

NEWCOMBO PROBLEMA IR AMERIKIETIŠKAS KLAUSIMAS*

Zenonas Norkus

Vilniaus universiteto Filosofijos katedra
Didlaukio g. 47–507, LT-2057 Vilnius
Tel. (370–2) 77 07 06; faks. (370–2) 76 18 90
El. paštas: zenonas.norkus@fsf.vu.lt

Straipsnyje supažindinama su Newcombo problema, kuri parodė klasikinės sprendimų teorijos vidinį prieštarumą. Ji iškyla, kai du racionalaus pasirinkimo principai – dominavimo taisyklė ir sąlyginės laukiamosios naudos maksimizavimo taisyklė – pateikia vienas kitam prieštaraujančius atsakymus į klausimą, kokio veiksmo pasirinkimas yra racionalus. Tai situacijos, kai veikėjas turi pasirinkti veiksmą, kuris yra gera naujiena, kad egzistuoja veikėjo trokštama pasaulio būseną – tokia, kad pasirinktas veiksmas jos atsiradimui negali turėti priežastinės įtakos; ar kitą veiksmą, kuris yra bloga naujiena, kad egzistuoja veikėjo nepageidaujama pasaulio būseną – tokia, kad jos veikėjas taip pat negali pakeisti, tačiau tas veiksmas turi ir nedidelės vertės pozityvią pasekmę. Praktiškai svarbiausias „njukombišku“ situacijų atvejis yra kalinio dilemos tipo situacijos. Svarbiausias diskusijų apie Newcombo problemą rezultatas yra kauzalinės sprendimų teorijos – klasikinės (evidencinės) sprendimų teorijos alternatyvos – atsiradimas. Dažniausiai nagrinėjami Newcombo situacijos pavyzdžiai paradoksalūs tuo, kad bendra nuo veikėjo nepriklausomų pasaulio būsenų ir jo pasirinkimų priežastis yra jo paties racionalumas. Amerikietiškas klausimas toks: kodėl veikėją, kuris nuolatos gauna prastesnius rezultatus, turime laikyti racionaliu?

Prasminiai žodžiai: Newcombo problema, amerikietiško klausimo argumentas, evidencinė ir kauzalinė sprendimų teorija.

Išvadas

Skaitytojau, įsivaizduokite, kad vienas iš Lietuvos televizijos kanalų rengia teležaidimą. Norint jame dalyvauti, reikia laimėti loterijoje, kurioje vieno bilieto kaina yra vienas litas. Laimėjusieji kviečiami į telestudiją, kur juos ištiria garso psichologas Mažeikis (gerai žinomo filosofo ir antropologo bendrapavardis, bet ne giminitis). Po savaitės jie vėl kviečiami į telestudi-

ją, – šį kartą dalyvauti tiesioginėje laidoje. Toje laidoje kita Lietuvos filosofijos garsenybė – Marijus Žiedas į sceną pirmiausia pakviečia Mažeikį. Mažeikis atsineša dvi dėžes – stiklinę ir medinę. Visų televizijos žiūrovų akivaizdoje Mažeikis į stiklinę dėžę įdeda 1000 litų. Jeigu Mažeikis numatė, kad laidos dalyvis (kuris tuo metu sėdi izoliuotoje kabinoje ir negali nei girdėti, nei matyti, ką daro Mažeikis) pasiims abi dėžes,

* Straipsnis parengtas Lietuvos filosofų draugijos ir Šiaulių universiteto Filosofijos katedros mokslinės konferencijos „Racionalumas ir laisvė: konfrontacija ar harmonija?“, vykusios 2002 m. gegužės 17–18 d. Šiaulių universitete, pranešimo pagrindu.

jis medinę dėžę palieka tuščią. Jeigu jis numatė, kad dalyvis pasiims tik medinę (nepermatomą) dėžę, Mažeikis į ją įdeda 10 000 litų. Po to Mažeikis išeina, o į jupiterių šviesą Žiedas pakviečia laidos dalyvį. Įsivaizduokite, kad tas dalyvis – Jūs. Draugas Jums padovanojo loterijos bilietą. Tas bilietas buvo laimingas, ir prieš savaitę Jūs jau apsilankėte Mažeikio kabinete. Tad dabar turite pasirinkti – ar paimsite tik medinę dėžę, palikdami ant stalo 1000 litų, ar imsite abi dėžes. Jūs esate jau 1000-asis laidos dalyvis ir žinote, kad Mažeikis numatydamas suklysta tik vieną kartą iš dešimties. Kaip Jūs pats galėsite įsitikinti žiūrėdamas laidą, tik kas dešimtas iš laidos dalyvių, kurie paėmė abi dėžes, galėjo pasidžiaugti radę medinėje dėžėje 10 000 litų. Ir tik kas dešimtas iš tų, kurie pasiėmė medinę dėžę, nuliūdo pamatę, kad ji tuščia. Kaip Jūs rinksitės – imsite abi dėžes ar tik medinę?

Jūs galite samprotauti taip: Mažeikis arba jau įdėjo į medinę dėžę pinigus, arba ne. Jeigu pinigai įdėti, būtų kvaila atsisakyti papildomų 1000 litų. 11 000 litų yra geriau už 10 000 litų, tiesa? Jeigu dėžė tuščia, būtų dar kvailiau pasiimti tik ją: 1000 litų geriau negu 0 litų! Taigi abiem atvejais Jums geriau imti abi dėžes.

Tačiau galite samprotauti ir kitaip. Mažeikis retai klysta. Labai nedaugelis iš tų, kurie pasiėmė abi dėžes, galėjo pasidžiaugti radę joje 10 000 litų ir labai nedaugelis, išsinešę tik vieną dėžę, rado ją tuščią. Tad ar ne protingiau būtų renkančias ne skaičiuoti grynuosius pinigus, bet pinigus, „pasvertus“ Mažeikio klaidos tikimybe? Šiuo atveju pasirinkimas priklausytų nuo tokio skaičiavimo rezultato. Kol to rezultato neturime, išpūdi, kurį tikriausiai padarė argumentai, išdėstyti ankstesnėje pastraipoje, gal kiek sušvelnins „amerikietiškas klausimas“ – jeigu Jūs toks protingas, tai kodėl turite mažiau pinigų? Kai laidos pabaigoje Žiedas vėl sukvies į sceną visus laidos dalyvius, ar Jūs, paėmęs dvi dėžes ir įsiti-

kinęs, kad medinė dėžė yra tuščia, nepavydėsite laimingiems „vienadėžininkams“, kurie džiaugsis savo 10 000 litų?

Kaip atlikti skaičiavimą, kuriuo būtų galima pagrįsti vienos dėžės pasirinkimą, išsiaiškinsime pirmoje straipsnio dalyje. Joje pateikiamos šiai užduočiai atlikti reikalingos elementarios sprendimų teorijos (*decision theory*) sąvokos. Pradžiai pakaks tik įvardyti problemą, kurią iliustruoja ką tik papasakota istorija. Tai yra vadinamoji Newcombo problema, pavadinta amerikiečių fiziko Williamo Newcombo, kuris pirmasis ją suformulavo, garbei. Kurį laiką ji cirkuliavo kaip JAV mokslininkų folkloro dalis. 1969 m. ją paskelbė, o kartu ir išgarsino amerikiečių filosofas Robertas Nozickas (Nozick 1985 (1969)). Šią problemą išdėstysiu lituanizuotu pavidalu, pritaikydamas prie Lietuvos telekanalų finansinių galimybių (pirminė amerikietiška versija numato, kad medinėje dėžėje yra 1 milijonas JAV dolerių, o numatymai yra dar patikimesni negu tie, kuriuos įstengė suformuluoti Lietuvos mokslininkas).

Antroje straipsnio dalyje aptarsime loginę problemos prigimtį. Be to, išsiaiškinsime, ar dažnai pasitaiko Newcombo tipo situacijos. Kitaip sakant, ar Newcombo problema yra „dirbtinė“, būdinga tik tam tikroms įsivaizduojamoms kontrafaktinėms situacijoms, ar su ja tenka susidurti ir realiai priimant sprendimus?

Trečioje straipsnio dalyje imsime labiausiai dominančio skaitytoją klausimo – kuris gi pasirinkimas yra „teisingas“? Aptarsime ir tą klausimą, kuris labiausiai domina sprendimų teoretikus: ką vienoks ar kitoks atsakymas į šį klausimą reiškia pačiai sprendimų teorijai? Kaip ji turi būti pakeista ar modifikuota, kad pasidarytų pajėgi įtikinamai pagrįsti ar pateisinti „dviejų dėžių“ (arba vienos) pasirinkimą? Šiuo – reikšmės sprendimų teorijos raidai – požiūriu Newcombo problemos svarbą sunku pervertinti. Būtent diskusi-

jose apie Newcombo problemą sprendimų teorija (o ją galima vadinti ir sprendimų logika arba formaliaja praktinio racionalumo teorija) išsiskyrė į du variantus – vadinamąją „evidencinę“ (ar duomeninę; iš angl. *evidence*) ir kauzalinę sprendimų teoriją. Taigi Newcombo problemos analizė yra trumpiausias kelias susiorientuoti šiuolaikinės analitinės praktinio racionalumo teorijos probleminėje situacijoje.

