

## Filosofija ir gamtos mokslai

### KOKIA YRA PROTINGŲ AUTOMATŲ EGZISTAVIMO GALIMYBĖ?

**Algis Davidavičius**

Vilniaus universiteto Filosofijos fakulteto  
Filosofijos istorijos ir logikos katedra  
Didlaukio g. 47, LT-2057 Vilnius  
Tel. (370 2) 66 12 16  
Faks: (370 2) 68 55 12  
El. paštas: dalgis@osf.lt

*Straipsnyje nagrinėjama, ar gali būti protingų automatų pačia radikaliausia prasme, t. y. ar tokios būtybės gali būti asmenys? Bandoma parodyti, kad tokia galimybė yra logiška, jei priimama ypatinga ontologinio monizmo ir epistemologiškai reikšmingo savybių dualizmo pozicijų dermė. Parodoma, kad tokios galimybės realizacija pirmiausia priklauso nuo ypatingo ontologinio įsipareigojimo šios galimybės tikrumui.*

*Prasminiai žodžiai: asmuo, protingas automatas, sąmonės tapatumas, ontologinis įsipareigojimas, ontologinis monizmas.*

#### Įvadas

Ar gali egzistuoti protingi automatai – asmenys, turintys fenomenalią patirtį ir galintys pranešti apie ją kitiems asmenims? Ar gali technologinės kilmės sistemos demonstruoti mums ne tik sudėtingą, bet ir autonomiņ elgesį, aprašyti savo patirtį iš pirmojo asmens perspektyvos? Ar galime tikėtis atpažinti automatą kaip asmenį taip, kaip kasdienybėje „aptinkame“ asmenis, pavyzdžiui, vaikuose – biologinės kilmės sudėtingos elgsenos sistemose – ir stebime jų „proto raidą“.

Šis atradimas priklauso nuo sudėtingo bendravimo, kai naudojama vienokio ar kitokio

pobūdžio (verbalinė arba ne) kalba, kuria yra keičiamasi pranešimais apie asmeninę fenomenalią patirtį. Kasdienybėje būtent šie pranešimai leidžia mums patikėti, kad konkretus automatas yra protaujantis asmuo, o ne „gudrus mechanizmas“.

Šiame straipsnyje savybė „būti asmeniu“ reikš būti „protingu“, t. y. šios savybės čia bus interpretuojamos kaip ekvivalencijos. Žinoma, galime kalbėti ir apie protingumą (pavyzdžiui, nuovokumo prasme) beasmenių sistemų atžvilgiu. Galime taip pat kalbėti apie asmenis, „nepasižyminčius dideliu protu“, tačiau šie ir panašūs atvejai čia nebus aptariami. Toks diskurso apribojimas reikalingas dėl pakankamai

didelio vertybinio krūvio, kasdienybėje skiriamo abiejų šių savybių tapatinimui. Argumentacijos ekonomiškumo prasme verta tirti šią savybę kaip vieną, o ne kaip sudedamąją ar kaip dvi atskiras savybes (sąvokų dvilypumas čia sąmoningai paliekamas, vartojant „asmeni“ ir „protingą būtybę“ kaip sinonimus – tai paprasčiausiai pagyvina kalbą).

Ar galimybė aptikti asmenis tokiose sistemoje yra tik pamąstoma, įmanoma vien suvokti, ar ji yra reali, t. y. priklauso nuo, pavyzdžiui, mūsų teorinių, technologinių ir kitokių praktinių pastangų, sudarančių pakankamas sąlygas tokiems asmenims būti? Analogiškai galime klausti, ar vaiko atveju galime aptikti asmenį dėl prokreacinių, „fizinio pradėjimo“ bei kitų pastangų (pvz., bendravimo verbaline ar kūno kalba, tėvų globos)? Gal ši galimybė nepriklauso nuo šios veiklos ir fizinių aplinkybių (prokreacijos akto ir fizinės, biologinės kūno konstitucijos), o yra metafiziškai nepriklausoma, turinti visai kitą ontologinį ir epistemologinį statusą ir nepamąstoma, pavyzdžiui, be nemirtingos sielos sampratos ir kreacionizmo teorijų?

Šiuo straipsniu bus mėginama pačiais bendriausiais bruožais parodyti, kad pats asmenų egzistavimas aptinkamas nepriklausomai nuo jų fizinės realizacijos pobūdžio, t. y. savybė „būti asmeniu“ gali būti priskirta įvairioms sudėtingoms fizinėms sistemoms neatsižvelgiant į jų kilmę, jei tos sistemos gali paliudyti pirmo asmens fenomenalią patirtį.

Automatai-asmenys yra galimi, jei: a) darome ontologinę prielaidą apie materialaus ir mentalaus egzistavimo aspektų tapatumą, b) epistemologiškai atskiriame asmens pažinimą nuo fizinės sistemos pažinimo. Tokią poziciją galima pavadinti savybių dualizmu arba natūralistiniu dualizmu, kaip jį vadina vienas atkakliausių šios problematikos tyrėjų Davi-

das Chalmersas. Šis autorius yra pateikęs (Chalmers 1996) išsamią natūralistinio arba savybių dualizmo analizę ir pritaikęs ją protingų automatų galimybei pagrįsti. Šiame straipsnyje daugiausia remsiuosi jo argumentais, parodydamas ryškesnes jų galiojimo ribas ir bandydamas juos praplėsti kai kuriais teoriniais elementais, kurių Chalmersas neįtraukė.

Natūralistinis dualizmas mums rūpimame kontekste galėtų atrodyti taip: jei epistemologiškai esiniai mums pažįstami per jų savybes, tai savybė „būti asmeniu“ kokybiškai skiriasi nuo savybės „būti fizine sistema“, net jei jos būtų būdingos tam pačiam esiniui. Jei suderintume šiuos du atžvilgius, kad sukurtume nuoseklią teoriją, galėtume aktualiai atrasti tokias būtybes įvairiose sudėtingose fizinėse sistemose. Pavyzdžiui, analogiškai protingų automatų paieškai žinome apie protingų primatų ar kitų aukštesniųjų žinduolių paiešką<sup>1</sup> – jie taip pat nėra žmogiškos būtybės, nors ir biologinės.

Tiksliai kalbant, empiriškai pažįstamoje vietoje tokių automatų, kaip ir, pavyzdžiui, protingų ir liudijančių savo protą embrionų, nėra. Beje, tai, kad protingi automatai galimi tik atrasti, taip pat nėra mums tiesiogiai akivaizdu. Sakau „atrasti“, nes, kaip vėliau bandysiu paaiškinti, „asmens“ negalime nei „išrasti“, nei „sukurti“. Šitokia vartoseną remiuosi dėl to, kad išvengčiau kreacionistinių aprašymų, kaip antai: „x sukūrė y“, „z padarė q“, implikuojamos

---

<sup>1</sup> Turiu omenyje plačiai diskutuojamą kogintyvinės filosofijos debatuose San Francisko zoologijos sode gyvenančios 29 metų gorilos patelės Koko, aukėjamos psichologijos daktarės Francine Patterson, atvejį (Patterson 1981). Koko, bendraudama su savo auklėtoja, išmoko per 1000 gestų kalbos ženklų, kiek prisitaikiusi juos pagal savo anatominę konstituciją. Ji naudoja juos ir pranešimams apie savo emocijų patirtį, antai nusako savo jausmus mirus jos globotai katei, o tai aiškiai liudija fenomenalios patirties elementus.

ontologinės hierarchijos kūrėjo ir kūrinio santykyje. Toks santykis nėra ontologiškai neutralus ir dėl to, mano manymu, žalingas kalbant apie autonomiškų, protingų būtybių santykius. Kitaip tariant, „kūrėjas“ turėtų didesnę autonomiją nei kūriny, šis būtų priklausomas nuo pirmojo, o tai trukdytų neutraliai įvertinti bendras asmens egzistavimo galimybes.

