas¹³, bet kaip tuo-

¹³ Kartais susidaro įspūdis, jog mūsų autorius, kaip ir ekonomistas Gary Stanley Beckeris, postuluoja visiems žmonėms vieną naudos funkciją, kurią kartu su Hartmutu Esseriu supranta kaip biosocialinio komforto siekimą, nors ir nemini šių pavardžių (plg.: Norkus 2001).

met tai suderinti su cituotu pavyzdžiu apie obuolius ir figos lapelius? „Mėgti obuolius“ tikrai nėra genetiškai nulemtas imperatyvas paties Binmore'o prasme, kadangi jis nevisuotinis ir individas gali jį prarasti, nebūtina, kad pasikeistų keletas kartų. Be to, švaraus vandens ir panašių išvardytų dalykų vertinimo biologinis sąlygotumas irgi labai abejotinas (esama individų, socialinių grupių ir kultūrų, rodančių ką kita). Kita vertus, *asmeninės* preferencijos pagal apibrėžimą negali būti vienodos, o jei būtų, tai konstruojant empatines nereikėtų jų pamiršti. Kaip tada sutikti su tvirtinimu, kad jos paveldimos karta iš kartos ir bendros skirtingų visuomenių nariams, kurie tik suteikia joms skirtingą pavidalą? Asmeninės pirmenybės nepriklauso nuo empatinių, tačiau jei įsivaizduosime, kad visuomenė obuolių valgymą paskelbia gėdingu ir baustinu poelgiu, ar dažnas obuolių mėgėjas netrukus nuoširdžiai neims manyti taip pat? Ar paties Binmore'o figos lapelio pasirinkimas nėra tokios kilmės? Pagaliau ar pavyktų „pirminėje pozicijoje“ valios pastangomis atsiriboti nuo *genetiškai* interpretuotų idiosinkrazijų?

Taigi ar įmanoma atskirti socialinius veiksmus nuo biologinių, o kartu empatines preferencijas nuo asmeninių? Ryžtingai ir radikalčiai „išlipti iš savo batų“? Neatrodo, kad Binmore'as tai daro labai nuosekliai. Be to, šiuolaikinės intelektualinės tendencijos tokioms pastangoms ne itin palankios.

5. Preferencijų kaita ir teisingumo kriterijus

RPT suponuoja preferencijų pastovumą, tačiau Binmore'as elgiasi ne visai ortodoksiškai. Jo manymu, tikslinga skirti tris temporalinius parametrus, kuriuos apibrėžia taip: „Trumpas periodas (*short run*) atitinka ekonominį laiką, per kurį rinka prisitaiko prie netikėtų naujovių; jis

matuojamas minutėmis ar valandomis. Vidutinis periodas (*medium run*) žymi socialinį laiką, trunkantį tol, kol kultūrinės normos ar socialinės konvencijos prisitaiko prie aplinkos pokyčių. Manau, kad jis matuojamas mėnesiais ar metais. Ilgą periodą (*long run*) visų pirma galima suprasti kaip biologinį laiką, per kurį genai prisitaiko prie naujo iššūkio. Jis matuojamas trukme, kurios reikia, kad užaugtų naujos kartos“ (ten pat: 226–7). Lošimų teorija galima pasinaudoti visuomenės modeliavimui tikrai trumpuoju periodu, kol abu preferencijų tipai išlieka stabilūs. Derybos „anapus uždangos“ negali užsitęsti tol, kol pasikeis jose vartojamas tarpasmeninio gerovės palyginimo standartas.

Šis standartas suteikia „pirminės pozicijos“ idėjai socialinį-istorinį kontekstą, kurio stygių ne kartą nurodė Rawlso kritikai: būsimą kontraktą lemia empatiniai vertinimai ir *status quo* (dabartinės Moralės lošimo taisyklės). Tai reiškia, kad „anapus uždangos“ tariamasi visai ne dėl teisingumo principų – taip jų apskritai neįmanoma sukurti, kadangi tam reikia iš anksto turėti palyginimo kriterijų: „Mes neturime kito pasirinkimo, kaip tik pasiremti tam tikru konsensu, *jau egzistuojančiu* tyrinėjamoje visuomenėje“ (1994: 300). Ieškoma ne principų, o juos tenkinančio kontrakto. Kriterijus susiformuoja vidutiniu periodu, kai visuomenė ir jos įsitikinimai stabilizuojasi – nusistovi vadinamoji *empatinė pusiausvyra* (*empathic equilibrium*)¹⁴, atspindinti realų veikėjų jėgų santykį (1998: 343). Tad teisingumas, kaip ir kitos moralės normos, galiausiai pasirodo esąs tik galios funkcija (čia Binmore'as labai nutolsta nuo Rawlso). Kokie vaidmenys, gebėjimai, poelgiai, asmeninės