1. Du praktinio racionalumo principai

Jeigu Jums, skaitytojau, jokio įspūdžio nepadarė amerikietiškas klausimas ir dviejų dėžių pasirinkimas atrodo visiškai neabejotinas, vadinausi, Jūs tvirtai ir nuosekliai teikiate pirmenybę racionalaus pasirinkimo principui, kuris vadinamas dominavimo taisykle. Tas principas teigia (žr., pvz., Eells 1982: 87), kad jeigu veikėjo veiksmai nedaro įtakos aplinkybėms, nuo kurių priklauso jo veiksmų rezultatai, tai iš kelių alternatyvių veiksmų racionalu rinktis veiksmą, kurio pasekmės yra geresnės už kitų veiksmų pasekmes visomis aplinkybėmis (tokia pasirinkimo alternatyva „stipriai dominuoja“), arba tokį, kurio pasekmės vienomis aplinkybėmis yra ne blogesnės už jo alternatyvų pasekmes, o kitomis aplinkybėmis – geresnės (tokiu atveju jis „silpnai dominuoja“). Trumpai tariant, dominavimo principas racionaliu skelbia stipriai arba silpnai dominuojančio veiksmo pasirinkimą. Jeigu taikytume šį principą mūsų pavyzdžiui, tai dviejų dėžių pasirinkimas yra stipriai dominuojantis, taigi ir racionalus pasirinkimas: šio pasirinkimo pasekmės yra geresnės už vien medinės dėžės paėmimo pasekmes ir tuo atveju, kai Mažeikis į ją neįdėjo pinigų ($1000 > 0$), ir tuo atveju, kai pinigus įdėjo ($11\,000 > 10\,000$).

Dominavimo principo trūkumas yra pernelyg siaura jo taikomumo sfera. Situacijos, kai

vienas veiksmas dominuoja kitų atžvilgiu, yra palyginti retos. Štai imkime tokią situaciją. Verslininkas Algirdas turi kuo greičiau nuvykti į Ivanovską, kur vietos pareigūnai sulaikė traukinį su jo nupirktu sunkiuoju metalu ir nori jį konfiskuoti. Algirdas gali važiuoti į Ivanovską geležinkeliu arba skristi lėktuvu. Pirmuoju atveju jis kelionėje sugaiš 12 valandų, antruoju – 2 valandas, jeigu bus giedras oras, arba 24 valandas, jeigu Ivanovske bus rūkas ir lėktuvas bus nutupdytas atsarginiame aerodrome. Šioje situacijoje dominuojančio veiksmo nėra – jeigu bus giedra, racionalu skristi lėktuvu, o jeigu bus rūkas, geriau važiuoti traukiniu. Ar tai reiškia, kad tokio tipo situacijose racionalus pasirinkimas neįmanomas?

Ne, juk egzistuoja dar vienas, plačiau taikomas racionalaus pasirinkimo principas, vadinamas laukiamosios naudos (toliau LN; angl. *expected utility*) maksimizavimo taisykle. Ši taisyklė nurodo pasirinkti veiksmą, kurio LN didžiausia. Veiksmo LN sudaro visų jo galimų rezultatų aibės (ta aibė vadinama veiksmo prospektu; žr.: Norkus 1998: 41) LN suma, o atskiro rezultato LN lygi jo naudai, pasvertai tikimybe (aritmetinis tokio svėrimo būdas yra daugyba). Kadangi Algirdui rūpi tik viena – kuo greičiau patekti į Ivanovską, tai rezultatų naudą (N) galima matuoti kelionėje sugaištomis valandomis (taigi tai bus nauda su minuso ženklu). Tarkime, Algirdas mano, kad giedros Ivanovske tikimybė $p = 0,7$, o rūko tikimybė $1-p = 0,3$. Tada visą Algirdo svarstymą, kuris grindžia laukiamosios naudos taisyklės pritaikymą, galima apibendrinti lentelę, techniškai vadinama sprendimo matrica (žr. 1 lentelę). Šis svarstymas parodo, kad racionalus pasirinkimas yra skrydis lėktuvu, nes $-8,6 > -12$.

Sprendimų teorijos ištakos siekia XVII amžiaus prancūzų filosofo ir matematiko Blaise'o Pascale'o darbus. Pažymint XVIII amžiaus bri-

Pasirinkimo alternatyvos	Aplinkybės, lemiančios pasirinktų veiksmų rezultatus		Prospektų LN
	Giedra virš Ivanovsko	Rūkas virš Ivanovsko	
Skristi lėktuvu	$N_{11} = -2; p_{11} = 0,7$ $LN_{11} = N_{11} \times p_{11} = -1,4$	$N_{12} = -24; p_{12} = 0,3$ $LN_{12} = N_{12} \times p_{12} = -7,2$	$LN_1 = LN_{11} + LN_{12} = -8,6$
Važiuoti traukiniu	$N_{21} = -12; p_{21} = 0,7$ $LN_{21} = N_{21} \times p_{21} = -8,4$	$N_{22} = -12; p_{22} = 0,3$ $LN_{22} = N_{22} \times p_{22} = -3,6$	$LN_2 = LN_{21} + LN_{22} = -12$

tų matematiko Thomaso Bayeso indėlių, laukiamosios naudos maksimizavimo principas kitaip dar vadinamas Bayeso taisykle. Užbaigtą pavidalą, kurį diskusijų apie Newcombo problemą retrospektyvoje galima pavadinti jau klasikiniu, sprendimų teorija įgijo britų filosofo ir matematiko Franko Ramsey'aus (1931 (1926)), amerikiečių matematikų ir ekonomistų Johno von Neumanno ir Oscaro Morgensterno (1944), amerikiečių statistikos teoretiko Leonardo Savage'o (1954), amerikiečių filosofo Richardo Jeffrey (1983 (1965)) darbuose. Pastarieji du teoretikai pateikia dvi įtakingiausias klasikinės sprendimų teorijos versijas.

R. Jeffrey sistemoje Bayeso taisyklė įgyja sąlyginės laukiamosios naudos (SLN; angl. *conditional expected utility*) maksimizavimo taisyklės pavidalą, skirtingai nuo Savage'o sistemos, kur veikėjas maksimizuoja „tiesiog“ laukiamąją naudą (LN). Veiksmo rezultatų laukiamoji nauda yra sąlyginė, jeigu ji nustatoma sveriant jų naudą vadinamąją sąlygine tikimybe. Sąlyginė tikimybė – tai aplinkybės (ar būsenos) b_j tikimybė esant sąlygai, kad veikėjas pasirinko veiksmą V_i , t. y. $P(b_j/V_i) = P(b_j \& V_i/V_i)$. Savage'o sistemoje, kur veiksmai suprantami kaip funkcijos, kurių argumentai yra aplinkybės, o reikšmės – rezultatai, aplinkybėms priskiriama tik besąlyginė tikimybė. Tokia tikimybė natūralu operuoti, kai veikėjo pasirinkimų rezultatai priklauso nuo aplinkybių, bet pačių aplinkybių tikimybės nepriklauso nuo to, kokį veiksmą vei-

kėjas pasirinko. Besąlygine tikimybe kelionėje sugaištas valandas svėrė Algirdas, nes jis neturėjo jokio pagrindo manyti, kad giedros arba rūko virš Ivanovsko tikimybės priklauso nuo to, kokią transporto priemonę jis pasirinks. Besąlyginės tikimybės atveju visi rezultatai tame pačiame stulpelyje (esant duotai aplinkybei) turi vis tą pačią tikimybę. Tačiau taip yra toli gražu ne visais atvejais. Pasitelksime kiek pakeistą R. Jeffrey (1983 (1965): 13–14) pavyzdį.

35 metų sulaukęs Jonas svarsto, ar mesti rūkyti, ar ne. Iš straipsnio žurnale „Mokslas ir gyvenimas“ jis sužino, kad iš visų 35 metų sulaukusių rūkančių vyrų, kurie rūko toliau, 41 proc. miršta nesulaukę pensijos (65 metų), o 59 proc. gyvena 65 metus ir ilgiau. Iš tų, kurie nerūkė ar metė rūkyti iki 35 metų, iki pensijos miršta 25 proc., o 75 proc. gyvena 65 metus ir ilgiau. Akivaizdu, kad šiuo atveju tikimybė, ar Jonas sulauks pensijos, nėra nepriklausoma nuo to, ką jis pasirinks – mesti rūkyti ar ne: $P(\text{nesulaukti pensijos/rūkyti}) = 0,41$, o $P(\text{nesulaukti pensijos/mesti rūkyti}) = 0,25$; $P(\text{sulaukti pensijos/rūkyti}) = 0,59$, o $P(\text{sulaukti pensijos/mesti rūkyti}) = 0,75$. Jonas labai nori sulaukti pensijos, tad įvertinkime šį rezultatą 100 balų, ar utilių. Rezultato, kai Jonas miršta iki pensinio amžiaus teberūkydamas, vertė tebūnie 0 utilių; to, kai jis meta rūkyti ir vis tiek miršta, vertė yra lygi –5 utiliams (abstinencijos kančios!); rezultatas, kai jis metęs rūkyti sulaukia pensininko amžiaus, jam vertas 95 utilių (100–5). Jono svarstymą,