Kad realizuotume galimybę atrasti sudėtinguose automatuose asmenis, mums reikia bent dviejų sąlygų: 1) ontologinio įsipareigojimo (angl. *ontological commitment* – šią sąvoką vartosiu Rosalindos Cartwright (1954) prasme) tokiai aktualaus pasaulio ontologijai, kur ši galimybė būtų reali; 2) teorinės tyrimų ir aiškinimų strategijos, kuri leistų aktualiai atrasti, atpažinti automatus kaip protingas būtybes. Iš esmės šis ontologinis angažuotumas ir atveria mums perspektyvą vieno, bendro, fundamentalaus teorinio aprašymo, neprieštarinčiai, nuosekliai, dideliu tikėtinumu atskleidžiančio ir paaiškinančio asmens ir fizinės sistemos, per kurią jį aptinkame, dermę.

Pirmiausia paaiškinsime asmens sampratą egzistavimo bei aptikimo būtinumo ir galimybės prasme, vėliau apžvelgsime ontologinio įsipareigojimo bei automato sampratą, istorinę pastarosios raidą. Galiausiai aptarsime šiuolaikinių autorių siūlomas teorines šios galimybės aiškinimų ir tyrimų strategijas.

### **Apie galimybę asmeniui egzistuoti ir galimybę asmenį atpažinti**

Čia aptarsime savybės būti asmeniu ontologinius ir epistemologinius aspektus. Pirmiausia: kas gi yra asmuo? Remiuosi Johno Locke'o apibrėžimu (Locke 1690: 280): t. y. „mąstanti protinga būtybė, pasižyminti protu ir refleksija ir galinti traktuoti save kaip save, tą pačią mąstančią būtybę [originale *thinking thing* –

A. D.<sup>2</sup>], skirtingu metu ir skirtingose vietose; tai įmanoma tik dėl sąmonės, kuri neatskiriamai nuo mąstymo ir, kaip man atrodo, esmingai jam būdinga, nes neįmanoma, kad kas nors suvoktų nesuvokdamas, jog suvokia.“

Nėra reikalo įrodinėti, kad Locke'o apibrėžimas neprarado savo poleminės vertės šių dienų anglosaksiškame moksliniame diskurse. Pagal šį apibrėžimą asmuo konstituojamas kaip *a priori* akivaizdi, atskira, tapati sau fenomenali patirtis (kitai, pirmojo asmens arba sąmonės patirtis).

Garreto teiginys iš esmės sutampa ir su Chalmerso (1996) pozicija, kai paprasčiausiai vietoje „asmens“ vartojama „fenomenali patirtis“ arba „sąmonės patirtis“, kuri negalinti būti griežta prasme redukuota nei į grynai metafizinę, nei į grynai fizinę plotmę. Pagal šį autorių, iš pateikto apibrėžimo panašūs apribojimai neplaukia – iš jo negalime išvesti, pavyzdžiui, ar būti asmeniu yra ekvivalentiška būti gyvūnu (biologine sistema), ar robotu (technologine sistema). Pirmąjį asmenį „As“ apibrėžia tik sąmonės tapatumas (Locke 1690 (2000): 293), tad negalime griežtai paneigti galimybės, kad atsirastų sąmonę įvairiuose kūnuose. Tokios sąsajos visuomet turės aposteriorinį pobūdį, t. y. jų tikroviškumas bus įmanomas įvertinti vien empiriškai, kaip kad tai, kad „vanduo yra H<sub>2</sub>O“ (pagal Saul Kripke 1972).

Kaip mėgina parodyti Chalmersas, lygiai taip pat neįmanoma išvesti iš fizinių faktų, ypač remiantis fizikalistiniais jų aprašymais (kaip

<sup>2</sup> Drįsčiau nesutikti su p. Ramutės Rybelienės „asmens“ (angl. Locke'o originale „person“, o ne „personality“ (Locke 1689: 2 (27): p. 9) sąvokos vertimu kaip „asmenybės“. Lietuvių kalboje „asmenybė“ kaip ir anglų „personality“, mano galva, yra daugiausia vartojama socialinei asmens saviraiškai nusakyti. Sakydami „šis žmogus – asmenybė“, turime omenyje, kad jo asmens saviraiška yra ženkliai pripažįstama ir vertinama socialinėje plotmėje, bendravimo požiūriu. Taigi „asmenybė“ – išvestinė iš „asmens“ sąvoka.

populiariausiu jų pateikimo pavidalu), nei fenomenalios sąmonės egzistavimo, nei jos griežtos, loginės ir kauzalinės priklausomybės nuo fizinių faktų: visuomet lieka klausimas, kodėl visi šie faktai lydimi „vidinio gyvenimo patirčių“, t. y. fenomenalios sąmonės? (Chalmers 1996: xii).

Pagal šį autorių, mes galime, pavyzdžiui, logiškai neprieštaringai postuliuoti pasaulį, kuris būtų fizinis mūsų pasaulio dublikatas, tačiau jame nebūtų fenomenalios sąmonės patirties. Tokiame pasaulyje gyventų savotiški zombiai, kurie atitiktų fizikalistinius žmonių aprašymus, tačiau nepatirtų jokių fenomenalių patirčių – nebūtų asmenys pirmiau minėta prasme (Chalmers 1996: 93–99).

Jei tokie zombių pasauliai yra logiškai įmanomi, tai „asmuo“ arba fenomenalios sąmonės patirtis negali logiškai priklausyti nuo fizinių faktų. Ar tai reiškia, kad asmuo niekaip nesusijęs su fizine sistema, per kurią reiškiasi? Chalmersas teigia (1996: 248–249), kad ryšys tarp sąmonės ir fizinių faktų egzistuoja – tai procesų koreliacija, kuri yra nomologinio pobūdžio. Jei sąmonė yra „ekstraingredientas“ fiziškai aprašomos pasaulio dalies atžvilgiu, tai turi būti fundamentalesni, platesni nei šiuolaikinės fizikos, neurobiologijos ir kt. fizikalistinių aprašymų formuluojami dėsniai, vadintini galbūt psichofiziniais. Tokie dėsniai paaiškintų sąmonės, kaip sudėtingos funkcinės šių sistemų savybės, atsiradimą fizinėse sistemose.

Kitaip tariant, fenomenali patirtis šiam autoriui yra „fizinės sistemos abstrakčios funkcinės sandaros kokybė“ (1996: 247). Ši funkcinė sandara gali būti pakartota įvairiose fizinėse sistemose, tad jei konkreiti fizinė sistema turi pakankamai sudėtingą funkcinę sandarą, tai joje gali atsirasti sąmonės patirčių. Chalmersas (*ibid.*) vadina šią funkcinę sandarą „abstrakčia kauzalinės sąveikos tarp įvairių fi-

zinės sistemos dalių struktūra“ (*abstract pattern of causal interaction between various parts of a system*).

