

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
«КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»**



Факультет гуманітарних наук
Кафедра філософії та релігієзнавства

В'ячеслав Циба

СУЧАСНА ФІЛОСОФІЯ СВІДОМОСТІ

**Навчально-методичний посібник з навчальної
дисципліни**

УДК 165.242.2+123+004.8(075.8)

Циба В. М. Навчально-методичний посібник з навчальної дисципліни «Сучасна філософія свідомості». – К.: НаУКМА, 2023. – 169 с. (7,5 др. арк.)

«Сучасна філософія свідомості» – авторська дисципліна, розроблена доцентом кафедри філософії та релігієзнавства НаУКМА, кандидатом філософських наук В. М. Цибою; викладається в НаУКМА з 2017 р. і входить до переліку дисциплін професійної та практичної підготовки студентів, які навчаються на першому (бакалаврському) рівні вищої освіти за спеціальністю 033 Філософія. Посібник є узагальненням досвіду автора щодо викладання сучасної філософії свідомості з урахуванням якісних змін у сучасній епістемології та новітніх освітніх підходів. Основну увагу приділено таким актуальним темам, як тенденції в дискусії між редуccionістськими та антиредуccionістськими концепціями, аналіз психофізичної проблеми, дослідження природи квалій, формальна реконструкція аргументів і контраргументів філософів свідомості, проблема штучного інтелекту та пов'язані з нею дилеми моральної філософії.

Автор:

Циба В'ячеслав Миколайович, кандидат філософських наук, доцент кафедри філософії та релігієзнавства Національного університету «Києво-Могилянська академія»

Рецензенти:

Кебуладзе Вахтанг Іванович, доктор філософських наук, доцент, професор кафедри теоретичної і практичної філософії Київського національного університету імені Тараса Шевченка

Менжулін Вадим Ігорович, доктор філософських наук, професор, завідувач кафедри філософії та релігієзнавства Національного університету «Києво-Могилянська академія»

*Ухвалено до друку Вченою радою факультету гуманітарних наук
Національного університету «Києво-Могилянська академія»
(протокол № 8 від 12.10. 2023 р.)*

Текст друкується в авторській редакції

© В.М. Циба, 2023

Зміст

Анотація.....	4
Загальний опис дисципліни	5
Тематичний план лекцій та семінарських занять	6
Лекційні заняття	6
Матеріали лекцій.....	9
Додаток 1. Список скорочень	132
Додаток 2. Короткий термінологічний словник	133
Додаток 3. Ілюстрації до лекцій	145
Семінарські заняття: тематика та література.....	152
Орієнтовний перелік тем для індивідуальної письмової роботи.....	157
Вимоги до оформлення індивідуального письмового завдання.....	158
ЛІТЕРАТУРА	159
Оцінювання знань і вмінь	165
Умови визначення навчального рейтингу	165
Результати і критерії оцінювання.....	166
Рекомендовані джерела інформації.....	168
Abstract.....	169

Анотація

Посібник є узагальненням досвіду автора щодо викладання сучасної філософії свідомості з урахуванням якісних змін у сучасній епістемології та новітніх освітніх підходів. Основну увагу приділено таким актуальним темам, як дискусії між редукціоністськими та антиредукціоністськими концепціями, аналіз психофізичної проблеми, дослідження природи квалій, формальна реконструкція аргументів і контраргументів філософів свідомості, проблема штучного інтелекту та пов'язані з нею дилеми моральної філософії.

Після того, як Л. Вітгенштайн розкритикував модель приватної мови, а його учень Г. Райл висловився про ум як про категорійну помилку, стало очевидним, що картезіанський суб'єктивний привілей втратив силу. Поява нових підходів, особливо фізикалістської теорії тотожності засвідчила, що ментальне є предметом поглибленого філософського аналізу. Як наслідок, за останні п'ять десятиліть філософія свідомості зробила стрімкий стрибок від проблеми квалій до проблеми штучного інтелекту. Було досліджено такі важливі для епістемології теми, як каузальність, автореферетність знання про себе, знання про світ, інтенційні та неінтенційні стани; модальний аргумент у логіці, на якому будувалися найвідоміші мисленнєві експерименти, а також протистояння детермінізму та індетермінізму стосовно вільного рішення. Становлення аналітичної філософії свідомості показує, що філософія в змозі подолати свою галузеву фрагментованість і знову повертається до системного підходу.

Ключові слова: сучасна філософія, аналітична філософія, ум, свідомість, психофізична проблема, квалії, мова, аргумент приватної мови, аргумент китайської кімнати, штучний інтелект, тест Тюринга, теорія тотожності, матеріалізм, дуалізм тіла і розуму, монізм, функціоналізм, філософський натуралізм, епістемологія, свобода волі, логічний бігейвіоризм, філософія.

Загальний опис дисципліни

Загальна кількість годин – 90 год.	Форма контролю – Залік
Самостійна робота – 62 год.	Кредитів ЄКТС – 3
Аудиторні заняття – 28 год.	Семестр викладання – V
А) Лекції – 16 год.	Галузь знань – 03 Гуманітарні науки
Б) Семінари – 12 год.	Спеціальність – 033 Філософія

Анотація дисципліни

Викладання дисципліни передбачає здійснення огляду й аналізу основних тенденцій розвитку сучасних течій філософії свідомості (як неklasичної європейської філософії самосвідомості, так і Philosophy of Mind), зокрема: співвідношення епістемології та когнітивних наук, теорії штучного інтелекту та теорій самосвідомості, філософії психології та логіко-семантичного аналізу, етики та соціальної філософії, дослідження структур інтелекту як природно-соціального симбіозу; з'ясування значення натуралізації епістемології для подолання проблем класичної метафізики самосвідомості; обґрунтування нормативності в людському мисленні та діяльності; адекватність словника філософії свідомості для розуміння єдності людської думки та діяння; аналіз ментальних експериментів («зомбі», «анти-зомбі», «китайська кімната»), що висвітлюють сутність соціального виміру ментального.

Мета дисципліни

Дисципліна «Сучасна філософія свідомості» є складовою блоку вибіркового дисциплін професійної та практичної підготовки студентів першого (бакалаврського) рівня вищої освіти за спеціальністю 033 Філософія у галузі знань 03 Гуманітарні науки в НаУКМА. Лекції передбачають систематизований і структурований виклад програмного матеріалу і мають на меті формування у студентів базового рівня знань з дисципліни, а також спрямування їх самостійних зусиль щодо засвоєння предмета. Семінарські заняття спрямовано на обговорення студентами питань, визначених робочою навчальною програмою, і мають на меті передусім ознайомлення з формами філософування про свідомість шляхом витлумачення та порівняльного аналізу низки модерних і новочасних філософських текстів, в епіцентрі яких перебуває тема свідомості як окрема галузь філософського знання. Знання та навички, набуті в процесі вивчення дисципліни «Сучасна філософія

свідомості», складатимуть підґрунтя для набуття комплексу фахових знань і розвитку професійних вмінь, необхідних для підготовки бакалаврів за спеціальністю 033 Філософія у галузі знань 03 Гуманітарні науки в НаУКМА.

Політика викладання дисципліни

Вивчення дисципліни має відповідати вимогам «Положення про Академічну доброчесність здобувачів освіти у НаУКМА» (затверджене наказом № 112 від 07.03.2018 року): https://www.ukma.edu.ua/index.php/about-us/sogodennya/dokumenty-naukma/doc_download/577-polozhennia-pro-akademichnu-dobrochesnist-zdobuvachiv-osvity-u-naukma

Відпрацювання пропущених семінарських занять можливе за умов наявності документального підтвердження того, що пропуск мав місце з поважної причини.

Тематичний план лекцій та семінарських занять

№	Тема	Лекції (кількість годин)	Семінари (кількість годин)
1	Витоки, специфіка та основні проблеми сучасної філософії свідомості	2	
2	Критика еґологічної теорії. Логічний біґевіоризм	2	2
3	Матеріалістичний редукціонізм. Фізикалізм і функціоналізм	2	2
4	Проблема квалій. «Кваліафіли» та «кваліафоби»	2	2
5	Дуалістичний антиредукціонізм (I). Дуалізм властивостей, інтераціонізм і натуралістичний дуалізм	2	
6	Дуалістичний антиредукціонізм (II). Емерджентизм. Епіфеноменалізм	2	2
7	Аномальний монізм. Елімінативізм. Гетерофеноменологія	2	2
8	Філософські проблеми штучного інтелекту	2	2
	Разом	16	12

Лекційні заняття

1. Мета і завдання лекційних занять

У лекціях здійснено систематизований і структурований виклад програмного матеріалу з метою формування у студентів базового рівня знань

з дисципліни, а також спрямування їх самостійних зусиль щодо засвоєння предмета. Семінарські заняття спрямовано на закріплення і поглиблення знань предмета, здобутих під час лекційних занять і самостійної роботи; прищеплення навичок аналізу конкретних тем і проблем; розкриття їх теоретичного змісту і практичного значення; ведення фахової дискусії тощо.

Дисципліна «Сучасна філософія свідомості» є важливою складовою підготовки бакалаврів за спеціальністю 033 Філософія до фахової діяльності в галузі освіти і гуманітарних наук.

2. План лекційних занять

Тема 1. (2 год.)

Витоки, специфіка та основні проблеми сучасної філософії свідомості

- 1.1. Зміст понять «сучасна філософія свідомості» та «philosophy of mind».
- 1.2. Характеристика філософії свідомості як сучасного теоретичного напрямку.
- 1.3. Вузлові питання сучасної філософії свідомості.

Тема 2. (2 год.)

Критика егологічної теорії. Логічний бігевіоризм.

- 2.1. Вітгенштайнова критика менталізму. Проблема приватної мови.
- 2.2. Логічний бігевіоризм Г. Райла. Ум як категорійна помилка.

Тема 3. (2 год.)

Матеріалістичний редукціонізм. Фізикалізм і функціоналізм

- 3.1. Фізикалістська теорія тотожності та її критика.
- 3.2. Модальний аргумент Кріпке.
- 3.3. Машинний функціоналізм і бігевіоризм. Тест Тюринга.
- 3.4. Аргументи проти функціоналізму. «Китайська кімната» і «китайська нація».
- 3.5. Функціоналізм і фізикалізм: порівняння підходів.

Тема 4. (2 год.)

Проблема квалій. «Кваліафіли» та «кваліафоби».

- 4.1. Проблема природи квалій.
- 4.2. Основні аргументи кваліафілів і кваліафобів.

Тема 5. (2 год.)

Дуалістичний антиредукціонізм (I): дуалізм властивостей, інтеракціонізм і натуралістичний дуалізм.

- 5.1. Філософський дуалізм, його принципи.
- 5.2. Дуалізм властивостей.
- 5.3. Інтеракціонізм.
- 5.4. Натуралістичний дуалізм Чалмерса, основні ідеї.
- 5.5. «Філософський зомбі»: боротьба аргументів.

Тема 6. (2 год.)

Дуалістичний антиредукціонізм (II): Емерджентизм і епіфеноменалізм.

- 6.1. Емерджентизм як натуралістична доктрина.
- 6.2. Критика емерджентизму.
- 6.3. Епіфеноменалізм.

Тема 7. (2 год.)

Аномальний монізм. Елімінативізм. Гетерофеноменологія.

- 7.1. Типи філософського монізму.
- 7.2. Аномальний монізм Д. Девідсона.
- 7.3. Аномальний монізм і проблема свободи волі. Експеримент Лібета.
- 7.4. Елімінативізм і розв'язання ним проблеми свідомості.
- 7.5. Гетерофеноменологія Д. Денета.
- 7.6. Квалії у суперечці інтерналістів і екстерналістів.

Тема 8. (2 год.)

Філософські аспекти штучного інтелекту

- 8.1. Машина Тюринга і тест Тюринга. Заперечення проти сильного штучного інтелекту.
- 8.2. Аналіз аргументів проти можливості сильного штучного інтелекту.
- 8.3. Новітні проекти штучного інтелекту. Проблема технологічної сингулярності.
- 8.4. Штучний інтелект: етичні виклики для людства. Дружній ШІ.

Матеріали лекцій

Лекція 1.

Витоки, специфіка та основні проблеми сучасної філософії свідомості

1.1. Зміст понять «сучасна філософії свідомості» та «*philosophy of mind*»

Під сучасною філософією свідомості (далі – СФС) зазвичай мають на увазі різноманітні теоретичні підходи, зайняті дослідженням проблеми свідомості та самосвідомості через аналіз семантики поняття «ментальне». Власне, відразу помічаємо ті концептуальні аспекти, без урахування яких розуміння мети й специфіки цієї навчальної дисципліни годі зрозуміти правильно й повно. Цими аспектами є: 1) формальний аналіз проблеми відношення між психічним (ментальним) і фізичним, яке прийнято називати психофізичною проблемою (*mind-body problem*) – переважна більшість дискусій між представниками СФС так чи так торкається саме цієї тематики; 2) спроби прояснити природу ментального, зокрема феноменальної даності ментального через індивідуальні нюанси його переживання – квалії (див. 4.1.); 3) порівняння пошукових стратегій для відповіді на питання про природу свідомості з позицій логіки та лінгвістики, з одного боку, та намагання побачити більш об'ємну картину ментальних феноменів через залучення інструментарію епістемології, біології та когнітивної психології, з іншого боку.

Між вказаними аспектами існує концептуальний взаємозв'язок, пояснюваний складним історичним контекстом постановки цього питання. Приміром, намагання запропонувати задовільний розв'язок психофізичної проблеми без порушення питання про сутність квалій або достатність аргументів матеріалізму залишиться практично нереалізованим, адже хоча квалії індивідуальних переживань вже постулюють особливий тип реальності, яку репрезентують, проте нефізичний вимір цієї реальності передбачає доступ до неї у хіба що гіпотетичних судженнях про власні феноменальні досвіди. Однак, як ми побачимо далі, одним лише стосунком між ментальними

актами, їхніми маркерами – кваліями та ментальними змістами діапазон обговорень не вичерпує СФС.

Психофізична проблема – це проблема співвідношення таких сутностей, як ум і тіло (душа/дух і тіло), передумов і способів їх можливої та реальної взаємодії. Її канонічна версія була сформульована ще Платоном, який стояв біля витоків субстанційного дуалізму, а в завершеному вигляді, з яким має справу сучасна філософія – Рене Декартом. Психофізичну проблему ще називають дилемою духу та тіла. Як ми побачимо з подальших лекцій, спроби запропонувати задовільне розв'язання цієї проблеми штовхали англomовну філософію від жорсткого матеріалізму до поміркованої натуралістичної програми. А проте, в останні десятиліття дедалі більшої популярності знову набуває дуалізм.

Зазначимо, що позаяк сучасна філософія свідомості ґрунтується переважно на дискурсі англomовної аналітичної філософії, нерідко цей теоретичний напрям в усьому ідентифікують з поняттям *philosophy of mind*. При цьому для українського читача пряме калькування цього терміна містить низку пасток, про які слід мати хоча б загальні уявлення. Передусім англ. іменник *mind* є багатозначним, входить у склад низки фразеологізмів і не відповідає точному змісту українського слова «свідомість». Якщо англomовний філософ висловлює свої припущення стосовно *mind*, він матиме на увазі *психіку* як замкнену систему образів та ідей, так і множину ментальних переживань в цілому. Ймовірно, керуючись саме цими міркуваннями, український філософ Д. Сепетий пропонує перекладати назву праці Д. Чалмерса *The Conscious Mind* як «свідома психіка». Крім того, *mind* визначають через такі характеристики, як «ум», «розум», «умова здатність», «внутрішній досвід», з якими цей термін етимологічно споріднений завдяки впливу латиномовної наукової традиції. Прямим еквівалентом українського поняття «свідомість» є не *mind*, а *consciousness*, себто індикатор рівня розпізнавання носієм ментального стану того, що з ним відбувається у цьому стані, і відрізнення від якісно схожих, але неідентичних, станів. Крім того,

mind і consciousness взаємодіють як своєрідні собі субстрат і властивість, або система й функція, що доволі важко відтворити за допомогою природної мови для семантики їх українського поняттєвого еквівалента – вона щоразу пред'являє вужчі конотації. Разом із тим, в англійській мові є чимало інших дієслівних позначників для акту усвідомлення свого переживання, нетотожних або непрямо-тотожних consciousness, напр., aware, apprehend, grasp тощо.

Либонь, найоптимальнішим відповідником для mind у межах української філософської лексики був би «дух», адже, по-перше, цей термін семантично ближче до англ. mind і, до того ж, охоплює собою множину нефізичних сутностей; по-друге, у ньому міститься багатий спектр проміжних явищ психічного, як-от «душі», «розуму», загалом «внутрішнього відчуття» суб'єкта на противагу сприйманню ним зовнішньої реальності. А проте, на перешкоді цій пропозиції стоїть відносно усталений узус поняття «дух», значно обтяжений релігійно-містичними конотаціями (духовність, святий дух тощо). З іншого боку, маємо тривалу історію вживання словосполук штибу «дух часу», «дух інновацій». До того ж, англійському еквіваленту слова «дух» мав би улягати вжиток spirit, який тепер уже в мові англо-американської філософії практично не задіяно (з тих самих міркувань його релігійної забарвленості). Щоправда, така нормативістська настанова знає винятки. Зокрема, праця Гегеля «Феноменологія духу» існує відразу у двох перекладах: *The Phenomenology of Spirit* (1977) і *Phenomenology of Mind* (1910). Компромісом у діапазоні розглянутих понять було би пристати на користь семантично ширшого слова «ментальне», якби цьому не стояла на заваді складність ужитку імплікованих його участю неоковирних конструкцій типу «індивідуальне ментальне» або «свідоме ментальне». Зважаючи на окреслений масив семантичних утруднень, я схиляюся до рішення про доцільність передання змісту поняття mind, керуючись виключно контекстом його використання та смисловим протиставленням суміжним із ним термінам.