2 lentelė

Alternatyvos	Būsenos		Prospektų SLN
	Nesulaukti pensijos (N)	Sulaukti pensijos (S)	
Rūkyti (R)	$N_{11} = 0$ $P_{11} = 0,41$ $SLN_{11} = 0$	$N_{12} = 100$ $P_{12} = 0,59$ $SLN_{12} = 59$	$SLN_1 = 59$
Mesti rūkyti (M)	$N_{21} = -5$ $P_{21} = 0,25$ $SLN_{21} = -1,25$	$N_{22} = 95$ $P_{22} = 0,75$ $SLN_{22} = 71,25$	$SLN_2 = -1,25 + 71,25 = 70$

3 lentelė

Alternatyvos	Būsenos				Prospektų LN
	RM/ NN	RM/ NS	RM/ SN	RM/ SS	
R	$N_{11} = 0$ $P_{11} = 0,1025$ $LN_{11} = 0$	$N_{12} = 0$ $P_{12} = 0,3075$ $LN_{12} = 0$	$N_{13} = 100$ $P_{13} = 0,1475$ $LN_{13} = 14,75$	$N_{14} = 100$ $P_{14} = 0,4425$ $LN_{14} = 44,25$	$LN_1 = 59$
M	$N_{21} = -5$ $P_{21} = 0,1025$ $LN_{21} = -0,5125$	$N_{22} = 95$ $P_{22} = 0,3075$ $LN_{22} = 29,2125$	$N_{23} = -5$ $P_{23} = 0,1475$ $LN_{23} = -0,7375$	$N_{24} = 95$ $P_{24} = 0,4425$ $LN_{24} = 42,0375$	$LN_2 = 70$

kuris pagrindžia SLN maksimizavimo taisyklės pritaikymą, vaizduoja 2 lentelė.

Savage'o sistemoje Jono svarstymą tektų modeliuoti matrica, kurioje skiriamos keturios pasaulio būsenos, kurių kiekvienos aprašymas yra hipotezė apie tai, kas bus, jeigu būtų pasirinkta viena iš veiksmo alternatyvų R arba M. Pirma hipotezė $RM|NN$ teigia, kad Jonas nesulauks pensijos ir tuo atveju, jeigu nemes rūkyti, ir tuo atveju, jeigu mes; antra – $RM|NS$ – teigia, kad jis nesulauks pensijos, jeigu rūkys, bet sulauks jos, jeigu mes rūkyti; trečia – $RM|SN$ – teigia, kad jis sulauks pensijos, jeigu rūkys, bet nesulauks, jeigu mes; ketvirta – $RM|SS$ – teigia, kad jis sulauks pensijos abiem atvejais – ir rūkydamas, ir metęs rūkyti. Pirmosios hipotezės tikimybę nustatome daugindami $P(N/R)$ ir $P(N/M)$, antros – $P(N/R)$ ir $P(S/M)$, trečios – $P(S/R)$ ir $P(N/M)$, ketvirtos – $P(S/R)$ ir $P(S/M)$.

Kaip rodo 3 lentelėje pateikta matrica, vaizduojanti Jono svarstymą pagal Savage'o sistemą, galutinė to svarstymo išvada yra ta pati, kaip

ir svarstant pagal Jeffrey sistemą (2 lentelė), nors pats svarstymas yra kur kas labiau komplikotas. Kita vertus, kai pagal Jeffrey analizuojame 1 lentelėje pateikto tipo situacijas, jokių komplikacijų nekyla: situacijas su besąlyginėmis tikimybėmis galima interpretuoti kaip atskirą situacijų su sąlyginėmis tikimybėmis atvejį. Būtent tai yra situacijos, kur $P(b_i/V_i)$ visiems V yra ta pati – $P(b_j)$. Šis ir daug kitų Jeffrey sistemos pranašumų pelnė jai plačiausią pripažinimą, o jo „Sprendimų logikai“ – kanoninio sprendimų teorijos teksto statusą.

Ja pasinaudosime ir teležaidimo dalyvio svarstymo rekonstrukcijai, kuri vaizduojama 4 lentelėje. Kaip ir Jonui, teležaidėjui tenka sverti medinės dėžės piniginių turinį sąlyginėmis tikimybėmis. Tai $P(\text{medinė dėžė tuščia/imti abi dėžes}) = 0,9$ ir $P(\text{medinė dėžė 10 000 Lt/imti abi dėžes}) = 0,1$; $P(\text{medinė dėžė tuščia/imti tik medinę dėžę}) = 0,1$ ir $P(\text{medinė dėžė 10 000 Lt/imti tik medinę dėžę})$. Kad skaiciuoti būtų paprasčiau, tarsime, kad teležaidėjo nuo-

Alternatyvos	Būsenos		Alternatyvų SLN
	Medinė dėžė tuščia	Medinėje dėžėje 10 000 Lt	
Imti abi dėžes (A)	$1000 \cdot 0,9 = 900$	$11\,000 \cdot 0,1 = 1100$	$900 + 1100 = 2000$
Imti tik medinę dėžę (B)	$0 \cdot 0,1 = 0$	$10\,000 \cdot 0,9 = 9000$	9000

stata rizikos atžvilgiu yra neutrali (jeigu teležaidėjas rizikos bijotų, 10 000 litų su tikimybe 0,1 nauda jam būtų mažesnė už garantuotų 1000 litų vertę, o jeigu jis mėgtų rizikuoti, tai ji būtų didesnė).

Taigi vadovaudamiesi SLN maksimizavimo taisykle, turėtume imti tik medinę dėžę, nes $9000 > 2000$. Dvi racionalaus pasirinkimo taisyklės – dominavimo principas ir SLN maksimizavimo principas – pateikia vienas kitam prieštaraujančius nurodymus.

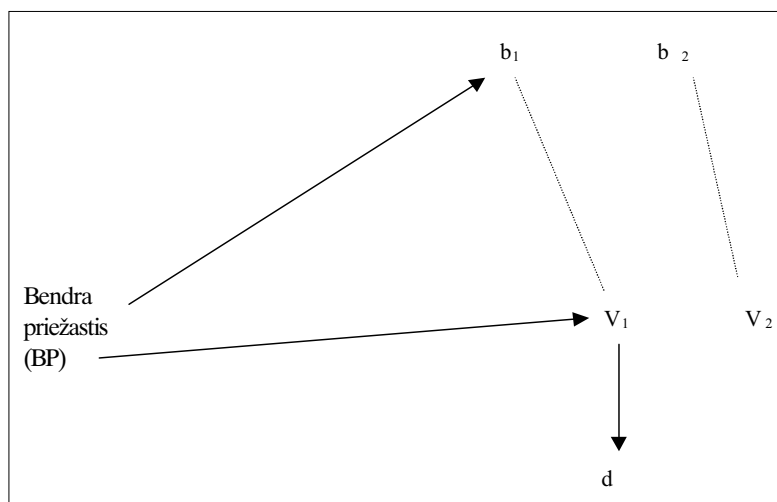
2. Problemos esmė ir paplitimas

Taigi Newcombo problema – tai dviejų racionalaus pasirinkimo principų – universalaus (SLN maksimizavimo principo), pritaikomo visose pasirinkimo situacijose, ir specialaus (dominavimo principo), pritaikomo siauresnei pasirinkimo situacijų aibei, prieštaravimo problema. Teorinė jos reikšmė yra ta, kad ji parodo klasikinės sprendimų teorijos vidinį prieštarumą. Sprendimų teorijos patobulinimo siūlymus, kurie buvo pateikti atradus Newcombo problemą, aptarsime kitame skirsnyje. Pirmiausia turime išsiaiškinti, kodėl abu principai kartais susikerta, ar dažni tokie atvejai? Būtent, ar tokie atvejai pasitaiko tik grynai hipotetinėse (mintinio eksperimento) situacijose, ar tai yra ir realios sprendimų praktikos problema?

Į pirmąjį klausimą būtų galima atsakyti taip: abu principai susikerta todėl, kad tikimybinių pasaulio būsenos priklausomybė nuo veikėjo pasirinkimo, su kuria susiduriame tada, kai tos bū-

senos besąlyginė tikimybė $P(b_i)$ ir jos sąlyginė tikimybė $P(b_i/V_j)$ nesutampa, gali reikšti du dalykus. Pirmia, ta priklausomybė gali egzistuoti dėl to, kad pasirinkimas priežastingai veikia pasaulio būseną. Tokiu atveju veiksmas yra priežastingai veiksmingas (*causally efficacious*). Antra, ta priklausomybė gali egzistuoti dėl to, kad ir veikėjo pasirinkimas, ir pasaulio būseną gali turėti tam tikrą bendrą priežastį BP, patys tarpusavyje būdami priežastingai nesusiję (koreliacija neimplikuoja priežastinio ryšio). Jeigu panaudotume tą patį pavyzdį, bendra priežastis gali būti tie veikėjo asmenybės, mąstymo ir pan. bruožai, kurie lemia ir veikėjo pasirinkimus, ir Mažeikio numatymų sėkmę (jis pagal tuos bruožus ir numato, kaip Jūs pasirinksite). Būtent kokie – pabandysime paspėlioti pačioje šio straipsnio pabaigoje.