Tačiau kiek toks nomologinis ryšys mums būtų pažinusi? Galbūt fenomenali patirtis yra neperteikiama, o tai liudytų jos uždaramą ir kauzalinio ryšio su „fiziniu pasauliu“ nebuvimą? Chalmersas teigia, kad tiek solipsistinės (1996: 85), tiek priešingos joms panpsichinės pozicijos (1996: 216) yra alogiškai prieštaringos. Tai, kad asmeninė fenomenali patirtis, fiksuojama savistabos (introspekcijos) būdu, kokybiškai skiriasi nuo bendravimo su kitomis būtybėmis ar jų elgsenos stebėjimo patirties, neįrodo, kad nėra kitų būtybių, turinčių fenomenalių, sąmonės patirčių, ar kad tokia patirtis nėra įmanoma nurodyti ir paliudyti.

Tą galime pailiustruoti Keitho Gundersono suformuluotu „tyrimo asimetrijos“ argumentu (1970: 112–127). Radikalesnis dualizmas<sup>3</sup> (arba, priešingai, reduktyvus monizmas, pavyzdžiui, fizikalistinis materializmas) epistemologiškai yra labai patogus, pripažįsta Gundersonas. Taip yra dėl tos priežasties, kad „Aš“, kaip bet kokių įvykių stebėtojas, akivaizdžiai negaliu būti dviejose vietose iš karto: save patį skanuojantis skaneris neįmanomas, kaip ir save pačią stebinti akis ar save patį pamatyti leidžiantis povandeninės valtys periskopas. Dėl to mes iš tikro negalime šnekėti apie „save“ kaip apie kitą asmenį ir turime vienskaitos pirmojo asmens kalbos formas. Čia galima prisiminti Jaakko Hintikka nurodytąjį asmenvardžio „Aš“, kaip neatsiejamai susiju-

---

<sup>3</sup> Be minėto savybių dualizmo, kitos žinomesnės ir paminėtinios dualizmo formos yra: sąveikos dualizmas (Descartes: fenomenali sąmonė ir kūnas ribotai sąveikauja, sąmonei niekaip kokybiškai nesikeičiant ir išliekant nematerialiai) arba lygiagretumo dualizmas (Leibniz: sąmonė ir kūnas niekaip vienas kito neveikia ir iš esmės yra du lygiagretūs pasauliai).

sio su bet kuriuo kalbėtoju, paradoksą, kai kuris nors „Aš“ pareiškia „Aš neegzistuoju“ – o kas gi tą skelbia? (1962: 332).

Individualaus proto / sąmonės / asmens koreliacija su individualiu kūnu kaip tik dėl to ir yra sunkiai apibrėžiama ir pagimdo keistas dualistines pozicijas (kaip Descartes'o). Apibendrinamas Gundersonas siūlo pripažinti, kad nustatę šį sunkumą galėtume galų gale priimti faktą, kad niekada nepajėgsime tiek pat išsamiai tirti save pačius, kaip tiriamo kitus asmenis, įvykius, daiktus ir t. t., tad, pripažįstant pirmojo asmens patirties aprašymus, dera pasikliauti neutraliais trečiojo asmens pozicija formuluojamais aprašymais. Svarbiausia, kad šiepiistemologinė asimetrija, pabrėžia Gundersonas, nesukuria ontologinio skirtumo tarp periskopo, stebinčio kitus periskopus, valtis, vandens gyvūnus ar pan., ir, kita vertus, negalinčio pamatyti paties savęs savo taikiklyje.

Vadinasi, jei aš turiu fenomenalią sąmonės patirtį, kuri koreliuoja su mano kūno procesais, pavyzdžiui, būdravimu, kurį galima aprašyti fiziologiškai kaip centrinės nervų sistemos būklę (Chalmers 1996: 225–229), tai pati ši dispozicija nesuteikia pagrindo neigti, kad fenomenali patirtis ir jos koreliacija nevyksta kitose autonominėse fizinėse sistemose, turinčiose panašią sudėtingą sandarą.

Tačiau kokiū būdu atpažįstame, kad kuri nors fizinė sistema turi fenomenalią „Aš“ patirtį? Kodėl galime teigti, kad esinio savybę būti asmeniu galima tik atrasti, o ne, pavyzdžiui, išrasti, lygiai kaip siaurąją prasme tėvai nepagimdo vaiko kaip asmens.

Tėvai gali skatinti ar slopinti vaiką kaip asmenį, gali neieškoti ir neaptikti jame asmens, bet vargu ar jie gali suteikti / sunaikinti pačią galimybę dėl jų prokreacijos akto atsiradusiai būtybei būti asmeniu, nebent jie sunaikintų šios būtybės fizinę sistemą. Kitaip tariant, ga-

lima neleisti fizinei sistemai būti aptiktai kaip asmeniui, tačiau negalima sąlygoti pačios galimybės būti asmeniu: sudarius palankias sąlygas, asmuo gali būti aptiktas. Mes tiesiog nežinome, ar asmuo aktualiai buvo fizinėje sistemoje, ar nebuvo ir ar jis bus joje bendravimui pasibaigus.

Aptikti asmenį (tai reiškia atpažinti jį atsitiktinai) ir atrasti jį (tikslingai ieškoti jo atpažintinų apraiškų) yra įmanoma, jei priimtume epistemologinę prielaidą, kurią Chalmersas (1998: 10–23) išreiškia „aktyvaus eksternalizmo“ pozicija. Pagal šią poziciją, mūsų sąmonės veikla – „kognityviniai procesai“ – vyksta aktyviai sąveikaujant su fizine asmens aplinka. Chalmersas čia remiasi Davido Kirsho ir Paulio Maglio (1994) „episteminių veiksmų“ sąvoka, pagal kurią ženkli kognityvinių veiksmų dalis iš tikro vyksta išorinėse žmogaus kūnu fizinėse struktūrose. Episteminiai veiksmai keičia fizinę pasaulio struktūrą, kad sukeltų kognityvinių sąmonės veiksmų eigą. Kaip an-tai: ką nors aiškindamiesi ar aiškindami pasitelkiame saviraiškai įvairius artefaktus ir „gamtinės kilmės“ sistemas, kad pasiektume rišlesnio mūsų fenomenalios patirties aprašymo. Logaritminės liniuotės, kalba, vaizdinės priemonės ir net atskiros „natūralios“ fizinės aplinkos dalys gali būti įtrauktos į episteminę veiklą ir atliktiveiksmus, įmanomus atliktivien fenomenalioje „vidinėje erdvėje“, tačiau tuo atveju pareikalasiančius daugiau veiksmų, nei įtraukiant „išorines“ fizines sistemas.

Tokia pozicija ne tik išvengia solipsizmo, bet taip pat leidžia teigti aktyvios sąveikos su kitais asmenimis galimybę be pavojaus apsiritoti panpsichiniais tokios sąveikos aprašymais.

Taigi aptartu atveju nesama griežtos priklausomybės tarp asmens egzistavimo galimybės bei jos fizinių sąlygų, tačiau nuo jų aiškiai priklauso asmenybės pažinimas ir sklaida tiek

introspektyviai, tiek bendraujant su kitais asmenimis. Čia „asmuo“ yra pirmausia epistemologinė, o ne ontologinė kategorija. Ontologiškai asmuo įgyja tą patį statusą, kaip ir kiekviena fizinė sistema, tačiau epistemologiškai jis negali būti redukuotas nei į fizikalistinius, nei į idealistinius esinijos aprašymus – yra tik įmanoma aprašyti jo koreliaciją su fizine sistema minėtos funkcinės sandaros prasme.