Отже, як бачимо, наш вибір на користь словосполучки «філософія свідомості» можна пояснити впливом кількох чинників, а радше обмеженнями лексично-термінологічного характеру. По-перше, попередньою унормованістю слововжитку в українській філософській практиці (вихід у світ книг, статей, антологій та інших публікацій, з якими слід рахуватися). По-друге, тим, що поняття «свідомість» та «ум» пов'язані не тільки структурно, а й функційно, себто взаємодія між ними відбувається у площині одного типу зв'язків. Загалом, з огляду на розподіл сфер аналізу, більш доречним, на мій погляд, виглядає вжиток словосполучки «філософія свідомості», коли йдеться про психологічний аспект дослідження ментального, і «філософія ума», коли обговорюють теми, дотичні до когнітивного стосунку розуміння і тлумачення щодо влаштованості людського суб'єкта. А проте, у робочому порядку ніщо не заважає послуговуватися й назвою «філософія ума» або, як це іноді роблять окремі дослідники, «філософія розуму».

1.2. Характеристика філософії свідомості як сучасного теоретичного напрямку

Дослідження проблеми свідомості в рамках цієї навчальної дисципліни ґрунтується на двох *методологічних принципах*: по-перше, ми проблематизуємо головні теми для сучасної епістемології, спрямовані на можливість пізнавати та несуперечливо описувати сферу ментального; по-друге, ми розглядаємо концептуальні труднощі, пов'язані з виконанням цього нетривіального завдання. З огляду на сказане, відправною точкою усіх подальших екскурсів є, з одного боку, виділення історичного тла практик інтерпретації питання про свідомість саме як *теоретичної* проблеми («як певні проблеми і способи їх розуміння поставали в межах тих чи тих філософських течій?»), а з другого, аналіз досягнутих на сьогодні результатів («які концептуальні новації у філософії тягли за собою спроби розв'язання цих проблем?»).

Отже, питання про сутність свідомості та самосвідомості безпосередньо стосується характеру філософського запитування і еволюції різноманіття його форм. Тим самим сучасна філософія свідомості (СФС) є не несподіваним якісним стрибком на шляху філософування, а перенесенням попередніх полемік у класичних західних теоріях пізнання на рівень з урахуванням революції у системі когнітивних наук. Це, своєю чергою, спонукає додатково вказати на міждисциплінарні аспекти й неоднозначний статус СФС, що коливається між проблемами традиційної метафізики та філософією науки.

Зауважмо, що від середини 1980-х років СФС стала провідною з філософських дисциплін в англomовному середовищі, інкорпорувавши у свої практики інші філософські тренди та підходи (епістемологію, етику, політичну філософію, філософію культури). Це, однак, не означає, що решта напрямків мислення академічно маргіналізувалися, а якраз те, що вони більше зважають на стрімкі зміни у СФС як у доволі продуктивному просторі розробки загальної філософської аргументації.

1.3. Вузлові питання сучасної філософії свідомості

Коли кажуть про обшир теоретичних проблем, якими переймається аналітична філософія свідомості, то мають на увазі ті, що їх виділяють на перетині трьох галузей думки: епістемологічної («Якими є підстави мого знання про стан X?»), логічної («Що робить судження про стан X істинним або хибним?») та лінгвістичної («Як моє знання про стан X може бути повідомлене?»). Отже, сучасна філософія свідомості – це синтез висновків у цих галузях знання, і саме на ґрунті цього сполучення варто шукати відповіді на її основні питання. Тим самим стосунок ментального до нементального можна сформулювати у вигляді питання: чи мають ментальні феномени матеріальне походження? Відповідно (якщо мають), чи означає це, що вони підлягають дії тих самих законів, що й решта матеріальних явищ? У разі, якщо відповідь є негативною, чи варто шукати інші, нематеріальні (скажімо, суто логічні) підстави пояснення природи ментального? Як бачимо, поняття

ментального виводить нас на ширший словник, основою якого є деривативи від іменників «суб'єкт»: суб'єктивність, суб'єктність, суб'єкція та ін.

Однак синтез згаданих перспектив не означає їх простої комбінації. До нього нерідко долучають результати емпіричних наукових досліджень, приміром, зі сфери психології, когнітивних наук та комп'ютерного моделювання. Отож, коли ставлять питання про природу ментального, на нього й відповідають по-різному. У цьому головно й полягає концептуальна трудність дискусій, до яких залучені представники різних течій СФС. Скажімо, коли обговорюють, чим є «біль», звертають увагу на те, що як ментальний стан можна описати психологічно (відчуття) або феноменологічно (безоб'єктний інтенційний стан), або когнітивно (перешкода в реалізації певного наміру чи акту), або епістемологічно (почуттєвий корелят переконання, що корегує наш стосунок щодо світу) тощо. Так чи так, проблема ментального охоплює собою й низку інших суміжних проблем: існування свідомості, статус суб'єктивності, умови самототожності суб'єкта, взаємостосунок ментального й тілесного, співвідношення свідомості та самосвідомості, проблему можливості машинного інтелекту, пізнання носіїв інших свідомостей тощо.

Джерела:

Комар О. (2016). Тесеєві мандри свідомості. *Філософська думка* 2: 57–70.

<https://dumka.philosophy.ua/index.php/fd/article/download/13/14/20>

Круглий стіл «Philosophy of Mind? – Mind the Gap!» (2016). *Філософська думка* 2: 6–29.

http://journal.philosophy.ua/sites/default/files/filefield_paths/5-philosophy_of_mind_-_mind_the_gap.pdf

Синиця А. (2017). *Сучасна аналітична філософія: від прагматики мови до концептуалізації свідомості*. Л.: ЛНУ ім. І. Франка, сс. 40–47.

Heil J. (2020). *Philosophy of Mind. A Contemporary Introduction*. New York; London: Routledge, pp. 1–28.

Запитання для самоперевірки:

1. Чим філософія свідомості відрізняється від інших сфер філософування?
2. Чи виправдане твердження, що сучасна філософія свідомості є комплексною дисципліною? Свою думку обґрунтуйте.
3. Чи стосуються терміни «душа», «свідомість», «ум», «суб'єктивність», «особа» однієї й тієї самої сутності чи різних сутностей? Якщо різних, тоді що їх різнить?
4. Поясніть 15-річній дитині сутність психофізичної проблеми.
5. Чи тотожний зміст понять «ментальне» і «психічне»? Якщо так, то чому, на Ваш погляд, філософи послуговуються різними термінами, а не одним?

Лекція 2.

Критика егологічної теорії. Логічний бігевіоризм.

2.1. Вітгенштайнова критика менталізму. Проблема приватної мови.

Біля витоків лінгвістичного аналізу ментального в аналітичній філософії ХХ ст. стояв австрійсько-британський філософ **Людвіг Вітгенштайн** (Wittgenstein). З ознайомлення з його теоретичним надбанням ми й починаємо свій екскурс в історію англосаксонських досліджень ментальної проблематики. У своїх працях «Філософські дослідження», «Синя книга» і «Філософська грамати́ка» (видані його учнями посмертно) Вітгенштайн зосередився на критиці так званої *егологічної теорії свідомості*. Стандартна модель цієї теорії заснована на «ярликовій теорії значення», згідно з якою дискретні лексичні одиниці мають значення єдино лише, якщо вони позначають нелінгвістичний референт (слово «яблуко» вказує – здійснює референцію – на яблуко як фізичний об'єкт). Проблема ярликової теорії полягає у відсутності багатьох класів об'єктів, з якими можна співвіднести слова чи вирази (напр., тисячокутник, квадратура кола чи ельфи). У рамках цієї візії посередником між словом-поняттям і його об'єктом є образ об'єкта.

Тож маємо дивну ситуацію: я вживаю слово «ельф», маю образ ельфа у своїй уяві, але не маю референта цього слова. Щось подібне відбувається і з поняттями, якими люди описують свої ментальні стани. Ба більше, одні й ті самі слова важко визначити двом співрозмовникам, як-от коли мовиться про «душу», «біль», «натхнення» і т. п. Власні образи при цьому з'являються у кожного зі співрозмовників, але щойно постає потреба визначити зміст ужитих слів, постає проблема пояснення (explanatory gap). І навіть тоді, коли двоє осіб інтуїтивно згодні, що говорять про одне й те саме, насправді вони можуть апелювати до цілком відмінних концептуальних схем тим паче різних сутностей.

На противагу ярликовій (образній) концепції Вітгенштайн висуває власну теорію *значення як ужитку* (meaning as usage). Ми всі граємо у своєрідні мовні ігри, переміщуючись між життєвими ситуаціями, кожна з яких вимагає врахування правил її опису та мовленнєвого вираження. Граючи у шахи (улюблений приклад Вітгенштайна), гравці співвідносять кожен наступний хід із алгоритмом гри; так само, розказуючи про свій досвід комусь іншому, ми вживаємо слова, адекватні контексту подібної розмови. Це означає, що одиницею мови є не слово і навіть не речення, а мовленнєві акти, завдяки яким розкривається семантичний потенціал мовних одиниць. Вирішальну роль у цьому відіграє співучасть мовців у виробленні конвенцій. Якщо звернутися тепер до такої мовної гри, як судовий процес, стає зрозумілим, що передбачені нею приписи, а отже, й їхні смисли, історично зазнають змін. Це доводить, що лінгвістичні значення, хай би яких об'єктів вони стосувалися, у природних мовах є *нестабільними*. Єдине, що здатне на деякий час їх урівноважити й цим самим універсалізувати – це мовчазна згода спільноти мовців послуговуватися ними за таких-то і таких-то обставин. Отож, з погляду Вітгенштайна, будь-яке висловлювання, надто ж про своє психічне життя, апелює до *правил спільновикористовуваної граматики*, при чому неясності й суперечності є виходом «за межі» цих правил.

Повертаючись до еґо-теорії, слід зазначити, що її засновком є твердження, що єдино істинним знанням є *суб'єктивне переконання* (belief). Тому опис особою своїх ментальних змістів може бути здійснений тільки у межах мови, зрозумілої цьому суб'єктові. Припускається, що словник даних описів є суб'єктивно верифікованим (адже його значення не є об'єктивними). Якщо екстраполювати ці положення на рівень споглядання, тоді головним способом самопізнання є інтроспекція. Отже, знання про себе в такому разі є привілейованим для того, хто має певні ментальні стани, і принципово неможливим для інших. Звідси й висновок, що знання про себе як сукупність самоописів мають пріоритет щодо об'єктивних умов знання. Підсумуймо еґо-теорію у вигляді умовиводу:

1. Обізнаність зі своїми ментальними станами тотожна моєму «Я».
2. «Я» дається мені з безумовною очевидністю.
3. Безумовна очевидність тотожна вичерпному розумінню.
4. Отже, моя обізнаність зі своїми ментальними станами є самоочевидною і змістовно вичерпною.
5. Джерелом опису моїх ментальних станів є я сам.
6. Отже, моє знання про себе є гранично вичерпним знанням.

У цілому, еґологічна теорія успадковує картезіанську модель істинності. Якщо критерієм знання є моя суб'єктивна певність, то й можливість описів своїх уявлень і переживань означає винятково *некорегованість* приватної мови, за допомогою якої я проваджу ці описи. У картезіанській версії даний висновок тримається на перформативній суперечності, що ототожнює мислення й буття (згадаймо Декартове *cogito, ergo sum*): я мислю про буття, в яке залучено моє мислення.

Якими є концептуальні наслідки еґо-теорії? Передусім, це скептичне оцінення пізнання як процесу, в якому не існує об'єктивно-нейтральних принципів судження, обов'язкових для дотримання всією спільнотою пізнавців. Крайньою версією такого скептицизму в епістемології є соліпсизм, а в етичному аспекті – моральний релятивізм. Світ у еґологіста ніби

стискається до цятки, поглиненої обширом його почуттів. «Я» переважає над світом і взалежнює його від хиткої динаміки своїх переживань. Зі сказаного явно випливає неможливість розбудови систематичного знання, чи то у вигляді філософських систем, чи то як можливість різноманітних наук. Загалом від приватної певності головно потерпає плюральність і варіативність досвіду.

Трудність, яку виявляє в егологічній теорії Вітгенштайн, полягає в тім, що їй імпліцитно притаманне верховенство *«приватної мови»*: значення висловлень у ній не має інтерсуб'єктивного виміру; воно є *тільки* моїм, тому позначає розрив між уявленням і знанням. Попри те, що для підходу еголога прикметним є нерозрізнення між індивідуальним і загальним (останнє акумулюється уявою індивіда), йому притаманне й те, що «приватник» недобачає розрив у часовій зміні його психологічних станів. Адже якщо не існує об'єктивних критеріїв відрізнити стан Y від стану X, тоді що мені дає підстави твердити, що я «знаю» стан Y як свій стан (що моє відчуття є «болем», а не «радістю»). Заперечення полягає в тім, що критерії розрізнення не є суб'єктивними і тим паче психологічними. У кожному разі теза про *безпомилність* тверджень захисника приватної мови відчуттів є Ахіллесовою п'ятою його аргументації.

А проте, цілком можливо виділити дві версії приватної мови: сильну та слабку. Сильну версію можна увиразнити у формі *неспроможності пізнати* чужі ментальні стани, а слабку – у вигляді спонтанної довіри до чужих фраз при одночасному визнанні *безпідставності* спільних для різних суб'єктів значень (смислів) висловлень. Розкол проходить, по суті, по лінії глибинності граматики: коли я кажу «я це знаю» чи «він не розумію цього» (себто тільки я маю необхідне «розуміння»), не йдеться про чийсь мозок або душу як інстанцію значущості. Замінімо місцями особові займенники «я/мій» і «він/його», й отримаємо самоспростування сказаного. Вітгенштайн пропонує ще радикальніше рішення: спробуймо (в експериментальних цілях) відмовитися від особових займенників першої особи – і виявимо абсурдність

наших тверджень. Порівняймо картезіанське «Я мислю, отже, існую» з реченням «Дещо (зараз) мислиться у тілі людини на ім'я Петро Сидоренко». Займенники першої особи однини потрібні мовцеві тільки вказування іншим мовцям на *інстанцію* поточного висловлювання, і аж ніяк не для того, щоб ізолювати від них його суб'єктність.

Ми здогадуємося, що значить «мати больовий стан», на основі власних відчуттів і що інші теж демонструють переживання, схожі з нашими. Але проблема ускладнюється тією обставиною, що біль є *неінтенційним* ментальним станом – йому бракує об'єкта, на який він спрямований. На відміну від страху (перед кимось), захвату (від чогось), біль просто «настає» і примушує центруватися тільки на самому його триванні. Якщо порівняти мовні ігри «Мені болить» і «Я сумніваюся, що...», то стане ясно, що їх визначено різною системою правил. І саме ці правила дають підказки для виявлення того, які значення тут грають «першу скрипку». Та й сама повсякденна мова *не приховує* значень і своїх правил. Варто лишень уважно стежити за тим, *як і за яких умов* ми кажемо те, а не те. Вітгенштайн доводив, що будь-яка мовна гра, як, зрештою, будь-яка ігрова активність, відбувається згідно з правилами. Вітгенштайн наголошував на імперативі: «Говоріть як завгодно, – головне, аби вас зрозуміли». Натомість правил для «Я»-граматики не існує. Позаяк існує больовий тип поведінки, він підказує нам, коли казати «моя рука болить» і коли недоречно вживати слово «біль». Для представника егологічної теорії одна з мовних ігор (сумнів) починає домінувати над іншими, руйнуючи їхню прозорість і розмаїття.

2.2. Логічний бігевіоризм Г. Райла. Ум як категорійна помилка.

Сучасник і послідовник Вітгенштайна, один із очільників оксфордської школи філософії мови **Гілберт Райл** (Ryle) обстежив проблему свідомості з позиції популярної в його часи бігевіористичної психології. Сутність цього методологічного підходу можна виразити так: про думки особи свідчать її *поведінка*, за якою ми висновуємо про наміри, що їй передували. Основним

критерієм виснування є встановлення зв'язку між думкою про дії, диспозицією до дії та власне виконанням самого акту. Якщо, приміром, людина прогулюється вулицею з парасолькою, коли немає дощу, наявність парасольки вказує на припущення носія парасольки про ймовірність дощу. Від покладення особою певності «сьогодні дощитиме» до її рішення взяти з собою парасольку бігевіорист робить логічний умовивід за правилом *modus ponens*:

1. А; 2. А, отже В; 3. В.

Базовою ідеєю бігевіоризму є уявлення про те, що будь-яку дію засновано на принципі «стимул-реакція-підкріплення». Якщо лабораторного щура привчити за запахом сиру знаходити захований корм, він сформує у себе звичку шукати корм після появи відповідного запаху в кожному новому приміщенні. Те саме відбувалося, коли Іван Павлов досліджував вироблення умовних рефлексів у собак. Пізніше, вже у 1960-х роках американський логік Вілард Квайн (Quine) розробив на основі бігевіористичної програми концепцію унітарної метамови науки.

Райлова теорія переймається тим, що менталістський словник перенасичений поняттями, для яких важко знайти відповідники у повсякденній мові і, що важливіше, яким бракує аналогів у публічно спостережуваній поведінці. Якщо займенник «Я» допомагає ідентифікувати актуального мовця, то, скажімо, іменники «душа» або «ум» взагалі позбавлені чітких семантичних маркерів. Цими словами можна назвати і щось абстрактне, і щось конкретне (пор. вислови: аристотелівський «душа світу», фразеологізм «вкласти душу в улюблену справу», антропологічний «шаман аборигенів виконав ритуал убивства душі злого демона» і повсякденний «після весілля Петрова душа радіє»). Однак вагомішим є Райлове міркування про те, що заборона на приватну мову відчуттів має поширитися й на сферу ментального. Британський мислитель вважав, що завданням філософії є головно здійснення *логічної картографії нашого мислення*. Пояснити цю ідею легко, залучивши метафору корабля, який, виходячи з порту А, вирушає у плавання до порту В. Плавання буде успішним за умови, що капітан судна має надійні лоції,

правильно прокладений маршрут і навички кермування. Аналогія з кораблем допомагає збагнути, що філософія потребує ясності, коли йдеться про висловлення, які містять труднощі й парадокси. Методом їх прояснення стає *reductio ad absurdum*, тобто тимчасове (у процесі спростування) прийняття істинності тези, яка імплікує суперечливість її висновків.