Vienoks ar kitoks veikėjo pasirinkimas jam pačiam yra tik simptomas ar indikatorius (žr. 1 pav.), kad pageidaujama pasaulio būseną egzistuoja ar neegzistuoja. Veiksmas yra pranašingas (*auspicious*), nebūdamas priežastingai veiksmingas (negali pakeisti tų sprendimų, kuriuos jau padarė Mažeikis savo numatymų pagrindu). Veikėjas, pasirinkdamas veiksmą, kuris yra ženklas, kad egzistuoja jam pageidaujama būseną (medinėje dėžėje yra pinigai), pats susikuria gerą naujieną (sen. graikiškai „evangeliją“), kad ta būseną egzistuoja, bet ne pačią būseną. Newcombo tipo situacija susidaro tada, kai veiksmas, kuris yra bloga naujiena (medinė dėžė tuščia), tuo pačiu metu yra priežastingai veiksmingas sukelti nedidelę, bet pageidautiną pasekmę (mūsų at-



1 pav.

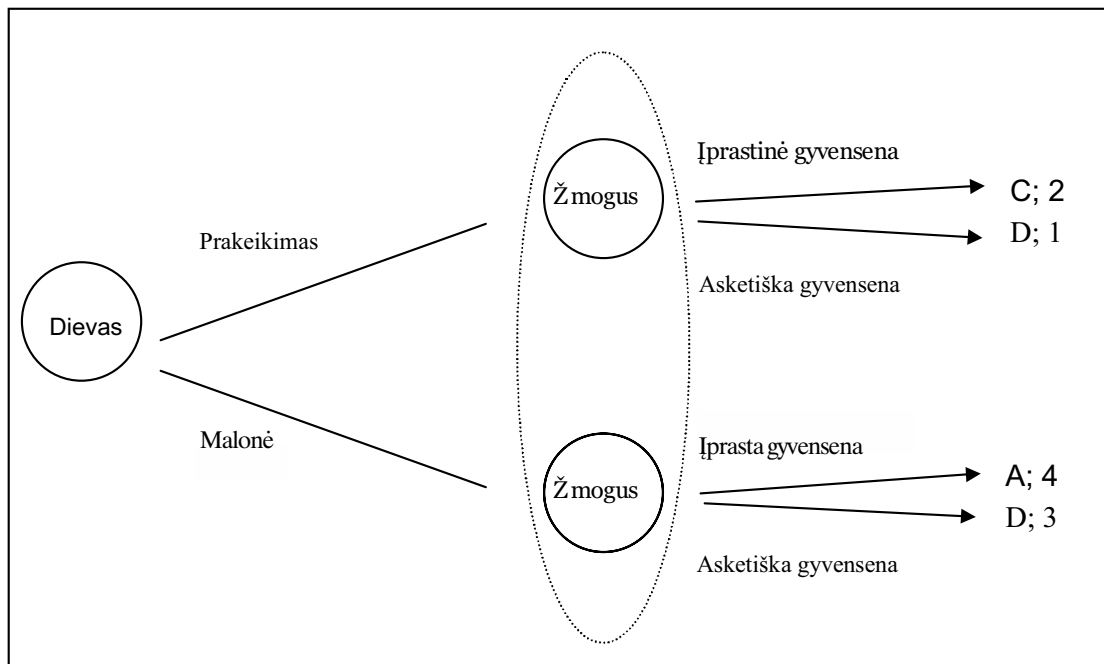
veju – papildyti veikėjo biudžetą 1000 litų). Taip būtų galima aprašyti ir bendrus skiriamuosius Newcombo situacijų bruožus: 1) veiksmai V_1 ir V_2 yra tikimybiškai koreliuoti, bet nesusiję priežastiniu ryšiu su pasaulio būsenomis b_1 ir b_2 , iš kurių viena (b_1) yra nepageidaujama, o kita (b_2) pageidaujama; 2) veiksmas V_1 , statistiškai koreliuotas su nepageidaujama pasaulio būsena b_1 , turi pageidaujamą šalutinę pasekmę d . Tokio tipo situacijose dominavimo principas (kuris akcentuoja priežastinę veiksmo ir pageidaujamos pasaulio būsenos nepriklausomybę) ir SLN maksimizavimo principas (kuris akcentuoja jų tikimybinę priklausomybę) teikia priešingus nurodymus; visais kitais atvejais, kuriais juos abu galima pritaikyti, jų nurodymai sutampa.

Galima sukonstruoti ir daugiau hipotetinių situacijų, kuriomis vienaip ar kitaip inscenizuojama ką tik aprašyta loginė struktūra. Autoriai, nagrinėjantys Newcombo problemą, ypač mėgsta pavyzdžius, kuriuose bendros priežasties vaidmenį atlieka tam tikras genetinis defektas; jis paaiškina ir kodėl žmonės serga tam tikra liga, ir kodėl jie linkę į tam tikrą elgseną, kuri yra simptomas ar indikatorius, kad jie gali susirgti ta liga, kartu nebūdama jos priežastimi. Nustatyta, kad tarp rūkymo ir plaučių vėžio egzistuoja statistinis ryšys. Šiuo metu vyrauja požiūris,

kad už šio statistinio ryšio slypi priežastinis (rūkymas sukelia vėžį). Tačiau kas (tokį klausimą iš tikrųjų kažkada kėlė vienas žymiausių statistikos teoretikų Ronaldas A. Fisheris (1959)), jei gu potraukis rūkyti ir polinkis susirgti plaučių vėžiu turi bendrą genetinę priežastį (paveldėjimas!)? Tokiu atveju asmuo, kuris svarstytų, mesti rūkyti ar ne, atsidurtų Newcombo tipo situacijoje. Nusprendęs mesti rūkyti, jis pasielgtų taip pat kaip teležaidėjas, kuris

atsisako 1000 litų (= rūkymo teikiamo malonumo) vardan tos žinios, kurią jis norėtų išgirsti – kad Mažeikis įdėjo į medinę dėžę pinigų (= jis neturi įgimto polinkio susirgti plaučių vėžiu).

Ar tokio tipo situacijų pasitaiko ir realioje sprendimų praktikoje? Taip. Vienas toks atvejis galbūt net turėjo istoriškai labai reikšmingų padarinių. „Njukombišką“ struktūrą turi Maxo Weberio išgarsinta kalvinisto, tikinčio dviguba predestinacija, problema. Ar racionalu pasirinkti „pasaulietiška asketišką“ gyvenimo būdą, jei gu esame įsitikinę, kad Dievas dar negimusiesiems jau skyrė rojų arba pragarą – tad jokiais savo poelgiais dieviškojo sprendimo pakeisti nebegalime? Būtų geriausia patekti į rojų po malonias ir patogias pragyvento žemiškojo gyvenimo (A; žr. 2 pav.). Antras savo gerumu likimas yra rojus po sunkaus darbo ir asketiško gyvenimo šiame pasaulyje (B). Toliau eina pragaros kančios po žemiško gyvenimo džiaugsmų (C). Blogiausias likimas, kokį tik galima įsivaizduoti, yra pragaros kančios nepaisant visų pasaulietiško asketizmo žygdarbių (D). Samprotaujant pagal dominavimo principą, racionalu gyventi kaip ir anksčiau. Jei gu Dievas mums jau skyrė pomirtinį gyvenimą rojuje, tai nebėra jokios prasmės atsisakyti žemiškųjų džiaugsmų. Jei gu jis mus jau prakeikė, tai asketizmas turi dar mažiau prasmės. Taigi įprastinė gyvensena yra do-



2 pav.

minuojantis pasirinkimas ir tuo atveju, jeigu Dievas mus jau išrinko išganymui, ir tuo atveju, jeigu jis skyrė pragarą. Bendra palaido gyvenimo ir pragaro kančių priežastis yra dieviškasis prakeikimas, o bendra pasaulietiška asketiško gyvenimo ir rojaus palaimos priežastis – dieviškoji malonė (išsamiau žr.: Norkus 1999¹).

Kita vertus, jeigu samprotaujame pagal SLN maksimizavimo taisyklę, turime nedelsdami stoti į doros kelią, nes $P(\text{pragaras/įprastinė gyvensena}) = 1$, o $P(\text{rojus/asketiška gyvensena}) < 1$. Palyginti su tipiniais Newcombo situacijos pavyzdžiais, kalvinisto atvejis ypatingas tuo, kad čia „dėžių tvarkytojas“ (Dievas) niekada neklysta. Todėl veikėjas neturi jokio pagrindo manyti, kad net nepabandęs asketiškai gyventi jis gali atsidurti rojuje (kita vertus, jis negali būti

visiškai tikras, ar jo pasiektas asketinio tobulumo lygis yra pakankamas, kad galėtų save laikyti priklausančiu išrinktiesiems, todėl $P(\text{rojus/asketiška gyvensena}) < 1$). Dėl šios ypatybės, viena vertus, kalvinisto atvejis yra ypač keblus dominavimo principo šalininkui (palyginimui: ar Jūs vis dar ryžtumėtės imti abi dėžes, jeigu žinotumėte, kad Mažeikis niekada neklysta?); kita vertus, galima abejojti, ar šią situaciją tikrai įmanoma subsumuoti Newcombo situacijos sąvokai. Tipinių pavyzdžių atveju dėžių tvarkytojas numato veikėjo pasirinkimus, bet nedeterminuoja tų pasirinkimų, o kalvinisto situacijoje Dievas ne tik numato, bet ir determinuoja jo būsimus pasirinkimus, pats būdamas bendra ir jo pomirtinio likimo, ir jo žemiško elgesio priežastimi. Ar kalvinisto pasirinkimas nėra tik tariamas pasirinkimas (jam atrodo, kad jis renka si, nors iš tikrųjų taip nėra)?