¹⁴ Netechninis paaiškinimas: situacija, kai visų individų empatinės preferencijos, stebint ir kopijuojant vieniems kitų elgesį, ilgainiui suvienodėja.

savybės *toje aplinkoje* labiau vertinami, koks elgesys su kokiais individualais viešai apbruojamas, priklauso nuo to, *kieno* poreikiai, patogumai, nuoskaudos ir gerovė laikomi svarbesniais. O tai savo ruožtu lemia lošėjo pajėgumas priversti save vertinti tame Moralės Lošimo variante. „Teisingumas“ sutampa su Nasho pusiausvyra, kuri ilgainiui, pakitus asmeninėms preferencijoms, virsta rinkos funkcionavimą aprašančia ekonomine Walraso pusiausvyra (*Walrasian equilibrium*)¹⁵. Socialinę elgseną galiausiai lemia rinkos dėsniai, „panaikinantys piliečių autonomiją renkantis savus tikslus“ (1994: 90).

Tačiau ši ciniška mintis neaktuali trumpuoju periodu, kai naudojamas „pirminės pozicijos“ modelis. „Teisingumas“ dar nėra spėjęs prisitaikyti prie naujos situacijos, jo padikuotas proporcingas derybų sprendimas (*proportional bargaining solution*)¹⁶ nurodo *ankstesnį*, ne dabartinį, jėgų balansą ir *kol kas* turi moralinę prasmę, todėl leidžia greitai, priimtiniu būdu, išvengiant tiesioginės fizinės konfrontacijos pasirinkti vieną iš daugelio Pareto nepalyginamų projektų. Pasirinkimas visada suponuoja šališkumą kieno nors interesų naudai, bet „teisingumo“ epitetas jį „nuskausmina“ ir taip leidžia išvengti nepasitenkinimo bei pavojaus, kad mėginimai rinktis dar kartą gali sukelti socialinę anarchiją.

Skirtingų laikų įvedimas būtinas siekiant įtvirtinti „pirminę poziciją“ socialiniame-istoriniame kontekste ir kartu laikinai išsaugoti „teisingumo“ sąvokos turinį bei preferencijų pastovumą, tačiau jis numato tam tikrą painiavą. Jeigu „pirminės pozicijos“ šaukiamasi tam, kad bū-

tų galima reformuoti visuomenę, tai kaip galima tikėtis susitarti dėl būsimo „socialinio kontrakto“ per kelias minutes ar valandas (tiek trunka trumpasis periodas)? Juo labiau kaip gali per tiek laiko įvykti socialinė transformacija? O jei gali, tai nebelieka pagrindo skirti ekonominį laiką nuo socialinio. Kita vertus, jei tada tik sudaromas projektas, kuriam realizuoti reikės vidutinio laiko, tai tuomet jis nebeatiks atsiradusių naujų empatinių preferencijų ir bus „neteisingas“, pasenęs, nebeįgyvendinamas arba neatpažįstamai iškreiptas. Ir kaip galima tikėtis praktiškai įtvirtinti teisingo bendrabūvio projektą, jeigu jis *ex definitio* neatitinka nei dabartinio, nei būsimo lošėjo jėgų santykio? Tad trumpo ir vidutinio periodų skirtis miglotai ir prieštarinčiai modeliuoja „socialinio kontrakto“ idėjos realų pritaikymą.

Be to, tampa neaiškus teisingumo standarto stabilumo imperatyvas. „Pakeisti visuomenę“ reiškia „pakeisti jos kultūrą“, jos konvencijas: „Jeigu nebūtų įmanoma pakeisti moralės taisyklių, tai visas mūsų vigiškas reformų troškimas būtų nieko nevertas“ (1998: 4). Kurti „socialinį kontraktą“ – tai kurti naujas socialines vertes (tik tai ir sudaro visuomenės substanciją ir apibrėžia konkrečias jos formas), vadinasi, ir jas fiksuojančias empatines pirmenybes kaip teisingumo kriterijų. Bet šito, pasak Binmore'o, kaip tik ir negalima padaryti, nes, pirma, reformos modeliuojamos *pagal* esamas empatines preferencijas, antra, šių preferencijų trumpu laikotarpiu pakeisti neįmanoma, o vidutiniu jos kinta greičiau savaime nei dėl sąmoningų pastangų. Kita vertus, neaišku, kaip gali nesikeisti empatinės preferencijos ir neatsirasti socialinių modifikacijų trumpu periodu, jeigu serija didesnės ar mažesnės trukmės trumpų laiko intervalų sudaro vidutinį periodą (metus ir mėnesius)? Jei preferencijos išliktų visiškai stabilios kiek-

¹⁵ Situacija, kai kainos yra tokios, kad kiekvienos prekės paklausa neviršija pasiūlos (Walras 1926).

