



Vilniaus universitetas

Filosofijos fakultetas

Filosofijos institutas

Marius Vaicieuskas

Filosofijos studijų programa

Magistro darbas

**Pirmojo ir trečiojo asmens perspektyvų sąveika integruotos  
informacijos teorijoje**

**The Interplay of First and Third Person Perspectives in  
Integrated Information Theory**

Darbo vadovas: prof. Marius Povilas Šaulauskas

Vilnius

2023

# TURINYS

SANTRAUKA .....	3
SUMMARY .....	4
ĮVADAS.....	5
1. INTEGRUOTOS INFORMACIJOS TEORIJOS PAGRINDAI.....	8
1.1. Pirmojo asmens perspektyva .....	9
1.2 Fenomenologinių aksiomų apodiktiškumo problema .....	14
1.3 Ontologinių postulatų išvedimo keblumas .....	17
1.4 Mintinių eksperimentų pirmumas ir informacijos samprata .....	20
1.5 Integruotos informacijos ir sąmoningumo tapatumo prielaida.....	23
2. IIT INTERPRETACIJOS IR JŲ SAŪVEIKA.....	27
2.1 Šešios IIT interpretacijos.....	27
2.1.1 Empirinis IIT pagrindumas.....	30
2.1.2 IIT sąmoningumo apibrėžimas.....	32
2.1.3 Realistinė IIT pozicija .....	33
2.1.4 Ontologinis IIT statusas .....	35
2.2 Argumentai, paremiantys minimaliąją IIT interpretacijų klasę.....	37
2.2.1 Išplėtimo grafiko argumentas ir smegenų plastiškumas.....	37
2.2.2 Informacijos integravimas ir nesąmoningi procesai.....	42
2.2.3 Minimalus integruotos informacijos kiekis ir numatantysis apdorojimas.....	47
IŠVADOS .....	51
LITERATŪROS SĄRAŠAS.....	53
PRIEDAI .....	59

## SANTRAUKA

Darbe analizuojama pirmojo ir trečiojo asmens perspektyvų sąveika Integruotos informacijos teorijoje (IIT). Tvirtinama, kad yra šešios pagrindinės IIT interpretacijos, skirtingai derinančios šias perspektyvas. Darbą sudaro dvi dalys. Pirmoje dalyje analizuojamos metodologinės IIT problemos, antroje dalyje pateikiama IIT interpretacijų klasifikacija, jų sąveikų analizė ir argumentai, nukreipti prieš integruotos informacijos ir sąmoningumo tapatumo tezę. Įrodinėjama, kad interpretacijos, atmetančios šią tezę, yra lengviau suderinamos su empiriniais tyrimais, išvengia kai kurių konceptualinių problemų, geriau dera su kitomis sąmoningumą tyrinėjančiomis teorijomis bei yra neutralios ontologinių problemų atžvilgiu.

**Pagrindiniai žodžiai:** integruota informacija, pirmojo asmens perspektyva, trečiojo asmens perspektyva, sąmoningumas

## SUMMARY

This paper analyzes the interplay of first and third person perspectives in Integrated information theory (IIT). It is argued that there are six main interpretations of IIT that combine these perspectives differently. Paper consists of two main parts. First part analyzes methodological problems of IIT, second is intended for the classification of IIT interpretations, for the analysis of their interplay and for the arguments against integrated information and consciousness identity thesis. This work demonstrates that interpretations that refuse identity thesis are better supported by empirical research, also they resist some conceptual problems, cohere better with surrounding theories and are neutral towards some ontological problems.

**Key words:** integrated information, first person perspective, third person perspective, consciousness

## ĮVADAS

Šiuolaikinės neuromokslinės sąmoningumo teorijos susiduria su pirmojo ir trečiojo asmens perspektyvų derinimo problema. Josephas Levine'as (1983) šių perspektyvų derinimo keblumą įvardino kaip aiškinamąją spragą (*explanatory gap*), vėliau Davidas Chalmersas (1995) pristatė sunkiąją sąmoningumo problemą (*hard problem of consciousness*). Trečiojo asmens perspektyvų duomenys apibūdinami kaip duomenys apie elgseną ir smegenų procesus, o pirmojo asmens duomenys suprantami kaip duomenys apie subjektyvų patyrimą (Chalmers, 2004). Spręsdamos šių perspektyvų derinimo problemą vienos teorijos pasuka reduktyvizmo keliu, kitos stengiasi jo išvengti. Reduktyvizmo išvengimo kelią renkasi ir Giulio Tononis pristatydamas Integruotos informacijos teoriją (Tononi, 2004; Tononi, 2012; Oizumi, Albantakis ir Tononi, 2014; Albantakis et al., 2022), kurioje mėgina pateikti savitą pirmojo ir trečiojo asmens perspektyvų derinimo būdą.

Darbe analizuojama pirmojo ir trečiojo asmens perspektyvų sąveika Integruotos informacijos teorijoje (IIT). Teigiama, kad priklausomai nuo to, kaip yra traktuojamas šių perspektyvų derinimas, galima išskirti šešias IIT interpretacijas.

### Tikslai

Galima išskirti tris pagrindinius darbo tikslus. Pirmasis tikslas yra esamų argumentų, paremiančių bei kritikuojančių IIT metodologiją, apžvalga. Antrasis tikslas yra IIT interpretacijų klasifikacijos sudarymas. Trečiasis tikslas yra papildomų argumentų pateikimas, siekiant apginti vieną IIT interpretacijų klasę, įvardinamą kaip minimalioji, atmetančią fenomenologinio pirmumo priegą bei sąmoningumo ir integruotos informacijos tapatumo prielaidą.

### Metodika

Analizuojant IIT, remiamasi pirmojo ir trečiojo asmens perspektyvų skirtimi (Chalmers, 2013). Darbe remiantis fundamentaliosios ir empirinės IIT (Michel ir Lau, 2020), stipriosios ir silpnosios IIT (Mediano, et al., 2021), ortodoksinės ir neutraliosios IIT (McQueen, 2018), aspiracinės ir inspiruotos IIT (Tsuchiya ir Leung, 2023), bei fenomenologinio ir eksperimentinio pirmumo teorijų (Negro, 2022) skirtimis, pateikiama nauja IIT klasifikacija, sudaryta iš šešių skirtingų IIT interpretacijų, besiskiriančių pirmojo ir trečiojo asmens perspektyvų derinimo būdais.

Pateikiant papildomus minimaliąją IIT interpretacijų klasę paremiančius argumentus remiamasi Kontroliuojamos haliucinacijos teorija (Seth, 2021), sąmoningo ir nesąmoningo informacijos apdorojimo skirtimi (Baars, 1997; Dehaene, 2001) bei neuroplastiškumo tyrimais (Frith, 2011; Askenasy ir Lehmann, 2013; Pyza, 2013).

## **Ginami teiginiai**

Magistriniame darbe teigiama, kad yra šešios pagrindinės IIT interpretacijos: pradinė, aspiracinė, eksperimentinio pirmumo, silpnoji, empirinė ir neutralioji. Šios interpretacijos klasifikuojamos pagal pirmojo ir trečiojo asmens perspektyvų sąveikos būdus, kurie gali būti apibūdinami dviem pagrindiniais aspektais: tuo, ar IIT tapatina pirmojo asmens perspektyvų duomenis su trečiojo asmens perspektyvos duomenimis ar ne, ir tuo, kaip šis tapatumas yra gaunamas.

Kalbant apie ontologinį statusą, tvirtinama, kad pateiktoje klasifikacijoje nurodytoms IIT interpretacijoms priskirtinas nevienodas ontologinis statusas. Be to, teigiama, kad šešios pagrindinės IIT interpretacijos priima realistinę poziciją fenomenaliojo sąmoningumo atžvilgiu.

Taip pat tvirtinama, kad eksperimentinio pirmumo, aspiracinė ir pradinė IIT interpretacijos nėra suderinamos su Kontroliuojamos haliucinacijos teorija, nepaaiškina sąmoningo ir nesąmoningo informacijos apdorojimo skirties bei nedera su smegenų neuroplastiškumo savybe.

## **Temos iširtumas**

IIT pateikė ir toliau plėtoja psichiatras ir neuromokslininkas Giulio Tononis (2004). Integruotos informacijos teorija susilaukia daug dėmesio ir yra laikoma viena iš pirmaujančių bei daug žadančių neuromokslinių sąmoningumo teorijų (Koch, 2019). Kita vertus, teorija neišvengia gausios kritikos, susijusios su fenomenologiniu metodu (Bayne, 2018; McQueen, 2018; Pokropski, 2019; Moon ir Pae, 2019; Negro, 2022), ontologiniu jos statusu (Searle, 2013; Grasso, 2013; Cerullo, 2015; Mørch, 2018; Grasso, 2019 ) ir empiriniu pagrindimu (Michel ir Lau, 2020; Seth, 2021; Mediano et al., 2022).

Galima išskirti tris pozicijas IIT atžvilgiu: teorijos proponentus (Tononi, 2004; Tononi, 2008; Tononi, 2012; Oizumi et al., 2014; Albantakis et al., 2022; Koch, 2019), atmetančius teoriją (Searle, 2013; Aaronson, 2014, Carruthers, 2019) ir tuos, kurie mėgina teoriją modifikuoti atmesdami dalį prielaidų (McQueen, 2018; Michel ir Lau, 2020; Mediano et al., 2022).

## Darbo struktūra

Magistrinį darbą sudaro dvi dalys. Pirmoje dalyje pristatoma IIT ir metodologinės problemos, su kuriomis ji susiduria. Pirmame skyriuje mėginama nustatyti IIT sąsajas su fenomenologine tradicija, vertinamas IIT metodologinis naujumas ir abejojama Tononio siekio konstruoti IIT pradedant nuo pirmojo asmens perspektyvos tyrimo sėkmingumu. Antrame skyriuje klasifikuojami fenomenologinio metodo pirmumą ir fenomenologinių aksiomų būtinumą bei pakankamumą kvestionuojantys argumentai. Trečiame skyriuje parodomas ontologinių postulatų išvedimo iš aksiomų problemiškas. Tvirtinama, kad postulatai tiesiogiai neplaukia iš aksiomų, jie nėra gaunami *a priori*, o yra įvedami atskirai nuo aksiomų ir sujungiami priimant tapatumo tezės prielaidą. Ketvirtame skyriuje pristatomi IIT naudojami mintiniai eksperimentai ir jų pagalba apibūdinama integruotos informacijos samprata. Akcentuojama, kad mintiniai eksperimentai yra pateikiami anksčiau nei fenomenologinės aksiomos. Penktame skyriuje parodoma, kad IIT tapatumo tezė yra ne teorijos išvada, bet priimama prielaida, susijusi su sąmoningumo laipsniškumo prielaida.

Antroje dalyje pristatoma IIT interpretacijų klasifikacija ir jų sąveika. Antrame skyriuje pateikiama IIT interpretacijų klasifikacija pagal tai, kaip toli nutolstama nuo pradinio metodologinio sumanymo. Pirmame poskyryje apžvelgiami IIT paremiantys empiriniai tyrimai, parodoma, kad jie nepakankami norint pagrįsti pilną IIT interpretacijų klasę ir geriau paremia minimalią IIT interpretacijų klasę. Antrame poskyryje aptariamas galimas IIT interpretavimas, priskiriant jai skirtingas sąmoningumo sampratas. Pabrėžiama, kad klasifikacija pagal metodologinį planą yra priimtinesnė tuo, kad iš principo įgalina modifikuoti teoriją, o ne tik priskirti kitą tyrimo objektą. Trečiame poskyryje išryškinama, kad penkios interpretacijos pagal metodologinį planą vadovaujasi bendra prielaida – realistine pozicija fenomenaliojo sąmoningumo atžvilgiu ir viena interpretacija šios prielaidos nepriima. Ketvirtame poskyryje parodoma, kad IIT interpretacijų derėjimas su skirtingomis ontologinėmis pozicijomis priklauso nuo to, kaip derinamos pirmojo ir trečiojo asmens perspektyvos. Antrame skyriuje pateikiami papildomi argumentai, kritikuojantys integruotos informacijos ir sąmoningumo tapatumo tezę ir tuo pačiu paremiantys minimalią IIT interpretacijų klasę. Pirmame poskyryje, sprendžiant Scotto Aaronsono ir Giulio Tononio ginčą, siūloma atsižvelgti į smegenų neuroplastiškumo fenomeną, kuris paremia Aaronsono poziciją. Antrame poskyryje parodoma, kad pilnoji IIT interpretacija nepriima sąmoningo ir nesąmoningo informacijos apdorojimo skirties ir dėl šios priežasties neišvengia prieštaravimų. Trečiame poskyryje parodoma, kad dalis IIT interpretacijų nėra suderinamos su Kontroliuojamos haliucinacijos teorija, besiremiančia numatančiojo apdorojimo principu.

# 1. INTEGRUOTOS INFORMACIJOS TEORIJOS PAGRINDAI

Integruotos informacijos teoriją pristatė psichiatras ir neuromokslininkas Giulio Tononis (2004). Vėliau pasirodė trys papildomos IIT versijos (Tononi, 2012; Oizumi et al., 2014; Albantakis et al., 2022). Vienas iš pagrindinių IIT tikslų yra pateikti patikimą būdą, kuris padėtų nustatyti ar tam tikra sistema yra sąmoninga ar ne. Šis tikslas neatskiriamas nuo siekio išsiaiškinti, kiek ji yra sąmoninga arba, kitaip tariant, koks jos sąmoningumo lygis. Kitas teorijos tikslas yra apibrėžti sąlygas, apsprendžiančias kokių sąmoningumu sistema pasižymi (Tononi, 2004). Šiame darbe daugiausia dėmesio skiriama pirmosios IIT problemos argumentų analizei. Taip yra todėl, kad pastaroji laikoma fundamentalesne už antrąją, t. y. neteisingai priskyrus sąmoningumą sistemai nelieka pagrindo tvirtinti, kad ji pasižymi specifiniu kokybiniu (*qualitative*) patyrimo turiniu.

Bandyamas sukurti teoriją, kuri nustatytų, ar sistema yra sąmoninga ar ne, vadinamas „gana sunkia problema“ (*the pretty-hard problem*) (Aaronson, 2014). Nors sąmoningumo priskyrimas sveiko, budraus suaugusio žmogaus atvejui yra mažiausiai problemiškas, net šiuo atveju nėra iki galo aišku, ar šis žmogus nėra tik „filosofinis zombis“, stokojantis sąmoningo patyrimo (Chalmers, 1996). Sąmoningumo priskyrimą kitam asmeniui galima paremti argumentu pagal analogiją (*argument from analogy*) arba geriausio paaiškinimo išvedimo argumentu (*argument from best explanation*) (Koch, 2019). Susidūrus su sudėtingesniu patologiniu atveju, pvz. akinetiniu mutizmu (*akinetic mutism*), kai žmogus yra nubudęs ir atmerktomis akimis, tačiau negali nei kalbėti, nei judėti ir beveik nereaguoja, nustatyti, ar jis yra sąmoningas ar ne, pasidaro dar kebliau. Problemiškų atvejų yra ir daugiau, neaišku, kiek sąmoningas yra lunatikuojantis ar komoje esantis žmogus, kurioje prenatalinės raidos stadijoje atsiranda sąmoningumas. Sudėtinga daryti išvadas apie gyvūnų ar kitoms karalystėms priklausančių gyvūnų organizmų bei dirbtinių mechanizmų sąmoningumą. IIT mėgina rasti patikimą būdą, kuris padėtų išspręsti tokius sudėtingus atvejus. Šios sąmoningumo matuoklio paieškos yra siekis išspręsti Aaronsono suformuluotą gana sunkią problemą. IIT taip pat sietina su kitų asmenų sąmonės, gyvūnų sąmoningumo bei dirbtinio sąmoningumo problemomis.

Be mėginimo pateikti patikimą būdą nustatyti, ar sistema yra sąmoninga ar ne, IIT dar mėgina išspręsti Chalmerso (1995) pateiktą sunkiąją sąmoningumo problemą. IIT proponentų ji priimama kaip realiai egzistuojanti problema, nors su tuo sutiktų ne visi (Dennett, 2018). Pasak Giulio Tononio ir Christofa Kocho (2015, p. 5) „kol pradedama nuo smegenų ir klausama kaip jose atsiranda patyrimas (*experience*), problema tampa ne tiek sunki, kiek beveik neįmanoma išspręsti“. Todėl siūlomas kitoks problemos sprendimo būdas – pradėti ne nuo smegenų tyrimo ir pereiti prie sąmoningumo, bet



atvirkščiai, pradėti nuo paties sąmoningumo, nustatant esmines jo savybes ir tuomet išsiaiškinant kokios fizinės struktūros gali jį atitikti. Tononis (2016) tvirtina, kad tai yra naujas būdas tyrinėti sąmoningumą. Šio naujumo būtinumas randasi dėl siekio išvengti reduktyvizmo. Pasak mokslininko, pradedama nuo fenomenologijos ir tik tuomet pereinama prie smegenų tyrimo.

IIT naudojamą metodologiją, taikomą nustatant sąryšį tarp sąmoningumo ir jo fizinio substrato, galima apibūdinti penkiais esminiais žingsniais. Pirmiausia atliekant pirmojo asmens perspektyvos tyrimą, pateikiamos penkios aksiomos, kuriomis mėginama apibūdinti esmines sąmoningumo savybes. Tada iš jų išvedami penki postulatai, kuriuos turi atitikti mechanizmai (tiek biologiniai, tiek nebiologiniai), pasižymintys sąmoningumu. Trečias žingsnis yra išvados, kad sąmoningumas yra integruota informacija, pateikimas. Ketvirtas žingsnis yra matematinis postulatų formalizavimas, atliekamas tam, kad būtų galima nustatyti sistemos integruotos informacijos kiekį ir įvertinti jos sąmoningumo lygį kiekybiškai. Po šių keturių žingsnių seka mechanizmų sąmoningumo lygio testavimas taikant trečiojo asmens perspektyvos tyrimo metodus. Keturiomis žingsniais mėginama *a priori* išvesti sąmoningumo teoriją, kurią būtų galima taikyti empiriškai. Tačiau galimos kitos IIT interpretacijos, nutolstančios nuo šio plano ir kitu būdu derinančios pirmojo bei trečiojo asmens perspektyvas. Toliau pateikiami argumentai, parodantys, kad IIT neišlieka nuosekli ir šis penkiais žingsniais paradigmiškai ją nusakantis planas nėra įgyvendinamas.

## 1.1. Pirmojo asmens perspektyva

Tononio siekis konstruoti IIT pradedant iš pirmojo asmens perspektyvos kyla dėl siekio išvengti reduktyvizmo. „Atlikti redukciją bendriausia prasme reiškia parodyti, kad (ir kaip) tam tikras reiškinys ar reiškinų grupė gali būti visiškai suvesta į kitą, skirtingą nuo pirmosios“ (Dagys, Dranseika, Grenda ir Vasilionytė, 2014, p. 141). Tononis siekia išvengti sąmoningumo redukavimo į nervų sistemos procesus. Kaip pažymi Davidas Papineau (2011, p. 6), įprasta manyti, kad šiuolaikinėse materialistinėse teorijose tapatumai tarp smegenų būvių ir sąmonės būvių yra nustatomi *a posteriori*, kitaip sakant, tapatumo nustatymas reikalauja empirinių tyrinėjimų, bet ne konceptualinių apmastymų. Taip pat, jo teigimu, plačiai priimta, kad aiškinamoji spraga kyla dėl negalėjimo išvesti faktų apie sąmoningumą iš fizinių faktų. IIT proponentai su tuo sutinka ir siekia atlikti atvirkščią veiksmą – išvesti teiginius apie fizinį sąmoningumo substratą iš teiginių, gautų naudojant pirmojo asmens perspektyvą. Tvirtinama, jog teorija pradedama nuo fenomenologijos (Oizumi et al., 2014). Šis teiginys, kartu su metodologinio naujumo pareiškimu, skatina klausti (1) kokios IIT sąsajos su

fenomenologine tradicija, (2) ar IIT pateikiama metodologija iš tiesų yra nauja ir (3) ar teorija tikrai konstruojama pradedant iš pirmojo asmens perspektyvos.

Abejotina, kad Tononio atliekamas pirmojo asmens perspektyvos tyrimas sietinas su fenomenologine tradicija nepaisant to, jog yra įvardijamas kaip fenomenologinis. Pirmiausia eksplicitiškai neišreiškiamas joks IIT santykis su fenomenologine tradicija ar jos atstovais. Teorijoje nėra nuorodų į gausius fenomenologinius tyrinėjimus, neužsimenama apie Edmundo Husserlio (2005) ar vėlesnių fenomenologų darbus, tad lieka ne iki galo aišku, kokia prasme Tononis vartoja sąvoką „fenomenologija“.

Žinant, kad ši sąvoka turi ne vieną reikšmę, galima klausti, kas sieja Tononio tyrimą su Husserlio fenomenologiniais tyrinėjimais. Vienas panašumas, pirmiausiai krentantis į akis, yra tai, kad Tononis, kaip ir Husserlis, pasiremdamas Rene Descartes'o meditacijų pavyzdiumi (žr. 2.1.3), siekia nustatyti apodiktines tiesas, apibūdinančias sąmoningumą. Straipsniuose, kuriuose pristatoma IIT, galima pastebėti iš Earnsto Macho (1914) tyrinėjimų paimtus elementus. Be konceptualinių elementų, panaudojamas Macho autoportretas, skirtas iliustruoti idėjoms apie vizualinio sąmoningumo savybes. Ne kiekvienas laikytų Machą fenomenologu, bent jau fenomenologinės tradicijos prasme, paprastai jis laikomas fenomenalistu. Tokio požiūriu prisilaikoma ir šiame darbe. Nors tiek Machas, tiek Husserlis laikėsi bendro įsitikinimo, kad tik tiesioginis patyrimas yra neabejotinas, bene svarbiausias skirtumas tarp Macho ir Husserlio metodų yra pastarojo atliekamas metafizinių prielaidų suskliautimas. Husserlis atskiria fenomenologiją nuo ontologijos, atlikdamas natūraliosios nuostatos apie pasaulį suskliautimą (*epoche*) ir tyrinėdamas tai, kas reiškiasi patirtyje (Mickūnas ir Jonkus, 2014). Tuo tarpu Machui, nors ir pradedančiam iš pirmojo asmens perspektyvos, rūpėjo išvesti „realaus pasaulio“ charakteristikas (McGinn, 1972). Šiuo klausimu Tononis artimesnis Machui nei Husserliui, nes jam rūpi išspręsti ontologinius klausimus, susijusius su sąmoningumu. IIT teorijoje suskliautimas arba nėra atliekamas visai, arba mėginama išvesti natūraliąją nuostata grįstus teiginius iš patyriminių teiginių. Bet kuriuo atveju IIT pagrindinis siekis yra susijęs su ontologiniais klausimais.