Kitas, praktiškai dar reikšmingesnis Newcombo situacijų atvejis yra pasirinkimo situacijos, kurios jau iki Newcombo problemos atradimo buvo labai nuodugnai išanalizuotos lošimų teorijoje (tai sprendimų teorijos atšaka, nagri-

¹ Nurodyto straipsnio p. 62 išibrovė klaidų. Dešiniaame stulpelyje pateikiamas lygybes ir nelygybes reikia skaityti: $P(\text{prakeikimas/palaida gyvensena}) = 1$; $P(\text{malonė/asketiška gyvensena}) = 1$; $P(\text{prakeikimas/asketiška gyvensena}) = 1 - P(\text{malonė/asketiška gyvensena})$; $P(\text{malonė/palaida gyvensena}) = 1 - P(\text{prakeikimas/palaida gyvensena}) = 0$.

nėjanti racionalaus pasirinkimo strateginėse situacijose problema) ir pavadintos kalinio dilemos lošimo vardu. Davido Lewiso (1979 (1985)) įžvalga, kad kalinio dilemos lošimas turi „njukombišką“ struktūrą (kitai sakant, kad kalinio dilemos situacijos yra atskiras Newcombo situacijų atvejais), yra vienas įdomiausių diskusijos apie Newcombo problemą rezultatų. Detalus šios tezės pagrindimas, be kurio keblu išsiversti dar ir todėl, kad kai kurie teoretikai siūlo jos pataisų teigdami, kad ne visos kalinio dilemos situacijos yra „njukombiškos“ (žr. Sobel 1985), deja, nebetilptų į ribotą šio straipsnio erdvę. Galbūt proga sugrįžti prie šios temos pasitaikys ateityje.

3. Nuo evidencinės link kauzalinės sprendimų teorijos

1973 metais Newcombo problema buvo plačiau aptarta žinomame JAV mokslo populiarinimo žurnale *Scientific American* (Gardner 1973). Redakcija gavo 148 laiškus. 89 skaitytojai manė, kad racionalu imti tik medinę dėžę, 37 manė, kad reikėtų imti abi dėžes, kiti ieškojo problemos formuluotės klaidų ar prieštaravimų (žr. Nozick 1997 (1974): 75–76).

Ką „njukombiškose“ situacijose pataria daryti sprendimų teorijos specialistai? Robertas Nozickas teigė (ne be svyravimų), kad tame problemos variante, kuriame elgesio numatytojas niekada neklysta, racionalu imti tik medinę dėžę. Tuo atveju, jeigu egzistuoja numatymo klaidos tikimybė, jis patarė imti abi dėžes (Nozick 1985(1969): 127–128). Terence Horganas yra bene vienintelis sprendimų teoretikas, kuris teigia, kad vienos dėžės pasirinkimas yra racionalus ir tais atvejais, kai numatytojas gali klysti (žr. Horgan 1985 (1981)). Jis samprotauja taip: „njukombiškoje“ situacijoje susikerta dvi intencijos ar ikinanalitinės įžvalgos. Pirmoji sako, kad

kvaila imti abi dėžes, jeigu manai, kad tai darydamas gausi tik 1000 litų; antroji – kvaila imti tik medinę dėžę, jeigu žinai, kad savo pasirinkimu tos dėžės turinio nepaveiksi. Horganas pabrėžia, kad ribiniu (visiško tikrumo, kad numatytojas neklysta) atveju tėra tik du, o ne keturi galimi jo pasirinkimų rezultatai. Veikėjas žino, kad jeigu ims abi dėžes, medinė dėžė bus tuščia, ir žino, kad jeigu ims tik medinę, joje bus pinigai. Tai, Horgano nuomone, reiškia, kad šiuo atveju egzistuoja loginė priklausomybė tarp medinės dėžės turinio ir veikėjo nuomonių, veiksmų. Ši loginė dėžės turinio priklausomybė nuo veikėjo nuomonių ir veiksmų turi „pragmatinę pirmenybę“, palyginti su kauzaline dėžės turinio nepriklausomybe nuo veikėjo nuomonių ir veiksmų. „Dviejų dėžių pasirinkimas yra akivaizdžiai iracionalus ribiniu atveju, nepaisant antrosios intencijos; čia loginei priklausomybei aiškiai atitenka pirmenybė, palyginti su kauzaline nepriklausomybe. Ir nėra jokio akivaizdaus teorinio pagrindo sakyti, kad racionalumas staugiai paliepia (*abruptly switches*) rinktis dvi dėžes, kai mes pereiname nuo ribinio atvejo prie atvejų su didelėmis tikimybėmis; vadinasi, pirmajai intencijai pirmenybė prieš antrąją intenciją turėtų atitekti ir neribiniais atvejais“ (Horgan 1985 (1981): 170).

Kiti teoretikai pirmenybę vis dėlto teikia antrajai intencijai. Tiesa, tik kai kurie iš jų pabrėžia, kad tai galioja ir ribiniam atvejui (žr.: Gibbard and Harper 1985 (1978): 151–152). Dauguma „dvidėžininkų“ savo svarstymus apriboja tuo atveju, kai numatytojas gali klysti. Neklystamo numatymo atveju susiduriame su klasikine valios laisvės problema, kurią sprendamos galvą laužė ištisos teologų kartos. Jeigu tariame, kad neklystamas numatymas galimas, tai turime abejoti, ar veikėjas, kurio veiksmai yra numatomi, tikrai renkasi. Ar tikrasis pasirinkimo subjektas tokiu atveju nėra pats numatytojas? Jeigu, pavyzdžiui,

Mažeikis gali hipnotizuoti teležaidėjus ir programuoti jų elgesį, tai jis gali ir tiksliai numatyti jų elgesį. Tačiau tokiu atveju teležaidėjams tik atrodo, kad jie renkasi. Renkasi ne jie, o Mažeikis (tas pat ir tuo atveju, jeigu tai ne Mažeikis, o kalvinistų Dievas). Kita vertus, jeigu primygtinai teigiame, kad renkasi būtent veikėjai, ar ši tezė nėra prasminga tik tuo atveju, jeigu kartu pripažįstame, kad numatytojas gali klysti? Taigi ribiniu Newcombo problemos atveju galime įtarti, kad problemos sąlygos slepia prieštaravimą (žr., pvz.: Mackie 1985 (1977)). Dėl šios aplinkybės sprendimo teorijos specialistai daugiausia dėmesio skiria tai šios problemos versijai, kuri telpa į „mokslinio realizmo“ rėmus. Tai versija, kai numatytojas kartais suklusta.

„Vienadėžiniam“ pasirinkimui pritaria mažai sprendimų teoretikų pirmiausia todėl, kad jiems labiausiai rūpi vieno ar kito pasirinkimo pripažinimo racionalių implikacijos teorinei problemai, kurią Newcombo problema iškelia sprendimų teorijai: kaip suderinti dominavimo ir SLN maksimizavimo principus? Pripažinę racionalių vienos dėžės pasirinkimą, turėtume atsisakyti dominavimo principo. Nors jis yra pritaikomas tik nedaugeliui pasirinkimo situacijų, ten, kur juo galima pasiremti, jis dėl savo paprastumo yra ne mažiau akivaizdus už tokias elementarias aritmetikos ir logikos tiesas kaip $2 \times 2 = 4$ ar „jeigu $A = B$, o $B = C$, tai $A = C$ “. Laikydami šį principą neliečiamu, sprendimų teoretikai turi atmesti „vienos dėžės“ intuiciją, o abiejų principų suderinimo problemą spręsti SLN maksimizavimo principo sąskaita. Pasirinkus šią problemos sprendimo strategiją, reikia tą principą modifikuoti arba pakeisti kitu, kuris jam prilygtų universalumu, o „njukombiškos“ situacijose turėtų tas pačias implikacijas, kaip ir dominavimo principas (nurodytų imti abi dėžes).