На думку Райла, ключовою підставою появи псевдосутностей на кшталт душі, свідомості та похідних від них понять менталістського словника, є вкоріненість у філософській традиції т. зв. картезіанського міфу. Суть цього міфу (Райл називає його *deus ex machina*, «дух з машини» або «привид у машині») полягає в тому, що у влаштованому за принципом механізму фізичному тілі існує таємне дещо, яке живе власним незалежним життям. Як і Платонів міф про печеру, Декартів міф про душу (свідомість), здатну аналізувати свої переживання самотужки не взоруючись на об'єктивні критерії, призводить до *категорійних помилок* у судженнях. Приклади здійснення цих помилок Райл наводить у своїй книзі «Поняття ума» (*The Concept of Mind*, 1949). Уявіть, каже він, що ви – іноземний турист, який приїздить до Оксфорду. Прагнучи ознайомитися з видовищами славетного університетського містечка, ви запитаете когось із перехожих, де розташовано університет. І почувте на свій подив: «А який коледж вам потрібен?» Мені потрібен Оксфордський університет, наполягаєте ви. І тут несподівано виявляється, що співбесідник не розуміє ваш намір. Але що йому згодом спадає на думку, так це те, що ви хибно уніфікуєте різні коледжі Оксфорду в *один* «університет» (який є для нього абстрактним поняттям). Так само і з ментальними явищами: замість вивчати їх порізно, а потому у можливих відношеннях, адепт картезіанської традиції дошукуватиметься тієї *сутності*, яка відповідає за *весь* клас ментальних процесів.

Таким чином картезіанська думка (услід за платонівською) вважає кожну людину комбінацією безсмертного духу й вмирущого тіла; одне відповідає за безпомилкове сприйняття речей і самого себе, а друге схильне помилятися у сприйманні через відсутність у тіла здатності рефлектувати. Дві сутності

протиставляються одна одній, що й веде до труднощів у пізнанні у цей спосіб подвоєного світу. Одначе Декарт постулює вищість духу стосовно тіла, а отже, й абсолютну цінність висновків у судженнях, які робить дух. Привілейований доступ до ментального (духу) виявляється інверсією приватної мови, через яку суб'єкт описує досвіди свого «внутрішнього» світу.

Менталістський лексикон є проблематичним ще й тому, що всі без винятку терміни, пов'язані з ментальними подіями, є важковизначуваними. Візьмімо хоча б дієслово «знати» і похідний від нього іменник «знання». Не так легко відрізнити знання про себе й знання про речі. Повторюючи тезу Вітгенштайна, Райл запитує: який статус має «свідомість» (mind) (якщо її існування прийняти на віру, адже безперечних доказів на її користь немає) – об'єкта чи суб'єкта інтроспективних самоописів? Якщо вона є суб'єктом, то чому у висловленнях про свої стани вона перетворюється на об'єкт аналізу? Чи підтверджує це відсутність у неї чіткого онтологічного статусу? Якщо ж її варто вважати об'єктом, тоді не зрозуміло, на яких підставах і за яких умов цей об'єкт однозначно трансформується у суб'єкта? І чи йдеться лише про граматичний суб'єкт речення (підмет), або ж, навпаки, про пізнавальний суб'єкт (Я)?

На ці питання неможливо дати однозначно позитивну відповідь. Свідомість, як і «Я», є невловимими для несуперечливого мислення. А отже, такий стан справ зайвий раз доводить, що свідомість або самість є ні чим іншим, як продуктом категорійної помилки мислення, взалежного від метафізики. *Категорійна помилка* – це хибний умовивід, в якому об'єкту приписують властивості, типові для зовсім іншого класу понять (категорій). Коли ми кажемо «я усвідомлював/-ла», «я розумів/-ла» тощо, ми кажемо про свої припущення як про надійні факти, але шлях до яких залишається неясним. «Самосвідомість» – поняття, похідне від «свідомості», проте в обох випадках його можна десубстантивувати, замінивши на дієслівні звороти «вдивлятися у» і «виявляти у собі». Але вдивляння спрямоване на зауваження одних подій і нехтування іншими (або виведення їх на периферію погляду), тоді як коли

йдеться про свідомість, у нас виникає враження повного і вичерпного огляду всіх подій, із них пов'язаних. Саме заміна виразів теоретичного дискурсу про ментальне на лексику повсякденної мови слугуватиме, на погляд Райла, проясненню наших знань про ментальну сферу. Як тут не згадати Вітгенштайнову пораду: звертай увагу на те, як функціонує певний вислів у повсякденній мові.

Джерело проблеми сплутування значень теоретичного й фактичного дискурсів полягає в абсолютизації слів, якими ми звикли описувати свої психологічні переживання. Їх поєднання у реченні називають *пропозиційними настановами*: це вираження бажань, намірів, думок, припущень, переконань, сприйняття, що напряду вказують на мовця. Частий ужиток цих слів навіює віру в те, що вони стосуються особливої сфери і особливого суб'єкта, який їх застосовує. Таке некритичне використання дієслів тягне за собою гіпостазування переживань мовця, який послуговується ними, перетворення їх на окремий «світ поза світом речей». Історія лінгвістики пояснює це явище як наслідок субстантивації іменників. Приміром, давньогрецькі мислителі субстантивували слово «бути» у «Буття», «діяти» у «діяльність» (*energeia*), римляни скалькували грецьке «те, що лежить в основі» як *subjectum* і т. п. Залежність (як і шкода) європейської культури від цих субстантивів, на погляд Райла, величезна.

Вихід Райл убачає у виконанні мінімальної двокрокової програми. Першим із них є розрізнення двох когнітивних модусів – «знання що» (*know what*) і «знання як» (*know how*). Коли ми повідомляємо про безсумнівні факти («за вікном дощить»), то оперуємо знанням-що; коли ж пояснюємо, як полагодити автомобіль, то звертаємося до знання-як. Адже не обов'язково знання про влаштованість автівки дає нам розуміння того, як усунути пошкодження. Нерідко, щоб зрозуміти причину теоретичних або практичних утруднень, ми курсуємо між першим модусом і другим, вважаючи, що вони є частинами одного пазлу. Так і відбувається, коли мова заходить про «свідомість». Можна задовільно вивчити іноземну мову або навчитися писати,

навіть не маючи уявлення про прийменники, дієприкметники та інші частини мови.

Другий крок стосується розрізнення психологічних *диспозицій до поведінки* (схильності чинити так, а не інакше за умови настання певних обставин) і зовнішніх *подій*. Констатація подій утворює клас нейтральних фактів, а опис диспозицій дає нам розуміння людських намірів, бажань і рішень. Разом із тим, події приналежні до царини знання-що, натомість диспозиції розкривають умови успішного набуття знання-як. Будучи бігевіористом, Райл кладе у підмурівок своєї концепції поняття регулярності та каузальності. Регулярність настання подій пояснює, чому вони є незалежними від особи і як саме їхня повторюваність уможлиблює отримання знання про фізичну реальність за допомогою прогнозування. Утім, регулярність подій створює необхідні умови для актуалізації людськими дієвцями (*agents*) диспозицій поводитися так, а не інакше за рівних однакових обставин. Регулярність подій, їхня визначуваність однорідним набором факторів показує, чому резонно казати про формування стабільного зв'язку між переконаннями й подальшими вчинками. Обидва кроки програми Райла передбачають захист «слабкого» бігевіоризму, заснованого на аналізі ужитку слів повсякденного мовлення. Прикметною є Райлова ремарка про етимологію засадничих понять менталіста: англ. *consciousness* походить від вигаданого протестантськими теологами лат. *conscientia* («сумління», «совість»), а *reflexion* – від *reflecto* («дзеркально відобразитися»). Тож у розмові про свідомість домінують оптичні та містичні метафори і саме як метафори вони ладні дезорієнтувати.

Зауважмо, що природа Райлового бігевіоризму є не психологічною, а логічною. Від психологічної течії, очільниками якої були видатні експериментатори Джон Вотсон і Берес Скінер, його відрізняє те, що Райл поділяє фундаментальну настанову багатьох аналітичних філософів – недовіру до психології як науки, зумовлену побоюванням перед *психологізмом*. Психологізм – це тенденція редукувати ментальні змісти до ментальних актів,

завдяки яким вони й уможливлені. Проте небезпека психологізму не так у редукції, як у навіюваних нею висновках. Психологіст вважає, що не існує відмінності між актом і змістом ні на універсальному, ні на індивідуальному рівнях. Візьмемо для розгляду популярне твердження «Логіка є наука про закони мислення». Воно є прикладом відвертого психологізму, адже виходячи з цього твердження неможливо провести демаркаційну лінію між логікою та психологією. Крім того, якби логіка вивчала закони мислення, а не закони істинності суджень (чим вона, власне, й займається), тоді довелося б визнати, що виведені Аристотелем логічні принципи мають чинність тільки для нього самого, або у кращому разі, для носіїв Античної грецької культури і в наш час втратили свою універсальність. Отже, логічний бігевіоризм виходить з того, що не тільки тілесна, а передусім *вербальна* поведінка (вона опосередковує механічні рухи тіла) дає нам доступ до наших ментальних станів, аби повідомити про них інших осіб.

Утім, первинний оптимізм Райлового проєкту наштовхується на критику, загальну щодо методології бігевіоризму. Виділимо ключові заперечення проти неї. По-перше, стимул-реактивна доктрина часто хибить. Згадаймо приклад з власником парасольки. Якщо людина несе її в руці, бігевіорист зробить висновок: вона ознайомилася з прогнозом синоптиків і готується захиститися від дощу. Насправді ж, парасолька може бути лише аксесуаром або полегшує ходу для кволого, або навіть річчю, яку людина несе до свого друга, який забув її вчора у неї в гостях. Тож просте спостереження за людською поведінкою насправді не дає підстав мати певність у своїх припущеннях.

По-друге, спробу транслювати всі «тьмяні» значення теоретичної мови на мову повсякденну навряд чи вдасться реалізувати, адже згадана лексика виконує принаймні важливу функцію економії мислення (пор. некоректну для Райла фразу «душевні муки» і коректну «надзвичайне переживання, що охоплює особу мовця, не пов'язане зі станом його тіла і виражене за допомогою таких дієслів...»).

По-третє, логічний бігевіоризм не заперечує існування ментальних явищ, а наголошує, що їх неправильно поійменовано. Таке корегування призводить, по суті, до заміни ментального словника на словник фізичний. Знання-що і знання-як так само вимагають верифікації за допомогою словника природознавства. Бігевіорист екстерналізує змісти, для яких відсутні фізичні відповідники.

По-четверте, аналіз граматики висловлювань про свідомість не пояснює функціонування самих ментальних станів. Поняття про переживання утворюються *post factum*, тож їх аналіз розповідає не про саме ментальне, а про те, як ми *звикли* говорити про події переживань. Це і є головний закид проти логічного бігевіоризму. У рамках цього підходу можна казати не про примноження знання, а радше про його (не завжди вдалі) шліфування й селекцію.

Наостанок додамо, що у середині 1970-х років, коли концепції Вітгенштайна і Райла вже перебували у статусі класичних, але не беззаперечних, доктрин, з'явилася теорія, що реабілітувала менталістський словник на противагу бігевіористському рішенню. Засновник психофункціоналізму **Джері Фодор** (Fodor) у книзі «Мова думки» (1975) стверджував, що завдяки своїй символічній структурі мова є вузловим посередником між комплексними змістами свідомості. Мова думки (*mentalese*) є універсальною знаковою системою, яка надає нам семантично пов'язані між собою ментальні поняття для самоописів. Отже, мова не передує мисленню, а опосередковує наявні в ньому концептуальні кластери через спільну для різних мовців лінгвістичну форму. Думки утворені обмеженим набором семантичних примітивів (лексемами), комбінації яких дозволяють створювати нові й нові фрази. Оскільки людський мозок не спроможний зберігати великі обсяги інформації про доступні семантичні комбінації, вони розміщуються у вигляді «підказок» у практиці самого мовлення.

Змісти наших ментальних станів є такими собі неартикульованими лексемами, які людина транслює за допомогою повсякденної мови. Коли носії

різних мов, спілкуються мовою-посередником (скажімо, англійською), обговорюючи предмет розмови, вони підбирають потрібні їм слова без особливих труднощів тому, що ментальні процеси є автономними стосовно процесу їх вербалізації. Отже, правило значення як ужитку і дотичний до нього поведінковий критерій не здатні пояснити, чому цілком відмінні індивідуальні досвіди вдається символічно увідповіднити. Саме наявність семантично однорідного внутрішнього мовлення уможливорює комунікацію між людьми. Таким чином нементальні слова і фрази конструюються на основі ментальних репрезентацій. Психофункціоналізм, з одного боку, прагне довести, що семантика підпорядкована ментальним репрезентаціям, а з другого боку, що несвідоме у нашій психіці становить її фундаментальний і найбільший пласт, тоді як наука має до діла лише з її верхнім, видимим шаром. Хай там як, символічна структура, з яким резонує активність нашого мозку, залишаються для Фодора ноуменальною і доконечно непізнаваною. У зв'язку з цим його підхід зазнавав неодноразових атак як від матеріалістів, так і від дуалістів, особливо за видову вродженість «мови думки» і тезу про безпомилковість її знаків.

Джерела:

Ryle, G. (1949). *The Concept of Mind*. Hutchinson House, pp. 11–23, 35–60, 154–198 (Ch. 1, 2, 6).

Wittgenstein, L. (1958). *The Blue and Brown Book*. Blackwell Publishing, pp. 53–74.

Wittgenstein, L. (1953). *Philosophical Investigations*. Double-Language Edition, trans, by G.E.M. Anscombe. Blackwell Publishing (§§ 241 – 420).

Синиця А. (2017). *Сучасна аналітична філософія: від прагматики мови до концептуалізації свідомості*. Л.: ЛНУ ім. І. Франка, сс. 91–112.

Запитання для самоперевірки:

1. Спробуйте стисло пояснити зміст егологічної теорії на прикладі.

2. Чому самоочевидність «Я» не є задовільним критерієм знання?
3. Як апологет картезіанства міг би відповісти на аргумент приватної мови?
4. Спробуйте обґрунтувати хибність егологічної теорії з позиції професійного юриста.
5. Картезіанське вчення твердить, що душа існує незалежно від тіла за своїми особливими законами. До яких утруднень веде така позиція?
6. Психологічний бігейвіоризм заявляє, що свідомості поза поведінкою не існує; логічний бігейвіоризм переконує, що свідомість (ум) є результатом категорійної помилки повсякденного мовлення. Чи справді цим двом підходам йдеться про різні речі?
7. Що таке привілейований доступ до свого «Я»? Як Декарт і Райл пояснюють сутність «привілейованого доступу»?

Лекція 3.

Матеріалістичний редукціонізм. Фізикалізм і функціоналізм.

3.1. Фізикалістська Теорія тотожності та її критика.

Матеріалізм – це вчення, що спирається на твердження про існування лише фізичних сутностей. Матеріалізм є теорією, що послуговується методом редукції одного типу сутностей до іншого типу сутностей (нематеріальних до матеріальних). Отже, редукціоністським є теоретичний підхід, який визначає невідоме через відоме або зводить поняття одного порядку до понять іншого порядку, заперечуючи самостійність перших. Наприклад, редукція психічного до фізичного означає, що психічне розглядається як несамостійне щодо фізичного або навіть як псевдосутність. Кажучи про редукціоністські теорії, що намагаються розв'язати психофізичну проблему, зосередьмося передовсім на двох його напрямках – фізикалізмі та функціоналізмі. Кожен із них є за визначенням *матеріалістичною концепцією*, себто заперечує існування нефізичних властивостей, сутностей або подій. Зауважмо, що сучасна

філософія свідомості є переважно матеріалістичною та сцієнтистською за своїм духом і водночас саме варіативне розмаїття в межах матеріалізму (від крайньої до поміркованої його форми) якраз і визначає простір дискусій про проблему свідомості та тіла.

Однією з головних редукціоністських течій у сучасній філософії свідомості є *фізикалізм*. Суть цієї доктрини можна звести до тези «існують тільки матеріальні явища, а існування нематеріальних довести неможливо». У 50-70-х роках ХХ ст. найвидатнішими представниками фізикалізму були австралійський філософ **Девід Армстронг** (Armstrong) (австралійська філософська традиція саме йому та його учням завдячує своєю появою) та британський мислитель **Джон Смарт** (Smart). На сьогодні фізикалістську теорію тією чи іншою мірою обстоюють Майкл Тай (Tye), Джері Фодор і подружжя Патрисія та Пол Чорчленди (Churchland). Аргумент фізикалістів можна викласти у вигляді полісилогізму:

Існує лише матеріальний світ.

Існування нематеріального світу є невідповідним.

Матеріальне утворення функціонує за фізичними законами.

Мозок є матеріальним утворенням.

Нефізичних властивостей мозок не має.

Отже, ум (mind) є мозком, бо працює як мозок.

Отже, думки і всю сферу ментального можна локалізувати у центральній нервовій системі живої істоти.

Скажімо, якщо людина випиває склянку води, до цього її спонукають не наміри чи ідеї, а виключно фізичні процеси у головному мозку (для роботи всього організму потрібна рідина, яка перетворюється за законами біохімії). Тож мозок фіксує проблему: «є нестача рідини», після чого (надшвидкісно) надсилає сигнал м'язам руки «налити склянку води і випити». Зважаючи на це, намір людини випити воду слід вважати не бажанням, а імпульсом, що вмикає систему дій, яку у вигляді єдиної схеми реалізують рука, стравохід і шлунок. Як бачимо, ідею фізикалістів можна виразити у принципі тотожності

фізичного й нефізичного. Нерідко фізикалізм зводять саме до *теорії тотожності* (далі – ТТ), яка, своєю чергою, поділяється на два окремі види. *Фізикалізм токєну* (token physicalism) заявляє, що кожна конкретна подія, що підпорядковується окремому типові психічних подій за певним маркером (токєном), також підпорядковується окремому типові фізичних подій. Наприклад, появі болу відповідає одночасне збудження С-волокон. У даному разі тотожність виявляють за окремим маркером, релевантним для обох явищ. *Фізикалізм типу* (type physicalism) стверджує, що певні типи психічних явищ корелюють з певними типами фізичних подій, а психічні властивості мають відповідники у фізичних властивостях. Приміром, кожен об'єкт, що має колір, завжди має також і форму, отже, ідентифікація об'єктів за кольорами веде до їх ідентифікації за формою. У цьому разі тотожність виявляють за окремим типом, релевантним для обох явищ.