Marekas Pokropskis (2019) išskiria du būdus kaip gali būti suprasta fenomenologija. Pirmasis siejamas su Husserlio iniciuota ir po jo sekusia fenomenologine tradicija. Pokropskis pabrėžia, kad vienas iš Husserlio fenomenologijos siekių buvo atitrūkimas nuo paprastų patyrimo tyrinėjimų introspekciškai, pateikiant sudėtingą sąmoningumo teoriją ir jos tyrinėjimo metodą. Antru atveju fenomenologija gali būti suprasta bendresne prasme ir siejama su patyriminiais sąmoningumo aspektais, kurie apibūdinami Thomaso Nagelio fraze „ką reiškia būti“. Nurodoma, kad šia prasme fenomenologija suprantama kaip introspektyvus patyriminių sąmoningumo savybių tyrinėjimas, kuris

susiduria su sunkiąja sąmoningumo problema, kai siekiama jį derinti su natūralistiniais tyrimais. Priėmus tokį fenomenologijos tradicine ir bendresne prasme atskyrimą, IIT pakliūna į antrąją kategoriją.

IIT proponentai (Oizumi et al., 2014, p. 1) eksplicitiškai nurodo, kad sąvokos „sąmoningumas“, „patyrimas“ ir „fenomenologija“ vartojamos kaip sinonimai. Taip pat, naujausioje IIT versijoje sąvoka „introspekcija“ vartojama kaip sinonimas sąvokai „fenomenologija“ (Albantakis et al., 2022). Toks sąvokų niveliavimas leidžia manyti, kad IIT neseka fenomenologine tradicija. Nors literatūroje galima rasti skirtingus požiūrius į introspekcijos ir fenomenologijos santykį, kad ir kokios būtų artimos, šios sąvokos nevertotinos kaip sinonimai. Griežtesniu atveju, tai yra visai skirtingi dalykai (Thomasson, 2003), tačiau yra tvirtinančių, kad fenomenologija yra siekis patobulinti introspekciją, padarant ją moksliska ir sistemiška (Gutland, 2018). Taigi, galima nurodyti, kad IIT vartoja sąvoką „fenomenologija“ platesne prasme, kaip tai apibūdinta anksčiau, arba norint išvengti nesusipratimo, kalbant apie IIT galima vartoti skėtinę sąvoką „pirmojo asmens perspektyvos tyrimas“, apimančią tiek fenomenologiją, tiek introspekciją, tiek kitus metodus (Lumma ir Veger, 2021). Toliau darbe analizuojant IIT bus vartojami abu variantai: tiek išraiška „pirmojo asmens perspektyvos tyrimas“, tiek sąvoka „fenomenologija“, suprantant ją bendresne prasme nei išreiškiama fenomenologinėje tradicijoje.

Chalmersas (2004) išskiria du būdus, kaip galima gauti duomenis iš pirmojo asmens perspektyvos – tyrinėti savo paties patyrimą arba rinkti kitų asmenų subjektyvius žodinius pareiškimus. Tononis, pateikdamas patyrimo analizę, naudoja pirmąjį būdą, t. y. jis analizuoja savo patyrimą iš pirmojo asmens perspektyvos. Šis analizė taikoma tik kuriant teoriją. Tai reiškia, ji pritaikoma kartą, apibūdinant sąmoningumo struktūrą, ir tuomet, pateikus teoriją, naudojami natūralistiniai metodai, mėginant tokią struktūrą identifikuoti įvairiuose potencialiai ją turinčiuose mechanizmuose. Tai, kas pasiekama naudojant vadinamąjį fenomenologinį metodą, yra veikiau apibrėžimas to, ką mėginama vėliau testuoti teorijos pagalba. IIT proponentai su šiuo teiginiu nesutiktų, tačiau, kaip parodoma toliau (žr. 2.1.2), nėra pagrindo manyti, kad fenomenologinis metodas, kaip jį naudoja IIT, pasitarnauja daugiau nei operacionalizuotas fenomenaliojo sąmoningumo apibrėžimas.

IIT mėginama derinti pirmojo ir trečiojo asmens perspektyvų tyrimo metodus. Galima pastebėti, kad mėginimas derinti fenomenologinius metodus kartu su natūralistiniais nėra naujas (Varela, 1999; Zahavi, 2005; Thompson, 2007). Kognityviniuose moksluose ši problema virsta fenomenologijos natūralizavimo projektu (Rimkevičius ir Gutauskas, 2014). Tačiau IIT fenomenologinių ir natūralistinių metodų derinimas interpretuotinas bent dviem būdais. Vienu atveju galima suprasti IIT kaip

pradedančią nuo fenomenologinės analizės, tam tikru būdu darančią išvadas apie fizinius mechanizmus, pasižyminčius sąmoningumu ir tik tuomet naudojančią natūralistinius metodus šiems mechanizmams įvertinti. Toks metodas, jei būtų realizuotas, skirtųsi nuo įprastai naudojamų kognityviniuose moksluose. Pagal kitą interpretaciją IIT tam tikru būdu derina pirmojo asmens duomenis su trečiojo asmens duomenimis, bet neišveda teiginių, nusakančių trečiojo asmens perspektyvą iš pirmų ir *vice versa*.

Niccolò Negro (2022) atskiria eksperimentinio pirmumo (*experiment-driven*) teorijas nuo fenomenologinio pirmumo (*phenomenology-first*) teorijų. Tvirtina, kad į antrąją kategoriją patenka IIT. Šis tvirtinimas atitinka įprastą teorijos interpretaciją, pagal kurią iš aksiomų *a priori* išvedami postulatai. Kadangi IIT proponentai tik tvirtina, kad atlieka tokią procedūrą, bet išvedimo nepagrindžia, IIT nėra fenomenologinio pirmumo teorija, o tik pretenduoja tokia būti.

Kraštutine požiūrių į pirmojo ir trečiojo asmens perspektyvų duomenų derinimą forma laikytinas eliminatyvizmas, atmetantis pirmojo asmens duomenų paaiškinimo reikalingumą. Įprastai teorijos derina pirmojo asmens perspektyvos tyrimo metodus su trečiojo asmens perspektyvos tyrimo metodais arba linksta į redukcionizmą, kuomet mėginama pirmojo asmens perspektyvą paaiškinti naudojantis vien trečiojo asmens perspektyvos teiginiais. Kai teorijos naudoja abiejų tipų metodus, juos derina paraleliai, dėl to jų naudojimo pirmumo klausimas lieka neaktualus. Tokio tipo metodų derinimo pavyzdys yra neuroninių sąmoningumo koreliatų (*neural correlates of consciousness*) paieškos. Šiuo atveju tarp skirtingais metodais gaunamų duomenų nustatomi koreliaciniai santykiai, bet nemėginama iš vieno tipo duomenų išvesti teiginių apie kito tipo duomenis. Fenomenologinio pirmumo teorija galėtų būti suprantama kaip tokia, kuri naudoja fenomenologinę analizę ir *a priori* daro išvadas apie fizinį sąmoningumo substratą. Nepriklausomai nuo to, sėkmingas ar ne, tokio tipo teorijos konstravimo pavyzdys galėtų būti Macho fenomenalizmas. McGinnas (1972, p. 156) apibendrina Macho atliekamą fenomenalistinę redukciją kaip tokią, pagal kurią mes nežinome nieko, tik savo pojūčių duomenis ir visi teiginiai apie numanomai realaus pasaulio objektus turėtų būti pilnai išverčiami iš pojūčiais gautų duomenų. IIT pretenduoja atlikti panašų išvedimą. Tačiau teorijoje implicitiškai glūdi prielaidos, gautos ne iš pirmojo, bet trečiojo asmens perspektyvos.

Eksperimentinio pirmumo ir fenomenologinio pirmumo skirtis gali būti naudinga pateikiant IIT interpretacijas. IIT gali patekti tiek į vieną, tiek į kitą kategoriją. IIT gali būti traktuojama kaip fenomenologinio pirmumo teorija, jei standartinė IIT interpretacija pasirodytų esanti teisinga. Negro pateiktame argumente priimamas IIT proponentų tvirtinimas dėl fenomenologinio pirmumo, neatsižvelgiama į tai, kad galima nevienareikšmė sąsaja tarp IIT pateikiamų fenomenologinių aksiomų

(žr. 1.2), ontologinių postulatų (žr. 1.3) bei mintinių eksperimentų (žr. 1.4) ir todėl IIT tapatumo tezė (žr. 1.5) laikoma gaunama išvada, bet ne priimama prielaida. Tačiau parodžius, kad IIT konstruojama naudojant prielaidas, gautas ne iš pirmojo asmens perspektyvos, galima pateikti tokią interpretaciją, pagal kurią IIT yra eksperimentinio pirmumo teorija.

Pirmasis argumentas galėtų būti siejamas su IIT teorijos geneze. Vien žvelgiant chronologiškai į pateikiamas IIT versijas galima pastebėti, kad pirmojoje versijoje pradedama ne nuo fenomenologinių tyrinėjimų ir fenomenologinių aksiomų pateikimo, o nuo mintinių eksperimentų, paremtų empiriniais duomenimis. IIT proponentai iki šiol pateikė keturias teorijos versijas ir fenomenologinės aksiomos buvo įvestos tik antrojoje versijoje, pirmojoje nieko apie jas neužsimenama. Toks argumentas parodo, kad IIT iš esmės buvo sukonstruota ne kaip fenomenologinio pirmumo, bet veikia kaip pirmojo ir trečiojo asmens perspektyvų tyrimo metodus derinanti teorija.

Kelvinas J. McQueenas (2018) taip pat abejoja, kad pagrindiniai IIT įrodymai yra gryni fenomenologiniai, o empiriniai yra tik antriniai. Pagal jį, pirminiai sąmoningumo kompleksškumo matavimo įrodymai yra gaunami iš smegenėlių bei smegenų žievės palyginimo. Fenomenologinės aksiomos atlieka antrinį vaidmenį, padedant tobulinti kompleksškumo matavimus ir grindžiant, kodėl konkrečiai turi būti matuojama integruota informacija. McQueenas neaptaria IIT pateikiamų mintinių eksperimentų, kurie gali sustiprinti IIT eksperimentinio pirmumo tezė. Pirmojoje IIT versijoje pateikiamas mintinis eksperimentas akivaizdžiai yra pagrįstas smegenėlių struktūros modeliu. Visai kaip smegenėlės sudarytos iš atskirų, tarpusavyje neintegruotų modulių, fotokameros sensorius yra sudarytas iš tarpusavyje atskirtų sesnsorių (žr. 1.4). Šis mintinis eksperimentas yra paremtas eksperimentiniais duomenimis, gautais tyrinėjant pacientus su pažeistomis smegenėlėmis (žr. 2.1.1). Fotokameros mintinio eksperimento naudojimas paremia McQueeno abejonę dėl IIT fenomenologinio pirmumo statuso.

Dar vienas argumentas, paremiantis tvirtinimą, kad IIT nėra fenomenologinio pirmumo teorija yra tai, kad IIT pagrindą, kaip pabrėžia Susana Blackmoore (2018), sudaro Dinaminės šerdies hipotezė (*dynamic core hypothesis*), pagal kurią neuronų grupių aktyvumas siejamas su sąmoningumu, jei jis yra Dinaminės šerdies dalis, kuri apibūdinama kaip „nuolat besikeičianti ir tuo pačiu stipriai integruota funkcinė grupė, apimanti labai didelius skaičius neuronų, plačiai pasisikirsčiusių smegenų žievėje ir gumbure, su stipriomis tarpusavio sąveikomis“ (p. 295). Galima išskirti du Dinaminės šerdies hipotezė (Tononi, Edelman, 1998) ir IIT vienijančius aspektus – pirma, abi tvirtina, kad nerviniai procesai, susiję su sąmoningu patyrimu, yra labai integruoti ir antra, šie procesai yra labai diferencijuoti. Šie nerviniai procesai siejami su sąmoningumo savybėmis, pagal kurias integruotumas reiškia, kad

kiekviena sąmoninga scena yra vientisa (*unified*), o diferencijavimas nurodo į tai, kad kaskart patiriamas vienas iš daugelio galimų sąmoningų patyrimų. Jau pateikiant Dinaminės šerdies hipotezę naudojamas fotodiodo mintinis eksperimentas, padedantis paaiškinti kas yra diferencijavimas. Fotodiodo mintinis eksperimentas taip pat pasitelkiamas pristatant pirmąją IIT versiją. Tačiau Dinaminės šerdies hipotezė nesieja sąmoningumo su mažu integruotos informacijos kiekiu ir fotodiodui nepriskiria minimalaus sąmoningumo kaip tai daro IIT. Todėl esminiu skirtumu tarp Dinaminės šerdies hipotezės ir IIT laikytina pastarosios priimama sąmoningumo ir integruotos informacijos tapatumo tezė.

Kaip parodo aukščiau pateikti argumentai, tvirtinimas dėl IIT fenomenologinio pirmumo bei metodologinio naujumo yra kvestionuotinas. Be to, atsižvelgiant į Papineau pastabą, kyla abejonė ir dėl IIT reduktyvizmo išvengimo sėkmingumo. Toliau pateikiami papildomi argumentai, paremiantys šias abejones.

## 1.2 Fenomenologinių aksiomų apodiktiškumo problema

Konstruojant IIT naudojama prieiga, apibūdinama kaip atvirkščia naudojamai eksperimentinio pirmumo teorijose. Vietoje to, kad pradėtų tirti sąmoningumą nuo smegenų, IIT proponentai tvirtina pradedantys nuo fenomenologinių aksiomų aprašymo (Tononi, 2012; Oizumi et al., 2014; Tononi ir Koch, 2015; Tononi et al., 2016; Massimini ir Tononi, 2018; Albantakis et al., 2022). Timas Bayne'as (2018) akcentuoja, kad aksiominiai metodai įprastai naudojami matematikoje ir logikoje, tad neįprasta juos aptikti gyvybės moksluose. IIT aprašomos sąmoningumo savybės, pavadintos aksiomomis, nes proponentų laikomos esminėmis (*essential*). Šis esmiškumas apibūdinamas kaip tiesioginis (*immediate*), neabejotinas (*indubitable*) ir teisingas (*true*) kiekvienam pamanomam patyrimui (Tononi et al. 2022). Kitur dar pabrėžiama, kad aksiomos yra tiesiog akivaizdžios (*immediately evident*) (Oizumi et al., 2014). Tononis aksiomomis siekia pateikti apodiktiškus teiginius, tačiau šis siekis yra problemiškas.

Pokropskis (2019) akcentuoja, kad IIT neišdėsto išsamios fenomenologinės teorijos, kuri pagrįstų kodėl pasirenkamos būtent šios aksiomos. Vietoje to, IIT tiesiog pateikia penkias aksiomas (Tononi, 2012). Pirmoji pavadinama egzistencijos (*existence*) aksioma (p. 296). Sąmoningumas yra patyrimas, kuris yra realus jį patiriančiam žmogui. Patiriami dalykai yra vienintelis faktas, kuris pasak Tononio yra neabejotinas. Descartes'o "maštau, vadinasi, egzistuoju" perfrazuojamas į "turiu patyrimą, vadinasi, egzistuoju". Antroji yra struktūros (*composition*) aksioma. Pagal pastarąją sąmoningumas yra struktūruotas – kiekvienas patyrimas turi vidinę struktūrą (p. 296). Trečioji yra informacijos aksioma.

Pastarąją Tononis pavadina centrine aksioma, pagal ją sąmoningumas yra specifinis: kiekvienas patyrimas tam tikru būdu skiriasi nuo kitų alternatyvių patyrimo (p. 297). Informacijos aksiomą padeda suprasti ankstesnėse IIT versijose pateiktas fotodiodo mintinis eksperimentas. Ketvirtoji vadinama integracijos (*integration*) aksioma. Pasak jos sąmoningumas yra vientisas, kitaip tariant, patyrimas neredukuojamas į atskirus komponentus – patiriamas pilnas vizualinis laukas, bet ne kairysis atskirai nuo dešiniojo, ar spalva atskirai nuo formos (p. 297). Integracijos aksioma aiškinama pasitelkiant fotokameros mintinį eksperimentą (žr. 1.4). Penktoji yra išskyrimo (*exclusion*) aksioma. Pagal ją kiekvienas patyrimas yra išskiriamas iš kitų ir vienu metu yra tik vienas patyrimas, turintis aiškias ribas, jis veikia tam tikru greičiu, jį sudaro tam tikras turinys (p. 297).

Pavienės aksiomos susilaukia kritikos. Pagal pirmąją aksiomą, sąmoningumo egzistavimas laikomas neginčijamu. Ši aksioma nėra suderinama su eliminatyvistine (Churchland, 1981) ir iliuzionistine (Frankish, 2016) pozicijomis (žr. 2.1.3). IIT priima, kad net fotodiodas yra minimaliai sąmoningas, todėl Bayne'as (2018) suabejoja struktūros aksioma, klausdamas kodėl patyrimas negalėtų būti nestruktūruotas tokiuose primityviuose mechanizmuose.

Įprastai traktuojama, kad fenomenalusis sąmoningumas apima pojūčių (*sensations*), jausmų ir suvokimų (*perceptions*) patirtines (*experiential*) savybes, be to yra tvirtinančių, kad jis apima dar ir mintis, norus bei emocijas (Block, 1995). Aprašant integracijos aksiomą, konkrečiai apibūdinamas vizualinis sąmoningumas, t. y. Tononis kalba apie vientisą vizualinį lauką, tačiau nepabrėžiama, kad vyksta integracija tarp skirtingų sensorinių modalumų, jausmų ir suvokimų, todėl šią aksiomą būtų galima išplėsti. Iš pirmojo asmens perspektyvos patiriamas vientisas visų juslių duomenis apjungiantis sąmoningumas, t. y. vienu metu matomas vaizdas, girdimas garsas, užuodžiamas kvapas, *etc.* Dėmesys gali krypti į vieną ar kitą pojūtį, tačiau jie sąmoningai patiriami kaip integruota visuma.

Penktoji aksioma kritikuojama Pokropskio (2018), kuris išskiria pavyzdžių, prieštaraujančių šiai aksiomai, kuomet patyrimas yra neryškus ar su persidengiančiomis ribomis, kaip pvz. periferinis matymas, apverčiamų figūrų ar daugiaveikos (*multitasking*) suvokimas. Jo teigimu, sąmoningas patyrimas neatrodo turintis aiškias, apibrėžtas, vientisas ribas, veikiau jis yra dinamiškas ir persiklojantis (*interrelated*) procesas. Kita problema, kurią įžvelgia Pokropskis, yra penktosios ir antrosios aksiomų nesuderinamumas. Kompozicijos aksioma, pagal kurią viename patyrimo galime turėti įvairų konceptualinį turinį, kaip pvz. iš kairės į dešinę judantį skritulį, kuris yra pusiau žalias, pusiau raudonas, tarsi leidžia turėti skirtingus patyrimus viename patyrimo, nors išskyrimo aksioma apibūdina sąmoningumą kaip vieną patyrimą su aiškiomis ribomis.

Aksiomų teisingumo priskyrimas kiekvienam įmanomam patyrimui, galbūt, yra teisingas kalbant apie įprastą atvejį – sveiką, suaugusį žmogų. Nors solipsistiškai nusiteikęs kritikas galėtų tvirtinti, kad tai yra nepagrįstas tvirtinimas, nes ką nors žinoti galima tik apie savo paties patyrimą, vis tik tokį teiginį galima pagrįsti panašumu (Russell 1948). Kalbant apie patologinius atvejus, gyvūnus ar nebiologinius mechanizmus, aksiomų teisingumas nėra akivaizdus. Kitokią smegenų struktūrą turintys gyvūnai gali pasižymėti kitokiomis patyrimo savybėmis. Toks pavyzdys yra balandžiai, kurių akys aprėpia atskirus regėjimo laukus, tačiau yra sujungtos tik su viena smegenų puse ir signalai gaunami iš akių nėra integruojami taip, kaip žmogaus atveju (Godfrey-Smith, 2021). Nėra aišku, ką reiškia būti balandžiu, analogiškai kaip ir nėra galimybės patirti tai, ką patiria šikšnosparnis orientuodamasis echolokacijos pagalba (Nagel, 1977). Todėl aksiomų teisingumas visais įmanomais atvejais atrodo abejotinas ir reikalauja įrodymo, bet IIT proponentai jo nepateikia, laikydami aksiomas teisingomis savaime.

Be šių aksiomų, kurias išskiria IIT, galima pateikti papildomas aksiomas. Pokropskis (2018) sąrašą papildė tokiomis: „subjektyvumo aksioma“, pagal kurią kiekvienas sąmoningas patyrimas yra subjektyvus, „intencionalumo aksioma“, pagal kurią sąmoningumas yra apie ką nors, „laikiškumo (*temporality*) aksioma“, kuri nurodo, kad sąmoningumas yra laikiškas procesas, ar „kvalia aksioma“, kuri sieja sąmoningumą su specifinėmis kokybinėmis patyrimo būsenomis, pvz. yra kažkas, ką reiškia patirti X. Tai, kad galima papildyti IIT pateikiamą aksiomų sąrašą, verčia abejoti pateiktų aksiomų pakankamumu.

Pagal Bayne'ą (2018) aksiomos gali būti interpretuojamos labai skirtingai. Interpretacijų įvairovė verčia abejoti jų akivaizdumu. Čia problema gali slypėti mėginime žodžiais išreikšti patyrimo savybes. Verbalinė išraiška gali būti netikslus patyrimo ar jo savybių apibūdinimas. Dėl to aksiomos tik apytiksliai nusako tai, ką siekiama jomis apibūdinti.

Jei fenomenologinės aksiomos sudaro teorijos pagrindą, bet, kaip parodyta, nėra pakankamos ir teisingos visais įmanomais atvejais, o IIT tapatumo tezė yra gaunama iš jų, tuomet IIT tapatumo tezės teisingumas yra abejotinas. Tačiau tapatumo tezei gauti reikalingas dar vienas žingsnis, ontologinių postulatų išvedimas, t. y. tam, kad galėtų tapatinti aprašytas sąmoningumo savybes su integruota informacija, IIT pirmiau turėtų parodyti kaip iš aksiomų yra išvedamos fizikinius mechanizmus apibūdinančios savybės, pavadintos postulatais.



### 1.3 Ontologinių postulatų išvedimo keblumas

Analogiškai aksiomoms, pateikiami penki postulatai, kurie apibrėžia kokios sąlygos turi būti patenkintos, kad fizinė sistema, kaip pvz. neuronai ir jų jungtys, generuotų sąmoningumą (Oizum et al., 2014). Postulatai, kitaip nei aksiomos, apibūdina fizines tikrovės savybes, tačiau tvirtinama, kad jie yra išvedami iš aksiomų. Nepaisant šio tvirtinimo, išvedimo procedūra pateiktoje teorijoje nėra pateikiama.

Postulatai apibrėžiami naudojant priežasties ir padarinio terminus, kurie nėra naudojami pateikiant aksiomas. Šiais postulatais nusakoma vidinė informacija suprantama kaip „skirtumai, darantys skirtumą“ (*differences that make a difference*) sistemos viduje (Oizumi et al., 2014, p. 3). Tononis šią frazę pasiskolina iš Gregory Batesono (1979). Pirmojoje IIT versijoje postulatai nepateikiami, kitose versijose pateiktų postulatų formuluotės kiek skiriasi. Darbe nesiekama detaliai išanalizuoti kiekvieno postulato ir skirtingų jų formuluočių pateiktų visose IIT versijose, o tik juos pristatyti ir pademonstruoti, kad ryšys tarp jų ir aksiomų yra keblus.