Klausimas, kaip tai padaryti, buvo ir tebėra probleminė sprendimų teorijos raidos ašis per

pastaruosius tris dešimtmečius. Šioje raidoje išryškėjo du požiūriai. Viena jų reiškia „evidencialistai“ (angl. *evidence* – duomenys). Tai pats R. Jeffrey (Jeffrey 1981; Jeffrey 1983 (1965)) ir Ellery Eelsas (1982). Jų požiūriu, dviejų dėžių pasirinkimas gali būti išvestas ir iš klasikinės (R. Jeffrey) sprendimų teorijos, ją kiek papildant tam tikromis naujomis sąvokomis (pvz., uždengimo ar *screening off*) ir analitinėmis perskyromis (pvz., tarp „ratifikuojamo“ ir „neratifikuojamo“ sprendimo), tačiau išsaugant principą, kad vertinant galimus alternatyvių veiksmų rezultatus juos reikia sverti sąlyginėmis tikimybėmis. Tačiau vis daugiau „dvidėžininkų“ pirmenybę teikia kitam – radikalesniam – problemos sprendimo būdui, kai SLN maksimizavimo principas pakeičiamas kitu – kauzalinės laukiamosios naudos (KLN) maksimizavimo principu. Kartu jie siūlo alternatyvą klasikinei (arba „bejėsiškai“) sprendimų teorijai. Tai Allanas Gibbardas ir Williamas Harperis (1985 1977), Jamesas M. Joyce'as (1999), Davidas Lewisas (1981), Brianas Skyrmsas (1980: 128–139).

Taip daryti „dvidėžininkus“ skatina ne tik aplinkybė, kad galima sukonstruoti tokias Newcombo problemos versijas, kurių taip pat patobulintas evidencinės sprendimų teorijos variantas galų gale nurodo rinktis vieną dėžę (žr. Joyce 1999:158–159). Daugiausia kritikos klasikinė teorija sulaukia dėl to savo bruožo, apie kurį jau užsiminėme 2 skirsnio pradžioje: sąlyginė tikimybė $P(b_j/V_i)$, kuria Jeffrey teorijoje yra sveriamą veiksmo rezultatų nauda, išreiškia statistinę priklausomybę tarp veiksmo ir pasaulio būsenos, o statistinė priklausomybė visai nebūtinai reiškia priežastinę priklausomybę (kaip skelbia vienas pamatinių daugiavariacinės statistikos principų, koreliacija neimplikuoja priežastingumo). Todėl veikėjas, kuris savo pasirinkimų alternatyvų vertę skaičiuoja pagal klasikinės sprendimų teorijos formules, nustato griežtai kalbant ne jų

laukiamąją naudą $LN(V_i)$, bet jų „naujieninė“ (ar „evangelinė“!) vertę (*news value*) $NV(V_i)$, kurią galimi veikėjo veiksmai turi kaip signalai ar simptomai, kad bus realizuoti tam tikri pageidaujami ar nepageidaujami rezultatai.

$NV(V_i)$ yra matas, kuris parodo, kiek veikėjas apsidžiaugtų (ar nusimintų), jeigu sužinotų, kad tai tiesa, jog jis padarė V_i . R. Jeffrey nelaiko esminių skirtumų tarp faktų, kurie priklauso nuo veikėjo (tokio tipo faktai ir yra veiksmai), ir tų, kurie nuo jo nepriklauso, traktuodamas veiksmus lygiomis teisėmis su kitomis pasaulio būsenomis, kurios viena kitos atžvilgiu yra duomenys ar naujienos (pvz., tamsūs debesys danguje reiškia naujieną, kad tikriausiai lis, o ta naujiena priklauso nuo situacijos ir veikėjo gali būti džiaugsminga arba liūdna). Taip pat ir veikėjas, kuris negali daryti įtakos įvykių eigai, skiria pageidaujamus ir nepageidaujamus tos eigos rezultatus, vienomis naujienomis džiaugdamsis (tomis, kurios yra duomenys, rodantys, kad ta eiga krypsta link rezultato, kurio jis pageidauja), o dėl kitų liūdėdamas. Tokioje padėtyje yra, pavyzdžiui, sporto rungtynių ar politikos įvykių raidos tolimoje šalyje stebėtojas, paralyžuotas žmogus ir pan. Klasikinė sprendimų teorija tokiu pat būdu – kaip geras ar blogas naujienas – traktuoja ir paties veikėjo veiksmus. Todėl, griežtai kalbant, klasikinės sprendimų teorijos suponuojamas ar implikuojamas veikėjas ne veikia (sukelia vienas pasaulio būsenas ir sutrukdo įvykti kitoms), o tik renkasi iš gerų ir blogų naujienų (žr.: Resnik 1987: 115–117).

Nurodydama maksimizuoti naujieninę vertę, evidencinė sprendimų teorija kartu nukreipia veikėją teikti pirmenybę tiems veiksmams, kurie yra geriausios naujienos, t. y. duomenys, kad pageidaujamas rezultatas įvyks. Kai priežastinė veiksmo reikšmė (angl. *causal import*) ir jo informacinė ar naujieninė reikšmė (*evidential import*) nesutampa, gauname Newcombo si-

tuacijas. Veiksmas ar sprendimas, kuris yra geriausia naujiena („aš paėmiau tik medinę dėžę“; „aš neprisipažinau“), nėra priežastingai veiksmingas, o veiksmas, kuris yra priežastingai veiksmingas („aš paėmiau abi dėžes“; „aš prisipažinau“), yra bloga naujiena. Nurodydama veikėjui rinktis veiksmus priklausomai nuo jų informacinės reikšmės, tokiose situacijose „evidencinė teorija liepia veikėjui geras naujienas vertinti labiau už gerus rezultatus (*to pursue of good news ahead of the pursuit of the good results*; Joyce 1999: 146). Toks patarimas, kauzalistų požiūriu, yra iracionalus: „racionalūs veikėjai renkasi veiksmus jų priežastinio veiksmingumo (*causal efficacy*), o ne jų pranašingumo (*auspiciousness*) pagrindu; jie veikia, kad pasiektų gerų rezultatų, net jeigu toks veikimas gali reikšti blogas naujienas (Joyce 1999: 146).

Kauzalinė sprendimų teorija nurodo veikėjui veiksmų rezultatų naudą sverti ne jų sąlygine tikimybe $P(s_i/V_i)$, bet kauzaline sąlygine tikimybe $P(s_i/V_i \& K)$, kuri reiškia tam tikros pasaulio būsenos s_i tikimybę, jeigu duota sąlyga, kad veikėjas padarė veiksmą V_i ir turi tam kauzalinį tikėjimą K . Newcombo problemos formuluotėje yra sąlyga, jog veikėjas žino, kad pinigai arba jau yra medinėje dėžėje, arba jų ten nėra ir kad jo veiksmai neturi jokios priežastinės įtakos jų turiniui. Jeigu veikėjo kauzaliniai tikėjimai yra tokie, tai

$$P(\text{medinė dėžė tuščia} / ((\text{paimtos dvi dėžės}) \& (\text{pasirinkimas nekeičia dėžių turinio}))) = P(\text{dėžė tuščia} / (\text{paimta viena dėžė}) \& (\text{pasirinkimas nekeičia dėžių turinio}))) = P(\text{medinė dėžė tuščia}) = p;$$

$$P(\text{medinė dėžė netuščia} / ((\text{paimtos dvi dėžės}) \& (\text{pasirinkimas nekeičia dėžių turinio}))) = P(\text{dėžė netuščia} / (\text{paimta viena dėžė}) \& (\text{pasirinkimas nekeičia dėžių turinio}))) = P(\text{medinė dėžė netuščia}) = 1-p.$$

Pasirinkimo alternatyvos	Sąlygos		Laukiamoji alternatyvų nauda
	Medinė dėžė tuščia	Medinė dėžė netuščia	
Imti abi dėžes	$1000 \times p$	$11\,000 \times 1-p$	$1000 \times p + 11\,000 \times 1-p$
Imti tik medinę dėžę	$0 \times p$	$10\,000 \times 1-p$	$10\,000 \times 1-p$

$(1000 \times p + 11\,000 \times 1-p) > 10\,000 \times 1-p$, kad ir kokios būtų p vertės.

Taigi kauzalistų požiūriu, Newcombo situacijoje veikėjas turi sverti galimus savo pasirinkimų rezultatus besąlyginėmis numatytojo prognozių teisingumo bei klaidingumo tikimybėmis. Nesvarbu, kad jis gali nežinoti ar neturėti konkrečios nuomonės, kokie jų dydžiai. Kaip rodo 5 lentelė, nepriklausomai nuo p (ir $1-p$) verčių, jeigu rezultatų nauda „njukombiškos“ situacijose sveriamė besąlygiškomis tikimybėmis, tai svarstymas, pagrįstas kauzalinės laukiamosios naudos maksimizavimo principu, visada baigiasi tokia pat išvada, kokią gautume taikydami dominavimo principą: turime imti abi dėžes.

Vietoje pabaigos: ir vėl amerikietiškas klausimas

Svarbiausias kauzalinės sprendimų teorijos (jos šalininkų akyse) pranašumas yra tas, kad ji niekada nepateikia nurodymų, prieštaraujančių dominavimo principui. Kita vertus, už tą privalumą tenka sumokėti tam tikrą kainą, kurią sureikšmina evidencinės teorijos šalininkai. Ir vienos, ir kitos teorijos idėjos čia buvo pristatytos neformaliomis kasdienės kalbos priemonėmis. Savo autentišku loginiu-matematiniu pavidalu jos yra sistemos aksiomų, kurios apibrėžia tų teorijų pamatinės sąvokas. Vertinant logikai ir matematikui ypač svarbiu loginio paprastumo kriterijumi, evidencinės teorijos pranašumas yra tas, kad ji paprastesnė. Jos turinį sudaro aksiomos, apibrėžiančios minimalias racionalios nuomonės ir racionalių pirmenybių sąvokas. Kauzalinė priežastingumo teorija, be to, dar turi ak-

siomatiškai apibrėžti racionalaus kauzalinio tikėjimo (*causal belief*) sąvoką.