Як фізикаліст, Смарт розглядав усі явища культури та, зрештою, класи імен речей як свідчення того, наскільки мова науки відрізняється своєю точністю від повсякденної мови. Такий підхід отримав назву *аргументу словника*. Наприклад, бачачи блискавку в небі, пересічний спостерігач скаже, що це «блискавка», а вчений уточнить, що об'єктом спостереження є «явище спалахів електричних розрядів у газоподібному середовищі». Хоча обидва описи вказують на один і той самий об'єкт, проте зміст цих описів суттєво різниться, адже слово «блискавка» є не в усьому тотожною визначенню, даному вченим (адже це слово можна вжити метафорично або іронічно). Точніше, простота поняття «блискавка» робить її вразливою до неточностей (пор. «блискавичний провал»). Смарт зауважує: всі ми живемо у просторі різних словників. Приміром, вчений послуговується і академічним, і посполитим словниками (один є чинним у спілкуванні з колегами, інший – з усіма ненауковцями), однак потрібна ієрархія словників за їхньою точністю. Отож вчений, розмовляючи «по-простому», санкціонує для себе тимчасовий відступ від наукового словника, розуміючи при цьому його недосконалість. Водночас ненауковець, навіть послуговуючись елементами академічної

лексики, не може вправно оперувати ними, тому залишається в полоні пасток повсякденної мови.

Одним із ключових заблудів цієї мови є її насиченість квазісутностями, для яких відсутні відповідники в академічному словнику. Авжеж, «свідомість», «психіка» та «ментальні явища» є прикладами таких сумнівних сутностей. Таку нестиковку задовго до Смарта описав німецький логік Готлоб Фреге. Він помітив, що одну й ту саму зірку, коли вона з'являється на небосхилі вранці і ввечері, люди називають по-різному («ранкова зірка» і «вечірня зірка»). Насправді, каже Фреге, в обох випадках референтом є один і той самий об'єкт – планета Венера. Виглядає так, що єдина сутність може мати щонайменше декілька семантично відмінних виразів, в яких обіжний ум не вбачає відмінностей. Це створює ілюзію множини речей, хоча насправді річ є одна й та сама. Постає питання: чи доцільно вважати ум або свідомість чимось окремим, як у випадку з ранковою зіркою, якщо їх можна вписати у більш синхронні нейрофізіологічні процеси? У такий спосіб фізикалізм відмовляє ментальному в самостійному онтологічному статусі.

Проти доктрини фізикалістів ледь не одразу пролунали заперечення. Для зручності їх можна згрупувати у кілька основних:

1. Якщо свідомість є ні що інше, як біохімічні процеси у мозку, тоді доводиться стверджувати, що такий результат актів мислення, як спеціальна теорія відносності Альберта Айнштейна, є всього лише реакцією натрію та хлору у мозку цього вченого. Однак у такому разі незрозуміло, чому типова хімічна реакція підштовхнула одну людину до відкриття, а інших подібних йому осіб навіть не наблизила до нього.
2. Теорія тотожності (ТТ) не здатна пояснити, чому людські істоти часто діють усупереч фізичним законам (завдають собі болю, входячи в охоплену вогнем будівлю; відмовляються від їжі, коли є голодними) або ж ризикують власним життям, вживаючи психотропні речовини чи жертвують собою заради «абстрактних цілей», оголошуючи, скажімо, політичне голодування. Стабільна робота мозку ніяк не пояснює ці

соціальні явища, що є надлишковими з точки зору нормального функціонування фізичного тіла.

3. До того ж, ТТ не може дати вдумливого пояснення того, чому певна серія імпульсів продукує саме таку реакцію, а не іншу. Приміром, чому збудження однієї групи С-волокон у мозку призводить до пригнічення, а не до радості або що.
4. Разом із тим, фізичні процеси є описово нейтральними, тоді як наші суб'єктивні переживання й досвіди зазвичай мають у собі потужне якісне та оцінкове забарвлення. Тим самим закид проти матеріалістів виводить на проблему *квалій* (докладніше див. 4.2.).

Ймовірний мотив цих опонувань криється в тім, що каузальність, покладена в основу процесів фізичного порядку, зумовлює дію нової причини, але міркування й переживання (особливо естетичні, політичні та моральні) самі по собі нічого не спричиняють у матеріальній реальності, а, навпаки, походять із реакції на неї. Як пояснити, що люди впродовж десятиліть малюють живописні полотна або студіюють твори Гегеля, які з погляду біолога не вкладаються у рамки необхідного ритму відтворення життя? Ці та інші заперечення проти узгодженості та задовільності ТТ значно підважують її доказову систему.

3.2. Модальний аргумент Кріпке.

Проте найбільшого удару по ТТ завдав модальний аргумент **Сола Арона Кріпке** (Kripke). Американський логік виходив з припущення, що концепція фізикалізму, аби бути переконливою, має спиратися на поняття *строгої необхідності*: якщо стверджують тотожність між фізичним і нефізичним явищем, то вона має бути не можливою, а *необхідною* тотожністю, себто ідентичність двох явищ не повинна передбачати жодних винятків. Два явища тотожні, тільки якщо вони *завжди й за будь-яких обставин* є тотожними. Свій аргумент філософ базує на теорії жорстких десигнаторів (однозначних позначників; rigid designators): поняття позначає об'єкт у такий спосіб, що в

цьому (фактичному) та інших (логічно можливих) світах зв'язок між поняттям і об'єктом не буде порушено. Скажімо, якщо взяти до розгляду власну назву «Еверест», то воно є жорстким десигнатором, адже йому відповідає вичерпне значення, яке неможливо змінити простою комбінацією слів «Гімалаї», «найвища вершина» і «виклик для альпініста». Загалом своє заперечення проти ТТ Кріпке сформулював так:

Будь-яка тотожність має бути необхідною.

Випадкова тотожність не є насправді тотожністю.

Необхідна тотожність чогось із чимось є тотожністю самою по собі.

Ум і мозок є нетотожними (I).

Мозок і ум взаємодіють каузально.

[Отже, існує вплив одного явища на інше.]

Отже, мозок і ум нетотожні (II).

Щоб відповісти на закид з боку Кріпке, фізикалісти мушили продемонструвати, що з логічним поняттям тотожності щось не гаразд і без нього можна обійтися у їхніх мисленневих побудовах. Головний контраргумент полягав у спростуванні зв'язку між тотожністю та необхідністю. Якщо Кріпке наполягав на тому, що це і є серцевина поняття «тотожність», його опоненти заявили, що необхідна тотожність – це суто метафізичне основоположення, застосування якого є малопродуктивним. І справді, строге тлумачення поняття «необхідність» Кріпке, ймовірно, запозичує у Ляйбніца, який вважав абсолютною тотожністю такий її вид, в якому *всі* без винятку властивості об'єкта X є тими самими, що й властивості об'єкта Y. Контраргументом проти цього твердження є той, що, покладаючи метафізичну тотожність, ми вже повинні знати *всі* умови необхідності цієї тотожності. Натомість прогрес знання, скінченність нашого розуму й відмінність фактичного й логічно можливих світів ставить перешкоду для виконання цього завдання. Якщо поняття тотожності, наполягають опоненти Кріпке, утверджує інваріантність, тоді метафізичний характер кріпкеанського аргументу веде ось до чого: будь-що, у чому наш розум виявить відношення

тотожності, вже міститиме повний обсяг поняття, яке дасть вичерпний опис самого явища, але це неможливо.

Крім того, продовжували фізикалісти, нездатність мислити складні комплекси понять, зокрема метафізичні поняття, не означає їх неможливості. Іншими словами: нестрога тотожність не виключає можливості певних явищ. Ми не можемо уявити собі тисячокутник, але це не є доказом проти його абсолютної неможливості. Так само і з тотожністю мозку та ума: хоча виразити це відношення технічно непросто, це не означає, що воно у принципі є неможливим. Якщо ми спроможні уявити людинолева чи однорога, що заважає нам мислити думки як точки, локалізовані в нервовій системі?!

Концептуальна трудність із тотожністю стосується не її природи, а способів її увиразнення в мові – природна мова завжди поступається в точності та адекватності формальній мові, проте відомо, що мова наук так чи так ґрунтується переважно на повсякденній мові і змушена враховувати це. Повністю формалізовані мови не придатні для спілкування. Коли постають проблеми з ідентифікацією політично значущих осіб, щодо яких застосовують різні описові конструкції («Джон Кеннеді» і «35-й президент Сполучених Штатів»), ці утруднення можуть бути розв'язані без заперечення інституцій, які продовжують діяти на тлі обговорення проблеми. Простіше кажучи, двоє людей, що сперечаються з приводу того, хто був 35-м президентом США, не ставлять під сумнів виборний (легітимний) статус об'єкта суперечки.

Те саме і з тотожністю мозку та ума: не потрібне додаткове знання, щоб казати, що я думаю «мозком», а не «умом», хоча і в першому, і у другому випадку йтиметься про процес мислення. Фактично цей контраргумент повторює Смартів аргумент словника – народна психологія спрощує складне з метою економії часу та зусиль (кажемо про «блискавку», а не про «електричний розряд у газоподібному середовищі»). Хай там як, але модальний аргумент не став смертельним цвяхом у домовину фізикалізму, хоча й суттєво послабив його позиції.

Підсумуймо основні заперечення проти фізикалізму. По-перше, стани свідомості не можуть бути одночасно станами мозку, адже стани мозку розміщені у каузально замкненому просторі, тоді як не всі ментальні стани мають очевидну причину своєї появи (образ спричинено сприйняттям, але спогад ним уже не спричинено). Тим паче, у деяких із них немає навіть конкретного об'єкта, як-от у депресії або стану тривоги. По-друге, дається ознаки відмінності феноменальних якостей. Якщо фізикаліст стверджуватиме (покликаючись на принцип нерозрізненості тотожного), що «гострий зубний біль» і «гострий стан мозку» є одним і тим самим, то нейрофізіологи парирують цю думку тезою, що набір тотожних ознак не викликає з необхідністю однакові стани свідомості. Активізація певних ділянок мозку призводить або до появи А, або до появи В, або В і С. По-третє, щоб довести тотожність якісно однакових станів, потрібно припустити існування феноменальних якостей (квалій), що їх фізикалісти рішуче відкидають. По-четверте, модальний аргумент Кріпке показує, що підмурівок ТТ погано справляється з поясненням природи тотожності. Якщо біль тлумачити як жорсткий десигнатор, тобто біль виникає *завжди*, коли збуджуються С-волокна, існують підстави вважати, що «збудження С-волокон» не є жорстким десигнатором і, отже, тотожність двох станів не є необхідною, а випадковою. «Слабка» версія тотожності ще більше підважує переваги цієї теорії і спонукає до її перегляду.

3.6. *Машинний функціоналізм і бігевіоризм. Тест Тюринга.*

На відміну від фізикалізму, цей напрям філософії свідомості вбачає вихід у розв'язанні психофізичної проблеми у відході від крайніх форм редукціонізму. Якщо сформулювати кредо філософа-функціоналіста, то воно звучатиме так: для кожної ментальної властивості існує ціле різноманіття способів її реалізації. Такий підхід зазвичай називають *принципом множинної реалізації*. Він і становить осердя функціоналістської доктрини. Поглянувши на природний світ, ми можемо помітити, що різні істоти – молюски, рептилії,

савці – мають структурно відмінний від людського мозок, а проте, вони так само, як і людина, переживають біль або задоволення, визначають для себе цілі й розпізнають джерела небезпеки. Отож варіативність способів задіяти свій «центр управління» не робить цих істот зовсім несхожими. Натомість наявність алгоритмічної подібності в їхніх діях дає підстави вважати, що слід зосередитися на вивченні не так будови цього центру, як моделей його конкретної алгоритмізації.

Отже, ум або психіку (mind) функціоналісти розуміють як посередник між вхідними даними (inputs) та вихідними даними (outputs). При цьому «посередник» не існує початково, а виникає *в результаті* відповідного відношення. У певному сенсі, наша психіка (для даної школи найбільше пасує саме такий переклад поняття mind) – це система, в якій злагодженість функцій веде до того, що ми не добачаємо її наявності, коли все працює гармонійно і, навпаки, схильні запитувати про психіку, коли в системі починаються нестиковки.

Вхідні дані – це каузально спричинені сигнали з зовнішнього середовища (хай би яким воно є), а вихідними даними вважаються реакції на подібні впливи (механічні дії, вербальні, як-от судження). Усю культуру таким чином можна назвати вихідними даними дій природи, спрямованими на нас та на інші біологічні організми. Як ми помітили, підґрунтя редукціонізму в такій функціоналістській теорії все-таки дає про себе знати. До того ж, роботу психіки нерідко – у згоді з запропонованою термінологією – функціоналіст схильний уподібнювати до роботи машини, скажімо, комп'ютерної програми, утвореної взаємодією матеріальних засобів (hardware, корпус, «тіло») і нематеріальних факторів (software, власне коди програми). Залучення метафор комп'ютерних систем дало нагоду прибічникам такої версії функціоналізму назвати його машинним функціоналізмом або машинним редукціонізмом. Представниками цієї течії у 1960-80-х роках було чимало мислителів, найприкметнішим з-поміж них вважають професора Гарвардського університету **Гіларі Патнема** (Putnam), видатного епістемолога і логіка.

Важливо вказати на залежність машинного функціоналізму від кібернетичної теорії **Алана Тюринга** (Turing), британського математика, який у 1940-х роках розробив протокомп'ютер – машину Тюринга (докладніше про це див. 8.1.). Тюринг вигадав тест, у разі проходження якого можна надійно визначити, хто його виконав – машина чи людина. Для цього членам журі достатньо просто «не розпізнати», де людська відповідь, а де зроблена машиною. Посиливши цю тезу й дещо осучаснивши умови самого тесту, можна стверджувати: якщо тест Тюринга буде пройдено, це засвідчить відсутність істотної відмінності між функціоналом людського інтелекту та інтелектом як таким. І якщо це припущення справді правильне, тоді вчені повинні будуть досліджувати принципи роботи ума незалежно від того, кому він належить, адже умові здатності є універсальними для будь-якого носія інтелекту. Тим самим функціоналізм прагне позбутися труднощів, що ставляться психофізичною проблемою, вкупі з менталістським словником. Заміною цього словника має стати термінологія кібернетичних і технічних наук. І це не дивно, адже комп'ютерна революція, як і стрибок у когнітивістиці, почалися саме у 1960-х і збіглися з періодом формування функціоналістської парадигми.

Патнем висунув низку контраргументів і проти ТТ, і проти її противника Кріпке. *Аргумент двійника Землі* (Twin Earth argument). Взявши за основу уже відомий нам аргумент словника, він запропонував здійснити мисленнєвий експеримент. Уявімо собі, що існує планета-двійник Землі і на нього висадилися астронавти з нашої планети. На ній все так само, як у нас, за винятком того, що воду ці двійняни називають не водою, а просто за відомою їм формулою рідини. Якщо на Землі формула води відповідає хімічній сполуці H_2O , то на двійнику Землі нею є XYZ. Спробувавши на смак цю рідину, астронавти-земляни виявлять, що вона нічим не відрізняється від нашої води. Однак, коли запитати місцевих мешканців, чи їхня рідина є «водою», вони заперечуватимуть. Тож хоча вода буде «водою», її назва створюватиме враження, що це інша субстанція, нетотожна земній воді. Мораль цього

експерименту в тім, що хай би як ми називали ментальне і свідомість, учасники дискусії про них навряд чи зійдуться у згоді про їхню природу; натомість, варто сперечатися не про букву, а про дух певного явища, або, кажучи інакше, про те, «як воно реально функціонує». Опис функцій у дії повідомляє про природу явища краще, ніж усі його формально-семантичні описи.

Аргумент «мізки у діжці» (Brains in a vat argument). Уявімо, що певні злі вчені вирішили прооперувати людину, помістивши її мозок у діжку, наповнену рідиною, яка підтримує його життєдіяльність. До мозку в діжці вчені приєднали дроти, які сполучають його з суперкомп'ютером. Таким чином на моніторі можна бачити «думки», які вигулькують у мозку, а з іншого боку, комп'ютер може передавати імпульси прямо у мозок для стимулювання його роботи в потрібному руслі, викликаючи в ньому образи. У цьому стані мозок начебто продовжує функціонувати так, ніби він перебуває у черепній коробці. Ментальним образам завжди відповідають слова, що їх промовляють, коли намагаються повідомити про образи. Мозок передає по датчикам сигнали, які суперкомп'ютер розкодує як слова чи фрази про те, що мозок зараз переживає. Однак проблема якраз у тому, що він «переживає» тільки те, що йому навіює сам комп'ютер. Відмінності між реальністю та вигадкою сам мозок встановити не спроможний.

Основа аргументу Патнема містить дві тези – скептичну й антискептичну. Скептичну тезу утворює таке твердження: можливо, світ і є суперкомп'ютером, а ми є «мізками в діжці», які не здатні усвідомити, що наші думки не є нашими – вони є продовженням певної раціональної програми, яка керує світом. Коли я кажу комусь що-небудь, я вважаю, що мене «чують», тоді як насправді цьому немає жодних підтверджень. Як і людина в описаному експерименті, ми не можемо побачити різниці між «реально» і «здається». Антискептичну тезу, яка спростовує попередню, формуючи такий собі двоступеневий антиномічний доказ, становить *перформативне* припущення: якщо я стверджую, що те, чим є реальність, неможливо з'ясувати, я вже

покладаю саме поняття «реальність», яке передбачає властивості, за якими його якраз і можна з'ясувати. Стосовно експерименту з мізками, це означає, що хоча може здаватися, що ми діємо як мізки в діжці, насправді ми ними не є. Адже бути реальним собою означає керувати своїм тілом, якого у діжці бракує. Тим самим Патнем доводить, що (а) бути реальним собою і (б) мислити згідно з певним алгоритмом не суперечить одне одному. У цьому й полягає головний доказ функціоналізму.