Tononio (2012) pateikti penkis postulatai pavadinti taip pat, kaip ir aksiomos. Pagal vidinės egzistencijos postulatą (1) „mechanizmai būsenoje egzistuoja (*in a state exist*)“ (p. 297). Mechanizmas suprantamas kaip tai, kas turi priežastinę funkciją sistemoje, pvz. neuronas smegenyse arba loginis elementas (*logic gate*) kompiuteryje. Sistema yra tų mechanizmų rinkinys. Vidinė egzistencija yra priešinama išorinei egzistencijai, pastarąją apibūdinat kaip priežastiniais ryšiais susijusių dalykų sąveiką (Koch, 2019). Tam, kad kažkas egzistuotų, jis turi daryti poveikį kažkam kitam, arba būti veikiamas, kitu atveju jis neturi priežastinės galios. Tam, kad sistema egzistuotų sau, ji turi daryti poveikį sau pačiai, kitaip tariant, dabartinė jos būseną turi būti paveikta jos praeities būsenos ir gebėti veikti ateities būseną. Struktūros postulatą (2) teigia, kad elementarūs „mechanizmai gali būti struktūruoti, sujungti į sudėtingesnius įvairiomis kombinacijomis“ (p. 297). Pagal informacijos postulatą (3) „iš vidinės sistemos perspektyvos, mechanizmas būsenoje generuoja informaciją tik tuomet jei jis turi ir atsirenkamas (*selective*) priežastis ir atsirenkamus padarinius sistemoje – tai yra, mechanizmas turi daryti skirtumus, darančius skirtumą sistemos viduje“ (p. 297). Pagal integracijos postulatą (4) „mechanizmas būsenoje generuoja integruotą informaciją tik tuomet, jei jis negali būti padalintas į nepriklausomus submechanizmus“ (p. 297). Pagal išskyrimo postulatą (5) „mechanizmas būsenoje generuoja integruotą informaciją tik apie vieną priežastį ir pasekmių rinkinį – tą, kuris yra maksimaliai neredukuojamas“ (p. 297). Mechanizmas gali turėti tik vieną priežastį ir vieną padarinį.

Anksčiau aptartas fenomenologinio pirmumo abejotinumą verčia susimąstyti apie IIT pateikiamų ontologinių postulatų ir fenomenologinių aksiomų sąsają. Galimi bent trys variantai, aiškinantys kaip šie postulatai siejasi su aksiomomis, – jie išvedami iš aksiomų, aksiomos išvedamos iš

jų, tai atskiri tvirtinimai, kurių tapatumui priimama papildoma prielaida. Kalbant apie aksiomų ir postulatų sąsają, aprašomą IIT, vartojami skirtingi terminai. Vienur tvirtinama, kad postulatai yra „išvedami“ (*inference*) iš aksiomų (Tononi, 2016), kitur teigiama, kad jie yra „išverčiami“ (*translate*) (Tononi, 2022), tačiau nei išvedimo, nei išvertimo procedūros nėra aptariamoms. Lieka neaišku, kokių būdu tai atliekama ir ką reiškia išvesti ar išversti postulatą iš aksiomų. IIT būtų fenomenologinio pirmumo teorija, jei iš fenomenologinių aksiomų *a priori* išvestų ontologinius postulatą. Toliau pateikiami argumentai, parodantys, kad taip nėra.

Visų pirma postulatai iš aksiomų nėra išvedami deduktyviai, nes iš vienos aksiomos išvedamas tik vienas postulatą. Tuo tarpu deduktyviame samprotavime norint gauti išvadą reikia bent jau dviejų prielaidų ar, šiuo atveju, aksiomų. Kitas argumentas, kodėl tai nėra deduktyvus išvedimas, yra tai, kad „postulatai yra informatyvesni už aksiomas“ (Negro, 2022). Postulatai apie fizinę tikrovę pasako tai, kas neslypi pateiktose aksiomose.

Pagal Kochą (2019, p. 75) postulatų išvedimui IIT naudoja abdukciją arba geriausio paaiškinimo išvedimą (*inference to the best explanation*). Tokio išvedimo procedūros Kochas nepateikia, tad lieka neaišku, kaip konkrečiai šis samprotavimo būdas yra pritaikomas. Be to, net jei postulatams gauti IIT naudoja abduktvų samprotavimo būdą, postulatai nebūtinai yra teisingi, nes abduktviviai gaunamos išvados yra tik tikėtinos (Mackonis 2009).

Negro (2022), priimdamas prielaidą, kad IIT pradeda nuo fenomenologinių aksiomų ir naudojantis geriausio paaiškinimo išvedimu gauna postulatą, pastebi, kad IIT susiduria su savęs pagrindimo problema (*self-evidencing problem*). Pagal jį postulatų išvedimas iš aksiomų yra loginė operacija, tačiau mes neturime instrumentų, kuriais galėtume patikrinti, ar postulatų išvedimas iš aksiomų yra sėkmingas, ir dėl to negalime nieko pasakyti apie IIT tapatumo tezės pagrįstumą. Iš pastarojo teiginio galima matyti, kad Negro, kaip ir IIT proponentai, tapatumo tezę laiko teorijos išvada, bet ne prielaida, tik abejoja šios išvados patikimumu.

Bayne'as (2018) išreiškia abejonę, kad abduktvivaus samprotavimo taikymo interpretacija atitinka tai, ką skelbia IIT proponentai pristatydami teoriją. Pagal jį IIT išreiškia savo metodą kaip dedukcinį, nes vartoja terminus „formalizavimas“ (*formalization*), „išvertimas“ (*translation*), „išvedimas“ (*derivation*), netinkamus išreikšti abduktvivaus samprotavimui. Be to, jei IIT naudoja abduktvų samprotavimą, jo manymu, lieka neaišku, kodėl neapsvarsto kitų galimų abduktvivaus samprotavimo išvadų ar įrodymų, patvirtinančių, kad išvesti postulatai yra geriausi.

Galima pamanyti, kad IIT naudoja ne vienos rūšies samprotavimo būdą visiems postulatams išvesti, bet skirtingiems postulatams naudojamos skirtingos strategijos. Moonas ir Pae (2019) detalai

analizuodami išskyrimo aksiomos ir išskyrimo postulato sąryšį prieina išvadą, kad jis yra neaiškus. Tačiau tokios problemos nemato su informacijos ir integracijos postulatų išvedimu, tvirtindami, kad juos galima tikėtina (*plausibly*) ar bent jau intuityviai išvesti iš aksiomų. Moonas ir Pae nelaiko tokių išvadų nei matematiniu įrodymu, nei deduktyviu išvedimu, o tik tikėtinomis arba intuityviomis išvadomis. Intuityvus išvedimas gali būti sukritikuotas dėl to, kad skirtingi žmonės naudodamiesi intuicija galimai išvestų skirtingus postulatus.

Galimas dar kitoks argumentas, pagal kurį iš aksiomų išvedant postulatus teorijoje implicitiškai glūdi papildomos prielaidos. Tokią mintį išreiškia McQueenas (2018). Jis pripažįsta, kad iš egzistencijos aksiomos galima išvesti egzistencijos postulatą priėmus papildomą prielaidą. Ši prielaida yra teiginys, kad jei egzistuoja sąmoningumas, jis turi fizinį substratą. Analogiška situacija yra ir su struktūros aksioma ir postulatu. Tačiau tvirtinama, kad informacijos postulatą nėra išvedamas iš informacijos aksiomos. Taip yra todėl, kad galima suformuluoti kitokį postulatą, suderinamą su informacijos aksioma. Jei iš vienos aksiomos galima išvesti skirtingus postulatus, tuomet galimos įvairios teorijos variacijos, kurios tarpusavyje nėra suderinamos ir galimai pateikia skirtingus rezultatus.

Kyla klausimas, ar apskritai įmanoma pirmojo asmens perspektyvos teiginius, koku nors būdu, išversti į trečiojo asmens perspektyvos teiginius. Pagal Chalmersą (2013, p. 26) „pirmojo asmens duomenys nėra redukuojami į trečiojo asmens duomenis ir *vice versa*“. Nors ne visi sutiktų, kad pirmojo asmens duomenų negalima paaiškinti trečiojo asmens duomenimis, vargu ar kas tvirtintų, kad iš trečiojo asmens duomenis išreiškiančių teiginių galima išvesti teiginius apie pirmojo asmens perspektyvą. Jei postulatai nėra išvedami iš aksiomų, tuomet reikia išsiaiškinti koku būdu jie yra gaunami.

Dar vienas galimas požiūris yra toks, pagal kurį postulatai apskritai nėra išvedami iš aksiomų. Bayne'as (2018) užsimena apie galimybę, kad aksiomos atlieka tik euristinį (*heuristic*) vaidmenį. Tokiu atveju nėra jokio įrodomojo ar pagrindžiančio vaidmens tarp postulatų ir aksiomų. Tokia interpretacija nėra detaliau eksplikuojama, nes manoma, kad ji yra mažiau tikėtina nei tvirtinimas dėl deduktyvaus ar abduktyvaus išvedimo taikymo. Panašiai svarsto ir Pokropskis (2019), pagal jį aksiomos nėra savaimė teisingos ir nėra pateikiamos prieš teoriją, o tik sistematizuoja jau esančią teoriją. Pratešiant šią mintį galima pabrėžti prieš tai minėtą IIT rėmimąsi Dinaminės šerdies hipoteze. Joje, dar iki IIT atsiradimo, priimtos dvi prielaidos, sutampančios su vėliau pateikiamais IIT postulatais. Todėl, priešingai nei teigia IIT proponentai, galima tvirtinti, kad bent du postulatai yra ne išvedami, o įvedami dar prieš aksiomų pateikimą ir yra remiami anksčiau buvusios teorijos. Tokiu atveju aksiomos yra pateikiamos po to ir

tiesiog sutapatinamos su postulatais. Pagal tokią interpretaciją IIT neatlieka jokios postulatų išvedimo iš aksiomų procedūros. Taigi, galima tvirtinti, kad aksiomos ir postulatai yra nepriklausomi vienas nuo kito teiginiai, sujungiami priimant jų tapatumo prielaidą.

#### 1.4 Mintinių eksperimentų pirmumas ir informacijos samprata

Be aksiomų ir postulatų IIT pateikia mintinius eksperimentus. Pirmiausia galima pabrėžti, kad IIT pateikiami mintiniai eksperimentai kompromituoja aksiomų akivaizdumą. Pagal IIT (Oizumi et al., 2014) pateiktos sąmoningumo aksiomos yra tiesiog akivaizdžios (*immediately evident*). Kaip pabrėžia Pokropskis (2018), ankstesnėje IIT versijoje eksplicitiškai pareikšta, kad informacijos integravimo savybė gali neatrodyti savaime akivaizdi (*self-evident*), dėl to kyla būtinybė pateikti mintinius eksperimentus, aiškinančius kas yra informatyvumas (*informativeness*) ar vienumas (*unity*). Kaip minėta anksčiau, mintiniai eksperimentai pateikiami dar iki fenomenologinių svarstymų ir aksiomų pateikimo, o tai nesuderinama su tvirtinimu, jog teorija pradedama nuo fenomenologijos. Labiau teoriją atitinkantis tvirtinimas būtų toks –pradedama nuo skirtingų mechanizmų, kurių fizinės struktūros yra gerai žinomos, palyginimo ir tik tuomet apsvarstomos potencialios šių mechanizmų fenomenologinės savybės.

Pirmame eksperimente mėginama paaiškinti sąmoningumą pasitelkiant informacijos sąvoką. Jame aprašomas fotodiodas – paprastas šviesai jautrus įtaisas. Siūloma įsivaizduoti, jog stovima priešais tuščią ekraną, kuris gali būti įjungtas arba išjungtas. Kai žmogus žiūri į tokį ekraną ir atskiria dvi būsenas „įjungta“ ir „išjungta“, jis turi subjektyvų šviesos ir tamsos patyrimą. Fotodiodas taip pat gali skirti įjungtą ir išjungtą ekraną, tačiau tai daro neturėdamas subjektyvaus šviesos ir tamsos patyrimo. Pagal IIT skirtumas tarp fotodiodo ir žmogaus yra skirtumas tarp to, kiek informacijos yra sugeneruojama atskiriant minėtas dvi būsenas.

Informacijos kiekis nusakomas informacijos talpos vienetais – bitais. Informacija dažnai apibrėžiama kaip neapibrėžtumo sumažinimas (*reduction of uncertainty*), t. y. kuo didesnis atmetų alternatyvų kiekis, tuo didesnis neapibrėžtumo sumažinimas ir kartu didesnis informacijos kiekis. Paprastai informacijos kiekis nustatomas naudojant entropijos funkciją, kas yra alternatyvų skaičiaus logaritmas, pvz. monetos mėtyme yra dvi alternatyvos, kas atitinka  $\log_2(2) = 1$  bitas, tuo tarpu lošimo kauliuko atveju yra šešios galimos baigtys, atitinkančios  $\log_2(6) = 2.59$  bitai (Tononi, 2008).

Atskirdamas dvi alternatyvas fotodiodas generuoja vieną bitą informacijos. Žmogaus atveju generuojamas žymiai didesnis kiekis informacijos, nes diskriminavimas vyksta tarp didesnio kiekio alternatyvų. Jei ekrane išsižiebtų raudona arba žalia spalva, arba būtų rodomas kino filmas, fotodiodas

tebeturėtų dvi būsenas, aplinkos šviesumas arba peržengtų fotodiodo aktyvavimo slenkstį, arba ne. Žmogaus atveju alternatyvų kiekis yra labai didelis.

Kaip minėta, pirmasis eksperimentas pristatomas dar Dinaminės šerdies hipotezėje (Tononi ir Edelman, 1998). Bet analogiškų svarstymų galima rasti ir anksčiau, fotodiodo mintinis eksperimentas yra analogiškas Chalmerso (1996) termostato mintiniam eksperimentui. Chalmersas pateikia ir analizuoja argumentus, kurie galėtų pagrįsti teiginį, kad termostatas yra sąmoningas. Galima pabrėžti, kad termostatas nuo fotodiodo skiriasi tuo, kad atskiria tris būsenas vietoje dviejų ir dėl to integruoja didesnę informacijos kiekį, kuris yra  $\log_2(3) \approx 1,585$  bito. Pagal IIT toks mechanizmas kaip termostatas turėtų būti daugiau nei pusantrą karto sąmoningesnis nei fotodiodas.

Antrame IIT mintiniame eksperimente aptariamas skaitmeninės vaizdo kameros pavyzdys. Siūloma įsivaizduoti vaizdo kamerą, kurios sensorių sudaro milijonas fotodiodų. Tokia kamera galėtų atskirti  $2^{1,000,000}$  alternatyvių būsenų ir generuoti milijoną bitų informacijos. Pagal IIT skirtumas tarp žmogaus ir skaitmeninės kameros yra susijęs su informacijos integravimu. Kamera iš pirmo žvilgsnio atrodo kaip viena sistema, atskirianti didelį kiekį galimų alternatyvų, tačiau iš tiesų ją sudaro atskiri fotodiodai, kurie tarpusavyje nesąveikauja, t. y. vieno diodo veikimas visiškai nepriklauso nuo to, ką daro kiekvienas kitas fotodiodas, ir kiekvienas diodas turi tik dvi alternatyvas. Kitaip nei kamera žmogus atskiria didelį kiekį alternatyvų kaip integruota sistema. Pasak IIT kiekvienas patyrimas yra integruota visuma, kiekvienas patiriamas vaizdas matomas kaip visuma ir negali būti išskaidytas į dalis, pvz. raudono kvadrato patyrimas negali būti išskaidytas į atskirą raudonos spalvos patyrimą ir atskirą kvadrato patyrimą. Pagal IIT vienintelis žinomas būdas kaip patyrimą perskelti į du atskirus patyrimus yra atskirti smegenų pusrutulius, perpjaujant didžiąją smegenų jungtį (*corpus callosum*). Norint nustatyti kiek sistemos informacijos yra integruota informacija, pirmiausia reikia nustatyti, kiek informacijos generuoja atskiros sistemos dalys, ir tuomet palyginti su visos sistemos generuojamos informacijos kiekiu (Tononi, 2008, p. 220). Integruotos informacijos kiekis žymimas simboliu  $\Phi$ . Vaizdo kameros atveju  $\Phi = 0$ . Iš esmės integruota informacija yra kompleksiško matavimas, parodantis kiek visa sistema integruoja informacijos daugiau nei jos dalys.

Integruotos informacijos samprata ir Claude'o Shannono komunikacijos teorijoje naudojama (sintaktinės) informacijos samprata pateikiamos kaip besiskiriančios tarpusavyje. Vienas iš IIT proponentų apibūdinamų skirtumų nurodo, kad informacija yra vidinė (*intrinsic*), tuo tarpu Shanono informacija yra išorinė (*extrinsic*). Aiškinant, kas yra informacija ir integravimas, pateikiami fotodiodo ir fotokameros mintiniai eksperimentai.

Trečiasis mintinis eksperimentas pristatomas tik antroje IIT versijoje. Šiame eksperimente (Tononi, 2012, p. 295) parodomas esminis skirtumas tarp interneto maršrutizatorių tinklo ir smegenyse esančio neuronų tinklo. Kitaip nei fotokameros sensorius, internetas pasižymi integruotumu, t. y. signalas gali keliauti iš bet kurio taško tinkle į bet kuriuos kitus taškus. Integruotumas, pasak Tononio, pasiekiamas dėka maršrutizatorių, kurie atlieka dinamiškų jungiklių (*dynamic switches*), sujungiančių vieną adresą tinkle su bet kuriuo kitu, funkciją. Pagal IIT neuronų tinklas, atsakingas už sąmoningumą yra sujungtas taip, kad galėtų pasiekti integruotos informacijos maksimumą (*maxima of integrated information*). Interneto atveju, toks reikalavimas nėra įgyvendinamas. Taip yra todėl, kad sąveikos vykstančios internete gali būti redukuotos į atskirus komponentus. Tai yra būtina norint išvengti nepageidaujamo kanalų persidengimo. Taigi, nors internetas yra pakankamai integruotas tam, kad galėtų vykti signalų mainai iš vieno taško į kitą, jis nėra maksimaliai integruotas.

Maksimalus informacijos integravimas susijęs su išskyrimo problema (*exclusion*). Iš pirmojo asmens perspektyvos nėra galimybės patirti to, kas vyksta tam tikrame smegenų modulyje. Galima girdėti ir suprasti garsiai ištartus žodžius, tačiau tiesioginė sąmoninga prieiga prie tokio tipo informaciją apdorojančios smegenų dalies nėra galima. Interneto atveju galima pilna prieiga prie atskirų jį sudarančių elementų. Tai reiškia, kad iš tam tikro taško galima patikrinti, kaip veikia paskiri tinklo elementai.

Išskyrimo problemą pritaikant sąmoningumui, nurodoma, kad tam tikru metu yra tik vienas sąmoningumas arba vienas maksimaliai integruotas subjektas, turintis vientisą patyrimą, bet ne keletą dalinių sąmoningumų. Kiekvienas patyrimas yra kompozicinis, sudarytas iš skirtingų aspektų įvairiomis kombinacijomis. Šį kartą Tononis (2012, p. 296) pateikia fenomenologinį patyrimo aprašymą: „Aš matau obuolio formą, aš matau, kad ši forma raudonos spalvos, aš matau poziciją erdvėje ir aš taip pat matau, kad obuolys yra raudonas ir užima šią poziciją“. Išskyrimas vyksta ir erdvėlaikiniu požiūriu, tai kas yra suvokiama, suvokiama kaip tam tikra erdvės ir laiko rezoliucija (*resolution*). Nėra galimybės tiesiogiai patirti procesus savose smegenyse, nes tai vyksta smulkesnėje erdvės rezoliucijoje, kaip kad molekulėse ir atomuose esančiuose neurono ląstelėje.

Searle'as (2013) kritikuoja IIT pateikiamą informacijos sampratą. Jo argumentas remiamas absoliučios ir priklausomos nuo stebėtojo informacijos skirtimi. Fotodiodas nėra sąmoningas, nes informacija fotodiode yra priklausoma nuo stebėtojo (*observer-relative*), žinančio, kaip fotodiodas funkcionuoja. Tuo tarpu sąmoningumas siejamas su absoliučia (*absolute*) informacija, kuri yra nepriklausoma nuo stebėtojo iš trečiojo asmens perspektyvos. Kochas ir Tononis (2013) sutinka, kad sąmoningumas yra nepriklausomas nuo stebėtojo. Tačiau, pasak jų, IIT pateikia kitokią informacijos

sampratą nei vartojama Shannono (1949) komunikacijos teorijoje. Pastaroji siekia nustatyti, kaip patikimai ir efektyviai informacija siunčiama ryšio kanalu. Joje informacija atskiriama nuo prasmės (*meaning*), t. y. informacijos, siunčiamos ryšio kanalu, prasmė glūdi ne pačiame signale, o subjektuose, kurie siunčia ir priima šį signalą. Informacija visuomet vertinama iš išorinės stebėtojo perspektyvos arba, kitaip tariant, ji yra priklausoma nuo stebėtojo, todėl Shannono informacija, pasak Tononio, gali būti vadinama išorine (*extrinsic*) informacija. Tuo tarpu IIT informacija įvardijama kaip vidinė (*intrinsic*). Norint nustatyti nuo stebėtojo priklausomą informaciją, pakanka stebėti, kaip sistema elgiasi per tam tikrą laiko tarpą. Tokia strategija nėra tinkama vidinės informacijos nustatymui.

Anilas Seth'as (2021) pabrėžia, kad norėdamas nustatyti vidinę informaciją stebėtojas turėtų žinoti visus galimus būdus, kaip sistema galėtų elgtis, net jei ji niekada taip nesielgs. Išskiriami du būdai, kaip galima pamatuoti vidinės informacijos kiekį. Vienas jų – sužadinti (*perturb*) sistemą visais įmanomais būdais ir stebėti, kas vyksta, taip išsiaiškinant visus įmanomus jos veikimo būdus. Antru atveju nurodoma, kad vidinės informacijos kiekį galima išreikšti žinant viską apie sistemos fizinį mechanizmą, jos priežasties–padarinio struktūrą (*cause–effect structure*). Nurodoma, kad tam tikrais atvejais, žinant mechanizmo fizinę struktūrą, galima numatyti visus jos veikimo būdus, net ir tuo atveju, kai sistema niekada jų nerealizuoja.

Pagal Seth'ą, pagrindinė IIT problema yra testuojamumas, nes sudėtingų sistemų, tokių kaip smegenys, atveju nei vienas iš minėtų vidinės informacijos nustatymo būdų nėra pritaikomas. Galima stebėti, kaip smegenys funkcionuoja, tačiau sužadinti jas visais įmanomais būdais nėra galimybės. Taip pat nėra priemonių gauti baigtines žinias apie jų fizinę struktūrą. Dėl šios priežasties Seth'as atmeta galimybę patikrinti, ar integruotos informacijos ir fenomenaliojo sąmoningumo tapatumo tezė yra teisinga. Seth'as neaptaria paprastų sistemų, tokių kaip fotodiodas ar termostatas. Šioje vietoje galima grįžti prie Searle'o argumento. Nepriklausomai nuo stebėtojo fotodiodas turi dvi būsenas, taip pat stebėtojas, kaip tvirtina Searle'as, žino, kad fotodiodas turi šias dvi būsenas, todėl fotodiodo atveju absoliuti ir nuo stebėtojo priklausoma informacija galimai persidengia.