Tačiau kaip galime šią koreliaciją nereduktyviai paaiškinti, nenaudodami nei materialistinio ar fizikalistinio reduktyvizmo, nei nuoseklesnio dualizmo pozicijų? Antai Terence Horganas (1987) mano, kad tai neįmanoma, ir kritikuoja Chalmerso poziciją kaip epifenomenalizmo išraišką. Horgano manymu, Chalmersas neturi galimybės paaiškinti, kad jei mentalinius įvykius, kurie konstituotų „fenomenalią asmens patirtį“, jungiame su fiziniais įvykiais tik nomenologiškai, tai tie mentaliniai įvykiai neturės jokio kauzalinio ryšio su fiziniais. Kitais žodžiais, mentaliniai reiškiniai liks fizinių epifenomenais. Tuo atveju žmogaus ar kitų pripažintų asmenimis būtybių elgesį bus lengviau aprašyti, pavyzdžiui, vien fizikalistiškai.

Chalmersas gina savo poziciją (1996: 155) gretindamas ją su savaip interpretuotu Bertrando Russello neutraliu monizmu<sup>4</sup>: esminės pasaulio savybės nėra nei mentalinės, nei fizinės, „fizinės“ ir „mentalinės“ savybės yra išvestinės iš bendresnių, fundamentalesnių.

Vienintelis sunkumas, priėmus tokią ontologiją, būtų epistemologinis triūsas „grįžti“ iš

---

<sup>4</sup> Pagal originalų Russello „neutralųjį monizmą“ (Russel 1921), protai ir kūnai susideda iš įvykių ir skiriasi tik jų kauzalumą tvarkančiais dėsniais: fiziniai įvykiai yra tvarkomi fizinių, o protai – „mneminių“ dėsnų, pagal kuriuos aktuali proto būklė yra lemiama anksčiausiai pasibaigusios proto būklės, tad čia protas ir kūnas turi skirtingas kauzales struktūras mums prieinamoje episteminėje esinijos srityje, tuo tarpu Chalmersas teigia kūno ir sąmonės „kauzalinės struktūros koherenciją“ (1996: 223).

partikuliarių esinijos domenų aprašymų lygio į vieningą fundamentalią, iš santykinų kauzaliųjų ryšių struktūrų – prie esminių savybių. Tad, pasak Chalmerso (*ibid.*), „kasdieniais aiškinimo sumetimais yra patogiau ir toliau mąstyti apie šią poziciją fizinių dėsnų sistemos kategorijomis, kur papildomi principai jungtų fizinių domeną su fenomenaliu“. Dera pabrėžti, kad čia pat, vengdamas kaltinimo reduktyvumu, šis autorius skelbia naujos, išplėstinės, „psichofizinės“ teorijos poreikį, siekia plėsti ir detalizuoti funkcinio psichofizinio kontinuumo aprašymą pagal pirmiau paminėtą šio autoriaus tokios sąmonės ir kūno vienovės apibrėžimą.

Tokiam epistemologiniam triūsui reikalingas – kaip ir kiekvienai mokslinei teorijai – išipareigojimas, nes nėra jokio griežto būtinumo atrasti ir paaiškinti šiuos fundamentaliuosius dėsnius. Galime sakyti, kad kiekvienas žmogus apsisprendžia (*commits her / him-self*) tam tikrai ontologijai: kai kurie žmonės tiki velniais, protėvių dvasiomis, agresyviais ir technologiškai pirmaujančiais visatoje ateiviais, bet yra žmonių, tikinčių neutraliu monizmu bei, žinoma, diena, kai turėsime protingus robotus, vadintinus lygiaverčiais mums asmenimis.

Čia ne tiek svarbu, kaip tie žmonės tuo patikėjo ar apsisprendė (pavyzdžiui, pažiūrėję D. Spielbergo filmą „Dirbtinis intelektas“), svarbu, kad jie pasišventė, pasak Rosalindos Cartwright (1954: 316), tokiai ontologijai, kurioje tokie dalykai yra realiai galimi. Kitais žodžiais, yra tikima, kad labai panorėjus galima tuos dalykus išvysti aktualiame pasaulyje. Tokios „neformalios ontologijos“, pasak Cartwright (*ibid.*), daro įtaką ar yra veikiamos formaliųjų, akademiškai diskutuojamųjų teorijų – kartais gana dideliu mastu.

Šiuo atveju galime teigti, kad formalizuotas Chalmerso ir kitų autorių, ginančių pana-

šias pozicijas, įsipareigojimas neutraliam monizmui arba kuriai kitai pozicijai yra tiesiogiai susijęs su neformaliu tikėjimu protingų automatų buvimu. Kad svarstytume, ar tokie automatai galimi, ar ne ir kokia teorija sėkmingiausiai tą paaiškintų, kontekstualiai mums reikia bent neformalaus tikėjimo teiginiu, kad tokie automatai neišvengiamai bus sukurti / atrasti.

Čia neketinu svarstyti sociologinių, psichologinių ir kultūrologinių šio konkretaus įsipareigojimo sąlygų, tačiau verta bent apgraužomis aptarti jo pagrindinio objekto – automato, kuriam galima priskirti proto ar sąmoningų patirčių savybę, sampratos raidą.

## Kas yra automatas?

### a) Kartezinė automato samprata

Reikia priminti, kad bent jau naujųjų amžių mechanizmo kontekste *automaton* (vedinys iš klasikinės graikų k. *automatos* – „savaime judantis“, „savaime nutinkantis“) reiškė bet kokį savarankiškai judėti galintį kūną, suvokiamą kaip iš mechanškai sąveikaujančių dalių sudarytą įrenginį.

Tačiau šiais laikais jau turime „dirbtinės gyvybės“, „įkūnytųjų robotų“ ir panašias sampratas, leidžiančias modeliuoti automatų ne kaip, pavyzdžiui, XVIII a. inžinieriaus Jacques Vaucansono mechaninė antelė, „pati“ besimaitinanti kruopomis ir besituštinanti savo sumaltomis liekanomis, bet kaip autonomiškas, informaciją apdorojančias mašinas.

Visa automatų teisės į protingą ir sąmoningą egzistenciją istorija remiasi pirmiausia Rene Descartes'u. Manau esant savai suprantama, kad be jo pastangų, nors ir nukreiptų kita linkme, tokie neformalaus ontologinio įsipareigojimo automatams pavyzdžiai kaip Isaco Asimovo fantastinės literatūros kūrinys „Aš – robotas“ arba populiarūs filmai „Matri-

ca“ bei „Dirbtinis intelektas“<sup>5</sup> nebūtų atsiradę, o ką jau kalbėti apie japonų buitinių prietaisų rinkoje graibstomus elektroninius naminius šunelius „Aibo“.

Kaip nurodo Brianas Baigrie ir Patricia Kazan (1997), pirmas Descartes'o nuopelnas automatų sampratos plėtrai ir jų technologijų raidai – tai žymus antikvinio esinių skirstymo į natūralius ir dirbtinius susilpninimas. Būtent jis atmetė dar nuo antikos likusią prigimties, valdančios materialius esinius, sampratą, tuomet vis dar gyvavusią Aristotelio formos ir telos distinkcijos pavidalu.

Būtent Descartes atpalaidavo materijos sampratą nuo paslaptinių tikslų, esmių ar besiaktualizuojančių formų, suteikdamas materijai inertiškos ir tolydžios bet kokio judėjimo ar kitų virsmų atžvilgiu „substancijos“ reikšmę. Po Descartes'o materija tapo galima įvairiai manipuluoti, žaisti jos dalimis, nebijant pažeisti jokio joje slypinčio *teloi*, esmių ar dar kažko, suprantamo tik išvalgiems teoretikams graikiška šio žodžio prasme.