Загалом аргументація Патнема (і ширше – функціоналістів) націлена на те, що зв'язок між репрезентацією і референцією до об'єктів світу порушується тільки в тому разі, якщо в їхнє опосередкування потрапляють сумнівні («магічні») поняття на кшталт «дух», «субстанція» тощо. Заперечуючи ум як сутність, незалежну від характеру відношення, в яке вона включена, функціоналісти не спотворюють структуру роботи ментального, адже воно є машинною програмою, а не самодостатнім чимось. Те, чим ця програма є, показують виконувані нею завдання, бо розглядати її саму по собі рівнозначно зазиранню у «темну скриньку» (black box) нашого розуму. Отож, питання «як воно працює» виявляється важливішим за питання «чим воно насправді є».

3.4. Аргументи проти функціоналізму. «Китайська кімната» і «китайська нація».

Американський філософ **Джон Сьорл** (Searle) у статті 1980 року «Розуми, мізки і програми» спробував спростувати аргументи Патнема за допомогою власного аргументу, що отримав назву *«китайська кімната»* (АКК). Сьорл пропонує альтернативний ментальний експеримент. Припустімо, існує кімната, в якій перебуває особа (за однією версією, це автор статті, за іншою – просто будь-який виконавець). У цієї кімнати є два отвори: в один потрапляють написані китайськими ієрогліфами тексти на табличках, з іншого отвору виконавець передає назовні їх англомовний переклад. Завдання журі полягає в тому, щоб підтвердити задовільність перекладу з китайської на

англійську мову. Важливий нюанс полягає в тому, що особа у самій кімнаті формально не володіє китайською мовою, а здійснює переклад, послуговуючись лише буквальними покроковими підказками з інструкції за типом «цей ієрогліф у мові К відповідає цій словосполучі у мові А». Питання, яке ставить перед читачем Сьорл, звучить так: чи можна успішно виконати подібний переклад, узагалі не знаючи оригінальної мови?

Аргумент Сьорла ґрунтується на припущенні, що знання мови відповідає вмінню користуватися нею не тільки на рівні синтаксису (у формі послідовної підстановки знаків К для мови А), а передусім на рівні семантики (тобто мовець мусить *розуміти*, що він перекладає). Без сполучення синтаксичних навичок із семантичними переклад не матиме успіху. Отже, Сьорл за допомогою цього експерименту доводить, що машині (або людині, яка діє на взірець машини) недостатньо оволодіти функціональними алгоритмами. З іншого боку, філософ переконаний, що здатність розуміти суть процесу для машинної програми теж недоступна. Отже, функціоналістська парадигма є початково хибною. Формальний характер комп'ютерної програми не повідомляє їй умов змістовності, тому, скажімо, завчання на пам'ять китайської абетки не засвідчує опанування китайської мови і не робить виконавця компетентним перекладачем.

В іншій своїй праці «Перевідкриття ума» Сьорл висміяв схематизм теорії функціоналіста Патнема. Пожартувавши про те, що вивчення «чорних скриньок» (алюзія на центр машинної системи) навряд чи задовольнить нас відповіддю про природу свідомості, так само, як «чорні дірки» не задовольняють астрономів, мислитель вигдав іронічний експеримент, скориставшись поясненням самого Патнема. Уявімо, міркує Сьорл, що вхідні дані, центр їх обробки і вихідні дані – це ланцюжок, який можна заповнити будь-якими об'єктами. Справді-бо, не важливо, як називати те, що просто «добре виконує свою функцію» у системі. Нехай ми маємо кілька порожніх пивних бляшанок, сполучених між собою і підвішених на дроті. Якщо вдарити по першій з них, коливання від неї перейде до наступної – таким чином увесь

ланцюжок увійде в резонанс і, за початковим припущенням, зреагує на фізичну реальність. Утім, цей ефект не виробить у вигляді інформації на виході нічого, крім шуму, хоча принципу функціональної взаємодії буде цілком дотримано. Це висміювання Сьорлом опонентів отримало назву *«аргументу пивної бляшанки»*. Як бачимо, його критику зосереджено на формальному боці функціоналістського пояснення роботи системи.

Деяко раніше (1978) схожий, і не менш іронічний, аргумент увів у філософський обіг Сьорлів колега **Нед Блок** (Block), поширивши його не на індивідуальний випадок застосування програми, а на дії великого колективу. Цей уявний експеримент отримав назву *«китайська нація»*, що за формою є оберненою версією «мізків у діжці». Блок пропонує уявити ситуацію, коли владу в Китаї (історично час написання статті збігся з періодом після смерті Мао Цзедуна) захопили вчені-функціоналісти. Їхня мета полягає у демонструванні цілому світові інтелектуальних потуг китайського народу. До всіх китайців країни приєднали датчики, які ведуть до надпотужного комп'ютера, кабелі якого сполучені з живим тілом, з якого вилучено мозок. Коли мільярд китайців одночасно виробляє думку, датчики показують на моніторі її візуальне зображення. Діяльність мозку піддослідних симулюється ззовні. Чи можна в такому разі вважати тіла китайців свідомими? Метафора нації як уособлення неперервного комунікативного обміну сигналами, що нагадують взаємодію нейронів у мозку, є доволі суперечливою.

Утім, Блок хотів сказати, що одного лише каузального впливу на вироблення певних функцій (створення образів) не буває достатньо, адже інші умови роботи свідомості не задіяні. Зображене на моніторі буде чим завгодно, крім подій свідомого досвіду. Де-факто експеримент Блока спрямовує Патнемів аргумент «мізки в діжці» проти нього самого. Тож не дивно, що Блок заявляє: аргумент проти функціоналізму можливий як аргумент з відсутніх квалій (див. 4.2). Водночас підкреслимо, що сам Блок визнавав незастосовність свого доказу проти специфічної версії функціоналізму, відомої як *психофункціоналізм* Джері Фодора. На відміну від машинного

функціоналізму, фодорівська доктрина переконує, що ототожнення ментальних станів з функціональними станами має передбачати коректний його опис мовою цілої низки наукових підходів, себто однаково залучати дані психології, лінгвістики, нейрофізіології, філософії. Тож якщо мати свідомі стани означає реалізувати психофункціональні стани, тоді система, подібна до «китайської нації», буде лише стимулювати, а не дублювати людські ментальні стани. А втім, навіть це заперечення не похитує тезу Блока, адже психофункціоналізм, як і функціоналізм загалом, не визнає *кваліативні* властивості і не може розрізняти «червоне», «бути схожим на червоне» і «почервоніння», себто апелює до аргументу з відсутності квалій.

Важливо підкреслити, що обидва докази проти обґрунтованості функціоналізму виходять з однієї настанови: самосвідомість є критерій наявності свідомості, а формальне приписування свідомості нічого не може ні підтвердити, ні спростувати її існування для систем, які не розуміють самих себе. Сьорл показав, що заперечення квалій, і зокрема усвідомлення себе в певному стані, є Ахіллесовою п'ятою функціоналізму (і бігевіоризму загалом). У випадку, коли це відбувається, машинний розум, або людський розум, витлумачений за аналогією з машинним, а) не усвідомлює і б) не розуміє, що він робить. Натомість люди, маючи суб'єктивність, отже й розуміння своїх дій, розуміють інших, оскільки ця здатність дає їм змогу розрізняти їх та себе. Відкидання суб'єктивності як «теоретичного надлишку» тягне за собою втрату людиною суб'єктності, себто самостійного статусу дієвця (actor). На прикладі болю це особливо примітно. Адже щоб відчувати біль, не достатньо перебувати у стані X, потрібно ще й усвідомлювати щось у цьому стані, інакше-бо така поведінка буде «граю у біль» з самим собою.

3. 5. Функціоналізм і фізикалізм: порівняння підходів.

Ми з'ясували, що програма функціоналістського розв'язання проблеми ума та свідомості має як переваги, так і недоліки. Серед переваг слід визнати *моністичний* характер пояснення того, як наочні прояви ментальної поведінки

є для нас доступними. Крім того, на відміну від фізикалізму (теорії тотожності), прибічники функціоналізму не обмежувалися зверненням лише до природничих наук, а й залучали широкий масив новітніх даних з когнітивістики та комп'ютеративних наук. Не меншою перевагою цієї течії була її компромісна модель опису ментальних процесів: замість говорити лише про матеріально-біологічну основу психіки, функціоналісти наголошували на тому, що психофізична дилема вимагає відмови від жорсткої опозиції «матеріалізм – іматеріалізм», отже, що саму проблему варто розглянути з альтернативного погляду. Хоча, по суті, функціоналізм тяжіє до матеріалістичної філософії, він узагалі не відкидав можливість аналізу матеріального в термінах нефізичних словників.

Утім, функціоналізму не вдалося вповні завершити побудову вичерпного пояснення проблеми свідомості. Чому цього не сталося? По-перше, будучи критикою фізикалізму, у певному сенсі функціоналістська школа залежить від логіки, обстоюваної ТТ. Приміром, природу ментального функціоналіст тлумачить як результат зовнішнього каузального втручання (inputs/outputs). Модель інформаційної інтерпретації психічних станів у вигляді «груп даних» не повинна вводити в оману: ідеться саме про каузальність щодо психіки, а це, авжеж, є доказом на користь впливу фізикалізму. Разом із тим, на відміну від ТТ, функціоналісти наголошують, що каузальний результат не є тотожним каузальній причині, позаяк у процесі переходу від «вхідних даних» до «вихідних» зазнає змін і сам цей вплив.

По-друге, функціоналізм, як і фізикалізм, ігнорує проблему квалій, внаслідок чого людська свідомість стає невідрізнуваною від машинних систем, хоча очевидно, що вони нетотожні. Якби було інакше, питання про *можливість* появи сильного штучного інтелекту не постало б. Та це й не дивно: місце для квалій у машинній парадигмі знайти важко. По-третє, формалізуючи всі процеси, що відбуваються у царині психічного, філософи-функціоналісти вимушені апелювати до специфічних критеріїв і метафор, заснованих на видимій поведінці. Тим самим найкращим помічником

функціоналіста стає та чи та версія бігевіористичної психології. Адже як дізнатися, що відбувається у «чорній скриньці» свідомості, як не шляхом розпізнання публічно видимих (поведінкових) актів її носіїв?! Утім, як було показано раніше, зі стеження за поведінкою можна зробити абсолютні протилежні висновки.

Нарешті, машинна термінологія функціоналізму втягує його у теоретичну пастку кривотлумачення: щоб пояснити ментальне, необхідно вживати словник, нейтральний щодо ментальних змістів, аби потім переписати добуті висновки звичною мовою менталістського дискурсу. Наприклад, щоб пояснити, чим є біль, фізикаліст спершу звертається по допомогу до нейрофізіології, тоді як функціоналіст ладен працювати з вокабуляром теорії систем, аби далі екстраполювати свої висновки на сферу ментального у термінах природної мови.

Попри сказане, функціоналістам явно імпонують окремі елементи теорії тотожності. Так принцип множинної реалізації добре улягає фізикалізму типу, оскільки якщо мозок вважати системою з подвійною структурою, тоді подвійність є не більш ніж інструментально зручний спосіб опису активності системи на різних рівнях. І функціоналіст, і фізикаліст наголошуватимуть на тому, що «свідомий досвід» годі зводити до безпосередніх відчуттів – він являє собою радше складне поєднання виснувань (*inferences*). Між перебуванням у збудженому/пригніченому стані активності і «болем» пролягає цілий спектр опосередкувань у вигляді «чорної скриньки», що оброблює вхідні дані, щоб виробити вихідні дані. В обох випадках логічна симетрія між станами виглядає привабливою з теоретичного погляду. Не зайве підкреслити, що такі філософи, як Ф. Джексон, Д. Льюїс і Е. Прайор, вбачали у функціоналізмі «обхідний шлях» у спробі оновити й посилити позиції фізикалізму.

У підсумку, як фізикаліст, так і функціоналіст говорять про один клас явищ, але це не усуває питання: а де ж тут, власне, ментальні процеси? Функціоналізм як філософський напрямок досі має чимало прибічників, головню через те, що альтернативи йому виявилися не набагато

продуктивними у розв'язанні психофізичної проблеми, а також через потужний розвиток комп'ютерних досліджень і прагнення наблизитись до психофізичної проблеми під іншим ракурсом – щоб розкрити суть ментального можна поглянути на те, як це можна симулювати ментальне у технологічних носіїв інтелекту. Але про це ми поговоримо докладніше у 8.1.

Джерела:

- Armstrong, D.M. (1968). *A Materialist Theory of the Mind*. Routledge & Kegan Paul, pp. 355–366.
- Block, N. (1978). Troubles with Functionalism, in C. W. Savage (ed.), *Perception and Cognition: Issues in the Foundations of Psychology*, Minneapolis: University of Minnesota Press, pp. 267–325.
- Lewis D. (2004). An Argument for the Identity Theory, In: Heil J. *Philosophy of Mind. A Guide and Anthology*. Oxford University Press, 2004, pp. 150–157.
- Патнем Г. (2003). *Розум, істина та історія*. Альтернативи, сс. 13–32.
- Серль Дж. (2014). Розуми, мозки і програми. *Антологія сучасної аналітичної філософії, або Жук залишає коробку*. – Л.: Літопис, сс. 229–255.

Запитання для самоперевірки:

1. Уявіть противника тези про те, що ум є мозковою діяльністю. Його аргумент міг бути таким: «Ум не є тотожним мозку, адже ви завжди обізнані зі своїми станами, натомість про те, що відбувається в мозку ви зазвичай не знаєте». Чи є такий аргумент переконливим?
2. Автор Теорії тотожності Дж. Смарт вважає переконливішим аргументом «лезо Окама» (принцип економії мислення: не потрібно примножувати сутності без нагальної потреби). Як на вашу думку, чи дійсно «лезо Окама» посилює аргументацію цієї теорії?
3. У чому полягає сила модального аргументу Кріпке? Чи спрацьовує цей доказ проти матеріалізму загалом?

4. Які підходи до ментального й тілесного можна виділити в межах функціоналізму?
5. Якщо вважати функціоналізм «слабкою» версією матеріалізму, чи міг би Декарт схвально сприйняти якийсь з аргументів функціоналістів?
6. Чи можна казати про те, що функціоналізм пропонує адекватне розв'язання психофізичної проблеми?
7. Чому виглядає аргумент «мізки в діжці» невдалим з точки зору фізикаліста?
8. Чи можна казати, що головна ідея кінофільму «Матриця» відображає аргумент Патнема? Якщо це так, що прихильник перебування «у» «матриці» міг би протиставити цьому аргументу?

Лекція 4.

Проблема квалій. «Кваліафіли» та «кваліафоби».

4.1. Проблема природи квалій.

Квалії (англ. мн. *qualia*, одн. *quale*) – поняття, що позначає сприймані особою чуттєві дані (запахи, кольори, смак, дотикові відчуття тощо), які специфікують його власний досвід взаємодії з фізичною реальністю. Ці відчуття повторюються у кожному наступному акті сприйняття, але щоразу по-новому. Тому, як зазначає у своїй праці «Ум і світовий лад» (*Mind and the World Order*, 1929) автор цього неологізму, американський прагматист Кларенс Льюїс (C.I. Lewis), можна твердити, що квалії не є «знаннями», але вони супроводжують знання. Вони є «розпізнаваними якісними характеристиками чуттєвого даного, що повторюються у різних досвідах» і саме в цьому сенсі Льюїс вважав їх універсаліями.

Корелятами квалій є об'єкти сприйняття, хоча самі сприйняття *не тотожні* своїм об'єктам (запах однієї й тієї ж квітки я сприймаю не так, як людина поруч зі мною, і не так, як я сам сприймав його вчора). Особливістю квалій є властивість бути *безпомільними* (не маючи змоги пояснити дантисту,

«як воно» відчувати зубний біль, я чітко відрізняю його від решти своїх відчуттів. Пор. відмінність між фактичними й фантомними болями). Ще одним прикладом квалій є наступне: у напівтемряві я бачу настінний годинник як такий, що має форму квадрату, а після ввімкнення світла – круглу форму. Якби у напівтемряві мене запитали про форму годинника, я з упевненістю відповів би, що вона є квадратною. Отже, природа квалій є предметом дискусії для сучасних філософів, які шляхом аналізу висловлювань про них намагаються наблизитися до розуміння того, як формуються змісти нашої свідомості.

4.2. Основні аргументи кваліафілів і кваліафобів.

Серйозну небезпеку для матеріалістичного заперечення квалій становив запропонований 1974 року доказ, авторство якого належить філософові **Томасу Негелу** (Nagel; іноді українською його прізвище транслітерують неправильно: «Нагель»). У статті «Як воно – бути кажаном?» він порушує питання: якщо вчені зможуть описати всі нейрофізичні властивості тіла кажана і навіть спробують змодельовати його поведінку, засновану на отриманні даних від ехолокаторів, що дасть нам змогу зрозуміти, як ця тварина переживає свої відчуття? Очевидно, що сукупна інформація спонукає хіба що наближено аналогізувати поведінку кажана з поведінкою інших ссавців або з людськими діями. Утім, відчути себе кажаном, щоб пояснити дії кажана, ми не здатні. Тож і наші власні суб'єктивні досвіди залишатимуться незбагненними для опису в нейтральних термінах науки. Негел має на увазі, що ТТ, попри вірність принципам наукової істини, не спроможна дати пояснення такому феномену, як ментальне на будь-якому з його рівнів.

Психічне життя є не сумою окремих процесів, які вчений узагальнює, а специфічним і недоступним для опису явищем, з чим і слід змиритися. Сутність ментального полягає не в його біологічній структурі або практичних функціях, а у кваліях – якісних фрагментах неповторності переживань і відчуттів, притаманних кожному суб'єктові досвіду, тваринному або людському. Те, як дитина відчуває радість або спортсмен – розпач після

поразки на змаганнях, годі схематизувати через відомі поняття; їхні переживання можна пояснити, але не відтворити у власному переживанні. Тим самим ближчим підступом до проблеми тілесного й ментального є аналіз квалій, а не дослідження роботи лише фізичного субстрату ментальних досвідів. Хоча мозок у всіх людей однаковий, кожен із нас має різні квалії, незвідні до яких-небудь окремих біологічних характеристик.