## **1.5 Integruotos informacijos ir sąmoningumo tapatumo prielaida**

Toliau tvirtinama, kad tapatumo tezė yra fenomenologines aksiomas ir ontologinius postulatus jungianti prielaida. Ši prielaida yra neatsiejama nuo sąmoningumo laipsniškumo prielaidos.

IIT aprašančioje literatūroje galima surasti penkias skirtingas tapatumo (*identity*) tezės formuluotes. Pagal vieną iš jų, (1) “sąmoningumas atitinka (*corresponds*) sistemos pajėgumą integruoti informaciją“ (Tononi, 2004, p. 1). Pagal griežčiausią formuluotę, (2),integruota informacija yra

sąmoningumas“ (Tononi, 2008, p. 216). Kita formuluotė nurodo, kad (3) „subjektyvus patyrimas yra vienas ir tas pats dalykas kaip (*is one and the same thing as*) sistemos pajėgumas integruoti informaciją“ (Tononi, 2004, p. 19). Būti „vienu ir tuo pačiu“ yra griežtesnė formuluotė nei „atitikti“, pirmu atveju išreiškiama ekvivalencija, antras atvejis gali būti suprastas kaip ne toks griežtas santykis. Kitur tapatumo tezė eksplicitiškai pateikiama kaip hipotezė: (4) „pagal centrinę hipotezę, tam, kad būtų fizinis subjektyvaus patyrimo substratas, fizinė sistema turi atitikti penkis reikalavimus (postulatus)“ (Mayner et al., 2018). Galima rasti dar kiek kitokią formuluotę, pagal kurią, (5) „fundamentalus IIT tapatumas gali būti išreikštas taip: patyrimas yra identiškas konceptualinei struktūrai, tai reiškia, kiekviena patyrimo savybė turi atitikti konceptualinės struktūros savybę ir *vice versa*“ (Tononi, 2016).

Nepaisant skirtumų, formuluotės turi vieną bendrą aspektą, integruota informacija laikoma būtina ir pakankama sąmoningumo sąlyga. Tai reiškia, kad tam tikrą informacijos kiekį integruojantis mechanizmas (tiek biologinis, tiek nebiologinis) yra sąmoningas. Pagal griežtesnę teiginį – integruota informacija ir yra sąmoningumas. Bet kuriai tapatumo tezės versijai esant teisingai, nustačius, kad mechanizmas generuoja bent minimalų integruotos informacijos kiekį, galima būtų konstatuoti, kad tas mechanizmas yra sąmoningas. Jei bet kuri iš formuluočių pasirodytų esanti teisinga, IIT išspręstų gana sunkią sąmoningumo problemą, nes galėtų nustatyti, kuri sistema yra sąmoninga, o kuri ne.

Nors nėra nustatytas tikslus integruotos informacijos kiekis, būdingas sąmoningo žmogaus smegenims, jis yra labai didelis. Šis kiekis sumažėja pasiekus būseną, kai sąmoningumo stokoja, pvz. besapnio miego ar bendrosios nejautos metu (Casali et al., 2013). Sąmoningumas koreliuoja su dideliu integruotos informacijos kiekiu, tačiau tai dar nereiškia, kad esant mažesniai integruotos informacijos kiekiui bus mažesnis sąmoningumo kiekis ir, kad integruota informacija yra būtina bei pakankama sąlyga sąmoningumui rasti. Tačiau būtent šias dvi prielaidas priima IIT.

Koreliacija nustatoma tarp didelio kiekio integruotos informacijos ir sąmoningumo. Būtent šiuo atveju galima apibrėžti esmines sąmoningumo savybes, kaip tai mėgina daryti Tononis. Besapniame miege integruotumas sumažėja ir nebelieka galimybės fenomenologiškai apibūdinti šią būseną. Todėl, ji įprastai ir laikoma kaip nesąmoninga. Nors integruotumas besapniame miege sumažėja, jis nėra lygus nuliui. Jei būtų įmanoma pateikti besapnio miego fenomenologinę analizę, tai būtų argumentas, paremiantis IIT tapatumo tezę. Tokiu atveju aptikus koreliaciją ne tik su dideliu integruotos informacijos kiekiu, bet ir su gerokai sumažėjusiu integruotumu, leistų manyti, kad bet koks integruotos informacijos kiekis sietinas su sąmoningumu. Yra mėginimų nusakyti besapnio miego patirtį kaip gryną laiko patyrimą (*pure temporal experience*) (Windt, 2015). Tačiau toks aprašymas nedera su IIT tvirtinimu, jog bendras kompleksas skyla į daug atskirų kompleksų su mažesniu



integruotos informacijos kiekiu. Dėl šio skilimo besapniame miege, remiantis IIT, turėtų atsirasti daug atskirų sąmoningumų (žr. 2.2.2), o ne vienas laiko patyrimas.

IIT suformuluotą tapatumo tezę reikėtų suprasti ne kaip pateikiamą išvadą, gautą taikant fenomenologinį metodą, bet kaip papildomą nepriklausomai nuo fenomenologinio metodo taikymo įvedamą prielaidą. Be to, ji yra tiesiogiai susijusi su fenomenaliojo sąmoningumo laipsniškumo prielaida. Integruota informacija yra įvertinama kiekybiškai ir yra laipsniška, tai reiškia, kad sistema gali pasižymėti didesniu arba mažesniu jos kiekiu. Priėmus tapatumo tezę, tenka priimti sąmoningumo laipsniškumo tezę ir todėl mažą integruotos informacijos kiekį sieti su kiekybiškai mažesniu sąmoningumo kiekiu. Analogiškai, didelį integruotos informacijos kiekį tenka sieti su dideliu sąmoningumo kiekiu. Jei sąmoningumo laipsniškumas būtų atmestas, tuomet tektų atmesti ir IIT priimamą tapatumo prielaidą. Fenomenologinėmis aksiomomis apibūdinamas fenomenalusis sąmoningumas, koreliuojantis su dideliu kiekiu integruotos informacijos. Integruotos informacijos ir sąmoningumo tapatumo tezės priėmimas lemia sąmoningumo laipsniškumo tezės priėmimą, tačiau nebūtinai sąmoningumas sietinas su mažu integruotos informacijos kiekiu.

Viena alternatyva būtų atsisakyti tapatumo tezės ir mėginti nustatyti, ar yra tam tikras integruotos informacijos slenkstis, kurį peržengusi sistema tampa sąmoninga. Pagal neutraliąją IIT versiją galima tokia situacija, kad sąmoningumas atsiranda tik tuomet, kai peržengiamas tam tikras aukštas integruotos informacijos slenkstis (McQueen, 2018). Tačiau esminė problema šiuo atveju yra testuojamumas, iki šiol nėra galimybės nustatyti, koks integruotos informacijos kiekis yra sietinas su sąmoningo žmogaus smegenimis. Kita vertus, kaip parodo besapnio miego tyrimai, integruotos informacijos kiekis yra sumažėjęs ir tuo pačiu stokojama sąmoningumo. Tononis ir Kochas (2015, p. 13) taip pat svarsto tam tikro slenksčio tikimybę, tačiau dėl tapatumo tezės priėmimo tvirtina, kad „galimai yra toks  $\Phi$  max slenkstis, žemiau kurio žmonės negali pranešti, kad ką nors jaučia, bet tai nereiškia, kad sąmoningumas pasiekė visišką minimumą, nulį. Net tuomet, kai panyrame į besapnį miegą ir nubudę negalime pasakyti nieko apie buvimo jame patyrimą, jo metu, tam tikrų mažų neuronų kompleksų  $\Phi$  yra didesnis už nulį...“. Tačiau neaišku, kodėl besapnio miego būsenos negalima būtų laikyti tiesiog nesąmoninga būseną (žr. 2.2.2).

Tikėtina, kad sąmoningumui reikalingas tam tikro integruotos informacijos slenksčio peržengimas, tačiau šio slenksčio peržengimas taip pat dar nereiškia, kad mechanizmas pasižymi sąmoningumu. Scottas Aaronsonas (2014) parodė, kad išplėtimo grafikas pasižymi didesniu integruotos informacijos kiekiu nei žmogaus smegenys, tačiau nėra sąmoningas (žr. 2.2.1). Taigi vien didelio kiekio integruotos informacijos kiekio nepakanka ir dėl to integruota informacija yra tik viena iš

sąlygų, be kurios sąmoningumo nebūtų. Veikiausiai, reikalinga papildoma sąlyga arba savybė, kurios nepaaiškina IIT (žr. 2.2).

Kita problema yra subjektyvus sąmoningumo laipsniškumo įvertinimas. Viena svarbi IIT praleista sąmoningumo savybė, pasak Peterio Carutherso (2019, p. 75) yra ta, kad „fenomenalusis sąmoningumas, žvelgiant iš pirmojo asmens perspektyvos, kategoriškai arba yra, arba ne“. Dėl to galima būtų tvirtinti, kad sąmoningumas bent jau neatrodo laipsniškas žvelgiant iš pirmojo asmens perspektyvos. Tononis (2008, p. 236) tvirtina kitaip, nors užmigimas laikomas staigiu procesu, tačiau prieš dingstant sąmoningumui galima fiksuoti progresyvų sąmoningumo kitimą savęs ir aplinkos atžvilgiu. Tačiau toks argumentas nepatvirtina, kad sąmoningumas yra laipsniškas. Žvelgiant iš pirmo asmens perspektyvos patiriamas pokytis, t. y. sąmoningumo turinys keičiasi, tačiau dingus sąmoningumui, dingsta bet koks turinys. Tol kol yra koks nors turinys, tol yra sąmoningumas. Jam dingus, turinys dingsta staiga. Pagal kitą Tononio argumentą (2008, p. 241) žmonių sąmoningumo lygio vertinimas yra prastas. Toks tvirtinimas remiamas analogija su temperatūros lygio vertinimu, kol temperatūra svyruoja mums įprastame diapazone, tol galime ją įvertinti gerai, tačiau ne tuomet, kai ji peržengia mums įprastas ribas.

IIT sąmoningumo laipsniškumo priėmimas iš pirmo žvilgsnio atrodo įtikinamas žvelgiant iš evoliucinės perspektyvos, nes atitinka Williamo Jameso suformuluotą „tęstinumo reikalavimą“, pagal kurį sąmoningumas neatsirado staiga (Godfrey-Smith, 2021). Kitaip tariant, sudėtingos patyrimo formos turėjo kilti iš paprastesnių. Tačiau galima skirti sąmoningumą nuo jautrumo (*sentience*), kaip tai daro Peteris Godfrey-Smith'as (2021). Jautrumas laikomas pirmesniu už sąmoningumą. Jis siejamas su jutimu ir veiksmu, kurie yra plačiai paplitę tarp biologinių organizmų, įskaitant bakterijas. Nors gyvi organizmai turi įvairius į aplinką skirtus reaguoti receptorius, dalis jų nėra sąmoningi. Tik tam tikrame evoliucijos etape kyla papildomų gabumų, dėl kurių atsiranda sąmoningumas. Todėl klaidinga manyti, kad paprastesnių gyvūnų subjektyvi patirtis yra tik silpnesnė žmogaus subjektyvios patirties versija. Godfrey-Smith'as remiasi ventralinio ir dorsalinio srautų hipoteze. Pagal ją ventralinis srautas siejamas su sąmoningumu, o dorsalinis srautas su nesąmoningu informacijos apdorojimu. Šią skirtį Milneris ir Goodale'as paremia pacientės, žinomos kaip D. F., atveju. Šiai moteriai buvo pažeista smegenų dalis, todėl ji buvo beveik akla, tačiau gebėjo apeiti kliūtis ar įmesti voką į pašto dėžutę, nors negalėjo įvardinti pašto dėžutės angos padėties. Tai parodo, kad gebėjimas jausti ir veikti galimi ir be sąmoningumo. Godfrey-Smith'o argumentas parodo, kad ne tik didesnis ar mažesnis integruotos informacijos kiekis yra esminė sąlyga sąmoningumui atsirasti, bet reikalingos ir papildomos savybės, kurios galėjo atsirasti vykstant evoliucijai.

## 2. IIT INTERPRETACIJOS IR JŲ SAŲVEIKA

### 2.1 Šešios IIT interpretacijos

IIT pateikta tapatumo prielaida susilaukia gausios kritikos, tad galimi du žingsniai: arba atmesti visą teoriją, arba kai kurias jos prielaidas, išsaugant tam tikrą teorijos dalį. Galima surasti tris tarpusavyje panašius mėginimus modifikuoti IIT (McQueen, 2018; Michel ir Lau, 2020; Mediano, et al., 2022). Šios trys interpretacijos remiasi bendra nuostata, jog reikia atsisakyti IIT priimamos tapatumo tezės.

Michelas ir Lau (2020) skiria fundamentalią ir empirinę IIT versiją. Fundamentalią vadinama originali teorija, empirinė yra pastarosios modifikacija. Norint suprasti skirtumą tarp jų svarbu skirti nervinius sąmoningumo koreliatus (*neural correlates of consciousness*) nuo sąmoningumo markerių (*markers of consciousness*) ir sąmoningumo konstituentų (*constituents of consciousness*). Pirmieji yra „minimalūs neuroninių įvykių ir mechanizmų rinkiniai, pakankami (*jointly sufficient*) specifiniam sąmoningam suvokimui“ (p. 2), pvz. raudonos spalvos patyrimui. Sąmoningumo konstituentais įvardinama „nervinė ar fizikinė būseną, tapati sąmoningumui“ (p. 2), kaip pvz. H<sub>2</sub>O yra vandens konstituentas. Pasak Michelo ir Lau, konstituentas yra būtina ir pakankama sąmoningumo sąlyga. Galiausiai, sąmoningumo markeriai apibūdinami kaip „bendrieji įrodymai, kuriuos galima panaudoti norint nustatyti, ar subjektai turi sąmoningas psichines būsenas ar ne“ (p. 2). Michelo ir Lau teigimu empirinė IIT versija yra tokia, kuri kompleksiskumą laiko sąmoningumo markeriu. Pagal ją smegenų neuronų tinklo kompleksiskumo matavimas gali būti panaudotas norint nustatyti, ar tiriamasis turi subjektyvią patirtį ar ne. Fundamentali IIT versija apibūdinama kaip požiūris, kad specifinė kompleksiskumo forma, t. y. integruota informacija, yra identiška sąmoningumui. Šiuo atveju kompleksiskumo forma yra sąmoningumo konstituentas. Taigi, esminis skirtumas tarp empirinės ir fundamentaliosios teorijos versijų yra požiūris į tapatumo tezę.

Pagal Michelą ir Lau, fundamentali IIT nėra pagrįsta. Pateikiami trys argumentai, nukreipti prieš fundamentalią IIT versiją. Pirmuoju argumentu parodoma, kad empirinė IIT yra geriau deranti su kitomis teorijomis, pvz. Globalaus darbalaukio teorija (GWT), o fundamentali IIT su ja nėra suderinama. Galima pastebėti, kad šis suderinamumo argumentas remiasi prielaida, kad IIT ir GWT vartoja tą pačią sąmoningumo sampratą. Tačiau bent jau konceptualiai galima atskirti fenomenalųjį sąmoningumą nuo prieigos sąmoningumo (Block, 1995, 2002, 2011). Yra manančių, kad IIT tyrinėja fenomenalųjį sąmoningumą, o GWT prieigos sąmoningumą (Seth, 2021), kita vertus, pagal Carruthersą

(2019), GWT tyrimo objektas yra fenomenalusis sąmoningumas. Jei šio konceptualinio skirtumo nėra, tuomet galima pritarti Michalui ir Lau, kad pagrindinis teorijas skiriantis aspektas yra IIT priimama tapatumo tezė. Antrasis argumentas siejamas su kompleksiskumo matavimais. Michelas ir Lau nurodo, kad pastarieji tiesiogiai nepagrindžia fundamentaliosios IIT, nes juos taikant realioms biologinėms sistemoms tenka naudoti labai apytikslius matavimus. Dėl to šie skaičiavimai nelaikytini sąmoningumo konstituentais, o tik markeriais. Nėgama to, galimi įvairūs apytiksliai išreikšti (*approximated*) integruotos informacijos matavimai. T. y. taikant vienus matavimus galima verifikuoti teoriją, taikant kitus, ją falsifikuoti. Trečiuoju argumentu kvestionuojamas kompleksiskumo matavimų validumas. Šis argumentas remiamas sąmoningumo ir kognicijos skirtimi. Didelė dalis kognityvinių gebėjimų skiriasi tarp budrumo ir nesąmoningos būsenos. Dėl to, kompleksiskumo matavimai galimai nustato šiuos skirtumus, o ne paties sąmoningumo buvimą ar stoką. Taigi, pasak Michelo ir Lau, norint paremti teiginį, kad sąmoningumas yra tapatus specifiniam kompleksiskumui, pirmiausia reikėtų išsiaiškinti, ar kompleksiskumas matuoja sąmoningumą, o ne kogniciją (*cognition*) ar kokį kitą faktorių, kuris skiriasi tarp budrių ir nesąmoningų asmenų. Pasak straipsnio autorių, šie trys argumentai yra nukreipti prieš fundamentaliąją, bet ne prieš empirinę IIT versiją. Teigiama, kad kompleksiskumo matavimai turėtų pasitarnauti kaip sąmoningumo markeriai ir tvirtinimas, kad kompleksiskumas turėtų būti tapatinamas su sąmoningumu, kaip tai daro IIT proponentai, yra nepagrįstas.

Dvejopa IIT interpretacija aptariama ir Mediano et al. (2022). Analogiškai Michalui ir Lau (2020) daromai skirčiai tarp empirinės ir fundamentaliosios IIT, Mediano analizuoja skirtumą tarp silpnosios ir stipriosios IIT versijų. Pasak jų, stipriosios IIT pagrindinė idėja yra ta, kad sąmoningumas yra tapatus sistemoje esančios vidinės integruotos informacijos savybėms. Kitaip yra su silpnąja IIT versija, pagal kurią integruota informacija yra aiškinamasis sąmoningumo koreliatas (*explanatory correlate of consciousness*). Pabrėžiamos dvi problemos, su kuriomis susiduria stiprioji IIT versija. Pirmoji problema apima IIT aksiomų teisingumo, nuoseklumo ir universalumo pagrindimą, antroji teorijos testuojamumą. Mediano et al. išvelgia dvi esmines problemas, susijusias su stipriosios IIT testuojamumu. Analogiškai Michalui ir Lau argumentuojama, jog galimi įvairūs integruotos informacijos matavimai. Kadangi nėra vieno universalaus integruotos informacijos empirinio matavimo, silpnosios IIT proponentai siekia paaiškinti, kaip skirtingi empiriniai matavimai siejasi tarpusavyje ir kaip siejasi su skirtingais sąmoningumo aspektais. Stiprioji IIT versija taikytina bet kokių fizinių sistemų sąmoningumo kiekiui nustatyti, tuo tarpu silpnoji versija apsiriboja neuronus turinčiomis sistemomis. Silpnosios IIT versijos tikslas yra įvairių sąmoningumo aspektų, išmatuojamų

empiriškai, ir stiprių aiškinamųjų korelacijų paieškos. Šios koreliacijos nelaikomos griežtai būtinomis ar pakankamomis sąmoningumo sąlygomis.

Michelas ir Lau nesistengia išsaugoti IIT pateikiamų fenomenologinių aksiomų. Kitaip daro Mediano et al., silpnojoje versijoje palikdami informacijos ir integracijos aksiomas. Pastarieji teigia, kad neuroniniai sąmoningumo substratai turėtų atspindėti šias dvi esmines fenomenologines savybes. Pagal pirmąją, kiekvienas sąmoningas momentas yra labai informatyvus, kitais žodžiais tariant, kad kiekvienas patyrimas yra vienas iš didžiulio galimų patyrimų repertuaro. Pagal antrąją, kiekvienas sąmoningas patyrimas yra integruotas arba, kitaip išreiškiant, jis yra patiriamas kaip koherentiška visuma. Silpnoji versija galėtų apsieiti ir nelaikydama šių savybių aksiomomis, nes integruotumas ir informatyvumas išreiškiami tiek IIT pateikiamuose mintiniuose eksperimentuose, tiek aptariami dinaminės šerdies hipotezėje.

McQueenas (2018) skiria ortodoksinę IIT nuo interpretaciškai neutralios (*interpretation-neutral*) IIT versijos ar trumpiau – neutraliosios IIT. Pastaroji yra tokia, kurioje siekiama minimalizuoti arba visai pašalinti kontroversiškas filosofines prielaidas, paliekama tik tai, kas svarbu neuromokslininkams ieškant nervinių sąmoningumo koreliatų. Iš esmės paliekamas matematinis ir empirinis teorijos turinys. Siūloma palikti tik matematinį formalizmą, skirtą integruotos informacijos kiekio nustatymui. Be to, neutralioji IIT dar turėtų apibrėžti prognozių rinkinį. Šiuo atveju, kaip ir silpnosios ar empirinės atvejais, taip pat siūloma atsisakyti tapatumo tezės ir apsiriboti smegenų tyrimais.

Tapatumo tezės atmetimas yra esminė empirinės, silpnosios ir neutraliosios IIT interpretacijų bendra nuostata, besiskirianti nuo pradinės teorijos. Dėl to tiek empirinė, tiek silpnoji, tiek neutralioji IIT interpretacijos patenka į vieną aibę. Norint išvengti painiavos, ši aibė toliau bus vadinama minimaliąja IIT. Pastaroji atmeta dalį prielaidų, geriau atitinka paprastumo kriterijų ir dera su empiriniais tyrimais. Teorijos interpretacija, išlaikanti tapatumo tezę, bus vadinama pilnąja IIT.

Nors eksplicitiškai neišreikšta, silnoji ir empirinė IIT interpretacijos, priima egzistencijos aksiomą, pagal kurią, sąmoningumas yra realus jį patiriančiam asmeniui, o neutralioji IIT interpretacija išlieka neutrali šiuo klausimu. Šią aksiomą galima apibūdinti kaip realistinės pozicijos sąmoningumo atžvilgiu priėmimą (žr. 2.1.2). Tokios prielaidos priėmimas nereikalauja atlikti pirmojo asmens perspektyvos tyrimo, bet yra būtinas ieškant nervinių sąmoningumo koreliatų ir mėginant nustatyti sąmoningumo markerius. Atmetus sąmoningumo realumą neliktų poreikio atlikti tokio pobūdžio tyrimus.