Materiali gamta, nepriklausoma nuo *telos*, tapo didžia inžinerinės veiklos scena. Tai jau nebebuvo *ars imitatio naturae* – prigimties / gamtos imitavimo „menas“ ar veikiau „technologinė kūryba“. Materialūs esiniai tapo įmanomi pagaldoti ir modeliuoti kaip iš tvarkingai sąveikaujančių dalių sudaryti agregatai, neturintys jokių vystymosi tikslų, išskyrus suteikiamus mūsų.

<sup>5</sup> Angliškas Davido Spielbergo filmo pavadinimas yra „Artificial Intelligence“. atitinkantis aktualios teorinių bei inžinerinių tyrimų srities pavadinimą. Kitur šiame tekste vartoju „dirbtinio proto“ sąvoką, kuri, mano nuomone, yra artimesnė originaliai šio angliško žodžių junginio reikšmei dėl to, kad „protas“ lietuvių kalboje skamba funkciškai neutraliau nei „intelektas“: „protas“ gali būti ekvivalentiškas bet kuriai aukštesnei mentalinei galiai ar jų kompleksui, o „intelektas“ labiau nurodo siauresnę proto galių sritį, susijusią pirmiausia su loginiu mąstymu ir nuovoka.

Vienintelis ryškus skirtumas tarp padirbto esinio (daikto) ir nepadirbto, kaip interpretuoja Descartes'ą Amy Schmitter (1996: 368–370), buvo tas, kad vadinamieji gamtos padirbti dalykai rečiau gsdavo: Descartes laiške Henry'ui More'ui sako, kad jei jau menai amatų prasme mėgdžioja gamtą kurdami įvairius automatus, atrodo pagrįsta, kad gamta turėtų kurti savo pačios automatus, daug tobulesnius nei padirbtieji (1649: 204).

Kitais žodžiais, mechanizmai ir ypač automatai esą „gamtos“ esinių modeliai, lygiai taip pat bedvasiai kaip vabalai, šunys ar žmogaus kūnas, jie paprastesni, jų dalys dažnai greičiau sudylančios arba jų dermė greičiau išyranti, tačiau, pavyzdžiui, šuns ar žmogaus kūno senatvė esanti ne kas kita, bet šių kūnų, šių natūralių automatų dalių susidėvėjimas.

Vis dėlto, šitaip aiškinant, padirbti automatai, gyvuliniai automatai ir dar viena privileijuota automatų rūšis – žmogaus kūnas nėra identiški žmogui kaip protaujančiai esybei, kuri esti nepriklausomai nuo *res extensa*. Kaip visi žinome, taip yra dėl to, kad šiam autoriui šalia kokybiškai homogeniškos materialios esinijos srities yra ir lygiai taip pat homogeniška, tačiau ontologiškai visiškai skirtinga ir nepriklausoma mentalumo sritis.

„Samprotavimo apie metodą“ 5 skyriaus pabaigoje išdėstyta pozicija, kad automatai negalėtų racionaliai mąstyti, t. y. metodiškai, taisyklingai manipuluoti mentalinių sistemų simboliais, nes jie paklūsta tik fizikos dėsniams. Kitais žodžiais, jie yra tik tįsiosios substancijos dariniai, menkai tegebantys mėgdžioti žmogaus protavimą (šiandien galėtume sakyti – kaip „Deep Blue“ mėgdžioja G. Kasparovą), bet niekada nesugebėsiantys kurti, pavyzdžiui, prasmingų atsakymų ar suprasti humoro.

Vaucansono antis gali būti puiki mechaninė anties replika, tačiau niekas, sekant Des-

cartes'u, niekada nebūtų galėjęs ir nebus pajėgus sukurti automatinį Descartes'o, Mersenne'o ar net kardinolo Masarini, kurių nebūtų galima visiškai patikimai atskirti nuo „originalo“. Skardinis medkirtys iš Herberto Franko „Ozo šalies burtininko“, pagal Descartes'o interpretaciją, nieko esmingo neįgytų, gavęs iš burtininko širdį (emocinius gebėjimus), kaip ir kaliausė – smegenų (racionalių gebėjimus): jie abu liktų vyrišką būtybę primenančiais agregatais, gebančiais tik mėgdžioti burtininką ar protiniais sugebėjimais pranašesnes šio pasakojimo veikėjas – mergaites.

### **b) Automatas kaip autopoetinė sistema**

Kaipgi automato samprata kito nuo naujųjų laikų? Paviršutiniškiausiai: nuo esinio, galinčio autonomiškai judėti, buvo pereita prie saviorganizuojančios ar autopoetinės „sistemos“ bei roboto sąvokų. Kaip matysime, „robotas“ apskritai reiškia sudėtingą, technologinės kilmės automatą, įvairiu mastu atkartojantį žmogaus fizinę, funkcinę, o gal ir mentalinę sandarą. Jau pakankamai aišku, kas yra automatai, vartojant šį žodį karteziškai. Ką gi dabar įvairūs dirbtinės gyvybės ir (ar) proto proponentai naudoja „automatui“ apibrėžti? Iš pradžių apžvelgsime „dirbtinės gyvybės“ tyrimų lauke relevantiškas automatizmo sampratas, o vėliau – naudojamas robotikoje.

Taigi, pirma, nuo Antrojo pasaulinio karo automatizmo gyvose būtybėse sampratos raidą lėmė evoliucinės biologijos ir šiuolaikinės fizikos tyrimų plėtra. Ši raida kulminuoja dirbtinės gyvybės (*artificial life*) tyrimų pavidalu. Antra, turime kibernetikos – kompiuteriais valdomų technologinių automatų – teorinį aparatą, kulminuojantį „pritaikytos robotikos“ tyrimais. Šie tyrimai kartu yra ir platesnių – dirbtinio proto tyrimų dalis ir yra ypač iliustratyvūs. Dirbtinio proto tyrimų konteksto ap-

tarimas dėl suprantamų vietos stokos sumetimų lieka už šio straipsnio ribų. Dera pabrėžti, kad šiedu laukai dėl neišvengiamo jų tarpdalykinio pobūdžio ženklumastu susisiečia. Kartu jie išlieka kaip atskiros tyrimų kryptys bendrame, taip pat tarpdalykiniame pažinimo mokslo ir filosofijos kontekste.

Kaipgi plėtojosi automatizmo dirbtinės gyvybės prasme samprata? Pirmiausia dėl evoliucinės biologijos (ypač neurologijos) ir fizikos pažangos pastarąjį pusšimtį metų pradėta kalbėti apie „saviorganizuojančias sistemas“. Ši sąvoka turi daugybę variacijų ir išoriškai iš pirmo žvilgsnio skirtingų vartosenų įvairių autorių grupių kontekstuose.

Vienas ankstyvesnių ir tinkamų visam šiam teoriniam kontekstui perprasti saviorganizuojančios sistemos apibrėžimų priskirtinas Herbertui Simonui (1962). Šis autorius, vadovaudamasis griežtu fizikalistiniu požiūriu, esinijoje atskiria dviejų tipų sistemas (t. y. sudėtingus esinius, kuriuose galime atpažinti bei modeliuoti tam tikrą juos sudarančių menkesnio lygio esinių, elementų, tvarką, arba „struktūrą“). Simonas skiria šiuos tipus pagal jų sandaros kitimą laiko atžvilgiu: taigi turime sudėtinės (*additive*) sandaros struktūras, kurios telkia į sistemą tam tikras elementarias dalis.