¹⁶ Geriausias kolektyviai pasiekiamas sprendimas (socialinis kontraktas) esant lošėjų naudos vienetų tarpusavy santykiui ir *status quo*.

vienu trumpu laikotarpiu, tai nesikeistų ir praėjus vidutiniajam; vadinasi, jos turi nuolatos po truputį keistis, o tai prieštarauja būtinai jų pastovumo prielaidai.

Išvados

Kaip matėme, Binmore'o sintetinei socialinio kontrakto koncepcijai nepavyksta išvengti tam tikro nenuoseklumo.

1. Sudaranti jos ašį Pareto pagerinimų atrankos kriterijaus, kaip „rolsiško“ teisingumo, sąlyga – griežta preferencijų distinkcija – yra ne visai aiški ir pagrįsta (*pirma*, nežinoma, kaip griežtai atskirti socialumą nuo biologijos ir asmeniškumo; *antra*, kaip nuo pastarųjų visiškai atsiriboti „pirminėje pozicijoje“, o atsiribojus išsaugoti empatinius gebėjimus; *trečia*, tiek empatinės, tiek asmeninės preferencijos skirtingose vietose apibrėžiamos nevienareikšmiškai).

LITERATŪRA

1. Binmore, K. 1994. *Game Theory and the Social Contract*. Volume 1. Cambridge, Mass.: The MIT Press.

2. Binmore, K. 1998. *Game Theory and the Social Contract*. Volume 2. Cambridge, Mass.: The MIT Press.

3. Hayek, F. A. 1978. *The Constitution of Liberty*. Chicago: The University of Chicago Press.

4. Norkus, Z. 1998. „Apie plonąją praktinį racionalumą ir jo pastorinimus“, *Problemos* 54: 39–52.

2. Dėl trumpo ir vidutinio periodų priešpriešos reformos realizavimas tampa dviprasmiškas (*pirma*, perėjimas prie suplanuotos naujos visuomenės būklės negali įvykti nė vienu iš šių periodų; *antra*, trumpu laiku sukurtas projektas, kaip neatitinkantis lošėjų galios balanso, negali būti stabilus).

3. Reikalavimas, kad teisingumo norma kuriant reformą būtų pastovi, susiduria su prieštaravimais (*pirma*, naują visuomenės būklę įkūnija ir išreiškia ne kas kita kaip naujos socialinės konvencijos, taigi *naujas* teisingumas, tad jį reikia iš naujo sukonstruoti; *antra*, kadangi vidutinis periodas susideda iš trumpų, tai teisingumo standartą fiksuojančios empatinės preferencijos negali visiškai nesikeisti trumpų intervalų metu).

Tikimės, kad pozityvūs siūlymai, kaip galimi šių problemų sprendimo metmenys, bus to tyrimo tema.

5. Norkus, Z. 2002. „Pirmenybių endogenizacijos problema racionalaus pasirinkimo teorijoje“, in *Seminari 2001 / Atviros visuomenės kolegija*; sudarė A. Jokubaitis ir T. Sodeika. Vilnius: Strofa, 29–67.

6. Rawls, J. 1972. *A Theory of Justice*. Oxford: Oxford University Press.

7. Skyrms, B. *Book Description*, 1998 07 15. <<http://hallgames.com/games/891.shtml>> (2002 01 20).

8. Walras, L. 1926. *Elements d'Economie Politique Pure*. Paris: Pichon et Durand-Auzias.

KEN BINMORE: THE SOCIAL CONTRACT WITHIN THE CONTEXT OF GAME THEORY

Nida Vasiliauskaitė

Summary

The article discusses Ken Binmore's conception of the social contract, which is an attempt to make political philosophy „more scientific“ by means of Rational Choice Theory. Binmore tries to reformulate the famous Rawlsian *original position* theorem in terms of game theory and for this purpose introduces a strict distinction between personal and empathetic preferences. This distinction and that between economic and social time are criticized for the lack of clearness and paradoxical implications. The aim of the article is to demonstrate that Binmore's conception is not as consistent as it claims to be.

ces. This distinction and that between economic and social time are criticized for the lack of clearness and paradoxical implications. The aim of the article is to demonstrate that Binmore's conception is not as consistent as it claims to be.

Keywords: social contract, fairness, rationality, game theory.