Yra manančių (Tsuchiya ir Leung, 2023), kad silpnoji IIT versija, kurią pateikė Mediano et al. (2022) yra pernelyg silpna. Todėl siūloma įvesti skirtį tarp aspiracinės (*aspirational*) ir inspiruotos

(*inspired*) IIT. Tvirtinama, kad silpnoji pozicija išlaiko informacijos ir integracijos idėjas, bet stokoja teorinio pagrindo, jungiančio jas su sąmoningumu. Tvirtinama, kad silpnoji IIT versija yra tik rinkinys koreliatų, inspiruotų IIT, dėl to nelaikytina viena iš IIT versijų. Tuo tarpu, aspiracinė IIT apibūdinama kaip tokia, kuri siekia empiriškai testuoti prielaidas, paaiškinimus ir spėjimus, pateikiamus IIT, tačiau pasiliekant kiek įmanoma arčiau filosofinio ir teorinio karkaso. Šios interpretacijos proponentai pabrėžia IIT testavimo keblumus, todėl siūlo kitokias testavimo alternatyvas. Teorine prasme aspiracinė IIT nedaug nutolsta nuo pilnosios IIT. Šios pozicijos atstovai siekia išsaugoti aksiomatinį pilnosios IIT metodą, todėl aspiracinėje pozicijoje implicitiškai glūdi požiūris, kad IIT yra fenomenologinio pirmumo teorija. Pagal šią poziciją IIT išveda postulatus iš aksiomų. Tačiau kaip parodyta anksčiau, toks išvedimas yra nepagrįstas. Aspiracinės pozicijos proponentai nepastebi, kad integracijos ir informacijos savybės, kurias išsaugo Mediano et al., aptinkamos ir kitais būdais.

Reziumuojant galima pateikti tokią IIT klasifikaciją, kurią sudaro šešios skirtingos IIT interpretacijos (žr. 1 priedas). Priklausomai nuo to, ar priimama tapatumo tezė, IIT gali būti klasifikuojama į minimaliąją arba pilnąją. Pilnoji IIT priima tapatumo tezę, o minimalioji ją atmeta. Į minimaliąją versiją patenka (1) empirinė, (2) neutralioji ir (3) silpnoji. Priklausomai nuo to, kaip tapatumo tezė yra gaunama, pilnoji IIT gali būti skiriama į (4) eksperimentinio pirmumo ir fenomenologinio pirmumo. Į šią klasifikaciją įtraukus (5) aspiracinę IIT, kaip fenomenologinio pirmumo teorijos poklasį, tenka ją atskirti nuo (6) pradinės IIT. Pradinė IIT yra pradinį metodologinį sumanymą atitinkanti teorija, kuri pretenduoja patvirtinti tapatumo tezę empiriniais metodais. Taigi IIT gali būti suprantama kaip pradinė, aspiracinė, eksperimentinio pirmumo, silpnoji, empirinė ir neutralioji. Taip pateiktas eiliškumas, žvelgiant iš kairės į dešinę, parodo, kaip stipriai IIT interpretacijos nutolsta nuo pradinio metodologinio sumanymo. Empirinė, neutralioji ir silpnoji interpretacijos yra labai artimos viena kitai, kaip ir pradinė yra artima aspiracinei. Eksperimentinio pirmumo interpretacija iki šiol nebuvo išskirta, tačiau ją paremia fenomenologinio pirmumo kritika, aptarta pirmoje darbo dalyje. Eksperimentinio pirmumo IIT interpretacija priima, kad iš aksiomų postulatai nėra išvedami ir tapatumo tezė priimama kaip prielaida, bet ne kaip išvada, kitaip nei tai, kas traktuojama fenomenologinio pirmumo interpretacijoje. Pateikta IIT klasifikacija naudojama likusioje darbo dalyje.

### **2.1.1 Empirinis IIT pagrįstumas**

Integruotos informacijos smegenyse koreliacija su sąmoningumu grindžiama smegenėlių (*cerebellum*), miego ir bendrosios neįtautos ir pacientų patyrusių smegenų pažeidimus tyrimais. Pirmu

atveju parodoma, kad vien aktyvių neuronų kiekis nėra esminė sąmoningumo sąlyga (Tononi, 2012, 2014). Smegenėlės, turinčios apie keturis kartus daugiau neuronų nei likusi smegenų dalis, nesukelia sąmoningumo ir prie jo neprisideda. Šis tvirtinimas grindžiamas patologiniu atveju, vadinamu smegenėlių ageneze (*cerebellar agenesis*). Turintys šį sutrikimą pacientai stokoja smegenėlių, bet yra pilnai sąmoningi. Smegenėles sudaro didelis kiekis funkciškai atskirų modulių, kurių įvesties ir išvesties signalai yra vienkrypčiai (*feedforward*), dėl to smegenėlės yra mažai integruotos.

Sąmoningumas susijęs su skirtingų smegenų dalių tarpusavio sąveika (komunikacija). Šios sąveikos nustatymui naudojama dviejų metodų – elektroencefalografo (EEG) ir transkranijinės magnetinės stimuliacijos (TMS) – kombinacija. Šie metodai pritaikomi tiriant skirtumą tarp tokių būsenų kaip budrumas, REM ir NREM miego stadijos (Massimini, 2005; Massimini et al., 2010; Lee et al., 2019). Supaprastinus šis metodas nusakomas taip – pirmiausia stimuliuojama tam tikra vieta smegenų žievėje TMS pagalba ir tuomet pasitelkiant EEG fiksuojamas aktyvumo sklidimas kitose žievės srityse (laike ir erdvėje). Gauti rezultatai parodo, kad nesąmoningose būsenose, tokiose kaip NREM (arba kitaip besapnis miegas) TMS sukelti aidai (*echoes*) yra labai silpni. Kitaip yra esant sąmoningoje būsenoje, TMS sukeltas aidas aprėpia visą smegenų žievės paviršių ir išnykdamas bei atsirasdamas sudaro kompleksiškas struktūras. Šis kompleksiškumas parodo, kad skirtingos smegenų dalys, ypač gumburas ir žievė (*thalamocortical system*), esant sąmoningoje būsenoje sąveikauja tarpusavyje sudėtingiau nei nesąmoningoje.

Analogiškas eksperimentas atliekamas ir su tiriamaisiais, kuriems taikoma bendroji nejautra (Sarasso, 2014). Šiuo atveju rezultatai panašūs į gautus tiriant NREM miego būsenoje esančius tiriamuosius. Kitaip yra sapnuojant, nors tiriamasis nereaguoja į aplinką, kaip ir NREM miego ar bendrosios nejautos atvejais, tačiau smegenų aktyvumas yra panašus į būdraujančio žmogaus. Šio tipo eksperimentai taikomi pacientams su smegenų pažeidimais, kuriems priskiriamos klinikinės diagnozės kaip vegetacinė būseną, minimalaus sąmoningumo būseną ar užrakinto viduje sindromas (*locked-in syndrome*) (Casali et al. 2013).

Visi minėti tyrimai grindžia koreliaciją tarp didelio integruotos informacijos kiekio ir fenomenaliojo sąmoningumo, tačiau ne priežastinį ryšį ar, juo labiau, tapatumą. Minimalioji IIT interpretacija atmeta tapatumo tezė ir integruotą informaciją laiko būtina, bet nelaiko pakankama sąmoningumo sąlyga. Todėl minėti argumentai paremia minimaliąją IIT interpretaciją. Kitaip yra su pilnąja IIT interpretacija – integruotos informacijos ir sąmoningumo tapatumo tezė yra priimama. Nors aptarti tyrimai neprieštaruoja, jų nepakanka norint pagrįsti pilnąją IIT interpretaciją. Veikiausiai,

integruotos informacijos būtinumo patvirtinimas nėra neįmanomas ir teorija iš principo yra falsifikabili, tačiau tiesioginis patvirtinimas susiduria su prieš tai aptarta testuojamumo problema (Seth, 2021).

Nepaisant to, vyksta bendras IIT ir GWT prognozių testavimas (Melloni et al., 2023). Šis testavimas parodys, ar penkios tarpusavyje besiskiriančios minėtų teorijų prognozės pasitvirtina ar ne. Tyrimų rezultatai parems arba atmes IIT pateikiamas prognozes, susijusias su sąmoningumu, taip paremdami ar susilpnindami IIT, tačiau net gavus patvirtinančius rezultatus, tai nebus pakankamas tapatumo tezės patvirtinimas. Taip yra todėl, kad tyrimų objektas yra vizualiojo sąmoningumo koreliatai žmogaus smegenyse. O vien koreliacinių ryšių patvirtinimas dar nereiškia tapatumo ar būtinumo patvirtinimo. Tiriant juos nustatoma sąsaja tarp didelio kiekio integruotos informacijos ir sąmoningumo, tačiau pilnosios IIT prognozės apima ir kitas biologines bei nebiologines sistemas su gerokai mažesniu integruotos informacijos kiekiu.

### **2.1.2 IIT sąmoningumo apibrėžimas**

Visose šešiose teorijos interpretacijose išskyrus neutraliąją laikomasi prielaidos, kad IIT tyrinėja fenomenalųjį sąmoningumą. Norint, kad teorija išliktų validi, tenka kitaip derinti pirmojo ir trečiojo asmens perspektyvas, nei tai daroma pirminėje IIT ir atmesti dalį jos prielaidų. Tačiau galimas ir kitas žingsnis, ne atmesti prielaidas susijusias su pirmo ir trečio asmens perspektyvų derinimu, bet išsiaiškinti, kokia sąmoningumo samprata teorijoje vartojama iš tiesų, nepriklausomai nuo to, kokia yra pateikiama eksplicitiškai.

Sudėtinga daryti tyrimus, susijusius su sąmoningumu, dėl šios sąvokos daugiaprasmiškumo. Nedas Blockas (1995) tvirtina, kad tai mišrus terminas, skirtingai vartojamas skirtingų autorių. IIT pateikia apibrėžimą: „kiekvienas žino kas yra sąmoningumas: jis yra tai, kas dingsta kasnakt kai panyrame į besapnį miegą ir pasirodo vėl, kuomet nubundame ar sapnuojame“ (Tononi, 2008, p. 216; Tononi et. al., 2016, p. 450). Tokio tipo apibrėžimą galima aptikti ir kitų autorių, pvz. Johno Searle'o (1990) tekstuose. Tačiau šis apibrėžimas nėra pakankamas dėl savo apimties. Jis yra per platus, nes apima įvairias sąmoningumo sampratas (pvz. prieigos sąmoningumą, fenomenalųjį sąmoningumą *etc.*).

IIT proponentų šis platusis apibrėžimas susiaurinamas iki fenomenaliojo sąmoningumo sampratos. Tai, kas pasiekama pateikiant fenomenologines aksiomas, yra mėginimas pateikti esmines ir pakankamas savybes, nusakančias fenomenalųjį sąmoningumą. Todėl kyla klausimai: ar taip susiaurinta sąvoka vis dar sietina su fenomenaliojo sąmoningumu ir ar atliekami matavimai yra validūs. Kaip parodyta 1.2 poskyryje, aprašant fenomenalųjį sąmoningumą galima pateikti daugiau



aksiomų nei pateikiama IIT, todėl tai yra nepilnas fenomenaliojo sąmoningumo apibrėžimas. Nors ir nepilnai apibrėžtas pirminis IIT objektas yra fenomenalusis sąmoningumas.

Dalis oponentų nesutinka su tuo, kad IIT tyrinėja fenomenalųjį sąmoningumą. Vieni (Carruthers, 2019) tvirtina, jog IIT tyrinėja būtybės sąmoningumą (*creature consciousness*), kiti (Cerullo, 2015), kad IIT tyrimo objektas yra protosąmoningumas (*protoconsciousness*), yra manančių (Marker et al., 2021), kad IIT tyrinėja ne sąmoningumą, bet tinklo efektyvumą (*network efficiency*). Šis mėginimas parodyti, kad IIT vartoja visai kitą sąmoningumo sampratą, nei tvirtina vartojanti, kyla dėl to, kad abejojama integruotos informacijos matavimo validumu. Manoma, kad integruota informacija matuoja kažką kitą ir mėginama išsiaiškinti, kas yra šio matavimo objektas.

Jei integruota informacija yra būtina, bet nepakankama sąlyga arba reikalingas tam tikras jos slenkstis tam, kad atsirastų sąmoningumas, t. y. jei tapatumo tezė yra klaidinga, tuomet nustačius, kad sistema integruoja informaciją, vienu atveju gali būti aptinkamas fenomenalusis sąmoningumas, kitu atveju nesąmoningas informacijos apdorojimas, arba dar kažkas, ką galima sieti su kitomis sąmoningumo sampratomis. Taigi tokia teorija nebūtų validi dėl to, kad skirtinguose kontekstuose matuotų skirtingus dalykus, bet ne tai, kas planuota ja pamatuoti. Šiuo atveju į minimaliąją IIT klasę pakliūnančios interpretacijos išvengia šios problemos, nes pagal jas integruota informacija koreliuoja su fenomenaliojo sąmoningumu, bet nėra jam tapati.

Svarstant, kokią sąmoningumo sampratą vartoja IIT, bandoma išsiaiškinti, koks yra teorijos tyrimo objektas, bet nesiūloma modifikuoti pačios teorijos taip, kad ji matuotų norimą tyrimo objektą. Tačiau atliekant svarstymą dėl IIT sąmoningumo sampratos ir mėginant nustatyti, kas yra matuojama IIT pagalba, kitos metodologinės teorijos problemos išlieka. Todėl priimtinesne laikytina ankstesniame skirsnyje pateikta klasifikacija, pagal kurią galimos skirtingos IIT interpretacijos, remiantis tuo, kaip stipriai modifikuojama teorija, tačiau išlaikant tą patį tyrimo objektą – fenomenalųjį sąmoningumą.

### 2.1.3 Realistinė IIT pozicija

Pagal fenomenologinio pirmumo IIT, aksiomos yra save grindžiančios tiesos (*self-evident truths*), apibūdinančios sąmoningumą (Oizumi et. al., 2014, p. 2). IIT proponentai mėgina sekti Descartes'o pėdomis ieškodami apodiktinių teiginių. Šio bandymo pavyzdys yra egzistencijos aksioma, Pagal ją, sąmoningumas yra patyrimas, kuris yra realus jį patiriančiam žmogui. Patiriami dalykai yra vienintelis faktas, kuris, pasak Tononio, yra neabejotinas, dėl to Descartes'o frazė "maštau, vadinasi, egzistuoju" pakeičiama į "turiu patyrimą, vadinasi, egzistuoju" (Tononi, 2012). Sąmoningumo egzistavimas Tononio yra laikomas neginčijamu.

Pokropskis (2018) sutinka su Tononiu ir tvirtina, kad egzistencijos aksioma nėra kontroversiška. Pagal šią aksiomą sąmoningumas egzistuoja, o tai, pasak Pokropskio, atrodo akivaizdu, nes norint tirti tam tikrą fenomeną reikia pripažinti jo egzistavimą.

Su pirmosios aksiomos apodiktiškumu nesutiktų iliuzionizmo atstovai (Frankish, 2016). Iliuzionizmas priima radikaliojo realizmo nuostatą apie fenomenaliosios sąmonės anomališkumą ir konservatyviojo realizmo radikalios teorinės inovacijos atmetimą. Apjungia šiuos tvirtinimus traktuodamas fenomenaliąsias savybes kaip iliuzines. Iliuzionistai atmeta tai, kad patyrimas turi fenomenaliąsias savybes ir koncentruojasi ties klausimu „kodėl atrodo, kad jis tokias turi?“ (Frankish, 2016). Iliuzionistai pripažįsta, kad žmonės introspekciškai suvokia (*introspectively aware*) savo sensorines būsenas, tačiau teigia, kad šis suvokimas (*awareness*) yra nepilnas bei iškreiptas, klaidinančiai pateikiantis sensorines būsenas kaip turinčias fenomenalias savybes. Pagrindinė Frankisho straipsnio tezė – fenomenalusis sąmoningumas neegzistuoja, mums tik atrodo, kad jis egzistuoja.

Siekiant paaiškinti fenomenalųjį sąmoningumą IIT priima realistinę poziciją. Ši pozicija alternatyvi iliuzionizmui bei eliminatyvizmui (Churchland, 1981). Pagal Frankisho klasifikaciją realizmas yra dviejų tipų – radikalusis ir konservatyvusis. Pirmasis traktuoja fenomenaliąją sąmonę kaip realią ir nepaaiškinamą be radikalios teorinės inovacijos. Į šią grupę patenka tokios metafizinės sąmonės filosofijos pozicijos kaip dualizmas, neutralusis monizmas, misticizmas (*mysterians*) ir apeliuojančios į naująją fiziką. Antroji pozicija priima fenomenaliosios sąmonės realumą, bet siekia paaiškinti ją fizikiniais terminais. Būtent šio tipo pozicija yra priimama IIT.

Chalmersas (2018) kritikuoja iliuzionizmą pats priimdamas realistinę poziciją fenomenaliojo sąmoningumo atžvilgiu. Jis teigia, kad iliuzionizmo problema yra susijusi su jo pristatoma meta-problema, tačiau yra siauresnė. Jo teigimu geriausi iliuzionizmo argumentai yra atskleidžiamieji argumentai (*debunking arguments*). Chalmersas šiuos argumentus apibūdina kaip tokius argumentus, kurie pradami nuo prielaidos, kaip įsitikinimai apie tam tikrą dalyką yra suformuoti, ir išvedimo, kad įsitikinimai apie tą dalyką nėra patikimi ar pagrįsti.

Pradinė, aspiracinė, eksperimentinio pirmumo, silpnoji ir empirinė IIT interpretacijos priima realistinę poziciją sąmoningumo atžvilgiu, o neutralioji stengiasi išlikti neutrali. Neutraliosios IIT proponentai (McQueen, 2018a) mėgina parodyti, kad ši interpretacija gali būti suderinta su Iliuzionistine pozicija. Empirinėje, silpnojoje ir eksperimentinio pirmumo IIT interpretacijose realistinės pozicijos priėmimas tėra prielaida, bet ne aksioma, gauta taikant fenomenologinį metodą.

Aspiracinė ir pradinė IIT interpretacijos taiko fenomenologinį metodą ir pateikia egzistencijos aksiomą, dėl kurios įgyja realistinės pozicijos sąmoningumo atžvilgiu statusą.

#### 2.1.4 Ontologinis IIT statusas

Nėra sutariama, kokią ontologinę poziciją atitinka IIT. Tvirtinama, kad IIT nesuderinama su epifenomenalizmu (E-tipo dualizmu) ir substanciniu dualizmu (D-tipo dualizmu) (Grasso, 2013), komputaciniu funkcionalizmu (Cerullo, 2015), Raseliškuoju panpsichizmu (Mørch, 2018), Raseliškojo monizmo kategorikalistine versija (*categoricalist version of Russellian Monism*) (Grasso, 2019). Kiti mėgina parodyti, kad IIT suderinama su viena ontologine pozicija, pvz. panpsichizmu (Searle, 2013) ar daliniu paneksperientalizmu (*partial-panexperientialism*) (Cerullo, 2015). Treti nurodo, kad IIT gali būti suderinama su keliomis pozicijomis (Grasso, 2013).

Darbe nesiekama nustatyti, su kuria ontologine pozicija geriausiai dera IIT, o tik parodyti, kad priklausomai nuo to, kaip IIT derina pirmojo ir trečiojo asmens perspektyvos duomenis, ją galima traktuoti skirtingai. Taigi tvirtinama, kad pateiktoje klasifikacijoje nurodytoms IIT interpretacijoms priskirtinas nevienodas ontologinis statusas.

Tononio ir Kocho (2015) teigimu, IIT turi bendrų sąlyčio taškų su panpsichistine pozicija. Vienas jų yra sąmoningumo laipsniškumo pripažinimas, taip pat galima pabrėžti, kad sąmoningumas pagal IIT yra plačiai paplitęs ir gali būti aptinkamas ne tik biologinėse sistemose. Pagal IIT visos sistemos, pasižyminčios gebėjimu integruoti informaciją, yra tam tikru laipsniu sąmoningos. Kadangi integruotos informacijos kiekis tarp sistemų skiriasi, viena sistema gali būti labiau sąmoninga nei kita. Šie tvirtinimai būdingi pilnajai IIT interpretacijai, kuri kaip ir panpsichizmas susiduria su kombinacijos problema (žr. 2.2.2). Tačiau į minimaliąją IIT interpretacijų klasę patenkančios interpretacijos išvengia kombinacijos problemos, nedera su panpsichizmu, nes netvirtina, kad sąmoningumas yra aptinkamas nervų sistemos stokojančiuose mechanizmuose ir yra plačiai paplitęs.

Matteo Grasso (2013), remdamasis Chalmerso ontologinių pozicijų taksonomija, pateikė argumentą, parodantį, kad IIT gali būti suderinama ir su B-tipo materializmu (arba fizikalizmu) ir su F-tipo monizmu. Pagal Grasso IIT priima natūralizmą, fizinio pasaulio priežastinį uždaramumą (*causal closure*), aiškinamąją spragą, tačiau ne iki galo aišku ar priima ontologinę spragą. Jei ontologinė spraga nepriimama, IIT suderinama su B-tipo materializmu. Jei IIT priima ontologinę spragą, ji yra suderinama su F-tipo monizmu. Grasso teigus, kad IIT gali būti suderinama su šiomis pozicijomis, tačiau jo argumentą galima praplėsti atsižvelgus į šešeriopą IIT interpretacijų klasifikaciją.

Chalmersas (2002, p. 14) nurodo, kad „dažniausiai pasitaikanti B-tipo materializmo forma laiko, kad fenomenalios būsenos gali būti tapatinamos su tam tikromis fizinėmis ar funkcinėmis būsenomis“. Šis tapatumas, pasak Chalmerso, nėra išvedamas vien konceptualinės analizės būdu, arba, kitaip tariant, nėra gaunamas *a priori*, o yra atrandamas empiriškai. Nors ir fenomenologinio pirmumo, ir eksperimentinio pirmumo IIT interpretacijos teigia esant tapatumui tarp sąmoningumo ir integruotos informacijos, jos išsiskiria požiūriu į tai, kaip šis tapatumo teiginys yra gaunamas. Pirmuoju atveju jis gaunamas *a priori*, antruoju atveju teiginys pateikiamas apibendrinant tam tikrų empirinių tyrimų rezultatus. Minimaliąją IIT interpretacijų klasę sudarančios interpretacijos apskritai nepriima tapatumo tezės. Todėl su B-tipo materializmu suderinama tik eksperimentinio pirmumo IIT interpretacija.

B-tipo materializmas apima ne vieną ontologinę poziciją, tad galima pabandyti kiek tiksliau išsiaiškinti su kokiomis pozicijomis geriau dera eksperimentinio pirmumo IIT interpretacija. Papildomai pabrėžiant, kad eksperimentinio pirmumo IIT yra fizikalizmo atmaina, galima nurodyti, kad ji remiasi Dinaminės šerdies hipoteze. Pastaroji yra fizikalizmo atmaina (Signorelli et al., 2021). Be to, eksperimentinio pirmumo IIT tapatumo tezė priima kaip prielaidą ir taikydama trečiojo asmens perspektyvos tyrimo metodus nustatytas fizines sistemų savybes, tapatina su sąmoningumo apibūdinimu, gautu tiriant pirmojo asmens perspektyvą. Apibūdinimas pasitarnauja kaip tiriamo objekto apibrėžimas. Ši IIT interpretacija nesiūlo išvesti teiginių apie fizines sąmoningumo savybes iš duomenų, gautų naudojant pirmojo asmens perspektyvos metodus, ir *vice versa*. Tačiau pagal ją, sąmoningumas pilnai paaiškinamas fizines sistemas apibūdinančiomis priežastinėmis sąvokomis ir dėl to eksperimentinio pirmumo IIT laikytina reduktyvistine. IIT galėtų būti laikoma tapatumo teorijos atmaina, bet IIT tapatumo tezė skiriasi nuo Smarto ir Feiglį tapatumo tezės, kuri „siūlo mentalines ar psichologines savybes laikyti tapačiomis fiziologiniams smegenų procesams...“ (Dagys, 2012, p. 23). IIT tapatina ne psichinius būvius su neurofiziologiniais smegenų procesais, bet psichinius būvius su specifiniu kompleksiskumu, pavadintu integruota informacija, kuri gali būti realizuota smegenyse.