Pridėtinės sandaros sistemos gali būti papildytos arba sumažintos pridėdant arba atimant jų elementarias dalis (dėl to ir „pridėtinės“), gali būti ir visiškai „išnarštytos“ dėl išorinių ir vidinių (sistemos atžvilgiu) galių poveikio. Tai, pabrėžia Simonas, galima palyginti lengvai stebėti ir aprašyti fizikos metodais.

Palyginti su kuo? Pagal Simoną, esama kitokių sistemų, kurios yra gerokai sunkiau papildomos ar sumenkinamos minėtos galių sąveikos. Šio tipo sistemos randasi spontaniškai bei demonstruoja didžiuliu mastu vieningą vidi-

nę / išorinę sąveiką, t. y. galime kalbėti apie šios sistemos vienovę ir jos, kaip atskiros nuo aplinkos, „elgseną“. Taip struktūruotų sistemų pavyzdžiai šiame kontekste varijuoja nuo savireguliuojančių, „pirminį būvį“ atkuriančių kristalų struktūrų iki ląstelių beisudėtingų organizmų.

Fizikalistiniame kontekste gyvas būtybes galime vadinti saviorganizuojančiomis sistemomis, kokybiškai neatskirdami jų nuo kitų fizinių sistemų. Tokios spontaniškai besioorganizuojančios sistemos, taikomos gyviems organizmams modeliuoti, iš esmės nesiskiria nuo Kartezijaus automatizmo sampratos – tėra tik išplėstos pasitelkiant šiuolaikinės fizikos ir biologijos sąvokynus.

Pagrindinis klausimas taip apibrėžiant gyvus organizmus yra šito spontaniško susitvarkymo sąlygos – dėl ko jis atsiranda? Daugelis autorių, kaip pabrėžia šios problematikos apžvalgininkė Margaret Boden (1999: 37), bando apeiti šį klausimą, akcentuodami spontanišką, savaiminį šio organizavimosi pobūdį. Kitais žodžiais, jis nėra sąlygojamas jokios išorinės sistemos galios ar „kūrėjo“, tačiau yra „lemiamas sistemos struktūros pobūdžio“, dėl kurio joje vyksta ne šiaip kaita, bet raida: primityvesnės struktūros keičia sudėtingesnės, aukštesnio lygio struktūros.

Taigi besioorganizuojančiose (*selforganizing*) sistemose aukštesnio lygio bruožai randasi dėl paprastesnių struktūrinių elementų sąveikos. Tačiau kaip vis dėlto atsiranda tokios aukštesnio lygio struktūrinės ypatybės kaip „autonmiškas elgsens“, „parazitinė / simbiotinė gyvensena“ ir pan.? Į šį klausimą ir bando atsakyti Dirbtinės Gyvybės (DG) tyrinėjimai. Patį vardą kryptčiai suteikė Christopheris Langtonas (1986 ir 1989). Pavadinimas leidžia manyti, esą šiais tyrimais siekiama kurti naujas gyvas būtybes ne biologinės prokreacijos būdu,

tačiau ne visi DG tyrėjai vieningai sutaria dėl tokio tikslo<sup>6</sup>.

Tiek Simono, tiek vėlesnės šios automatizmo sampratos interpretacijos buvo inicijuotos matematikų Alano Turingo ir Johno von Neumanno, pasiūliusių pirmuosius teorinius saviorganizavimo modelius, rodančius, kaip paprasti vyksmai galėtų sudaryti sudėtingas sistemas su besirandančia tvarka. Juos tad galima istoriškai laikyti ne tik Dirbtinio Proto, bet ir Dirbtinės Gyvybės tyrimų pradininkais, o tai dar kartą paliudija šių kryptų lygiagretumą ir sąsają.

Antai Turingas (1952) parodė, kaip sąveikaujantys cheminės difuzijos gradientai gali sąlygoti aukštesnio lygio (be kita ko, periodines) struktūras iš pradinės vientisos medžiagos. Kaip detalai nurodo von Neumanno mokinys Burksas (1970: 3–64), jo pirmtakas dar iki DNR atradimo aprašė bendriausias replikavimosi struktūras. Jis taip pat sukūrė „universalus replikatoriaus“ modelį: ląstelinių automata (*cellular automaton*), gebanti kopijuoti bet kurią sistemą, įskaitant ir save patį<sup>7</sup>. Von Neumannas taip pat nurodė, kad tokio automato atliekamų kopijų klaidos galėtų paskatinti jo evoliuciją, kuri vėliau tapo kertine va-

---

<sup>6</sup> Be radikalių nuostatų, numatančių dirbtinės kilmės gyvybę, tarkime, kompiuterinių reprezentacijų erdvėje („kibererdvėje“ *cyberspace*), yra nuosaikesnių, modeliuojančių vadinamuosius „animatus“ – fizinius arba kompiuteriškai simuliuotus robotus, tam tikru mastu atkartojančius realias gyvas būtybes. Kitais žodžiais, animatai būtų „mažiau“ gyvi nei jų biologiniai prototipai.

<sup>7</sup> Ląstelinis automatas yra skaičiavimų erdvė, sudaryta iš daugelio atskirų ląstelių, kai kiekviena ląstelė gali būti vienos iš kelių būklių. Ši būklė kinta ar išlieka pagal ypatingas, dažniausiai tik tai ląstelei ar ribotai jos aplinkai būdingas taisykles (Von Neumann 1951: 1–31). Net gana paprastuose ląsteliniuose automatuose aukštesnio lygio struktūros gali pasirodyti po daugybės iteracijų (matematiškai aprašomų vyksmų kartojimosi). Tokie modeliai reikalauja didelės galios modernių kompiuterių, tad von Neumano ir Turingo hipotezės bei modeliai yra tinkamu mastu tikrinami tik dabar.

dinamosios evoliucinės kompiuterikos (evoliucinio programavimo, genetinių algoritmų ir pan.) prielaida.

Remdamasis šiomis sampratomis, Stuartas Kauffmanas (1992) yra bandęs įrodyti, kad von Neumanno ląstelinių automatų savireguliacija yra pakankamas saviorganizuojančių sistemų modelis. Kitą požiūrį gina Humberto Maturana ir Fracisko Varela (Maturana ir Varela 1980). Pagal juos, yra aiškus skirtumas tarp paprastų susitvarkančių (pavyzdžiui, kristalinių) ir „autopoetinių“ sistemų. „Autopoezė“ čia reiškia gyvų būtybių gebą keisti savo aplinką ir kurti tiek tos aplinkos, tiek savęs pačių reprezentacijas. Autopoezės kulminacija, žinoma, būtų sąmonės patirtis.

Kaip pažymi Bodeną (1996), nesutariama keletu svarbių dirbtinės gyvybės, lygiai kaip ir dirbtinio proto, klausimų: ar gyvybė, kaip ir protas, turi būti būtinais įkūnyti ar gali tvyroti kažkokioje nematerialioje, abstrakčioje erdvėje? Be to – ar gyvybė, kaip sąmonės patirtis, gali būti laipsniuojama savybė („mažiau“ ar „daugiau“ gyva)?

Tačiau čia slypi tam tikras konceptualus nesusipratimas, galimas išryškinti dėl kai kurių Chalmerso pozicijų (1996: 53–54). Būtent: jei aukštesnio lygio savybė (pavyzdžiui, būti gyvam arba būti sąmoningam) turėtų griežtai priklausyti nuo įvairių žemesnio lygio savybių, pavyzdžiui, „turėti medžiagų apykaitos sistemą, veikiančią 75 nuošimčių efektyvumu“, tai nebūtų sunku atskirti gyvą daiktą nuo negyvo. Tačiau „gyvybė“, kaip ir „sąmonė“, negali būti tiksliai apibrėžtos: neaišku, ar DNR molekulė arba kompiuterių virusai yra gyvi, tačiau, pavyzdžiui, aišku, kad šuo yra gyvas.