Patenkinama šios sąvokos eksplikacija turi pateikti tokį aksiomatinį šios sąvokos apibrėžimą, kuris leistų susigaudyti tokio tipo situacijose. Rezultatams, pasiektiems sprendžiant šį uždavinį, aptarti reikėtų dar vieno atskiro straipsnio, tuo labiau kad įvairūs kauzalinės sprendimų teorijos šalininkai jį sprendžia nevienodai. Pačius didžiausius keblumus, su kuriais susiduria tokia analizė, atskleidė dar Davidas Hume'as savo klasikinėje empiristinėje priežastinio ryšio sąvokos analizėje. Pagrindinė šios analizės išvada ta, kad nėra jokio esminio skirtumo tarp „priežastinio ryšio“ ir „koreliacijos“ sąvokų. Prisimindami Hume'o pamokas, kai kurie šiuolaikiniai sprendimų teoretikai skeptiškai tebežiūri į kauzalinės sprendimų teorijos šalininkų pastangas pateikti šios sąvokos analizę. Dažniausiai tai daroma kontrafaktinių sąlyginių teiginių reikšmę analizuojant galimų pasaulių semantikos terminais. Evidencialistų nuomone, tokioje analizėje – vienos neaiškos sąvokos („priežastinis ryšys“ apibrėžiamas vartojant dar tamsesnes – „galimas pasaulis“ ir pan.). Klasikinė sprendimų teorija šio trūkumo neturi, kadangi iš principo („hžiūmiškai“ ar nuosekliai empiristiškai) ignoruoja statistinių ir priežastinių ryšių skirtumą. Kaip jau minėjome, kiek pako-reguota ji gali būti (jeigu tikėtume evidencialistais) suderinta su dominavimo principu.

Tačiau kiek tvirta intuicija, kad racionalus yra būtent dviejų dėžių pasirinkimas, kuriai pirmenybę teikia ir kauzalistai, ir absoliuti dauguma

šiuolaikinių evidencialistų? Geriausias būdas ją išsklibinti yra jau minėtas „amerikietiško klausimo“ argumentas. Tarkime, Jus įtikino kauzalistai, ir pasiėmėte abi dėžes. Kaip turėtumėte jaustis po televizijos laidos laimingų „vienadėžininkų“ („dešimtūkstantininkų“) su 1000 litų apsuptyje? Ar Jus patenkintų savoji pranašesnio racionalumo pajauta? Ar tikrai turite pagrindą jaustis už juos racionalesnis, praradęs tiek pinigų? Ar ji Jus tebepatenkintų ir tuo atveju, jeigu praradote 100 000 litų? 1 000 000? 5 000 000? Ar Jūs pasiimtumėte stiklinę dėžę ir tuo atveju, jeigu ten tebūtų vienas litas, kitoms sąlygoms likus toms pačioms (plg. Nozick 1993: 52–53)?

Įsivaizduokime dviejų teležaidėjų – Onos ir Nerijos – pokalbį po tiesioginės laidos. Ona (nereformuota evidencialistė) pasirinko reiškiantį „geras naujienas“ veiksmą (imti vieną dėžę) ir laimėjo 10 000 litų. Nerija (kauzalistė) pasirinko priežastingai veiksmingą veiksmą (imti abi dėžes) ir turi 1000 litų.

Nerija. Ona, kodėl palikai Mažeikiui 1000 litų?

Ona. Jeigu tu tokia gudri, tai kodėl išlošei mažiau pinigų?

Nerija. Todėl, kad aš nesu tokio tipo žmogus, apie kuriuos Mažeikis mano, kad jie ims dvi dėžes?

Ona. Argi nenorėtum būti tokia kaip aš? Juk būtum dabar turtinga kaip ir aš! Argi man nepavydi?

Nerija. Aš pavydžiu tau tų pasirinkimo galimybių, kurias turėjai. Tu galėjai rinktis imti arba 11 000 litų arba tik 10 000, o aš – arba imti 1000 litų, arba tuščią dėžę. Aš išnaudojau savo situaciją maksimaliai, o tu – ne. Pasielgei kvailai, Ona.

Ona. Jeigu tu tokia protinga, tai kodėl vis dėlto turi mažiau pinigų? Jeigu būčiau mąščiusi kaip tu, tai Mažeikis būtų tai numatęs ir nesudaręs man tokių galimybių, kurių man pavydi!

Nerija. Taip, bet kai jis tas pasirinkimo galimybes sudarė, vis dėlto tu jomis iki galo nepasinaudojai. Kitaip nei aš, tu nepadarei to, kas tavo situacijoje, turint tavo galimybes, buvo geriausia. Todėl tu pasielgei kvailai, o aš – protingai.

Ona. Bet kodėl tu tada turi mažiau pinigų?

Čia verta įsiterpti į Onos ir Nerijos dialogą ir aptarti klausimą, kuris kažkodėl retai svarstomas Newcombo problemą nagrinėjančių autorių darbuose: koks Mažeikio numatymų sėkmės aiškinimas yra pats geriausias? Kokio tipo veikėjų pasirinkimus jam būtų lengviausia numatyti? Atsakymas į antrąjį klausimą gana akivaizdus: toks būtų racionalių veikėjų elgesys. Kad numatytume tokių veikėjų elgesį, pakanka įsitikinti, ar jie suprato užduoties sąlygas ir ar tikrai jie nori laimėti kuo daugiau pinigų. Jeigu tuo esame tikri, tai pakanka žinoti, kokia pasirinkimo taisyklė jie vadovaujasi situacijose, analogiškose teležaidimo situacijai. Mažeikis tai galėtų nustatyti duodamas jiems spręsti testus, kuriuose modeliuojamos tokio tipo situacijos. Geriausiai tam tiktų išbandymas, kaip veikėjas elgiasi kalinio dilemos situacijose, kurios, kaip jau nurodėme, yra labiausiai paplitusi „njukombišku“ situaciją atmaina. Jeigu jis nustatytų, kad veikėjas X „njukombiškos“ situacijose nuosekliai vadovaujasi SLN maksimizavimo taisykle, galėtų ramia širdimi krauti į medinę dėžę pinigus, būdamas beveik tikras, kad veikėjas X stiklinę dėžę paliks ramybėje. Aptikęs, kad veikėjas Y nuosekliai renkasi pagal dominavimo principą, jis nesuklystų palikęs medinę dėžę tuščią. Mažeikiui dar būtų galima patarti pasidomėti, ar teležaidėjai studijavo sprendimų teoriją. Tokiems tikrai nevertėtų dėti į medinę dėžę pinigų.

Tarkime, sprendimų teoretikai teisūs ir dominavimo principas turi absoliučią pirmenybę. Tada tas, kas to nesupranta, turėtų būti laikomas neracionalia būtybe, arba lietuviškai – kvai-

liu. Tada galime sakyti ir taip: kad numatytų, kaip teležaidėjas rinksis, Mažeikiui pakanka nustatyti, tas veikėjas protingas ar ne. Jeigu protingas, tai Mažeikis nesuklys palikdamas medinę dėžę tuščią, o jeigu kvailas, – pridėdamas į ją pinigų. Juk tik kvailys neims stiklinės dėžės.

Kadangi tai kam nors vis dar gali būti neakivaizdu, trumpam sugrįžkime į telestudiją. Tarkime, Jūs – ne laidos dalyvis, o televizijos žiūrovas, kartu – dalyvio draugas, žmona ar brolis, sesuo ar vyras. Jūs ką tik matėte, ką padarė su antra dėže Mažeikis. Jeigu matėte, kad jis padėjo pinigus į medinę dėžę, ir jeigu turėtumėte bent menkiausią galimybę padėti apsispręsti Jums artimam žmogui, ką jam norėtumėte „pasufleruoti“? Aišku, imti abi dėžes! Argi turėtumėte pagrindą patarti kitaip, jeigu matote, kad medinė dėžė liko tuščia, ir Jūsų artimasis liks be pinigų, jeigu paims tik medinę dėžę? O jeigu tas lošėjas yra visiškai pašalinis žmogus, argi Jums jis neatrodys kvailys, jeigu paliks stiklinę dėžę, kai Jūs ką tik matėte, kaip Mažeikis deda pinigus į medinę dėžę? Ar Jums lošėjas neatrodys dar kvailesnis, jeigu ką tik matėte, kad Mažeikis nedėjo į medinę dėžę pinigų, o lošėjas ją vieną tik ir pasiima?