Chalmersas nurodo, kad F-tipo monizmo versija yra jo paties pateikta teorija (Chalmers, 1996), t. y. Dvigubo aspekto teorija (*Double-Aspect Theory*). Pagal Grasso, IIT gali būti suderinama su F-tipo monizmu, tačiau reikėtų pabrėžti, kad su F-tipo monizmu dera tik fenomenologinio pirmumo IIT interpretacija. Dvigubo aspekto teorijoje (Chalmers, 1996) informacija yra tarpinis dėmuo tarp sąmoningumo ir jo fizinio substrato. Pasak Chalmerso informacija yra fundamentali ir ji turi du aspektus – atitinka fizines ir fenomenalias pasaulio savybes. Esminis skirtumas tarp Chalmerso ir Tononio pozicijų yra toks, kad pastarasis tiksliau apibrėžia informacijos sampratą. Tai leidžia pateikti

tikslesnę teoriją ir padaryti ją falsifikabilesne. Dėl šio informacijos sampratos patikslinimo IIT gali būti laikoma Dvigubo aspekto teorijos tąsa.

Minimaliąją IIT interpretacijų klasę sudarančios teorijos atmeta sąmoningumo ir integruotos informacijos tapatumo prielaidą ir lieka neutrali ontologinių klausimų atžvilgiu. Galima tvirtinti, kad pilnoji IIT koncentruojasi ties „sunkiąją sąmoningumo problema“ (Chalmers, 1996), o minimalioji IIT sprendžia „realiąją sąmoningumo problemą“. Realioji problema nukreipta į įrankių, skirtų paaiškinti, nuspėti ir kontroliuoti sąmoningumą, paieškas (Seth, 2021, p. 30).

Tapatumo tezė, kurią išveda fenomenologinio pirmumo IIT arba kaip prielaidą priima eksperimentinio pirmumo IIT, laikytina esminiu teorijos trūkumu, lemiančiu tiek vidinių teorijos prieštaravimų atsiradimą, tiek nesuderinamumą su kitomis labiau pagrįstomis teorijomis. Toliau išskiriami argumentai, nukreipti prieš IIT tapatumo tezę ir tuo pačiu paremiantys minimaliąją IIT interpretacijų klasę.

## **2.2 Argumentai, paremiantys minimaliąją IIT interpretacijų klasę**

Ankstesniuose skyriuose parodyta, kad tapatumo tezė, kitaip nei teigia fenomenologinio pirmumo IIT, tiesiogiai neplaukia iš fenomenologinių tyrinėjimų, todėl tvirtinama, kad tai yra papildomai įvedama prielaida, kurią priima eksperimentinio pirmumo IIT. Šioje dalyje aptariami argumentai, grindžiantys teiginį, kad vien integruotos informacijos nepakanka tam, kad sistema būtų sąmoninga. Minimalioji IIT interpretacijų klasė nepriima integruotos informacijos ir sąmoningumo tapatumo tezės, todėl argumentai nukreipti prieš šią tezę gali parodyti, kad šią klasę sudarančios IIT interpretacijos yra priimtinausia iš pateiktų.

Norint falsifikuoti IIT pateikiamą tapatumo tezę, reikia surasti tokį mechanizmą, kuris integruoja informaciją, bet nėra sąmoningas. Šis bandymas problemiškas tuo, kad susiduria su papildomų kriterijų poreikiu, t. y. jei IIT tvirtina, kad tam tikras mechanizmas yra sąmoningas, reikia kito kriterijaus, kuris prieštarautų teorijos tvirtinimui. Toliau aptariami trys kriterijai, kurie nederą su IIT pateikiama tapatumo teze: kliovimasis intuicija, IIT išvadų apie sąmoningumą nedarnumas, papildomos sąmoningumu pasižyminčių mechanizmų savybės ir kitos teorijos.

### **2.2.1 Išplėtimo grafiko argumentas ir smegenų plastiškumas**

Vienas iš būdų atmesti tapatumo tezę yra parodymas, kad egzistuoja sistema, pasižyminti dideliu integruotos informacijos kiekiu, bet stokojanti sąmoningumo. Scottas Aaronsonas (2014) parodė, jog priėmus IIT tapatumo tezę, galima daryti išvadą, kad išplėtimo grafikas (*expander graph*) yra

sąmoningas, tačiau, pasak jo, ši išvada yra klaidinga. Aaronsonas nustatė, kad išplėtimo grafikas, pagamintas iš paprastų loginių elementų (*logic gates*), gali integruoti didelį kiekį informacijos. Tokia, atitinkamai sukonstruota, sistema gali turėti didesnę kiekį  $\Phi$  nei žmogaus smegenys. Pagal IIT sistema, integruojanti didelį informacijos kiekį, yra sąmoninga, todėl taip sujungti loginiai vartai yra sąmoningi. Išplėtimo grafikas, Aaronsono teigimu, akivaizdžiai nėra sąmoningas, todėl daroma išvada, kad IIT yra neteisinga.

Tononis (2014) pripažįsta, kad loginiai vartai integruoja didelį informacijos kiekį, tačiau atmeta intuicija paremtą tvirtinimą, kad jie nėra sąmoningi. Pagal Aaronsoną loginių vartų tinklai yra nesąmoningi, nes intuityviai atrodo, kad jie nėra sąmoningi, taigi teiginys apeliuoja į intuiciją. Pasak Tononio, jei teisinga teorija pateikia kontraintuityvias išvadas, reikia kliautis teorija, o ne intuicija.

Pagal Aaronsoną intuicija yra svarbus kriterijus testuojant sąmoningumo teorijos teisingumą, todėl teorijos išvados neturėtų būti jai prieštaraujančios. Toks poreikis grindžiamas sąmoningumo eksperimentiniu neprieinamumu (*experimental inaccessibility*). Aaronsonas teisus dėl eksperimentinio neprieinamumo, arba, kitaip tariant, patyrimo iš pirmojo asmens perspektyvos privatumo. Tačiau žmonių gebėjimas priskirti sąmoningumą tam tikroms sistemoms yra nepatikimas. Žmonės linkę priskirti sąmoningumą būtybėms, pasižyminčioms savybėmis, tokiomis kaip akys, savita judėjimo trajektorija ir priklausoma sąveika (*contingent interaction*) (Arico et al., 2011). Šiomis savybėmis gali pasižymėti stokojantys sąmoningumo objektai, pvz. geometrinės figūros animaciniame filme atliekančios socialinę interakciją primenančius judesius (Heider ir Simmel, 1944). Vien judesiu pakanka, kad žmonėms kiltų noras priskirti joms psichines būsenas (*mental states*). Išplėtimo grafikas stokoja visų šių savybių, todėl mažiau tikėtinas intuityvus sąmoningumo priskyrimas jo atžvilgiu. Tačiau tai dar nereiškia, kad jis negali būti sąmoningas.

Tononio pateiktą argumentą galima panaudoti ir prieš IIT. Mėginant aprašyti esmines sąmoningumo savybes, fenomenologinio pirmumo IIT pradeda iš pirmojo asmens perspektyvos, toks mėginimas taip pat tėra paremtas intuicija. Kaip parodyta pirmame skyriuje, abejotina, kad aprašytos sąmoningumo savybės yra pakankamos, esminės ir teisingos visais atvejais. Žmonių intuicija dėl esminių sąmoningumo savybių išsiskiria. Jei kalbėsime apie eksperimentinio pirmumo teoriją, joje tapatumo tezė nėra pagrįsta, o tik intuityviai priimama, šiuo atveju intuicijos teisingumą būtų galima testuoti, tik, deja, tiesiogiai to padaryti iki šiol nepavyksta.

Vienas iš būdų išsklaidyti įtampą tarp Aaronsono ir Tononio požiūrių – išsiaiškinti ar jų tyrimo objektas yra tas pats. Michelas Cerullo (2015) tvirtina, kad IIT ir Aaronsono vatomas sąmoningumo sampratos nesutampa. Pasak jo, IIT yra paneksperientalistinė teorija, kurios tyrimo objektas yra

subjektyvi patirtis bendresne forma, ir todėl sietina su protosąmoningumu, kuris šiame kontekste suprantamas kaip patyrimas, atskirtas nuo kognicijos. Aaronsono sąmoningumo samprata laikoma labiau atitinkančia kognityviniuose moksluose aptariamą subjektyvų patyrimą, sietiną su žmonėmis ar gyvūnais. Kaip parodyta anksčiau, IIT tyrinėja fenomenalųjį sąmoningumą ir nėra paneksperimentalinė teorija. Atmetus tvirtinimą, kad Tononis ir Aaronsonas vartoja skirtingas sąmoningumo sampratas, jų ginčas lieka neišspręstas.

Nors Aaronsonas kritikuodamas IIT pirmiausia apeliuoja į intuíciją, tačiau tai leidžia iškelti prielaidą, kad integruota informacija yra tik viena iš savybių, kuriomis turi pasižymėti už sąmoningumą atsakingas mechanizmas. Parodžius kokių savybių nepaiso IIT galima paremti Aaronsono poziciją. Pasak jo, IIT problemiška dėl to, kad vertina sistemos sąmoningumą pagal tos sistemos integruotos informacijos kiekį ir nepaiso struktūrinių ir funkcinių tos sistemos ypatumų.

Toks Aaronsono spėjimas nesusilpnina tvirtinimo, kad integruota informacija yra būtina sąmoningumo sąlyga, abejojama tik jos pakankamumu. Integruota informacija koreliuoja su sąmoningumu, jei teiginys, kad patyrimo grafikas, pasižymintis dideliu integruotos informacijos kiekiu, nėra sąmoningas, yra teisingas, tai dar nereiškia, kad visos sąmoningumu pasižyminčios sistemos neturi atitikti šio kriterijaus (integruoti informaciją). Tikėtina, kad integruota informacija yra vienas iš būtinų, bet nepakankamų kriterijų. Todėl išplėtimo grafiko argumentas negalioja minimaliajai IIT klasei, nes ji netvirtina esant tapatumo ar būtinumo tarp sąmoningumo ir integruotos informacijos, tik nustato koreliacinius ryšius.

Vienas iš galimų kriterijų yra sistemos aktyvumas. Išplėtimo grafikas yra neaktyvi sistema. Pagal IIT neaktyvios sistemos gali būti sąmoningos (Oizumi et al., 2014, p. 17). Pasyvumas šiame kontekste gali turėti dvejopą prasmę, sistema gali būti atskirta nuo informacijos patekimo iš išorės, t. y. ji gali stokoti įvesties ir išvesties signalų, arba sistema gali būti neaktyvioje būsenoje. Pirmojo pasyvumo pavyzdys galėtų būti patyrimo salos (*islands of awareness*), kurios yra atskirtos nuo išorinės informacijos (Bayne et al., 2020). Keliamo hipotezė, kad patyrimo salos galimai atsiranda iš kūno išimtose smegenyse (*ex cranio brains*), po neurochirurginės procedūros, vadinamos hemisferektomija (*hemispherotomy*), ir dirbtinai išaugintuose smegenų organoiduose (*cerebral organoids*). Remiantis IIT tokios patyrimo salos yra įmanomos. IIT priskiria sąmoningumą ir gerokai paprastesniems mechanizams už tokias patyrimo salas ir net už anksčiau minėtą išplėtimo grafiką. Pasak Tononio (2014) net nedarantis visiškai nieko 2D tinklelis (*grid*), su visais esančiais įėjimais (*gates*) išjungtais, gali būti sąmoningas. Taip suprantamas pasyvumas prieštarauja eksternalistinei pozicijai, pagal kurią sąmonė yra ne tik smegenų produktas, bet ir priklauso nuo išorinių veiksnių. Kognityvinių mokslų

kontekste, įprastai eksternalistinės pozicijos atstovai koncentruojasi ties kognityviniais aspektais (Clark ir Chalmers, 1998), tačiau yra ir siejančių fenomenalųjį sąmoningumą su dinamiška sąveika, vykstančia tarp smegenų, kūno ir pasaulio (Kirchhoff ir Kiverstein, 2018).

Antru atveju, pasyvumas suprantamas kaip sistemos elementų neaktyvumas. IIT numato, kad „net jei visi dominuojančio komplekso neuronai yra neaktyvūs, jie vis tiek gali generuoti sąmoningumą tol, kol jie yra pasiruošę sužadimui“ (Tononi ir Koch, 2015). Teigiama, kad tokia neurofiziologinė būseną galimai pasiekama tam tikromis meditacinėmis praktikomis, kurių metu siekiama patirti gryną suvokimą be turinio. Tai prieštarauja įsitikinimui, kad neuronai gali būti atsakingi už sąmoningumą tik tuomet, kai yra aktyvūs ta prasme, kad jie gali pernešti (*broadcast*) informaciją ir sužadinti priekinę parietalinę (*frontoparietal*) smegenų dalį. Tokio įsitikinimo laikosi GWT proponentai. Taigi IIT ir GWT išvados nesutampa mechanizmų aktyvumo klausimu.

Galima apsvarstyti ir trečią aktyvumo aspektą – sistemos struktūrinius pokyčius. Geraldo Edelmano pateikiamoje Nervinio Darvinizmo teorijoje išryškunami trys atrankos principai – raidos atranka (*developmental selection*), patyriminė atranka (*experiential selection*) ir pakartotinis įėjimas (*re-entry*) (Blackmore, 2018). Raidos atranka vyksta tuomet, kai smegenys vystosi ir neuronai šakojasi įvairiomis kryptimis, sudarydami labai sudėtingas struktūras. Tuomet vyksta genėjimas, po kurio išlieka daugiausiai naudotos struktūros, kurios sudaro ilgai išliekančias funkcinės grupes. Patyriminė atranka, kurios metu tam tikros neuronų sinapsės yra sustiprinamos arba susilpninamos, vyksta viso gyvenimo metu. Tiek vykstant raidos, tiek patyriminei atrankai, keičiasi neuronų tarpusavio ryšiai, tai reiškia, kad smegenys keičiasi struktūriškai. Pakartotinio įėjimo procesas yra aiškinamas jau minėtoje Dinaminės šerdies hipotezėje, kuri yra sudėtinė Nervinio Darvinizmo teorijos dalis. Tiek raidos atranka, tiek patyriminė atranka nevyksta kai kuriose sistemose, kurias IIT laiko sąmoningomis. Nervinio Darvinizmo teorija sąmoningumą tyrinėja iš biologinės perspektyvos, o pagal IIT „jei mechanizmas gali generuoti integruotą informaciją, nepriklausomai nuo to, ar jis yra organinis ar ne, nepriklausomai nuo to, ar sudarytas iš neuronų ar silicio lustų, ir nepriklausomai nuo jo gebėjimų pateikti žodinį pareiškimą, jis turės sąmoningumą“ (Tononi, 2008, p. 237). IIT priskiria sąmoningumą ne tik biologiniams nervų sistemą turintiems mechanizms, bet ir tokiems mechanizms kaip jau minėtas išplėtimo grafikas ar fotodiodas. Šių mechanizmų struktūra, integruojant informaciją, išlieka stabili. Iš esmės smegenų integruotos informacijos kiekis gali kisti, o fotodiodo ar išplėtimo grafiko atvejais jis lieka pastovus.

Fotodiodo atveju galimos dvi skirtingos būsenos, žmogaus atveju yra labai daug būsenų (Tononi, 2004). Nors neaišku, koks tikslus galimų būsenų skaičius būdingas žmogaus smegenims, galima



pamanyti esant begalinį skaičių galimų variantų. Taip yra dėl to, kad patyrimo turinys niekada nebūna toks pat, bet yra nuolat besikeičiantis. Kad ir koks panašus, jis niekuomet nebūna identiškas. Galima teigti, kad smegenys iš esmės gali patirti neribotą kiekį skirtingų patyrimų. Ši patyrimų įvairovė veikiausiai galima todėl, kad „smegenys be paliovos keičiasi sinaptiniame lygmenyje su kiekvienu patyrimu (Askenasy ir Lehmann, 2013). Tokia smegenų gebėjimo keistis ir persitvarkyti tiek struktūriškai, tiek funkciškai savybė vadinama neuroplastiškumu. Kaip parodyta, neuroplastiškumas yra svarbi Nervinio Darvinizmo teorijos dalis, vykstanti tiek raidos atrankos, tiek vėliau patyriminės atrankos metu.

Smegenų plastiškumas galimai yra viena iš kliūčių matuojant integruotą informaciją. Kaip minėta anksčiau, norint nustatyti vidinę informaciją, reikia žinoti visus galimus jos veikimo būdus, net jei sistema niekada tam tikru būdu nesielgs, bet potencialiai galėtų. Net jei sistema yra labai sudėtinga, bet nekintanti, kaip Aaronsono aptartas išplėtimo grafikas, iš esmės įmanoma nustatyti jos integruotos informacijos kiekį. Tačiau problema kyla tada, kai sistema yra nuolatos besikeičianti, būtent tokios yra smegenys, pasižyminčios neuroplastiškumu.

Neuroplastiškumo ir sąmoningumo ryšį galima suprasti dvejopai. Vienu atveju vyksta diskusija dėl sąmoningumo priežastinio ryšio (Askenasy ir Lehmann, 2013). Kitu atveju neuroplastiškumas gali būti suprantamas kaip svarbi smegenų savybė, dėl kurios atsiranda sąmoningumas. Šiame argumente aktuali antroji diskusija.

Pirmiausia neuroplastiškumas plačiai paplitęs gyvūnų karalystėje, tiek stuburinių tarpe, tiek kai kurių bestuburių (Pyza, 2013). Galima pastebėti, kad ši smegenų savybė koreliuoja su sąmoningumu, t. y. gyvūnai, pasižymintys sąmoningumu, turi daugiau ar mažiau plastiškas smegenis. Taigi tikėtina, kad vienas iš mechanizmų, generuojančių sąmoningumą, kriterijų yra plastiškumas.

Neuroplastiškumas gali būti siejamas su sąmoningumo turiniu tiesiogiai. Ši sąsaja išryškėja eksperimentuose, parodančiuose, kad neuroplastiškumo dėka gali keistis sąmoningumo turinys. Tokiuose eksperimentuose parodoma, kad įmanoma išmokyti patirti naujus sąmoningus patyrimus, pvz. atskirti kvapus, kurie anksčiau atrodė vienodi, arba pakeitus genus, lemiančius naujų receptorių susidarymą, beždžionių ir pelių smegenys prisitaiko skirti naujas spalvas (Frith, 2011). Nors neuroplastiškumas koreliuoja su įvairiomis smegenų funkcijomis, šie eksperimentai demonstruoja, kad jis tiesiogiai susijęs su sąmoningu patyrimo turiniu. Tad tikėtina, kad be neuroplastiškumo sąmoningumas nebūtų įmanomas. Jei neuroplastiškumas yra būtina fizinio mechanizmo savybė tam, kad jis galėtų būti sąmoningas ir pilnoji IIT teigia, kad sistemos, kurios nepasižymi neuroplastiškumu, yra sąmoningos, tai pilnosios IIT išvados yra klaidingos.

Taigi, IIT tvirtinimas, kad neaktyvios sistemos yra sąmoningos, gali būti suprantamas trejopai. Priklausomai nuo to, kuris neaktyvumas pabrėžiamas, pilnoji IIT yra nesuderinama su GWT išvadomis, taip pat nesuderinama su eksternalistinėmis sąmoningumo teorijomis ir neatsižvelgia į smegenų plastiškumo savybę.

## 2.2.2 Informacijos integravimas ir nesąmoningi procesai

Šioje dalyje parodoma, kad fenomenologinio pirmumo IIT interpretacija nepriima skirties tarp sąmoningų ir nesąmoningų procesų ir dėl to susiduria su problemomis. Viena jų yra vidiniai teorijos prieštaravimai, siejami su fotodiodo mintiniu eksperimentu ir gilaus miego būsenų traktavimu. Be to, IIT susiduria su panpsichizmui priskiriama kombinacijos problema (Chalmers, 2017), neišvengia nesudedamumo principo (Schwitzgebel, 2014) pažeidimo ir susiduria su sąmoningumo laipsniškumo paaiškinimo problema.

Davidas Rosenthalis (1993) skiria būsenos sąmoningumą (*state consciousness*) nuo būtybės sąmoningumo (*creature consciousness*). Ši skirtis naudinga apibūdinant IIT. Pastaroji tyrinėja skirtumą tarp nubudusio ir komoje ar gilaus miego būsenoje esančio žmogaus arba, kitaip tariant, mėgina paaiškinti, kas skiria būdraujančią būtybę nuo stokojančios sąmoningumo būtybės. Tai eksplicitiškai išreikšta plačiajame sąmoningumo apibrėžime. Tačiau teorija susiduria su būsenos sąmoningumo aiškinimo sunkumais, nes neskiria sąmoningų būsenų nuo nesąmoningų.

Didžioji dalis procesų smegenyse vyksta nesąmoningai (Baars, 1997). Taikant pirmojo asmens perspektyvos tyrimo metodus, pvz. introspekciją ar atliekant fenomenologinę analizę, nesąmoningi procesai nėra pasiekiami (Doerig et al., 2020). Todėl nesąmoningi procesai turi būti tyrinėjami iš trečiojo asmens perspektyvos. Teorijos pagrindu pasirenkant tyrimą iš pirmojo asmens perspektyvos, kaip tai daro fenomenologinio pirmumo IIT, skirtis tarp sąmoningų ir nesąmoningų procesų, vykstančių smegenyse, iš principo negali būti įvedama.

Skirtis tarp sąmoningo ir nesąmoningo informacijos apdorojimo grindžiama gausiai replikuotais regimosios užuominos (*priming*) eksperimentais ir neurologinių sutrikimų turinčių pacientų tyrimais (Dehaene 2001). Chalmerso (2004) teigimu, sąmoningumo teorija turėtų paaiškinti šią skirtį.

Sąmoningo ir nesąmoningo informacijos apdorojimo skirtumas išsamiai aptariamas Globalaus darbalaukio teorijoje, pristatytoje Bernardo Baarso (1988) ir toliau plėtojamoje Stanislaso Dehaene'o (1998) pakeitus pavadinimą į "Globalaus nervinio darbalaukio teorija". Šiame darbe globalaus nervinio darbalaukio teorija traktuojama kaip globalaus darbalaukio teorijos tąsa, papildanti pastarąją neuromoksliniais duomenimis. Globalus darbalaukis yra informacijos keitykla (*exchange*), leidžianti

specializuotiems nesąmoningiems procesams, vykstantiems nervų sistemoje, sąveikauti tarpusavyje (Baars, 1988, p. 61). Vienas pagrindinių GWT sprendžiamų klausimų – kodėl nesąmoningas informacijos apdorojimas yra toks platus, o sąmoningi procesai tokie riboti (Baars, 1997). GWT mėgina nustatyti su sąmoningumu susijusias funkcijas. Pagrindine sąmoningumo funkcija laikoma prieigos (*access*) funkcija.