Šio straipsnio temai aktuali Dirbtinio Proto teorijų, kurių raidos ir įvairovės čia dėl vietos stokos nenagrinėsime, taikymo sritis yra

pritaikytoji arba pagrįsta elgesiu, robotika (*situated / behaviour based robotics* – Matarič 1999). Robotai čia suprantami kaip gyvų būtybių – nuo vabzdžių iki žmogaus – elgesio modeliai, koreguojami „iš apačios aukštyn“, t. y. sąveikaujant su aplinka. Taip norima simuliuoti gyvų fizinių sistemų evoliuciją iki, galimas daktas, sąmonės atsiradimo. Tokia strategija liudija bihevioristinį ir vienokią ar kitokią prasmę materialistinį šių projektų vykdytojų požiūrį į sąmonės ir kūno santykį.

Kaip jau buvo minėta, iš esmės dirbtinės gyvybės, dirbtinio proto ir pritaikytosios robotikos tyrimų laukai persikloja. Humanoidinių ar nehumanoidinių robotų kūnai, jei pakankamai evoliucionuos Chalmerso siūloma kryptimi (t. y. turės gana sudėtingą funkcinę sandarą), galbūt galės pateisinti pačių tyrėjų ir šių tyrimų mėgėjų ontologinį išipareigojimą.

### **Kokia yra galimo protingų automatų atradimo strategija?**

Galima tik sutikti su Chalmersu, kad kol toks modeliavimas bus fizikalistinis, t. y. iš principo reduktyvus fenomenalios patirties atžvilgiu, tol nebus įmanoma aptikti sąmonės, kad ir kokie sudėtingi ir autonomiški tie modeliai būtų. Nereduktyvumas sąmonės atžvilgiu čia reikštų jau minėtos funkcinės sandaros išryškinimą. Kaip tą daryti, bent jau kol kas menkai teaišku.

Dar vienas apribojimas čia būtų mastelis, kurį pasiekus tokios abstrakčios struktūros atskirtotų fizinėse sistemose. Kokio masto turėtų būti „pakankamas“ tikrovės simuliuojimas, kad dirbtinė fizinė reprodukcija „nesiskirtų“ nuo biologinio originalo taip, kad galėtume joje laukti asmens apraiškų?

Chalmersas (1996: 331), atsakydamas į fiziko Rogerio Penrose'o<sup>8</sup> skeptinę ataką, teigia, kad jei reprodukcija pasiekia funkcinį ir architektūrinį tapatumą originalui 10<sup>-12</sup> masteliu, mes galime kalbėti apie pakankamą ir lygiavertę repliką. Tolimesniame kvantų lygyje neapibrėžtumo principas replikoje būtų tolygus „natūralios“, t. y. biologinės, sistemos neapibrėžtumui ir apie jį nieko pasakyti tiesiog nebegalėtume, nors jis reikštų neatskiriamą replikos savitumą, leidžiantį minimaliai atpažinti ją kaip „kitą originalo variantą“. Toks variantas turėtų turėti visas autonomiškos, maksimaliai sąmoningos veiklos galimybes, nors, kaip jau esame aptarę, nebūtinai jas parodys.

Kaip galėtume to pasiekti? Kaip jau matėme, neįmanoma stebėti įvairių fenomenalių patirčių kitimo priklausomai nuo fizinių procesų, nes šių dviejų plotmių dermė yra visuomet aposteriorinė. Galbūt čia net neįmanoma rasti fundamentalių bendrų dėsnių mokslinė prasmė ir teks visą laiką tenkintis pseudomokslu, remiamu vien ontologinio išipareigojimo?

Chalmersas (1996: 247), nuosekliai remdamasis savuoju Russello monizmo variantu, siūlo tris principus, kurie turėtų padėti kurti naująją psichofiziką, atitinkančią neutralųjį monizmą, arba „natūralistinį dualizmą“, kaip šis autorius pats vadina tokią poziciją.

1. Struktūrinės darnos principas: jau seniai empiriniais tyrimais pastebėta struktūrinė sąmonės patirties ir informacijos, apdorojamos smegenyse, atitiktis (pagal Muellerio (1986)

---

<sup>8</sup> Penrose'as (1987) teigia, kad kvantinių lygiu mes negalime išskirti jokios aiškios struktūros. Šis esinijos lygmuo, jo manymu, ir yra spontaniško sąmoningo elgesio laidas: jo negalima apskaičiuoti, kompiuteriškai ar kaip nors kitaip simuliuoti. Chalmerso kontrargumentas nurodytoje ištraukoje pateikia šio lygmens įvykių neapibrėžtumą kaip epistemologiškai nesvarbų ir niekaip nepraskaidrinantį „psichofizinių dėsnių“.

„psichofizines aksiomas“ ir Kohlerio (1947) „struktūrinį izomorfizmą“). Pavyzdžiui, įsisa-  
moninimas (*awareness*), kaip jį interpretuoja  
Chalmersas, yra smegenų vykdomas informa-  
cijos apdorojimas, užtikrinantis „visuotinę pro-  
cesų kontrolę“ (*global control*) ir bendriausia  
prasme struktūriškai izomorfiškas sąmonės pa-  
tirčiai.

2. Sandaros pastovumo principas, nusakan-  
tis, kad sąmoninga būtybė turi vieną funkcinę  
struktūrą, bendrą tiek mentaliniams, tiek fizi-  
niams įvykiams toje būtybėje.

3. Dvejybinio atžvilgio informacijos reali-  
zuotumo sąmoningose būtybėse principas rei-  
škia, kad jei aptinkame fenomenaliai realizuo-  
tą informaciją, aptinkame ir analogišką fizinį  
jos realizuotumą: pavyzdžiui, fenomenalispal-  
vos patirtis trijų matmenų informacinėje erd-  
vėje atitiks centrinėje nervų sistemoje neuro-  
nų koduotą trijų matmenų reprezentaciją re-  
gėjimo centre (Chalmers, 1996: 285).

Chalmersas mano, kad pakankamai at-  
skleisti šie principai padėtų pakartoti funkci-  
nę sandarą, reikalingą sąmonei tarpti įvairio-  
se fizinėse sistemose.

Iš to išeitų, teigia šis autorius, kad jei būtų  
sukurta jo, Davido Chalmerso, funkciškai ly-  
giavertė reprodukcija iš metalo ir silikono (t. y.  
tokio dirbtinio kūno funkcionavimo struktūra  
būtų izomorfiška dirbtinio proto tyrinėtojiui  
Chalmersui), tai tokia metalinė-silikoninė šio  
autoriaus reprodukcija galėtų demonstruoti ne  
tik psichologiškai ekvivalentišką sąmoningu-  
mą (budrumą), bet ir bylotų mums apie lygia-  
vertes fenomenalias patirtis. Lygiavertes, bet

ne tapačias: mat tai jau būtų „visai kito“ as-  
mens patirtys. Tai liudija atkaklų Davido Chal-  
merso ontologinį įsipareigojimą panašioms  
perspektyvoms. Deja, šis įsipareigojimas au-  
toriaus nėra atskirai reflektuotas ir dėl to kar-  
tais sudaro perdėto optimizmo įspūdį.