Taigi ta bendra priežastis, kuri lemia ir Mažeikio numatymus, ir teležaidėjo pasirinkimus, yra protas ar jo nebuvimas. Jeigu Mažeikis gali nustatyti, veikėjas protingas ar ne, jis gali numatyti jo pasirinkimus. Kita vertus, būtent veikėjo protas verčia jį teikti pirmenybę dominavimo taisyklei ir imti abi dėžes. Jeigu veikėjas neįstengia nuslėpti nuo Mažeikio, koks jis protingas, tai jis teležaidime neturi šansų laimėti 10 000 (beje, jeigu Mažeikis pastebės, kad teležaidėjas bando nuslėpti, koks jis protingas, ir apsimesti kvailiu, tai tikrai pripažins jį protingu).

Čia situacija tokia pat, kaip ir rusų liaudies pasakose, kuriose Dievas ar dar kažin kas proteguoja kvailius. Imdamas abi dėžes, teležaidėjas sužino gerą naujieną, kad jis yra protingas, ir

blogą – kad jis negaus pinigų. Imdamas vieną dėžę, teležaidėjas sužino blogą naujieną, kad jis kvailas, ir gerą – kad jis bus turtingas. Ką jam daryti, jeigu jis iki imdamas tas dėžes jau žino, kad jis yra protingas? Tada jis gali būti tikras, kad ir Mažeikis tai suvokia, nes veikėjas juk žino, kad Mažeikis yra tikrai geras psichologas. O tokio juk neapgausi. Tokiam veikėjui tikrai belieka imti abi dėžes ir liūdėti, jog pasaulis yra toks blogas, kad jame protas atneša vien vargą. Jeigu teležaidėjas dar nežino, jis protingas ar kvailas, tai jis gali sužinoti pasiimdamas arba abi dėžes, arba vieną. Jeigu jis sužino, kad yra kvailas, praturtėja; jeigu sužino, kad yra protingas, tai įsitikina, kad dideli pinigai – ne jam.

Kaip sakoma, „realiame gyvenime“ yra kitaip. Veikėjas sužino, kad yra protingas, iš turto ir kitokios sėkmės, kuri lydi jo veiksmus. Newcombo problema parodo, kaip ir kokio tipo situacijose gali būti atvirkščiai. Jose susiduriame su tuo pačiu paradoksu, kaip ir kalinio dilemos situacijose². Jeigu Jūs įsivėlėte į kalinio dilemos situaciją, tai neturite jokių šansų išvengti kolektyviai suboptimalaus rezultato, jeigu kiti tokie pat protingi, kaip ir Jūs. Geriausia naujiena visiems būtų ta, kad visi yra kvaili. Bet ar žodžiai „protingas“ ir „kvailas“ tebereikštų tą patį, jeigu tokios situacijos būtų dar dažnesnės, negu jos iš tikrųjų pasitaiko? Problema ypač dramatiška ribiniu atveju, kurio nemėgsta „dvidėžininkai“. Calvinistas, nusprendęs nekeisti įprastinio palaido gyvenimo, sužino gerą naujieną, kad jis yra protingas, ir blogą – kad jis atsidurs pragare. Calvinistas, nusprendęs pakeisti savo gyvenimo būdą, sužino gerą naujieną, kad pateks į rojų, ir blogą – kad jis yra kvailas. Kuris jų iš tikrųjų yra kvailas ir kuris – protingas?

² Kadangi kalinio dilemos situacijos (visos ar kai kurios) yra „njukombiškos“, tai yra, tiesą sakant, vis tas pat paradoksas.

LITERATŪRA

1. Gardner, M. 1973. „Free Will Revisited, With a Mind-Bending Prediction Paradox by William Newcomb“, *Scientific American* 229(1): 104–109.
2. Eells, E. 1982. *Rational Decision and Causality*. Cambridge: Cambridge UP.
3. Fisher, R. A. *Smoking: The Cancer Controversy*. Edinburgh and London: Oliver and Boyd.
4. Jeffrey, R. C. 1983 (1965). *The Logic of Decision*. 2nd ed. Chicago: University of Chicago Press.
5. Jeffrey, R. C. 1981. „The Logic of Decision’ Defended“, *Synthese* 48: 473–492.
6. Joyce, J. M. 1999. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge UP.
7. Horgan, T. 1985 (1981). „Counterfactuals and Newcomb’s Problem“, in *Paradoxes of Rationality and Cooperation. Prisoner’s Dilemma and Newcomb’s Problem*, eds. R. Campbell, L. Sowden. Vancouver: The University of British Columbia Press, 159–182.
8. Lewis, D. 1985 (1979). „Prisoner’s Dilemma Is a Newcomb’s Problem“, in *Paradoxes of Rationality and Cooperation. Prisoner’s Dilemma and Newcomb’s Problem*, eds. R. Campbell, L. Sowden. Vancouver: The University of British Columbia Press, 251–255.
9. Lewis, D. 1990 (1981). „Causal Decision Theory“, in *Rationality in Action*, ed. P. K. Moser. Cambridge: Cambridge UP, 235–263.
10. Mackie, J. L. 1985 (1977). „Newcomb’s Paradox and the Direction of Causation“, in J. L. Mackie. *Selected Papers*, ed. by Joan Mackie and Penelope Mackie. Vol. 1. *Logic and Knowledge*. Oxford: Clarendon Press, 145–158.
11. Neumann, J. von, Morgenstern O. 1944. *Theory of Games and Economic Behaviour*. Princeton: Princeton UP.
12. Norkus, Z. 1998. „Apie plonąjį praktinį racionalumą ir jo pastorinimus“, *Problemos* 54: 39–53.
13. Norkus, Z. 1999. „Weberio tezė ir Newcombo problema“, *Filosofija, sociologija* 4: 58–65.
14. Nozick, R. 1969 (1985). „Newcomb’s Problem and Two Principles of Choice“, in *Paradoxes of Rationality and Cooperation. Prisoner’s Dilemma and Newcomb’s Problem*, eds. R. Campbell, L. Sowden. Vancouver: The University of British Columbia Press, 107–133.
15. Nozick, R. 1993. *The Nature of Rationality*. Princeton, N.J.: Princeton UP.
16. Nozick, R. 1997 (1974). „Reflections on Newcomb’s Problem“, in R. Nozick. *Socratic Puzzles*. Cambridge, Mass.: Harvard UP, 74–84.
17. Ramsey, F. 1931 (1926). „Truth and Probability“, in F. P. Ramsey. *The Foundations of Mathematics and Other Logical Essays*, ed. by Braithwaite. London: Routledge, 156–198.
18. Resnik, M. D. *Choices: An Introduction to Decision Theory*. Minneapolis: The University of Minnesota Press.
19. Savage, L. J. 1954. *The Foundations of Statistics*. New York: Wiley, 1954.
20. Skyrms, B. 1980. *Causal Necessity. A Pragmatic Investigation of the Necessity of Laws*. New Haven and London: Yale UP.
21. Sobel, H. J. 1985. „Not Every Prisoner’s Dilemma Is a Newcomb Problem“, in *Paradoxes of Rationality and Cooperation. Prisoner’s Dilemma and Newcomb’s Problem*, eds. R. Campbell, L. Sowden. Vancouver: The University of British Columbia Press, 263–274.

NEWCOMB’S PROBLEM AND THE AMERICAN QUESTION

Zenonas Norkus

Summary

The article introduces Lithuanian reader into the work related to the Newcomb’s problem in the decision theory made famous by R. Nozick. The problem is described as that of the conflict between the dominance principle and the maximization of the conditional expected utility (CEU) principle. This conflict arises in the situations where an actor has to choose between the action bespeaking good news that a desirable state of world obtains upon which the chosen action has no

causal influence, and the action bespeaking bad news but causally efficient to produce a small benefit. The decision theorists are almost unanimous in the opinion that the contradiction between both principles of the rational choice must be solved at the expense of the CEU maximization principle. However, some of them (evidentialists – E. Eells, R. Jeffrey) maintain that the classical decision theory if properly applied recommends the correct – „bad news, small benefit“

– choice, whereas others (causalists – J. Joyce, D. Lewis, B. Skyrms at al.) plead for its replacing by the causal decision theory. After introducing the reader into the present problem situation in the decision theory, the author concentrates on the version of the Newcomb’s problem involving the Predictor and discusses the neglected question what is the best explanation for the Predictor’s stunning success? To explain this success, the assumption is sufficient that the Predictor is able to discover whether an actor is bright (= is prone to choose in the Newcomb-like situations according to dominance principle) or is stupid or only boundedly rational (including the case where he is prone to reason in the straightforward CEU-maximization way). So the actor’s rationality or its absence is the common cause of actor’s „bad news, small

benefit“ choices and of Predictor’s predictions who is discriminating against the bright actors (including all causal decision theorists, of course) by providing for them no opportunity to avoid undesirable state. Therefore, the version of Newcomb’s problem involving the Predictor is about the situations where actor’s dispositional rationality is not a bonus but the plague causing him to fail (the Prisoner’s Dilemma situations are of the same kind). The American question argument in favor of „good news“ choice draws its persuasive force from the implausibility of the instrumental rationality dooming its possessors to be permanent failures and of unreason vouching for constant success.

Keywords: Newcomb’s problem, American question argument, evidential and causal decision theory.

Iteikta 2002 06 05