GWT pateikia teatro metaforą (Baars, 1997), nusakančią kaip funkcionuoja sąmoningumas. Metaforoje aprašomą teatrą sudaro penkios dalys. Teatro scena – tai darbinė atmintis. Dėmesį reprezentuoja sceną apšviečiantys prožektoriai, pasak Baarso, tik įvykiai esantys ryškiajame prožektoriaus apšviestame lauke yra sąmoningi. Aktoriai, t. y. potencialios mintys, vaizdiniai ar pojūčiai, mėginantys patekti į apšviestą lauką scenoje, yra sąmoningumo turinys. Aktoriai pasirodo po vieną ir yra keičiami kitų aktorių. Aktoriai kalba publikai, kuri nėra apšviečiama, t. y. lieka nesąmoninga. Publika suprantama kaip specializuoti nesąmoningi moduliai, atliekantys specifinę funkciją smegenyse. Taip pat ir vykdančiąją (*executive*) funkciją atliekantis režisierius nėra apšviečiamas, tačiau jis formuoja tai, kas vyksta prožektoriaus apšviestame lauke. Pagal GWT kontekstas, kuris nėra sąmoningas, daro įtaką sąmoningumui. Teatro metaforoje siejamas sąmoningumas ir kitos kognityvinės funkcijos. Metafora atspindi ribotą sąmoningumo dalį kognityvinių procesų visumoje – tik maža dalis to, kas vyksta smegenyse, vyksta sąmoningai. Sąmoningo ir nesąmoningo informacijos apdorojimo skirtis yra esminis GWT ir IIT skiriantis aspektas.

IIT netraktuoja sąmoningumo kaip informacijos apdorojimo, o traktuoja kaip priežastinį sistemos pajėgumą (*causal power*) keisti save (*make a difference to itself*). Fenomenaliojo pirmumo IIT pradeda nuo fenomenaliojo sąmoningumo aprašymo, siekia nustatyti esmines jo savybes, bet šių savybių nesieja su kognityvinėmis funkcijomis. IIT implicitiškai glūdi kognicijos ir sąmoningumo skirties prielaida.

Aprašyti tam tikrą neurofiziologinį reiškinių galima tiek vartojant sąvokas „informacijos integravimas“, tiek „informacijos apdorojimas“. Jei sąmoninga būseną skiriasi nuo nesąmoningos būsenos tuo, kad pirmuoju atveju informacija yra integruojama, o antruoju atveju integruotumo stokojama, tuomet kyla problemų aiškinant nesąmoningą informacijos apdorojimą. Nustačius, kad sistema, vykdanči nesąmoningą informacijos apdorojimą, generuoja integruotą informaciją, toks procesas, remiantis IIT, turėtų būti sąmoningas. Priimant šiuos teiginius vienu metu, gaunama kontradikcija, nes nesąmoningas informacijos apdorojimas laikomas minimaliai sąmoningumu. Sprendžiant šią dilemą arba reikia atsisakyti skirties tarp sąmoningo ir nesąmoningo informacijos apdorojimo ir priimti, kad visas informacijos apdorojimas smegenyse vyksta daugiau ar mažiau sąmoningai, arba pripažinti, kad integruota informacija nėra pakankama sąlyga sąmoningumui. Tokiu

atveju tektų sutikti, kad įmanomi mechanizmai, integruojantys informaciją, bet stokojantys sąmoningumo. Iš principo šios sąvokos yra nesuderinamos tik tuo atveju, jei priimama tapatumo tezė.

Dėl sąmoningų ir nesąmoningų procesų skirties praleidimo IIT neišvengia ir vidinio prieštaravimo. Pirmiausia galima pastebėti neatitikimą tarp mintiniuose eksperimentuose gautų išvadų bei išvadų priėmus tapatumo tezę. Pirmame mintiniame eksperimente pabrėžiant fotodiodo ir žmogaus skirtumą nurodoma, kad fotodiodas nėra sąmoningas: „fotodiodas taip pat gali atskirti išjungtą arba įjungtą ekraną, bet, veikiausiai, jis „nemato“ šviesos ir tamsos sąmoningai“ (Tononi, 2004, p. 2). Taip pat, lyginant fotodiodą su akies tinklainėje esančiais receptoriais, tvirtinama, kad „akies tinklainėje esančių neuronų aktyvumas nėra tiesiogiai asocijuojamas su sąmoningu šviesos ir tamsos patyrimu – jie elgiasi tarsi biologiniai fotodiodai, siunčiantys signalus į aukštesnius centrus“ (Tononi, 2004, p. 20), bet priėmus tapatumo tezę jam yra priskiriamas minimalus sąmoningumas „IIT suponuoja, kad net binarinis fotodiodas nėra visai nesąmoningas, bet veikia patiria lygiai vieną bitą sąmoningumo“ (Tononi, 2008). Kitoje vietoje tvirtinama, kad „net tokios paprastos schemos kaip fotodiodas, sudarytas iš sensoriaus ir atminties elemento, gali turėti nedidelį kiekį patyrimo“ (Tononi ir Koch, 2015, p. 11). IIT proponentai galėtų atsakyti, kad mintiniame eksperimente daroma išvada yra ikiteorinė ir todėl tik paremta intuicija. Tačiau šis mintinis eksperimentas yra teorijos dalis ir išreiškia esminę sąmoningumo savybę – informatyvumą. Mintiniu eksperimentu pabrėžiama didelio informatyvumo sąsaja su sąmoningumu ir tik įvedus tapatumo tezę bet koks kiekis integruotos informacijos pradedamas sieti su sąmoningumu. Įvedus sąmoningų ir nesąmoningų procesų skirtį, fotodiodą, integruojantį minimalų informacijos kiekį, būtų galima traktuoti kaip nesąmoningą.

Pagrindinį smegenyse esantį kompleksą IIT sieja su sąmoningumu (Tononi, 2012). Smegenyse be pagrindinio aptinkami ir mažesni kompleksai. Pastarieji atlieka nesąmoningo informacijos apdorojimo funkciją. Mažesni kompleksai veikia kaip pagrindinio komplekso įėjimo ar išėjimo kanalai. Tuo tarpu jų vidinė informacinė struktūra išlieka izoliuota nuo pagrindinio komplekso. Didesnis kompleksas nuo mažesnių skiriasi tuo, kad jo integruotos informacijos kiekis yra didesnis. Kadangi IIT priima fenomenaliojo sąmoningumo ir integruotos informacijos tapatumo tezę, galima daryti išvadą, kad mažesni kompleksai taip pat yra sąmoningi. Tokiu atveju tektų pripažinti, kad smegenyse yra daug atskirų sąmoningumų. Tačiau žvelgiant iš pirmojo asmens perspektyvos, patiriamas tik vienas sąmoningumas, o kiti sąmoningumai iš pirmojo asmens perspektyvos lieka neprieinami.

Neuronų tinklai su dideliu struktūriniu modularumu generuoja mažą integruotos informacijos kiekį (Toker ir Sommer, 2019). Tokios struktūros pavyzdys yra smegenėlės (*cerebellum*). Nors jos sudarytos iš didesnio kiekio neuronų nei smegenų žievė, jos nėra atsakingos už sąmoningumą. Pagal

IIT, tai lemia jų anatomicinę struktūrą, besiskirianti nuo smegenų žievės, sudarytos iš specializuotų elementų, kurie tarpusavyje greitai ir efektyviai sąveikauja, taip generuodami aukštą integruotos informacijos kiekį, tuo tarpu smegenėlės yra sudarytos iš mažų modulių, turinčių atskiras įvestis bei išvestis ir veikiančių nepriklausomai vienas nuo kito (Tononi ir Koch, 2015, p. 10). Smegenėlių struktūrą galima sulyginti su IIT aptariamąs fotokameros struktūra. Fotokamera yra sudaryta iš atskirų, tarpusavyje nesąveikaujančių, paraleliai veikiančių fotodiodų. Tiek smegenėlės, tiek ir fotokamera, pagal IIT, nėra sąmoningos, tačiau jei IIT priskiria minimalų sąmoningumą fotodiodui, tuomet analogiškai sąmoningumas priskirtinas ir smegenėles sudarančiam atskiram moduliu, kuris integruoja tam tikrą nedidelį kiekį informacijos. Taigi, jei IIT tapatumo tezė yra teisinga, tuomet nesąmoninga sistema gali būti sudaryta iš sąmoningų dalių. Galima suabejoti ar vertinant kokio nors žmogaus sąmoningumo būklę yra svarbus tik pagrindinis kompleksas ar ir „kiti“ sąmoningumai yra svarbūs.

Panaši problema kyla ir aiškinant besapnio miego atvejį. Nors jo metu integruotumas ženkliai mažesnis nei būdraujant, vis dėl to jis nėra nulinis. Pagal IIT pateikiamą platų sąmoningumo apibrėžimą besapnis miegas laikytinas nesąmoninga būseną (Tononi, 2008; Tononi et al., 2016). Kaip minėta anksčiau sapnų tyrinėjimai padeda pagrįsti sąmoningumo ir didelio integruotos informacijos kiekio koreliaciją. Tokiu atveju problemų neįkyla. Tačiau priėmus tapatumo tezę, atsiranda prieštaravimas: viena vertus besapnis miegas yra nesąmoninga būseną, kita vertus jo metu integruotos informacijos kiekis nėra nulinis (Tononi ir Koch, 2015) ir tai reiškia, kad tokia būseną laikytina sąmoninga. Norint išvengti prieštaravimo tektų atmesti platųjį sąmoningumo apibrėžimą arba tapatumo tezę. Tapatumo tezę ginantis argumentas galėtų būti toks: besapnio miego atveju fenomenalusis sąmoningumas išlieka, tik nubudus nėra galimybės prisiminti, kas buvo patiriama. Tačiau šis tvirtinimas yra problemiškas tuo, kad jo nėra kaip patikrinti. Taip pat, jei pagrindinis kompleksas nebėra integruotas ir lieka tik tam tikras kiekis mažesnių kompleksų (Tononi, 2015), neaišku, ar besapnio miego metu smegenys generuoja vieną patyrimą ar tiek, kiek yra atskirų kompleksų.

Dar viena problemiška situacija yra su perskirtų smegenų (*split brain*) pacientais, IIT prognozuoja, kad smegenų pusrutulių atskyrimas (tokios operacijos atliekamos epilepsija sergantiems pacientams) veda prie sąmoningumo sudvejinimo, t. y. tiek viename, tiek kitame pusrutulyje atsiranda atskiras sąmoningumas (Tononi ir Koch, 2015). Replikuojant tyrimus (Pinto et al., 2017a; Pinto et al., 2017b) gaunamos išvados neatitinka ankstesnių tyrimų išvadų ir paremia teiginį, kad smegenų pusrutulių atskyrimas neveda prie sąmoningumo sudvejinimo. Pagal šiuos tyrimus pacientams su perpjauta didžiąją smegenų jungtimi buvo atskirtas suvokimas, bet ne sąmoningumas. Suvokimas, šiuo atveju, suprantamas kaip nesąmoningai galintis vykti procesas. Kadangi diskusija dėl atskirtų smegenų

tebevyksta ir galutinis sutarimas vis dar nepasiektas (de Haan et al., 2020), šiuo klausimu vieni nesutartę su IIT išvadoms, tačiau atsirastų sutinkančių su jomis.

Nepaisant to, kad galutinių išvadų nėra, argumentą galima pritaikyti kritikuojant IIT laipsniškumo tezę. Žinoma, kad perskirtų smegenų pacientai elgiasi įprastai ir nurodo nejaučiantys jokio skirtumo tarp būsenos iki operacijos ir po jos (de Haan et al., 2020). Tai verčia suabejoti sąmoningumo laipsniškumu. Jei tokie pacientai turi du atskirtus kompleksus su mažesniu integruotos informacijos kiekiu, tai pagal IIT dėl to kiekvienas iš kompleksų yra mažiau sąmoningas. Kadangi jie yra mažiau sąmoningi, remiantis IIT, tektų tvirtinti, kad tokių pacientų sąmoningas patyrimas prieš ir po operacijos turėtų labai stipriai skirtis. Tačiau taip nėra, tad pasidaro neaišku, ką reiškia būti mažiau sąmoningu, bet jaustis taip pat.

Be atskirtų smegenų pusrutulių atvejo, apsvarstomi hipotetiniai scenarijai, kuomet atliekamas atvirkščias veiksmas – atskirų smegenų apjungimas. Kritikuodamas, kaip įvardina prieštaringas materializmo išvadas, Ericas Schwitzgebelis (2014) pateikia nesudedamumo principą (*anti-nesting principle*). Pagal šį principą „negali būti jokių sudėtų fenomenaliojo sąmoningumo formų (*nested forms*) ir, dėl to, vientisas sąmoningumas negali būti sudarytas iš dalių, kurios ir pačios yra sąmoningos“. Nepritaikius šio principo tenka pripažinti, kad Jungtinės Valstijos, sudarytos iš sąmoningų asmenų, pačios yra sąmoningos. Schwitzgebelis pastebi, kad IIT šis principas išreiškiamas išskyrimo postulatu. Pagal IIT Jungtinės Valstijos yra mažiau integruotos nei jas sudarančių piliečių smegenys. Kitais žodžiais tariant Jungtinės Valstijos kaip visuma yra mažiau nei ją sudarančios dalys (piliečių smegenys), nes neuronų ir jų sinapsių kiekis smegenyse pranoksta tarpusavyje sąveikaujančių Jungtinių Valstijų piliečių skaičių.

Tačiau nors IIT nelaiko Jungtinių Valstijų sąmoningomis, nesudedamumo principas yra pažeidžiamas. Kochas (2019) pateikia mintinį eksperimentą, kuriame apibūdinama hipotetinė situacija, kuomet sujungiamos dvi atskiros smegenys ne tik įvesties ir išvesties kanalais kaip prieš tai pateiktame pavyzdyje, bet tiesiogiai. Glaustai jį galima pristatyti taip: po truputį kuriant ryšius tarp dviejų atskirų  $\Phi$  įvyksta lūžis, kuomet bendra  $\Phi$  viršija atskirose smegenyse esančią  $\Phi$ , įvykus šiam lūžiui, du atskiri sąmoningumai staiga tampa vienu sąmoningumu. Jei toks smegenų sujungimas, kokį aprašo Kochas, būtų realizuotas, pagal IIT du atskiri sąmoningumai taptų vienu bendru. Taigi, išskyrimo postulatą nepadeda išvengti nesudedamumo principo.

Nesudedamumo principas taikomas materialistinėms teorijoms, o panpsichizmas susiduria su kombinacijos problema. Nors pateikiamos skirtingos kombinacijos problemos formuluotės (Chalmers, 2017), bendrais bruožais ją galima išreikšti klausimu – kaip daugybė subatominių dalelių, turinčių savo

sąmoningumą, susijungia, kad sudarytų sąmoningas gyvūno smegenis? Nors IIT nepriskiria sąmoningumo konkrečiai subatominiams dalelėms, tačiau neišvengia kombinacijos problemos, nes pagal ją sąmoninga sistema gali būti sudaryta iš sąmoningų dalių. Kadangi pavieniai neuronai integruoja informaciją, jie pagal IIT laikytini minimaliai sąmoningais. Šie sąmoningi neuronai yra sąmoningų smegenų sudedamoji dalis. Vadinasi, pagal IIT sąmoningas mechanizmas gali būti sudarytas iš sąmoningų dalių. IIT galėtų išvengti šios problemos priimdama sąmoningų ir nesąmoningų procesų skirtį, bet tai pavyktų tik atmetus tapatumo tezę. Tokiu atveju pavieniai neuronai nebūtų laikomi sąmoningais.

Su šiomis problemomis susiduria fenomenologinio pirmumo IIT, nes priima tapatumo tezę. Tačiau ji nepriimama minimaliosios IIT interpretacijos. Kadangi pastaroji atmeta tapatumo tezę, ji gali priimti sąmoningų ir nesąmoningų procesų skirtį ir išvengti aptartų problemų.

### **2.2.3 Minimalus integruotos informacijos kiekis ir numatantysis apdorojimas**

Šioje dalyje parodoma, kad IIT nėra suderinama su Kontroliuojamos haliucinacijos (*Controlled hallucination*) teorija (Seth, 2021), pritaikančia numatančiojo apdorojimo principą specialiai fenomenaliojo sąmoningumo aiškinimui.

Skirtumas tarp Kontroliuojamos haliucinacijos teorijos ir numatančiojo apdorojimo (*predictive processing*) teorijos yra tyrimo objektas. Numatančiojo apdorojimo teorija (Clark, 2013) yra platesnė teorija, aiškinanti smegenyse vykstančius procesus, atsakingus už suvokimą (*perception*), kogniciją (*cognition*) ir veiksmus (*action*), o Kontroliuojamos haliucinacijos teorija siekia paaiškinti fenomenaliąsias sąmoningo suvokimo savybes.

Anilas Seth'as (2021) lygindamas sąmoningumo tyrinėjimus su gyvybės tyrimais biologijoje argumentuoja, kad sąmoningumas nėra vienas fenomenas ir reikia išskirti skirtingus tyrimo būdus, tokius kaip sąmoningumo lygis, sąmoningumo turinys, savastis (*self*). Seth'as išsamiai analizuoja IIT sąmoningumo lygio matavimo kontekste, tačiau svarų argumentą prieš IIT tapatumo tezę galima įžvelgti ir sąmoningumo turinio kontekste. Tam galima pasitelkti paties Seth'o plėtojamą Kontroliuojamos haliucinacijos teoriją ir palyginti jos išvadas su IIT daromomis išvadomis. Toks palyginimas iki šiol nėra atliktas, veikiausiai, dėl iš pažiūros skirtingų sąmoningumo aspektų, ties kuriais koncentruojasi teorijos. Tačiau toks bandymas jas palyginti gali padėti surasti papildomą sąmoningumo kriterijų ir geriau paaiškinti anksčiau aptartus kontroversiškus atvejus, tokius kaip išplėtimo grafikas ar fotodiodas.

Remdamasis Bayeso smegenų hipoteze, pagal kurią smegenys yra „prognozavimo mašina“ (*prediction machine*) ir viskas, ką mes matome, girdime ar jaučiame yra smegenų „geriausias spėjimas“ (*best guess*), Seth'as (2021) grindžia tvirtinimą, kad sąmoningumo turinys yra kontroliuojama haliucinacija. Sąmoningumas tuo pat metu yra ir daugiau, ir mažiau nei tikrovė iš tiesų yra. Subjektyvus patyrimas, apspręstas „iš viršaus į apačią“ (*top-down*) spėjimais, bet ne „iš apačios į viršų“ (*bottom-up*) sensoriniais signalais. Tai reiškia, kad patiriami ne patys sensoriniai signalai, o tai, kad patyrimas susijęs su šių signalų interpretacijomis. Tokia laikysena priešinama „kaip dalykai atrodo“ (*how things seem*) požiūriui. Šis požiūris yra vienakryptis, t.y. stimulai, pvz. šviesa ar garsas, paveikia sensorinius organus, kurie siunčia elektrinius signalus, apdorojamus atskiruose moduluose. Kontroliuojamos haliucinacijos teoriją galima priešinti su naiviojo realizmo pozicija.

Seth'as išskiria tris pagrindines Kontroliuojamos haliucinacijos teorijos sudedamąsias dalis. Pagal pirmąją, „smegenys nuolatos daro spėjimus apie sensorinių signalų priežastis, spėjimus, besileidžiančius iš viršaus į apačią smegenų suvokimo hierarchijos kryptimi“ (Seth, 2021, p. 85). Pagal „antrąją, sensoriniai signalai, patenkantys į smegenis iš apačios į viršų, arba iš išorės į vidų, palaiko spėjimus, naudingai susietus, su jų priežastimis: šiuo atveju kavos puodeliu“ (Seth, 2021, p. 85). Šie signalai įvardinami kaip spėjimo klaidos (*prediction errors*), registruojančios skirtumą tarp to, ko smegenys tikisi ir to, ką gauna kiekviename apdorojimo lygyje. Svarbiausią mintį Kontroliuojamos haliucinacijos teorijoje išreiškia teiginys, kad „patyriminė patirtis (*perceptual experience*) – šiuo atveju subjektyvus puodelio patyrimas – yra sąlygotas turinio, gauto iš viršaus į apačią spėjimų, bet ne iš apačios į viršų sensorinių signalų“ (Seth, 2021, p. 86). Pabrėžiama, kad patys sensoriniai signalai nėra patiriami, o patiriamos tik jų interpretacijos. Suvokimas (*perception*) laikomas kontroliuojama haliucinacija, o haliucinacija laikoma nekontroliuojamu suvokimu. Laikantis tokios pozicijos, smegenyse visuomet kuriamas tam tikras pasaulio modelis ir pasaulis niekada nėra suvokiamas toks, koks yra iš tiesų.

Lyginant Kontroliuojamos haliucinacijos teoriją su IIT galima pabrėžti, kad pirmoji nieko netvirtina apie sąmoningumo lygį, tik mėgina paaiškinti, kaip atsiranda sąmoningumo turinys, tuo tarpu IIT sąmoningumo lygio ir turinio tyrimai persikloja. IIT siekis nustatyti, kurios sistemos yra sąmoningos, yra lydimas sąmoningumo turinio tyrinėjimo. Sąmoningumo turinio skirtumas tarp žmogaus ir fotodiodo IIT nusakomas kiekybiškai. Fotodiodo mintiniame eksperimente nurodomas skirtumas tarp potencialiai gausaus fenomenaliojo sąmoningumo turinio žmogaus atveju ir minimalaus turinio, sudaryto iš dviejų galimų variantų, fotodiodo atveju.



Kitas svarbus aspektas yra tai, kad tiek Kontroliuojamos haliucinacijos teorijos, tiek IIT (išskyrus neutraliosios interpretacijos) tyrimo objektas yra fenomenologinis sąmoningumas. Dėl bendro tyrimo objekto teorijas galima lyginti tarpusavyje. Kontroliuojamos haliucinacijos teorija subjektyvų patyrimą sieja su „iš viršaus į apačią“ prognozėmis, bet ne su „nuo apačios į viršų“ sensoriniais signalais. Jei Kontroliuojamos haliucinacijos teorija yra teisinga, tuomet visi sąmoningi mechanizmai turėtų veikti pagal šį principą. Todėl norint išsiaiškinti, ar integruota informacija yra būtina ir pakankama sąmoningumo sąlyga, reikia išsiaiškinti, ar visais atvejais nustatys, kad sistema integruoja informaciją, ji tuo pačiu neprieštarauja „iš viršaus į apačią“ veikimo principui.

Minimalioji IIT interpretacijų klasė koncentruojasi ties smegenų tyrimais. Tokiu atveju didesnių problemų nekyla, smegenys tuo pat metu ir integruoja didelį kiekį informacijos, ir, Andy Clarko (2013) terminais tariant, yra „prognozavimo mašina“ (*prediction machine*). Taigi koreliacija tarp sąmoningumo ir didelio integruotos informacijos kiekio smegenyse suderinama su numatomojo apdorojimo teorija, konkrečiau, su Kontroliuojamos haliucinacijos teorija.