## Išvados

1. Visos Davido Chalmerso ir kitų šio dis-  
kusijų lauko dalyvių teorinės pastangos pagrįsti  
asmenų technologinės kilmės kūnuose atradi-  
mo galimybę liudija ne tiek tokio atradimo rei-  
kalingumą ar neišvengiamumą, kiek specifinį,  
dažnai neformalios kilmės ontologinį įsiparei-  
gojimą šitokiai galimybei.

2. Toks ontologinis įsipareigojimas nors ir  
turi aiškių trūkumų, pirmiausia teigiamai vei-  
kia Vakarų filosofijos diskursą, ryškiau tema-  
tizuodamas „asmens“, „sąmoningos patirties“  
ir jos koreliacijos su fiziniais procesais bei  
struktūromis problemą, kurią Levine'as (1983)  
yra pavadinęs „aiškinimo spraga“ (angl. *expla-  
natory gap*): mes nežinome, kodėl viena ar ki-  
ta fizinė sistema paįžymi fenomenalia, sąmo-  
ninga patirtimi, net jei turime nepaprastai iš-  
samius šios sistemos fizikinius aprašymus.

3. Ontologinis įsipareigojimas ieškoti po-  
zityvaus, „aiškinimo spragą“ užpildančio apa-  
rašymo praturtina ir etinius svarstymus: as-  
mens neredukuojamumo derinimas su onto-  
loginiu monizmu leidžia išplėsti esinių, gali-  
mų vadinti asmenimis, klasę, o taisyteikia per-  
spektyvą pripažinti asmenimis iki šiol „nežmo-  
gišką“ statusą turėjusias būtybes, potencialiai  
galinčias liudyti sąmonės patirtis.

## LITERATŪRA

1. Baigrie, B. S., Kazan, P. J. (1997). „Technology  
and the Creation of Health Biotechnology and the Cre-  
ation of Health Care Needs“, *Techné: Journal of the  
Society for Philosophy and Technology*, Vol. 2, Nr. 3–4.

2. Boden, M. A., ed. (1996). *The Philosophy of Ar-  
tificial Life*. Oxford: Oxford University Press.

3. Burks, A. W. (1970). *Essays on Cellular Automa-  
ta*. Urbana, Illinois: University of Illinois Press.

4. Cartwright, R. L. (1954). „Ontology and the Theory of Meaning“, *Philosophy of Science* 21: p. 316–325.
5. Chalmers, D. (1996). *The Conscious Mind*. Oxford University Press.
6. Chalmers, D.; Clark, A. (1998). „The Eütended Mind“, *Analysis* 58: p. 10–23.
7. Descartes, R. (1694). *OEuvres*, tome X, p. 204: *Réponse de M. Descartes a M. Morus*.
8. Franklin, S. (1997). *Artificial Minds*. Cambridge, MA; London, UK: MIT Press.
9. Gunderson, K. (1970). „Assymetries and Mind-Body Perplexities“, in D. M. Rosenthal (ed.) *Materialism and the Mind-Body Problem*. Hackett Publishing, 1987, p. 112–127.
10. Hintikka, J. (1962). „Cogito, Ergo Sum: Inference of Performance?“, *The philosophical Review*, vol. LXXI, No. 1, Whole No. 397, p. 3–32.
11. Horgan, T. (1987). „Supervenient qualia“, *Philosophical Review*, No 96: p. 491–520.
12. Kauffman, S. A. (1992). *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford: Oxford University Press.
13. Kirsh, D. & Maglio, P. (1994). „On Distinguishing Epistemic from Pragmatic Action“, *Cognitive Science* 18: p. 513–549.
14. Kohler, D. (1947). *Gestalt Psychology*. New York: Liveright Publishing Corporation.
15. Kripke, S. A. (1972). „Naming and Necessity“, in G. Harman and D. Davidson (eds.), *The Semantics of Natural Language*. Dordrecht: Reidel.
16. Langton, C. G. (1986). „Studying Artificial Life with Cellular Automata“, *Physica D* 22: p. 1120–1149.
17. Langton, C. G. (1989). „Artificial Life“, in C. G. Langton (ed.) *Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems (held September 1987)*. Redwood City, CA: Addison-Wesley, p. 1–47. (Reprinted, with revisions, in M. A. Boden (ed.) *The Philosophy of Artificial Life*. Oxford: Oxford University Press, p. 39–94.)
18. Levine, J. (1983). „Materialism and Qualia. The Explanatory Gap“, *Pacific Philosophical Quarterly*, No 64: p. 354–361.
19. Locke, J. (1690). *Esé apie žmogaus intelektą*. Iš angl. k. vertė Ramutė Rybelienė. Vilnius: Pradai, 2000.
20. Locke, J. (1689). *An Essay Concerning Human Understanding*, ed. J. Yolton. London: Dent, 1961.
21. Matarič, M. (1999). „Behavior-Based Robotics“, in *The MIT Encyclopedia of the Cognitive Sciences*, ed. by Robert A. Wilson, Frank C. Keil. MIT Press, p. 74–76.
22. Maturana, H. R., and Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. London: Reidel.
23. Mueller, G. E. (1986). „Zur Psychophysik der Gesichtsempfindungen“, *Zeitschrift fuer Psychologie und Physiologie der Sinnesorgane* 10: p. 1–82.
24. Patterson, F., Linden, E. (1981). *The Education of Koko*. New York: Holt, Rinehart & Winston.
25. Penrose, R. (1989). *The Emperor's New Mind*. Oxford: Oxford University Press.
26. Russell, B. A. W. (1921). *The Analysis of Mind*. London: Routledge, 1992.
27. Schmitter, A. M. (1996). „Formal causation and the explanation of intentionality in Descartes“, *Monist*, Jul96, vol. 79 Issue 3, p. 368–388.
28. Simon, H. A. (1962). „The Architecture of Complexity“, in *Proceedings of the American Philosophical Society* 106: p. 467–482.
29. Turing, A. M. (1952). „The Chemical Basis of Morphogenesis“, *Philosophical Transactions of the Royal Society: B* 237: p. 37–72.
30. Van Gelder, Tim (1993). „The Distinction Between Mind and Cognition“, in *Mind and Cognition: Collected Papers from the 1993 International Symposium on Mind and Cognition*; ed. by Y.-H. Houg, p. 1–29.
31. Velmans, M. (1995). „The Relation of Consciousness to the Material World“, *Journal of Cosciousness Studies*, 2: p. 255–265.
32. Von Neumann, J. (1951). „The General and Logical Theory of Automata“, in L. A. Jeffress (ed.), *Cerebral Mechanisms in Behavior – The Hixon Symposium*, September 1948, Pasadena, CA. New York: Wiley, p. 1–31.

## IS IT POSSIBLE FOR INTELLIGENT AUTOMATA TO EXIST?

**Algis Davidavičius**

### Summary

This article deals primarily with the logical and the factual possibilities for automata to become persons and in that sense intelligent beings. It is argued here, that there are no strong logical arguments against conceivability of possible existence of such entities. This existence is logically possible, if we rely on special kind of epistemologically relevant property dualism, coupled with a broader ontological monism and a functional definition of a person. It is argued

later, that factual realizability of those possible technological persons is strongly dependent on special ontological commitment to the realization of such possibility. The arguments and analysis provided are mostly based on David Chalmers's views and their direct critique by other authors.

**Keywords:** person, intelligent, automata, identity of consciousness, ontological commitment, ontological monism.

*Įteikta 2001 10 11*