А проте, існує ключ до тлумачення кваліативних станів. Ним Негел вважає **інтенційність** – спрямованість ума (mind) на певний об'єкт, який корелює з тими чи тими переживаннями. Іntenційним об'єктом не обов'язково є предмет зовнішнього світу (інтенційним об'єктом спогаду є обличчя людини, що встигла зістаритись). Щоправда, можливими є й неінтенційні стани, наприклад, біль або спалах гніву. Якщо я боюсь когось конкретно (хижака у лісі чи сусіда напідпитку), тоді мій ментальний стан є інтенційним, адже він пред'являє конкретний об'єкт, але коли я просто холону від жаху, не знаючи чому, такий стан є безоб'єктним, а отже, неінтенційним. Ідея Негела полягає в тому, що усвідомлювати щось означає перебувати у певному психічному стані, про який ми не в змозі висловитися вичерпно. Отож трудність, яку демонструють квалії, виявляє неусувну перешкоду для матеріалізму: ототожнення фізичного й ментального відбудеться лише за умови ігнорування квалій; однак ігнорування квалій позбавить ментальне його основоположних властивостей; отже, фізикалізм є хибним і нічого не пояснює.

Ще одним доказом проти матеріалізму став **аргумент знання** (the knowledge argument; далі – АЗ), висунутий **Френком Джексоном**. У статті «Епіфеноменальні квалії» (1982) австралійський мислитель прагне спростувати редукцію ментального до фізичного і тим самим елімінацію квалій як підґрунтя суб'єктивних переживань. Пояснюючи свою думку, Джексон пропонує мисленневий експеримент. Уявімо, що в чорно-білій кімнаті постійно перебуває вчена Мері (образ фізикаліста), яка «знає все» про кольори, хоча вивчає їх виключно за допомогою зображень на чорно-білому моніторі. Припустімо, одного ранку Мері вийде за межі кімнати у справжній

кольоровий світ і побачить, скажімо, червоний колір, з яким вона доти не стикалася. Чи дізнається вона щось нове для себе? Навіть якщо Мері концептуалізує свої відчуття червоного, це засвідчить, що її «повна» інформація про світ була неповною. Тож хоча квалії переживаються виключно суб'єктивно, вони розширюють у сприймачів просту фізичну картину реальності. Тому фізикалізм є хибним, бо ніколи не дає повноти картини світу.

Проти аргументу знання його опоненти підготували низку заперечень. Приміром, американський мислитель **Деніел Денет** (Dennett; про його концепцію див. 7.5) навіть переписав уявний експеримент Джексона, аби спростувати цей доказ. Денет знаний відмовою визнавати ідею свідомості та спростуванням наявності психофізичної проблеми. Хід його міркування розгортається так: якщо Мері є вченою, тоді вона знає «все» про кольори і нашу систему зору, а отже, нездатна набувати новий досвід і навички. Якби Мері позбулася ізоляції у монохромному приміщенні й вийшла назовні, вона все одно не набула б жодних кваліативних змістів: нові сприйняття (нове знання загалом) лише уточнює окремі нюанси попередньо засвоєної інформації. Якщо, наприклад, дорослу людину з хатців Амазонії перевезли до мегаполісу, то після періоду соціальної акліматизації вона продовжуватиме сприймати гігантське місто як «виняток» з того світу, в якому зростала.

На посилення своєї тези Денет розвиває експеримент про *РобоМері*. Якби на місці Мері у чорно-білій кімнаті перебувала РобоМері, вона могла би перемикатися між монохромним і кольоровим зображеннями, і тільки в такому випадку зрозуміла б умови функціонування зору. Денет виходить з презумпції, що обробка візуальних досвідів людей і їхнє відображення у кібернетичній матриці нічим структурно не відрізняються. Тож власне сприймання і запис програмою звіту про сприймання принципово нерозрізнявані. Ясна річ, такий хід розкриває засадничу настанову Денета, нічого не повідомляючи про специфіку самих квалій. До того ж, він недобачає те, що роботизовані механізми можуть зазнавати змін у процесі обробки інформації. І справді, коли говорять про квалії, вказують не на рефлексію про

них, а на безпосередність набутих вражень, а цю умову в експерименті Денета проігноровано. Його заперечення проти АЗ видається, отже, не настільки руйнівним для аргументу, як про це заявляє філософ.

Окреме заперечення проти АЗ знаходимо й у філософа-елімінітивіста **Пола Чорчленда** (Churchland). У статті-відповіді Джексону під назвою «Пізнаючи квалії» (1992) він намагається переконати, що перебування у чорно-білій кімнаті не сформувало б у вченої певного знання про світ, а навпаки, її досвід ізоляваності призвів би до протилежного – деградації її пізнавальних здібностей. Нові сприйняття (той таки червоний колір), отримані назовні, не були б належно засвоєні, тож поза кімнатою вчена не дізналася б нічого нового, що суперечило б закріпленій у неї картині світу. Ба більше, знання зародило б у неї сумнів у тому, що вона звикла вважати обґрунтованими висновками про реальність. Однак звідки взятися сумніву, якщо світ Мері однозначно монохромний, а система її понять є «вичерпною»? Квалії, каже Чорчленд, не мають жодного стосунку до формування знань.

До аргументу знання Джексон додає свою версію *модального аргументу*. Цей доказ передував відомому на сьогодні аргументу філософського зомбі (докладніше див. 5.5.) і за формальною структурою продукує аналогічний умовивід. Суть модального аргументу Джексона полягає в тому, що весь обсяг фізично доступної інформації дає нам уявлення про певну істоту, однак не здатен відповісти на питання, чи є ця істота свідомою. Квалії в такому разі виявляються ключем, що відкриває доступ до проблеми ментального. Крім того, логічна модальність вказує на діапазон можливостей, які аж ніяк не ідентичні фактичному стану справ. Ідеться лише про ймовірнісні показники, але й цього буває достатньо для підваження сильної версії матеріалізму. На думку Джексона, обидва докази – і аргумент знання, і модальний аргумент – можна сполучити у вигляді довго аргументу проти кваліафобів. Додамо, що Негелів аргумент кажана, дещо доопрацьований, теж улягає вказаній стратегії аргументації на користь свідомого досвіду суб'єктивності. Адже, попри те, що

ми ніколи не дізнаємося, «як воно» бути кажаном (або Ернестом Гемінгвеєм), у нас залишається шанс постійно дізнаватися щось нове про себе.

Які висновки випливають з аргументу знання і аргументу кажана? По-перше ментальні події не можуть спричинювати фізичні події, проте одні ментальні події спроможні породжувати інші. Приміром, усвідомлення мною спраги може змусити мене пити воду і змусити не пити її допоки для цього не настане слухна нагода. Керування своєю волею доводить безпідставність лише фізичної причини тамування спраги.

По-друге, якщо припустити, що квалії ніяк не впливають на наш фізичний світ, тоді стає незрозумілим, навіщо розуму потрібен такий надлишок, який ще слід обробити, перетворивши на спогад, судження-проекцію абощо. Бо ж наявність або відсутність квалій не позначається на процесі нашого біологічного виживання. Отже, якщо квалії і загалом елементи психіки не сприяють появі більш докладного знання про природу, ймовірно, вони виникають як побічний наслідок продуктивної діяльності розуму зі створення образів (а це важливо для функціонування мозку). Коли довіритися епіфеноменалістам (див. 6.3.), то моє бажання читати книжку не є ані моїм рішенням, ані усвідомленим бажанням, а просто реєстрацією отриманого мозком сигналу, кінцеву мету якого важко оприявнити.

По-третє, одні лише фізичні характеристики речей і явищ не дають нам належної інформації ні про них, ні про нас. У такому разі все, що ми знатимемо про склянку води, можна категоризувати до поняття «будь-яка склянка води» або «наполовину заповнена склянка води». Поміж тим, сам по собі опис фізичних і хімічних особливостей процесу пиття не дасть збагнути, чим є «пиття води зі склянки» (або «болючість» болю чи «задоволення від аромату квітів»). І якщо фізичне не відкриває змоги для розуміння нефізичного, матеріалізм слід визнати помилковою теорією.

У полеміці з антиредукціоністами **Сидні Шумейкер** (Shoemaker) запропонував *аргумент з відсутності квалій* (the absent qualia argument, далі – АВК). У ньому він так само акцентує увагу на критерії осмисленості:

Квалії можна мислити.

Якщо можна мислити відсутність квалій, тоді їх відсутність є можливою.

Отже, відсутність квалій є можливою.

Якщо відсутність квалій є можливою, тоді функціоналізм є хибним.

Отже, функціоналізм є хибним.

Інакше кажучи, якщо ми здатні помислити відсутність у нас квалій, тоді безкваліативний стан можна визнати нашим нормальним станом. Але якщо відсутність квалій можлива, це вказує на хибність функціоналізму. Однак довести мисленність певної гіпотези не означає спростувати (чи підтвердити) цю гіпотезу. У цьому полягає обмеженість будь-якого модального аргументу. АВК демонструє, що кожному доказові «кваліафіла» можна протиставити контрдоказ «кваліафоба». Якщо кваліафіл переконує, що інтелектуальна діяльність неможлива без квалій або осмислення свого досвіду, то кваліафоб доводитиме, що значеннями можна оперувати й неосмислено, не здогадуючись, чим вони є. Вказана дихотомія кваліафобів і кваліафілів має на меті поділити всю систему філософських аргументів стосовно квалій на дві незвідні позиції: однозначно «за» або однозначно «проти». У питанні про природу квалій, якщо йдеться не про полеміку між вченими, філософи обстоюють одну з цих позицій, навіть тоді, коли визнають саму проблему псевдопроблемою.

В іншій статті «Функціоналізм і квалії» (1975) Шумейкер вдався до стратегії нейтралізації аргументів кваліафілів. Він припускав, що попри можливість існування квалій, їхнє об'єктивне пізнання є невідним. Справді-бо, як верифікувати той факт, що пацієнт відчуває біль у кінцівці, яку йому давно ампутували?! Критерій очевидної каузальної дії, що виконує головну роль в ідентифікації ментальних станів, у випадку кваліативних подій не спрацьовує. Отже, твердити, що квалії, як нередуктивні властивості, якимось чином впливають на нас, немає підстав, тим паче залучати їх до філософського обговорення. Висновок Шумейкера відтворює скептичну настанову античних

софістів: квалії або є, або їх немає; а якщо вони є, вони є непізнаними; тож яка різниця, існують вони чи ні – ця проблематика не наближає нас до розуміння ментального.

Відомий сучасний філософ Девід Чалмерс (Chalmers; див. 5.4.) піддав АВК послідовній критиці. Його точка зору заснована на тезі про темпоральну відмінність у процесі відчуження. Чалмерс припускає, квалії не є просто психічними властивостями – вони є комплексами невідомих *протоквалій*, феноменальних і водночас нефеноменальних, які, згідно з гіпотетично висуваними фундаментальними законами взаємодії фізичного та психічного, лежать в основі більш зрозумілих феноменальних властивостей. На переконання Чалмерса, існують два типи квалій: ті, які ми маємо безпосередньо в моменті перцепції, і ті, які починають згасати після акту перцепції. Ми ще якийсь час утримуємо їх у своїй пам'яті, але ці спогади є нічим іншим, як лише проблисками колишніх відчужень (він називає їх «мерехтливими кваліями»). По суті, ця теза передбачає, що поступове зникнення квалій (або, можемо сказати, квалія-афектів) дуже подібне до визнання їх відсутності: те, що може зникнути, насправді, не є чимось стійким і очевидним. В обох випадках квалії не слід вважати «власністю» сприймача, а хіба що елементом системи сприйняття загалом, якою вона «чіпляється» за предметний світ. Тим не менш, запропонувавши уявний експеримент із заміною людської нейрофізіологічної системи на кремнієві чіпи (в яких фіксуватиметься кваліативне відчуження і алгоритм яких не відрізняється від передачі сигналів у нейронному середовищі), Чалмерс висновує, що остаточно позбутися квалій неможливо. Отже, аргумент з відсутності квалій є помилковим, хай навіть його результат можна помислити. Якою мірою Чалмерсова пропозиція є переконливою – це ключове питання досі живої дискусії.

Згадаймо іще про один відомий доказ проти заперечення квалій з боку Гіларі Патнема – *аргумент «суперспартанці»*. Зауважмо, що в ньому прямо не мовиться про квалії – хід думки в даному разі є кружним. Доказ проти

неможливості квалій представлено як оспорення залежності між відчуттям і його публічним вираженням. Предметом аналізу Патнема є безумовність вираження відчуття болю. Біль не є однорідним відчуттям – у ньому безліч чуттєвих і смислових відтінків; це так зване *кластерне* поняття. Коли я кажу про своє відчуття болю, словом «біль» я намагаюся зафіксувати стан, якому завжди бракує повної сукупності індикаторів. У розмові про певний предмет ми можемо вдатися до прямого (остенсивного) вказування на нього, натомість у розмові про біль цей шлях є для нас закритим. Звісно, можна вказати пальцем на зуб, який болить, однак ця вказівка локалізує джерело болю, а не сам біль. Отож такі кластерні поняття, як «біль», охоплюють симптоми різних фізичних відчуттів. До того ж, клас «больових» понять має відношення не до каузальної основи відчування, а до спектру самозвітів того, кому болить.

Ілюстрацією думки Патнема є логічний експеримент, відомий під назвою «суперспартанці» або «суперстойки». Уявімо, що існує спільнота осіб – суперспартанці, яких з дитинства привчили управляти вербальним вираженням своїх переживань і водночас придушувати будь-які зовнішні прояви больових відчуттів. Навіть зазнаючи ударів і відчуваючи біль, ці люди мають роками натреновану волю, яка дає їм сили витерпіти гострий біль. Незалежно від того, якими мотивами скеровані суперспартанці, їхня поведінка доводить на практиці, що зв'язок між ментальним змістом і його публічною фіксацією *не є* безумовним. Експеримент можна скорегувати, якщо висловити припущення, що внаслідок тривалої еволюції та штучного добору ця спільнота навчилася народжувати дітей (суперсуперспартанців), які сприймають себе як «дорослих невеликого зросту» і поводяться аналогічно їм. Такі особи позбавлені спроможності виражати свій біль, бо не завчили ні вигуків типу «йой» чи «ай» і лексем або ідіом, споріднених з дієсловом «боліти». Якби представник цієї спільноти опинився у нашому світі і розповів про свої реакції на те, що ми звично називаємо болем, ми постали би перед дилемою: або вважати його самозвіти спотвореними, або взагалі нерелеватними щодо больових відчуттів. Не чуючи від суперсуперспартанців розповідей про

больовий досвід, ми змушені були б повірити, що вони й справді його не мають. З експерименту Патнема можна зробити висновок, що причина та опис ментальної події *не* залежать від способів реагування на неї, тим паче від того, як її відображено мовою. Аргумент суперспартанців є виразно антиредукціоністським. Однак парадокс полягає в тому, що спростування редукції ментального до фізичного лише посилює тезу про неповідомлюваність відчуттів.

На завершення слід зазначити, що проблема квалій є недостатньо вивченою з наукового боку, а також є складною з погляду її семантичної неоднозначності. Попри те, що ми запропонували компромісне визначення квалій, натепер не існує якої-небудь згоди щодо його несуперечливого визначення як у середовищі кваліафілів, так і поміж кваліафобами. Боротьба аргументів триває. І це означає лише те, що формулювання проблеми квалій і спроб її розв'язання, окреслили межу, яку сучасна філософія виявилася неспроможною подолати. Утім, без обговорення природи квалій неможливо сподіватися як на прояснення позицій щодо психофізичної проблеми між представниками основних течій СФС, так і на відокремлення психічного від фізичного в аналізі різних станів свідомості.

Джерела:

Churchland, P.M. (1992). Knowing Qualia: A Reply to Jackson. In Churchland, P.M. *A Neurocomputational Perspective. The Nature of Mind and the Structure of Science*. Cambridge: MIT Press, pp. 67 – 76.

Dennett, D. (2006). What RoboMary Knows. In Torin Andrew Alter & Sven Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press, pp. 14–31.

Shoemaker, S. (1975). Functionalism and Qualia. *Philosophical Studies* 27: 291–315.

Деннет Д. (2017). Що знає РобоМері. *Антологія сучасної філософії науки, або Усмішка ASIMO*. Л.: ЛНУ ім. І. Франка, С. 414–446.

Джексон Ф. (2014). Епіфеноменальні квалія. *Антологія сучасної аналітичної філософії, або Жук залишає коробку*. Л.: Літопис. С. 260–272.

Запитання для самоперевірки:

1. Чим є *квалії*? Наведіть зрозумілі приклади з повсякдення.
2. Поміркуйте про слова Вітгенштайна: «Якби лев міг розмовляти з нами, ми не зрозуміли б його». Як це прояснює проблему квалій?
3. Негел стверджує, що нам годі збагнути, як воно – бути кажаном. Що аргумент Негела прояснює в питанні про свідомість? Як він стосується відношення між свідомістю та самосвідомістю?
4. Як Ви розумієте аргумент з відсутності квалій? Якою є мета використання цього доказу?
5. Чи виправданою є думка про те, що розв'язання проблеми квалій є кроком до розв'язання психофізичної проблеми? Чому?
6. Якщо думки існують як ментальні змісти, чи можна собі уявити думки «без змісту»? Якими вони будуть? Спробуйте навести приклади.
7. Уявімо, що існування квалій доведено не лише для людини, а й для всіх вищих живих істот. Чи дасть це підстави стверджувати, що свідомість є універсальною властивістю будь-чого живого?

Лекція 5.

Дуалістичний антиредукціонізм (I):

дуалізм властивостей, інтеракціонізм і натуралістичний дуалізм.

5.1. Філософський дуалізм, його принципи.