Kitaip yra su pilnąja IIT interpretacija, ji neapsiriboja vien smegenimis, ir laiko sąmoningais bet kokius mechanizmus, integruojančius informaciją. Tokia išvada nėra suderinama su numatomojo apdorojimo šalininkų požiūriu. Tiksliau tariant šis požiūris nesuderinamas su IIT pateikiama tapatumo teze. Jei būtina sąmoningumo turinio sąlyga yra smegenų gebėjimas atlikti iš viršaus žemyn prognozes, tuomet tokie mechanizmai kaip fotodiodas nėra sąmoningas. Taip yra todėl, kad fotodiodo veikimas paremtas iš apačios į viršų principu. Kitaip tariant, fotodiodas atlieka sensoriaus funkciją, o tai nėra pakankama sąmoningumo sąlyga pagal Kontroliuojamos haliucinacijos teoriją. Tuo tarpu integruotos informacijos ir fenomenaliojo sąmoningumo tapatumo tezė leidžia fotodiodui priskirti sąmoningumą. Galima pridurti, kad fotodiodas yra pernelyg paprastas mechanizmas ir, nors atlieka tam tikrą funkciją, stokoja specifinių savybių, kurios leistų kurti vidinius modelius, būtinus sąmoningumui.

Kita problema derinant IIT ir Kontroliuojamos haliucinacijos teoriją yra anksčiau minėtas sistemų aktyvumas. Pagal IIT sąmoningos sistemos neprivalo turėti įvesties ir išvesties kanalų bei gali nedaryti nieko, kaip patyrimo salų ar išplėtimo grafiko atvejais. Pagal Kontroliuojamos haliucinacijos teoriją, sistemos sąmoningumas yra aktyvus procesas. Pilnoji IIT netapatina sąmoningumo su konkrečiu fiziniu mechanizmu, todėl, pagal ją, ne tik smegenys, bet ir nebiologiniai mechanizmai gali būti sąmoningi. IIT prognozuoja, kad fotodiodas ir išplėtimo grafikas yra sąmoningi. Šios dvi prognozės nėra suderinamos su numatančiojo apdorojimo hipotezę priimančiomis teorijomis. Fotodiodas yra pernelyg paprastas mechanizmas ir stokoja specifinių savybių, kurios leistų kurti vidinius modelius, būtinus sąmoningumui pagal Kontroliuojamos haliucinacijos teoriją.

Taigi, galima pabrėžti, kad dar vienas potencialus papildomas kriterijus, padedantis nustatyti, ar sistema yra sąmoninga ar ne, yra išsiaiškinimas, ar ji gali veikti pagal iš „viršaus žemyn“ principą. Sistemoms, veikiančioms vien pagal „iš apačios į viršų“ principą, laikytinos stokojančiomis sąmoningumo. Taigi, numatančiojo apdorojimo paradigma, besiremianti Kontroliuojamos haliucinacijos teorija, nėra suderinama su pilnąja IIT versija, nes pastarosios prognozės neatitinka minėto kriterijaus. Kitaip yra su minimaliąja IIT interpretacijų klase sudarančiomis teorijomis, kurios galėtų derinti integruotos informacijos matavimus kartu su apibūdintu kriterijumi nustatant sąmoningumą abejotinuose atvejuose.

## IŠVADOS

1. IIT pateikta metodologija gali būti išreikšta penkiais žingsniais: fenomenologinių aksiomų pateikimu, postulatų išvedimu, tapatumo išvados pateikimu, postulatų formalizavimu ir testavimu.

1.1 IIT derina pirmojo ir trečiojo asmens perspektyvų duomenis. Tononis tyrinėja savo paties patyrimą iš pirmojo asmens perspektyvos. IIT sąvoka „fenomenologija“ vartojama platesne prasme nei ją vartojo Husserlis. IIT gali būti interpretuojama tiek kaip fenomenologinio pirmumo, tiek kaip eksperimentinio pirmumo teorija.

1.2 IIT siekis pateikti apodiktines aksiomas yra nesėkmingas, nes jų teisingumas visais įmanomais atvejais nėra pagrįstas, pateiktos aksiomos nėra pakankamos, nes galima suformuluoti papildomų aksiomų, taip pat jos gali būti interpretuojamos skirtingai.

1.3 Ontologinių postulatų išvedimas iš aksiomų nėra pagrįstas, todėl jie laikytini atskiromis IIT priimamomis prielaidomis, bet ne iš fenomenologinių aksiomų gautomis išvadomis.

1.4 Mintiniai eksperimentai pateikiami anksčiau nei fenomenologinės aksiomos. Mintiniais eksperimentais apibūdinamos jau žinomos fizinių sistemų ypatybės.

1.5 IIT pateikiama tapatumo tezė traktuotina kaip prielaida, bet ne kaip išvada, gauta taikant fenomenologinį metodą. Tapatumo tezė pasitarnauja kaip fenomenologines aksiomas ir ontologinius postulatus sujungianti prielaida. Tapatumo tezė yra tiesiogiai susijusi su sąmoningumo laipsniškumo pripažinimu. Alternatyva laipsniškumo pripažinimui yra tam tikro integruotos informacijos slenksčio, kurį peržengus atsiranda sąmoningumas, nustatymas.

2.1 Priklausomai nuo to, ar priimama tapatumo tezė ir ar IIT traktuojama kaip fenomenologinio pirmumo, ar kaip eksperimentinio pirmumo, galima išskirti šešias IIT interpretacijas: pradinę, aspiracinę, eksperimentinio pirmumo, silpnąją, empirinę ir neutraliąją.

2.1.1 Empiriniai tyrimai pagrindžia integruotos informacijos ir sąmoningumo koreliaciją, bet ne būtinumą. Šie tyrimai paremia empirinę, neutraliąją ir silpnąją IIT interpretacijas. Norint pagrįsti likusias IIT interpretacijas aptartų tyrimų nepakanka.

2.1.2 Pradinės, aspiracinės, eksperimentinio pirmumo, silpnosios ir empirinės IIT interpretacijų tyrimo objektas yra fenomenalūs sąmoningumas, o neutralioji IIT interpretacija siekia likti neutrali šiuo klausimu.

2.1.3 Realistinės pozicijos fenomenaliojo sąmoningumo atžvilgiu (priešingos iliuzionizmui ir eliminatyvizmui) priėmimas yra būdingas visoms išvardintoms IIT interpretacijoms išskyrus neutraliąją IIT interpretaciją.

2.1.4 Pateiktą klasifikaciją sudarančios IIT interpretacijos suderinamos su skirtingomis ontologinėmis pozicijomis. Eksperimentinio pirmumo IIT interpretacija gali būti suprantama kaip fizikalizmo atmaina. Fenomenologinio pirmumo IIT interpretacijos gali būti suderinamos su Dvigubo aspekto teorija. Minimaliąją IIT interpretacijų klasę sudarančios teorijos yra neutralios ontologinio klausimo atžvilgiu.

2.2. Tapatumo tezės kritika parodo, kad minimalioji IIT interpretacijų klasė, kuri ją atmeta, yra geriau suderinama su kitomis teorijomis, nesikerta su sąmoningų ir nesąmoningų procesų skirtimi ir tyrimų duomenimis.

2.2.1 Pilnąją IIT interpretacijų klasę sudarančios teorijos nesutinka su GWT sąmoningų mechanizmų aktyvumo klausimu, taip pat nėra suderinamos su eksternalistinėmis sąmoningumo teorijomis ir neatsižvelgia į smegenų plastiškumo savybę. Minimaliąją IIT interpretacijų klasę sudarančios teorijos išvengia išvardintų problemų, nes atmeta tapatumo tezę.

2.2.2 Į minimaliąją IIT interpretacijų klasę įeinančios interpretacijos galėtų priimti sąmoningų ir nesąmoningų procesų skirtį. Pilnąją IIT interpretacijų klasę sudarančios interpretacijos su sąmoningų ir nesąmoningų procesų skirtimi nėra suderinamos.

2.2.3 Minimalioji IIT interpretacijų klasė yra geriau suderinama su Kontroliuojamos haliucinacijos teorija nei pilnoji IIT interpretacija.

## LITERATŪROS SĄRAŠAS

- Aaronson, S. (2014). Why I Am Not An Integrated Information Theorist (or, the Unconscious Expander). Prieiga internetu: <https://scottaaronson.blog/?p=1799>
- Albantakis, L., Barbosa, L., Findlay, G., Grasso, M., Haun, A.M., Marshall, W., Mayner, W.G., Zaeemzadeh, A., Boly, M., Juel, B. E., Sasai, S., Fujii, k., David, I., Hendren, J., Lang, J. P., Tononi, G. (2022). Integrated information theory 4.0: Formulating the properties of phenomenal existence in physical terms. *arXiv:2212.14787*. <https://doi.org/10.48550/arXiv.2212.14787>
- Arico, A., Fiala, B., Goldberg, R. F. ir Nichols, S. (2011). The folk psychology of consciousness. *Mind & Language*, 26(3), 327–352. <https://doi.org/10.1111/j.1468-0017.2011.01420.x>
- Askenasy, J. ir Lehmann, J. (2013). Consciousness, brain, neuroplasticity. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00412>
- Baars, B. J. (1997). In The Theatre of Consciousness Global Workspace Theory, a Rigorous Scientific Theory of Consciousness. *Journal of Consciousness Studies* 4: 292–309. <https://psycnet.apa.org/record/1998-00855-001>
- Bayne, T. (2018). On the axiomatic foundations of the integrated information theory of consciousness. *Neuroscience of Consciousness*, 2018(1). <https://doi.org/10.1093/nc/niy007>
- Bayne, T., Seth, A. K. ir Massimini, M. (2020). Are there islands of awareness? *Trends in Neurosciences* 43(1):6–16. <https://doi.org/10.1016/j.tins.2019.11.003>
- Bateson, G. (1979). *Mind And Nature A Necessary Unity*. New York: Dutton.
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18, 227–247. <https://doi.org/10.1017/S0140525X00038188>
- Block, N. (2002). *Concepts of Consciousness. Philosophy of Mind Classical and Contemporary Readings*. Oxford: University Press.
- Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences* 20: 1-9. <https://doi.org/10.1016/j.tics.2011.11.001>
- Carruthers, P. (2019). *Human and Animal Minds: The Consciousness Questions Laid to Rest*. Oxford: University Press.
- Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., Casarotto, S., Bruno, M. A., Laureys, S., Tononi, G., Massimini, M. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Science Translational Medicine* 5(198):198ra105. <https://www.science.org/doi/10.1126/scitranslmed.3006294>

- Cerullo, M. A. (2015). The Problem with Phi: A Critique of Integrated Information Theory. *PLOS Computational Biology*, 11(9), e1004286. <https://doi.org/10.1371/journal.pcbi.1004286>
- Chalmers, D. J. (1994). How Can We Construct a Science of Consciousness? In Gazzaniga, M. (ed.) *The Cognitive Neurosciences III* (pp. 1111–1120). Cambridge, MA: MIT Press.
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *J. Conscious.Stud.* 2, 200–219. <https://doi.org/10.1093/acprof:oso/9780195311105.003.0001>
- Chalmers, D. J. (2016). The Combination Problem for Panpsychism. In Brüntrup, G., Jaskolla, L. (eds.) *Panpsychism* (pp.179 – 214). Oxford University Press.
- Chalmers, D. J., (1996). *The Conscious Mind: In search of a fundamental theory*. Oxford University Press.
- Churchland, P. M. (1981). Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, 78 (2), 67–90. <https://doi.org/10.2307/2025900>
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci.* 36(3):181-204. <https://doi.org/10.1017/S0140525X12000477>
- Clark, A. ir Chalmers, D. J. (1998). The extended mind. *Analysis.* 58 (1): 7–19. <https://doi.org/10.1093/analys/58.1.7>
- Dagys J., Dranseika V., Grenda V., & Vasilionytė I. (2014). Reduktyvizmo raiška šiuolaikinėje analitinėje filosofijoje. *Problemos*, 85, 141-152. <https://doi.org/10.15388/Problemos.2014.0.2915>
- Dagys, J., (2012). Šiuolaikinė psichofizinio reduktyvizmo ir antireduktyvizmo polemika ir jos metodologinės išdavos. *Problemos* 2012 priedas: 22–36. <https://doi.org/10.15388/Problemos.2012.0.1875>
- de Haan, E. H. F., Corballis, P. M., Hillyard, S. A., Marzi, C. A., Seth, A., Lamme, V. A. F., Volz, L., Fabri, M., Schechter, E., Bayne, T., Corballis, M., Pinto, Y. (2020). Split-Brain: What We Know Now and Why This is Important for Understanding Consciousness. *Neuropsychology Review* 30:224–233. <https://doi.org/10.1007/s11065-020-09439-3>
- Dennett, D. C. (2018). Facing up to the hard question of consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373. <https://doi.org/10.1098/rstb.2017.0342>
- Doerig, A., Schurger, A. ir Herzog, M. H. (2020). Hard criteria for empirical theories of consciousness. *Cognitive Neuroscience*, 1–22. <https://doi.org/10.1080/17588928.2020.1772214>
- Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23. <https://www.ingentaconnect.com/contentone/imp/jcs/2016/00000023/f0020011/art00002>

- Frith, C. D. (2011). What Brain Plasticity Reveals About the Nature of Consciousness: Commentary. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00087>
- Godfrey-Smith, P. (2021). *Kiti protai Apie aštuonkoji, jūrą ir sąmonės kilmę*. Klaipėda: Liūtai ne avys.
- Heider, F. ir Simmel, M. (1944). *The American Journal of Psychology*, 57(2): 243-259. <https://doi.org/10.2307/1416950>
- Husserl, E. (2005). *Karteziškosios meditacijos*. Vilnius: Aidai.
- Kirchhoff, M. D. ir Kiverstein, J. (2018). *Extended consciousness and predictive processing: A third wave view*. New York: Routledge.
- Koch, C. (2019). *The Feeling of Life Itself: Why Consciousness Is Widespread but Can't Be Computed*. The MIT Press.
- Lee, M., Baird, B., Gosseries, O., Nieminen, J. O., Boly, M., Postle, B. R., Tononi, G. ir Lee, S. W. (2019). Connectivity differences between consciousness and unconsciousness in non-rapid eye movement sleep: a TMS–EEG study. *Scientific Reports* 9, 5175. <https://doi.org/10.1038/s41598-019-41274-2>
- Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly* 64:354-61. <https://doi.org/10.1111/j.1468-0114.1983.tb00207.x>
- Lumma, A. L. ir Veger, U. (2021). Looking from within: Comparing first-person approaches to studying experience. *Current Psychology*, 1-17. <https://doi.org/10.1007/s12144-021-02277-3>
- Mach, E. (1914). *The analysis of sensations and the relation of the physical to the psychical*. Chicago, London: The Open Court Publishing Company.
- Mackonis, A. (2009). Geriausio paaiškinimo išvedimas. Tarp dedukcijos, indukcijos ir abdukcijos. *Problemos* 76, 150-161. <https://doi.org/10.15388/Problemos.2009.0.1936>
- Mayner, W. G. P., Marshall, W., Albantakis, L., Findlay, G., Marchman, R. ir Tononi, G. (2018). PyPhi: A toolbox for integrated information theory. *PLOS Computational Biology*, 14(7). <https://doi.org/10.1371/journal.pcbi.1006343>
- Massimini, M., Ferrarelli, F., Murphy, M., Huber, R., Riedner, B., Casarotto, S. ir Tononi, G. (2010). Cortical reactivity and effective connectivity during REM sleep in humans. *Cognitive Neuroscience* 1(3):176-183. <https://doi.org/10.1080/17588921003731578>
- Massimini, M., Ferrarelli, F., Nuber, R., Esser, S. K., Singht, H. ir Tononi, G. (2005). Breakdown of Cortical Effective Connectivity During Sleep. *Science* 309, 2228–2232. <https://www.science.org/doi/10.1126/science.1117256>

- McGinn, C. (1972). Mach and Husserl. *Journal of the British Society for Phenomenology*, 3(2), 146–157. <https://doi.org/10.1080/00071773.1972.11006249>
- McQueen, K. J. (2018). Interpretation-Neutral Integrated Information Theory. *Journal of Consciousness Studies*, 26 (1-2):76–106. <https://philpapers.org/rec/MCQIII>
- McQueen, K. J. (2018a). Illusionist Integrated Information Theory. *Journal of Consciousness Studies* 26 (5-6):141-169. <https://philpapers.org/rec/MCQIII-2>
- Mediano, P. A., Rosas, F. E., Bor, D., Seth, A. K. ir Barrett, A. B., (2022). The strength of weak integrated information theory. *Trends in Cognitive Sciences* 26, 646–655. <https://doi.org/10.1016/j.tics.2022.04.008>
- Melloni, L., Mudrik, L., Pitts, M., Bendtz, K., Ferrante, O., Gorska, U., Hirschhorn, R., Khalaf, A., Kozma, C., Lepauvre, A., Liu, L., Mazumder, D., Richter, D., Zhou, H., Blumenfeld, H., Boly, M., Chalmers, D. J., Devore, S., Fallon, F., de Lange F. P., Jensen, O., Kreiman, G., Luo, H., Panagiotaropoulos, T. I., Dehaene, S., Koch, C., Tononi, G. (2023). An adversarial collaboration protocol for testing contrasting predictions of global neuronal workspace and integrated information theory. *PLoS One*. 18(2). <https://doi.org/10.1371/journal.pone.0268577>
- Merker, B., Williford, K. ir Rudrauf, D. (2022). The integrated information theory of consciousness: A case of mistaken identity. *Behavioral and Brain Sciences*, 45, E41. DOI:10.1017/S0140525X21000881
- Michel, M. ir Lau, H. (2020). On the dangers of conflating strong and weak versions of a theory of consciousness. *Philosophy and the Mind Sciences*, 1(II). <https://doi.org/10.33735/phimisci.2020.II.54>
- Moon, K. ir Pae, H. (2019). Making Sense of Consciousness as Integrated Information: Evolution and Issues of IIT. *arXiv:1807.02103*. <https://doi.org/10.48550/arXiv.1807.02103>
- Nagel, T. (1977). *What does it like to be a bat*. Oxford: Oxford University Press.
- Negro, N. (2022). Can the Integrated Information Theory Explain Consciousness from Consciousness Itself?. *Rev.Phil.Psych.* <https://doi.org/10.1007/s13164-022-00653-x>
- Oizumi, M., Albantakis, L. ir Tononi, G. (2014). From the Phenomenology to the Mechanisms of Consciousness; Integrated Information Theory 3.0. *PLoS Comput Biol* 10(5). <https://doi.org/10.1371/journal.pcbi.1003588>
- Pinto, Y., de Haan, E. H. F. ir Lamme, V. A. F. (2017a). The Split-Brain Phenomenon Revisited: A Single Conscious Agent with Split Perception. *Trends in Cognitive Sciences* 21(11):835-851. <https://doi.org/10.1016/j.tics.2017.09.003>



- Pinto, Y., Neville, D. A.; Otten, M.; Corballis, P. M. ; Lamme, V. A. F.; de Haan, E. H. F.; Foschi, N.; Fabri, M., (2017b). Split Brain: Divided Perception but Undivided Consciousness. *Brain* 140: 1231–1237. <https://doi.org/10.1093/brain/aww358>
- Pyza, E. M. (2013). Plasticity in invertebrate sensory systems. *Frontiers in Physiology*, 4. <https://doi.org/10.3389/fphys.2013.00226>
- Pokropski, M. (2018). Commentary: From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *Frontiers in Psychology* 9:101. <https://doi.org/10.3389/fpsyg.2018.00101>
- Pokropski, M. (2019). Phenomenology and mechanisms of consciousness: Considering the theoretical integration of phenomenology with a mechanistic framework. *Theory & Psychology* 1–19. <https://doi.org/10.1177/09593543198687>
- Rimkevičius, P. ir Gutauskas, M. (2014). Fenomenologijos ir natūralizmo suderinamumo klausimas Shauno Gallagherio filosofijoje. *Problemos* 86: 120–126. <https://doi.org/10.15388/Problemos.2014.0.3950>
- Rosenthal, D. M. (1993). State Consciousness and Transitive Consciousness. *Consciousness and Cognition*, 2(4), 355–363. <https://doi.org/10.1006/ccog.1993.1029>
- Russell, B. (1948). *Human Knowledge: Its Scope and Limits*. London: George Allen & Unwin
- Sarasso, S., Rosanova, M., Casali, A. G., Casarotto, S., Fedchio, M., Boly, M., Gosseries, O., Tononi, G., Laureys, S., Massimini, M. (2014). Quantifying cortical EEG responses to TMS in (un)consciousness. *Clinical EEG and Neuroscience* 45(1):40-9. <https://doi.org/10.1177/1550059413513723>
- Searle, J. (2013). Can information theory explain consciousness? Prieiga internetu: <https://www.nybooks.com/articles/2013/01/10/can-information-theory-explain-consciousness/>
- Seth, A. (2021). *Being You: A New Science of Consciousness*. New York: Dutton.
- Shannon, C. E. (1948). The mathematical theory of communication. *Bell System Technical Journal*. 27 (3): 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Signorelli, C. M., Szcotka, J. ir Prentner, R. (2021). Explanatory profiles of models of consciousness-towards a systematic classification. *Neuroscience of Consciousness*, 7(2), 1–13. <https://doi.org/10.1093/nc/niab021>
- Toker, D. ir Sommer, F. T. (2019). Information integration in large brain networks. *PLOS Computational Biology*, 15(2), e1006807. <https://doi.org/10.1371/journal.pcbi.1006807>

- Tononi, G. (2008). Consciousness as Integrated Information: a Provisional Manifesto. *Biological Bulletin* 215 (3): 216-42. <https://doi.org/10.2307/25470707>
- Tononi, G. (2012). Integrated information theory of consciousness: an updated account. *Archives Italiennes de Biologie*. 150, 56–90. <https://pubmed.ncbi.nlm.nih.gov/23165867/>
- Tononi, G. (2014). Why Scott should stare at a blank wall and reconsider (or, the conscious grid). Prieiga internetu: <http://www.scottaaronson.com/blog/?p=1823>
- Tononi, G. ir Koch, C. (2015). Consciousness: here, there and everywhere? *Philosophical Transactions of The Royal Society B Biological Sciences* 370(1668). <https://doi.org/10.1098/rstb.2014.0167>
- Tononi, G., Boly, M. ir Grasso, M. 2022. IIT, half masked and half disfigured. *Behavioral and Brain Sciences* 45. <https://doi.org/10.1017/S0140525X21001990>
- Tononi, G., Boly, M., Massimini, M. ir Koch, C. (2016) Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience* 17, 450–461. <https://doi.org/10.1038/nrn.2016.44>
- Tononi, G., Edelman, G. M. (1998). Consciousness and complexity. *Science* 282(5395):1846-51. <https://www.science.org/doi/10.1126/science.282.5395.1846>
- Tsuchiya, N. ir Leung, A. (2023). Separating weak integrated information theory (IIT) into IIT-inspired and aspirational-IIT approaches. *PsyArXiv*. <https://psyarxiv.com/kxywt/>
- Windt, J. M. (2015). Just in time – dreamless sleep experience as pure subjective temporality – a commentary on Evan Thompson. *Open MIND*: 37(C). <https://openmind.net/DOI?isbn=9783958571174>

# PRIEDAI

1 priedas. *Integruotos informacijos teorijos interpretacijų klasifikacija*