Аби спрощено викласти сутність психофізичного дуалізму, уявімо собі дуб, на корі якого вирости гриби-трутовики. Вони призвичаюються співіснувати з деревом, живляться його соками і гинуть без нього. Але ці гриби-нарості є самостійними сутностями, нетотожними поняттю «дуб». Схожим чином ментальне та нементальне *якось* співіснують, при чому у їх

взаємозв'язку можна запримітити риси структурної взаємодії, коли тісний симбіоз дає плоди, непритаманні біологічним можливостям кожного з них порізно. Ризиковано твердити, що це самостійні субстанції (хоча і цей вид дуалізму має свою традицію), а проте, звести одне до іншого з метою прояснення вказаної взаємодії так само не вдається без втрат. Отже, дуалізм є відповіддю редукціоністській парадигмі, яка у вигляді жорсткого матеріалізму зазначала невдачі, принаймні на етапі свого становлення.

Джерела дуалістичної парадигми явно простежуються вже у картезіанському *субстанційному дуалізмі*. Нагадаймо, що, згідно з Декартом, протяжна й мисленна субстанції є внутрішньо замкненими, отже самодостатніми, а проте, вони якимось дивним чином взаємодіють. Декарт був переконаний, що будь-яке живе тіло є складним механізмом, принципи роботи якого відображаються у принципах мислення. Перший уявний експеримент («Декартів робот») з «Міркування про метод» (1637) показує, що між живою людиною і автоматом різниця полягає не у алгоритмі, а радше в походженні. Декарт пропонував дві моделі тлумачення способів цієї взаємодії: з одного боку, йому йшлося про втручання божественної сили, з другого боку, він «з'ясував», що місцем причинового впливу є тілесний орган, що поєднує у собі обидві функції – шишкоподібна залоза мозку (епіфіз), яка буцімто продукує в тілі розумність. Позаяк це пояснення наукою не підтвердилося, картезіанці схилилися до першої моделі: існує *непричинова кореляція* душі й тіла, за яку відповідає Бог. У такому разі, вона є різновидом дива, а отже, її слід просто прийняти на віру. Але навіть якби дива були можливі, вони руйнували б закони природної причиновості, що, ясна річ, не надає проблемі взаємодії розв'язання. Картезіанське рішення є типово метафізичним і саме цього прагне уникнути сучасна аналітична філософія.

5.2. Дуалізм властивостей

Указаний тип дуалізму як доктрина спирається на тезу про те, що будь-який опис реальності має враховувати відмінність у властивостях між

сторонами взаємодії. Приміром, коли йдеться про таке складне явище (а воно завжди є комплексним у випадку інтеракції), як робота автомобільного двигуна, варто взяти до уваги, що внутрішнє згоряння палива утворене послідовністю чотирьох окремих процесів. Інший приклад – грозові дощі, в яких молекули води сполучаються згідно з фізичними законами. Ми наразі кажемо про події у матеріальній реальності, проте самі по собі ці процеси не відбуваються ізольовано або спонтанно. Те саме, міркує дуаліст, є релевантним і для психічних явищ. Для їх опису послуговуються символізацією (згадаймо Вітгенштайнове «мислення є свідоме оперування знаками»), питання лише в її характері та інтенсивності. Між двома різними явищами має існувати посередник, яким у даному випадку стає енергія або життєва сила абощо. Ці явища не є субстанційно відмінними, адже належать одному середовищу – природі, але вони різняться своїми властивостями. На відміну від субстанційного, цей дуалізм властивостей оминає утруднення, знайомого Декарту, і зберігає безсумнівну необхідність каузальності. Насправді, дуалізм властивостей не погоджується з думкою редукціоністів, що однієї лише фізичної онтології недостатньо для розуміння психічного, але заперечує й тим, хто вважає, ніби словник фізики є «всього лише» витвором людського розуму. Крім того, змінюється сенс поняття «детермінізм», який набуває менш жорсткого (односпрямованого) звучання, лишаяючи в силі принцип каузальності. Чинність фізичних законів не скасовується, але у процесі взаємного впливу вони здатні породжувати нові явища. Популярній в останні десятиліття синергетиці відоме явище флуктуацій, себто непередбачуваних і нестандартних виходів зі станів кризи системи. Наприклад, людина, що розв'язує математичне рівняння, у стані крайньої втоми спроможна дійти простого й оригінального рішення. Схоже пояснення пропонує дуаліст властивостей.

5.3. Інтераціоналізм

Одним із прикладів дуалізму властивостей є інтераціоналізм. Інтераціоналістський дуалізм є своєрідною інверсією епіфеноменалізму (див. 6.3.), що стверджує, що свідомість (як психіка) є нефізичною, хоча і здатна впливати на фізичне (мозок). Як це відбувається, інтераціоналісти не пояснюють, покликаючись на той факт, що у науці і філософії є багато досі непояснених феноменів. Але попри наявність подібних «прогалин у поясненні», ми змушені запитувати не про граничні підстави – це прерогатива метафізики, а задовольнятися вже відкритими наукою взаємозв'язками між явищами, які можна почасти екстраполювати на інші явища *за аналогією*. Власне, такий вердикт влаштовує інтераціоналіста передовсім через побоювання дурної нескінченності в доведенні. Авжеж, подібний підхід не долає питання про умови чинності психофізичних законів взаємного впливу духу та тіла. Але, як і у випадку з іншими різновидами дуалізму, при обговоренні проблеми психофізичних законів на допомогу приходять принципи логічної можливості.

Припущення про інтераціоналізм може теж виглядати по-різному; його описують як контактне або як безконтактне. Приміром, геомагнітні бурі на поверхні Сонця, як вважають, впливають на електромагнітне поле Землі. З цього факту 1) деякі вчені висновують погіршення здоров'я у людей із серцево-судинними захворюваннями, а 2) деякі псевдонауковці (астрологи) переконують, що у подібні періоди може фатально змінюватися життя окремих людей і навіть цілих суспільств. Відсутність доказів безконтактного впливу між тілесним і нетілесним ще не означає краху дуалістичної аргументації. Відомо, що батько сучасної фізики Айзек Ньютон переймався незадовільністю свого пояснення взаємодії тіл на відстані, хоча це не змусило його загалом відмовитися від теорії гравітації. Уже у XX ст. філософ науки Карл Поппер (Popper), услід за видатним математиком Куртом Геделем (Gödel), знайшов рішення для подібної проблеми: будь-яке обґрунтоване судження ґрунтується на аксіоматичному положенні, далі за яке розум не має виходити,

адже це зруйнує вже злагоджену систему доказів. Так, математик приймає аксіому для доведення теореми, а не намагається довести аксіому. Цим методологічним орієнтиром послуговуються не тільки філософія, а й окремі науки.

Отже, слід «просто» довести можливість безконтактного впливу (у фізичному світі), натомість його *необхідний* характер вже виходить за рамки даного обґрунтування. Особа цілком спроможна керувати автівкою, не порушуючи законів механіки і водночас не знаючи, як їх правильно сформулювати. Але при цьому використовує авто (і сили природи) для досягнення своєї цілі (розбити його, скоївши самогубство). У такий спосіб детермінізм спрямовується на подолання детермінізму. Як одного разу сказав Попер, «піднімаючись зі стільця, ми штовхаємо Землю у протилежному напрямку». Одначе сьогодні можна протиставити детермінізми класичної і некласичної фізики (у квантовій теорії чинною є зовсім інша, багаторівнева причиновість). У царині знання, де жорстка каузальність була неухильною, тепер панує всесилля множини ймовірностей, кожна з яких претендує на свою перевагу, – і це є певним порятунком для доктрини інтеракціонізму. Звична вимога про те, що для каузальності потрібні дві сутності якісно однакового роду, вже не є переконливою. Враховуючи сказане, свідомість *може* впливати на тіло з огляду на нерівновагу в природі об'єктивних дієвих сил. У своїй теорії трьох світів Попер доводив, що між фізичною і ментальною реальністю має бути посередник («третій світ») – сфера принципів, законів, абстрактних сутностей на кшталт атомів, аргументів і універсальних понять, що є незвідними ні до умовиводів окремого індивіда, ні до «образних копій» речей. Як бачимо, ця стратегія інтеракціонізму залежить від реінкарнації платонівської доказової бази.

Що можна закинути інтеракціонізму як підвиду психофізичного дуалізму? *По-перше*, взаємодія між носіями цілковито відмінних властивостей виглядає доволі гіпотетичною і засновується лише на припущенні про *логічну*, а не фактичну можливість цього. Протилежності не можуть утворити новий

життєздатний симбіоз. Відповіддю на цю заувагу може бути існування ГМО, в яких генна модифікація переналаштовує організм на існування в нетипових для нього умовах, не змінюючи його базових ознак. *По-друге*, не до кінця ясно, на якому етапі і чи закономірно виникає суб'єктивність як імовірний результат каузалізації психічного фізичним і навпаки. Якщо припустити факт автономії суб'єктивності, тоді дуалізм має поступитися або емерджентизму, або епіфеноменалізму. *По-третьє*, дуалізм не пояснює «привілейованість самодоступу» до ментальних станів. Адже відомі кейси, коли свідомість функціонувала у стані неактивного мозку (хоча інтерпретації лікарями випадків клінічної смерті загалом неоднозначні). *По-четверте*, незрозумілими є підстави самототожності суб'єкта, сформованого двома класами властивостей. Більш того, навіть якщо якось довести, що самототожність має місце, це штовхає до наступного питання – про умови індивідуалізації свідомості (чому дві однакові сили по-різному формують ментальні змісти певної особи). *По-п'яте*, якщо не брати до уваги Поперову теорію трьох світів, важко дати позитивну відповідь на питання про природу абстрактних сутностей: або це розум «вигадує» закони, які нав'язує природі, або ці закони існують «до» появи людського розуму і спричинюються в ньому природою. Кожен із варіантів є проблематичним, адже втягує нас у метафізичну суперечку про ідеалізм і матеріалізм, об'єктивне й суб'єктивне начала. Ці закиди можна множити, але головним є ось що: ментальні явища (як і фізичні) дуаліст розглядає як елементи складного діяльного комплексу, натомість проблема полягає саме в обґрунтуванні умов можливості *єдності*, що зветься «особистість».

5.4. Натуралістичний дуалізм Чалмерса, основні ідеї.

Новаторством Чалмерса є розгляд проблеми свідомості як багаторівневої. З погляду філософа, відповідаючи на питання про природу свідомості, слід брати до уваги, що сама ця проблема може мати відносно легкорозв'язний характер або ж, навпаки, містити такі труднощі, які вимагають розгляду ряду

суміжних із нею проблем. Перший тип запитування про свідомість Чалмерс вважає «легкою проблемою свідомості», другу – «важкою проблемою». До легкої проблеми (далі – ЛПС) належать запитання, що постають у стосунку до диференціації нами своїх ментальних станів (як розрізняють сон і бадьорість; як ми повідомляємо про те, що з нами відбувається; на чому ґрунтується наша певність у переживанні саме цього, а не іншого відчуття тощо), натомість важку проблему свідомості (далі – ВПС) унаочнює запитання про умови принципової можливості свідомих досвідів.

Позаяк між словником фізики і менталістськими поняттями існує прогалина в поясненні (*explanatory gap*) – логічно й змістово неможливо однозначно редукувати друге до першого, тож немає підстав вичерпно пояснити свідомості у термінах самого лише природознавства. Саме тому ВПС і є важкою проблемою. Якщо терміни двох різних словників годі ототожнити, тоді для усунення прогалини в поясненні необхідно припустити специфічне відношення ментального й фізичного. У цьому разі постає проблема доведення психофізичних законів, що обумовлювали б вказану взаємодію. Утруднення полягає в тому, що подібні «закони» являють собою погано замасковану екстраполяцію принципів природознавства на сферу ментального і, отже, ідея опосередкування двох сфер знов опиняється під питанням. Тим самим ні редукціонізм (якого дотримуються матеріалістичні моністи), ні теорія рівноправного відношення, що її обстоюють представники інтеракціонізму, виявляються непродуктивними.

Тому варто сфокусуватися на структурі ума. У ній Чалмерс виокремлює два основоположні аспекти: феноменальний і ментальний. Якраз напруження між способом пояснити феноменальне (своєрідне «як» кожного психічного переживання) і на основі цього збагнути сутність ментального (своєрідне «що» сукупності переживань як рамки цілого досвіду) утворює ядро психофізичної проблеми (*mind-body problem*). Повертаючись до диференціації ЛПС та ВПС, можна стверджувати, що шлях до розв'язання першої пролягає через прояснення наших феноменальних даних, а шлях до розв'язання ВПС має

передбачати розуміння того, як наш ум або свідомість включено у спричинення нашої поведінки. Простіше кажучи, феноменальний вимір свідомості пояснює, як переживання й відчуття відбуваються у фізичних тілах, тоді як ментальний вимір «відповідає» за те, щоб конкретизувати, чому ці переживання стають свідомими і проходять концептуалізацію. Для феноменального характерне повідомлення у судженнях першої особи однини (за допомогою займенника «я»), а критерієм достовірності є власна певність мовця у правдивості своїх слів. Для обговорення змістів ментальних феноменів ми звертаємося до суджень третьої особи однини або множини (за допомогою займенників «він», «вона», «вони» і власних імен), де критерієм є підтвердженням мого судження слугують судження інших людей. Приміром, феноменальний вимір свідомості розкривається іншим, коли я стверджую: «я вважаю, що...», «я маю намір...» тощо, а ментальне (краще сказати рефлексивно) висловлюванням про когось (Степан, Олена) із залученням пропозиційних настанов (переконаний, знає, збирається, почуває себе зобов'язаним тощо). Як бачимо, критерій для феноменального не тотожний критерію для визначення ментального попри те, що формально вони мають до діла з доволі схожим епістемічним каркасом речення.

Чалмерсове рішення передбачає дуалістичне розв'язання психофізичної проблеми і загалом проблеми свідомості. Своє власне вчення Чалмерс називає *панпротопсихізмом*. Воно постулює, що вторинні якості (такі як «зеленість» зеленого або «яблуковість» яблука) даються у сприйнятті фізично (через візуальні та тактильні перцепції), проте цих сприйнятів не могло би бути, якби сприймачі не усвідомлювали їх. Отож фізична доступність якостей речей невіддільна від психічного оформлення їх як окремих «речей». При цьому самі ці якості не є ані фізичними, ані психічними – вони мають нейтральний статус. Прикладом співвідношення фізичного й нефізичного у досвідах сприймання і розуміння є процес, який Чалмерс називає *супервентністю* (*supervenience*). Спроби перекласти його українською видаються малопродуктивними, хоча в обігу перебувають щонайменше два терміни – «надвідповідність» і калька

«супервенція». Зміст цього терміна можна визначити через відношення між вищим класом В-властивостей (напр., процесом мисленням) та нижчим класом А-властивостей (напр., нейрофізіологічною структурою мозку), коли *В-властивості є супервентними на/надвідповідають (supervene on) А-властивостям, причому А- і В-властивості якісно відрізняються між собою.* Чалмерс вказував, що йдеться про особливий тип зумовлення одного іншим, але при цьому уточнював, що має на увазі не однорівневе спричинення і не жорстку каузальну залежність між А і В. Слід відрізнити логічну супервентність, що може мати чинність у будь-якому можливому світі, від причинової (природної) супервентності, що не є логічно необхідною, а фактично необхідною. Ця супервентність чинна лише у такому фізичному світі, як наш, і описується мовою фізичної науки. Єдиною реальністю, яка не є логічно супервентною щодо фактів фізичної реальності, є свідомість, адже вона не є протяжною, отже не є фізичною. Відношення між природної і логічною супервентностями полягає в тому, що певній множині фактів $M(A)$ на мікрорівні з *необхідністю* відповідає факт вищого рівня $\Phi(B)$, де M позначає ментальні, а Φ – фізичні характеристики.

Введення поняття супервентності скероване на досягнення двох цілей: 1) уникнути редукціонізму ментального до фізичного, і 2) довести стійке корелятивне співвідношення фізичного та ментального, спираючись на доробок сучасної науки. Супервентність дає змогу уявити співіснування однакових фізичних структур у тілах при неоднаковій нефізичній структурі їх свідомості і цим слугує обґрунтуванню дуалістичної позиції Чалмерса.

Підсумуймо представлену мислителем доктрину у вигляді полісилогізму:

1. Будь-яка *система*, що оброблює потік інформації, має здатність усвідомлювати.
2. Свідомість не залежить від матеріального носія, тобто є автономною.
3. Якщо всі інформаційні системи мають свідомість, тоді відмінність між людиною та термометром визначена суто *кількісними* величинами.
4. Свідомість не тотожна самосвідомості.

5. Термометр не має самосвідомості, але він відчуває зовнішній світ поза собою й тому виконує певну дію (включає механізм нагрівання ртуті при зміні температури).
6. У термометра, як і в «будь-якої системи», є «суб'єктність» (його «Я»), тобто здатність діяти самостійно згідно з певними принципами.
7. Самосвідомість наявна лише у досконалих систем (людина, вищі примати).
8. Отже, попри те, що термометр не має самосвідомості, він має суб'єктність.
9. Отже, будь-яка система, що має суб'єктність, має й свідомість.

Така версія моністичної доктрини заснована на твердженні, що матеріальне й нематеріальне є лише двома аспектами цілісної структури світу. Вони є аспектами «для нас», оскільки наш розум не спроможний охоплювати цю цілісність моментально. Відповідно і свідомість (ум), і тіло *за властивостями* подібні, а за просторовими параметрами протилежні, через що між ними стає можливою взаємодія. Як ми пам'ятаємо, це бачення нагадує двоаспектну теорію, в межах якої когнітивний аспект (доступні нам способи пізнання), епістемічний аспект (багатоманіття того, що ми вважаємо, що «знаємо») і онтологічний аспект (те, як речі існують незалежно від нашого знання про них) взаємодіють на двох рівнях – рівні понять і рівні відчуттів. Адекватні поняття дають змогу узгодити способи пізнання зі способами даності нам явищ. У процедурному сенсі це доволі очевидна реставрація кантівської моделі пізнавця. Чалмерс розуміючи її труднощі, каже про awareness gap: пізнавцеві доводиться припускати те, існування чого він довести гранично не в змозі (атоми, кварки, регулярні причини тощо). Відмінність між натуралістичним дуалізмом і Кантовою епістемологією лише в тім, що перший відкидає трансцендентальні засади знання, а критерії достовірності ґрунтує на конвенційній практиці науки.

Розгляньмо, які критичні атаки приймає на себе натуралістичний дуалізм. *Перш за все*, непізнавані сутності (як сказав би Кант, ноумени або «речі самі

по собі») ми лише беремо за основу своїх подальших розміркувань, аби відкрити простір для можливості концептуалізації сприйняття. Але якщо такі сутності є невідними і певною мірою «нейтральними» чи «автономними» щодо нашого розуму, їх треба визнати й автономними щодо нашого досвіду. Тим самим буде обґрунтовано їхній об'єктивний, хай і незбагнений, статус. Отже, в узагальненні своїх досвідів ми спираємося на щось, з чим сама наука має проблеми, і що, зрештою, якраз і становить каркас нашого уявлення про світ. Себто ми одночасно не «знаємо», чим це є, і «знаємо» завдяки припущенню про це дещо.

По-друге, свідомість (або суб'єктність) завжди є *числюсь*. Отже, перш ніж приписувати суб'єктність будь-яким системам, слід зауважити, що вони є витворами нашого ума. Роздуми про кварки, нейрони, атоми нічого не повідомляють про *мою* свідомість. Варто запитати: чи справді структурно подібні «свідомість *цього* клена» і «свідомість Андрія»? Якщо так, тоді з імовірністю має постати наступне питання про причини їхньої відмінності: одна й та сама модель функціонування з одним принципом дії не потребує безлічі форм її реалізації; якщо ці форми різні й актуалізуються по-різному, принцип реалізації повинен містити модель тлумачення таких трансформацій.

По-третє, будь-яке поняття, зокрема «свідомість» або «ум», є продуктом діяльності розуму, а не навпаки. Свідомість не є продуктом різних понять. І іншому разі необхідно пояснити джерела походження понять (звідки у нас поняття?) та мету цього (навіщо нам поняття?). Та обставина, що ми самі їх генеруємо, ще не дає ясної відповіді на питання про ума. Назва головної праці Чалмерса («Свідомий ум») дає підтвердження цій думці: можна довести, що ми усвідомлюємо свої стани або стани справ у реальності, але, врешті, що таке сам «ум»? І чи справді можливі свідомості (суб'єктності) поза умом? Якщо звернутися до прикладів у природі, то ситуація тут виявляється складнішою, ніж коли ми вивчаємо роботу термометру. Скажімо, тіло інфузорії дає їй здатність усвідомлювати зовнішню небезпеку, у тім числі через припущення про притаманні їй квалії, але це навряд чи буде строгим доказом на користь

наявності в неї ума. Але якщо ум і свідомість відокремити, чи працюватиме цей підхід у випадку з механічними системами?

Загалом концепція Чалмерса балансує між поняттям про феноменальний і психологічний аспекти свідомості. Водночас те, що філософ обстоює ідею дуалізму, вириваючи з рук редуccionістів натуралістичну доктрину, робить його підхід теоретично більш прийнятним. Чалмерс висловлюється доволі обережно і заявляє, що його твердження є попередніми і мають не стільки вигляд концепції, скільки зауваг для появи більш узгодженої доктрини. Це певною мірою забезпечує йому маневр в опонуванні своїм критикам. Та як філософ Чалмерс здобув визнання передовсім завдяки своїй винахідливій системі доказів у полеміці про «філософського зомбі». Розгляньмо її зміст докладніше.

5.5. «Філософський зомбі»: боротьба аргументів.

У своєму творі «Свідомий ум (The Conscious Mind)» (1996) Чалмерс звертається до критики ідей епіфеноменалізму та інтеракціонізму за допомогою аргументу про філософського (або феноменального) зомбі (АФЗ). Уперше аргумент з аналогічною структурою і відповідним терміном запропонував Роберт Кірк у статті «Зомбі проти матеріалістів» (1974). Згідно з Кірком, уявлюваність того чи іншого явища засвідчує його логічну неспростовність. Якщо матеріалісти можуть припустити існування істоти, що підлягає модальному критерію, вони мусять визнати нефізичне частиною фізичного світу.

Отож уявімо собі, що поруч із нами існують подібні до нас істоти, вони виглядають як ми – з виду люди, а всередині роботи, сприймають, розуміють, говорять, діють як ми – але нічого не відчують і не усвідомлюють себе (позбавлені квалій). Таких істот можна назвати феноменальними зомбі. Вони мають певні ментальні стани, але не рефлексують про них. Їхній психіці бракує інтенційності: вони не здатні пізнавати власні наміри й учинки. Вони кричать від болю, не відчуючи її – просто тому, що «так потрібно». Своім

існуванням ці зомбі засвідчують, що свідомість не є обов'язковим наслідком нашої тілесної організації. За умови, що описувати свідомість як дещо фізичне є хибним, свідомість не є чимось фізичним. Звідси й висновок, що ми – не зомбі, хоча задовільний опис наших «зомбі»-станів є теоретично допустимим.

Звернімо увагу на формальну структуру цього аргументу:

Матеріалістична картина світу описує все, що існує, у термінах фізики.

Те, що вона не описує фізично, не є можливим.

Поняття феноменального зомбі є логічно *несуперечливим*.

Отже, існування феноменального зомбі є логічно можливим.

Феноменальний зомбі *фізично тотожний* людині.

Свідомий досвід не можна описати в термінах фізики.

Людина відрізняється від зомбі наявністю свідомості (суб'єктивності).

Між людиною й зомбі існує *нефізична відмінність*.

Свідомість (суб'єктивність) існує.

Отже, матеріалізм хибний.

Як бачимо, епіцентром аргументу є теза про модальність, яку філософ перекваліфіковує в термінах тотожності. По суті, Чалмерс твердить: за умови, що наш світ обмежено описом лише через фізикалістський словник, існування феноменальних зомбі спростувати неможливо; тому припущення про розмаїтість станів світу (а отже і типів його опису) підважує злагодженість фізикалістського пояснення.

Значимо, що Чалмерсів доказ споріднений із аргументом з відсутності квалій, спрямованого проти ідей Патнема (див. 3.3). Фактично, Чалмерс репродукує його і доповнює міркуванням Кірка, застосувавши замість «квалій» поняття «зомбі» як втілення суб'єкта, позбавленого неповторності власного перцептивного досвіду. Але на відміну від від автора АВК, Шумейкера, Чалмерсів аргумент виводить дискусію у нове для неї русло. Аргумент Шумейкера можна або прийняти, або відкинути, натомість теоретичну вибудову, зроблену Чалмерсом, можна розвивати далі – навіть усупереч початковій гіпотезі Чалмерса.

Скажімо, обстоюючи позицію функціоналізму і прийнявши загальну логіку модального аргументу, є всі підстави казати, що можна уявити собі наміри чи бажання без феноменального змісту, а це дає можливість прийняти АФЗ і водночас (sic!) відстояти послаблену версію матеріалізму. Або ж, навпаки, з позиції елімінативістів стверджувати, що у разі, якщо ми є описаними вище феноменальними зомбі, а наука допомагає нам це зрозуміти, тоді наш стан не є для нас проблемою, мовляв, якщо «буття зомбі» нічим не відрізняється від «буття не-зомбі», окрім як специфікацією окремих переживань, то й проблема є радше формальною – це проблема для філософів, а не для самих «зомбі».

Теза про мислимість / уявлюваність (conceivability thesis), на якій тримається процедура АФЗ, є доволі хитким підґрунтям, і саме їй довелося стати епіцентром критики з боку опонентів Чалмерса. Загалом суму критичних закидів щодо аргументу зомбі можна звести до чотирьох головних пунктів:

1. АФЗ спростовує лише логічну можливість людей як зомбі. Це не означає, що «в одному з логічно можливих світів» (в одній із можливих галактик) зомбі з нефеноменальним досвідом справді неможливі. Тоді пафос антиматеріалізму буде зруйновано. Як і у випадку з модальним аргументом Кріпке проти теорії тотожності, АФЗ спирається на метафізичний засновок, який не підлягає обґрунтуванню і є предметом епістемічної довіри.

2. Поняття «зомбі» є самосуперечливим: неможливо бути «як люди» і не мати «людських властивостей», себто свідомості. Отже, зомбі або існують – а це не так, або не існують. У кожному разі АФЗ є нерелевантним.

3. *Заперечення стосовно точності* (Сьорл): коли ми кажемо про зомбі «такі як ми істоти», ми маємо на увазі наші точні копії. Але – наші точні копії мали би квалії.

4. *Заперечення стосовно предикації* поняття «зомбі»: у зомбі, якщо вони взаємодіють, мала би виникнути своєрідна «колективна свідомість» – тоді вони були би людьми, але це означає, що зомбі не були би зомбі. Крім того, будь-яка спільнота, і тут зомбі не є винятком, вербально взаємодіють, себто

оперують спільною знаковою системою знаків. Така система може бути тільки інтерсуб'єктивною, тож користуватися її правилами й не розуміти їх – це логічна помилка в один крок. Отже, АФЗ містить суперечність у своїй ключовій тезі.

Як зауважив іще **Марвін Мінські** (Minsky), аргументація, подібна до АФЗ, містить ознаки *циркулярного доказу*, тобто хибного кола в доведенні: тезу, що її потрібно довести чи спростувати, покладено в основу самої процедури доведення. Умовно кажучи, якщо філософ має на меті довести існування несвідомих тіл, таких «якими є ми», він передовсім ухиляється від виправдання коректності фрази «такі, як ми», хоча аргумент вказує на вимогу протилежного. Інакше кажучи, щоб показати, що можливі феноменальні зомбі, в усьому схожі на людей, доводиться прийняти на віру, що вони *eo ipso* «схожі на людей». Так чи так, метафізичний засновок стає ключовим опертям для цілої споруди доказу. І, відповідно Чалмерс доводить те, що мав би спростовувати з очевидності.

Філософ **Кіт Франкіш** (Keith Frankish) добре вловив багатозначність вказаного доказу і створив натомість *аргумент анти-зомбі*. Він намірявся продемонструвати, що Чалмерс, а з ним і критики фізикалізму, непримітно для себе цілять в осердя власної теорії. В його основі інверсія головної Чалмерсової тези при дотриманні формальної структури його аргументу. Отже, Франкіш розмислює:

Антизомбі є мислимими.

Якщо антизомбі є мислимими, тоді вони можливі.

Якщо антизомбі можливі, тоді свідомість є суто фізичною*.

Отже, свідомість є фізичною.

Якщо антизомбі можливі, тоді мікрофізичні властивості світу є *метафізично достатніми* для існування свідомості у фізичному світі.

(*Фізичні властивості не заперечують можливості метафізичних (якихось нами неврахованих) передумов появи свідомості. Отже, фізичні закони *не є*

перешкодою для існування ментального: фізична природа свідомості може вказувати й на її нефізичні властивості).

Франкіш наполягає: «В аргументі антизомбі ми уявляємо відсутність наших... нефізичних властивостей за умови збереження нашої свідомості, і висновуємо, що свідомість є фізичною». Аргумент антизомбі скерований переконати, що свідомість є не необхідно фізичною, а «насправді фізичною». Проте Франкіш не уточнює, який варіант пояснення «фізичності» потрібен для свідомості: ним може бути і теорія тотожності, і функціоналізм, і емерджентизм. Отож контрдоказ Франкіша показує, що «аргумент зомбі» є самоспростовним, а отже, хибним. Що може би на це заперечити Чалмерс? *По-перше*, що антизомбі є *немислимими*, адже спершу ми маємо помислити зомбі. Немисливість антизомбі підважує їх автономний статус як абстрактних сутностей, а це, погодьмося, руйнує всю подальшу конструкцію аргументу. *По-друге*, якщо антизомбі є суто фізичними сутностями, тоді свідомість є надлишковим феноменом у діапазоні решти властивостей фізичних істот. Одначе свідомість як проблема якраз і виявляє, що редукувати її до чогось питома фізичного не вдається: проблема не зникає сама собою.

Зазначу, що модальний аргумент у принципі не ладен розв'язати проблему посутньо. Згадаймо Ансельма Кентерберійського, який розвинув онтологічний доказ буття надприродної істоти. Чи сама поява його доказу кладе край скептицизму щодо божественної сутності? Авжеж, ні. Та й не спроможна це зробити. І в цьому пункті варто погодитися з критиками АФЗ.

І все ж, евристичний характер АФЗ з-поміж інших ментальних експериментів аналітичної філософії є доволі продуктивним. Крім Франкіша, до цієї теми зверталися чимало мислителів. Принагідно розгляньмо її тлумачення Деніелом Денетом. Якщо Франкіш у своєму контраргументі доводив логіку міркувань Чалмерса до її самозаперечення, то Денет рішуче наполягає, що висновок, який випливає з нього, є прямо протилежним: зомбі є можливими, натомість суб'єктивність як синонім свідомості є химерою.

Логіка його розмислу є такою: квалії не мають автономії щодо фізичної природи нашого мозку, тому всі люди з мізками є зомбі особливого типу, Денет називає їх «зімбо» (zimboe), тобто такими зомбі, які внаслідок самоконтролю мають несвідому обізнаність вищого порядку про змісти своїх станів нижчого порядку. Висновок очевидний: мати квалії та суб'єктивний досвід – не одне й те саме. Адже всі ми маємо чуттєві переживання, які відрізняють нас поміж собою, однак твердити про те, що хтось має свій унікальний і неповторний досвід означало б наполягати, що існує якийсь принципово непізнаваний досвід, а це є суперечністю.

У стилі властивій йому іронії Денет заявляє, що його, професійного філософа, та автомеханіка, який лагодить йому автівку, ніщо посутньо не відрізняє. У сенсі Чалмерса вони обоє є феноменальними зомбі. Денет як не-філософ (у період життя, коли він іще не став філософом) і Денет як філософ, або Денет як яхтсмен (він бере участь у спортивних змаганнях) є різними моделями цього зомбі: перший є автентичною несамосвідомою істотою, а другий його тип унаочнює істоту, яка намагається навчити себе самосвідомого досвіду, а третій коливається між двома попередніми (не всі мудреці вміють кермувати яхту). Іронічне, а нерідко й скептично-саркастичне обігрування Денетом патерну, запропонованого Чалмерсом, має продемонструвати, що колотнеча відбувається довкола пустопорожньої проблеми, розв'язку якої не існує.

Денетова теза значно радикалізує критику щодо висновків Чалмерса. Продовження дискусії підхоплює *аргумент зомбі* (zombie argument), запропонований **Ричардом Брауном** (Brown). Можна, міркує Браун, припустити існування істот, в усьому подібних до людей за *нефізичними* ознаками, але таких, які не мають феноменального досвіду. Такі істоти – назвемо їх «зумбі» – все розуміють, не усвідомлюючи цього, говорять, не рефлектуючи про сказане (Варто поспостерігати за собою, аби погодитися з тим, що нерідко ми самі перебуваємо у схожих станах). Якби вони існували, це спростувало би дуалізм (теорію, що свідомість за своєю природою є не є

тільки фізичною) і не суперечило б емпіричній реальності. Отже, Браун акцентує на тому, що психічні властивості теж не повідомляють достатніх критеріїв ототожнення об'єктів за нефізичними ознаками. Що можна відповісти на це? Чалмерс ладен погодитися: дійсно, зумбі можуть існувати, а дуалізм є хибною теорією. Проте огріхом аргументу Брауна є його логічні наслідки для психофізичної проблеми загалом: на модальному рівні – а ми маємо до діла лише з мисленнєвими експериментами та аргументацією з можливих станів справ – розв'язати дилему «фізикалізм чи дуалізм» виявляється неможливим. Тож навіть якщо «зумбі» існують, це мало що змінює у змісті теоретичної проблеми.

Хоч як це виглядає дивним, варіації з власною назвою «зумбі» у спрямованих проти концепції Чалмерса альтернативних доказах не є просто іронічним обігруванням його припущення. Їхнім мотивом є передусім фокус на тому, що імена як такі не прив'язані до сутності їхніх носіїв; під «зумбі/зімба» можна розуміти будь-яку істоту лише в тому разі, якщо її існування та її природа не є тотожними. Адже видимість і справжня сутність цих істот виявляються підставово розірваними. Завдання полягає у виявленні того, чим є згадана підстава і до якого типу вона належить (онтологічна, психологічна, концептуальна, методологічна і т.п.). Певною мірою, такий логічний хід імітує характер психофізичної проблеми за формою – подвійність у межах однієї поняттєвої рамки. Тож розгляд АФЗ задає нове спрямування для осмислення цього утруднення. Суперечка не вщухає.

Джерела:

Brown, R. (2012). Zombies and Simulation. *Journal of Consciousness Studies* 19 (7-8): 21-25.

Shoemaker, S. (1975). Functionalism and Qualia. *Philosophical Studies*, 27: 291–315.

Джексон Ф. (2014). Епіфеноменальні квалія. *Антологія сучасної аналітичної філософії, або Жук залишає коробку*. Л.: Літопис, сс. 260–272.

Франкіш К. (2014). Аргумент анти-зомбі. *Антологія сучасної аналітичної філософії, або Жук залишає коробку*. Л.: Літопис, сс. 340–360.

Чалмерс Д. (2015). Аргумент 1: Логічна можливість зомбі. *Філософська думка*, 5: 60–67. <http://journal.philosophy.ua/article/argument-1-logichna>.

Запитання для самоперевірки:

1. Як Ви можете пояснити сутність інтеракціонізму?
2. Що спільного у таких доктринах, як дуалізм властивостей та інтеракціонізм? Своєю думки підкріпіть прикладами.
3. Що таке натуралістичний дуалізм? Що відрізняє цю теорію від решти типів психофізичного дуалізму?
4. Чим різняться «важка» й «легка» проблеми свідомості?
5. У чому полягає explanatory gap, з погляду Чалмерса?
6. Чи приховано внутрішню суперечність прихована в аргументах «зомбі» й «анти-зомбі»? Яку?
7. В чому Дж. Сьорл міг би погодитися з аргументацією Д. Чалмерса про природу свідомості?
8. Чи є успішною спроба Д. Денета показати зомбійність людського ума? Чи можна (і якщо можна, то яким чином) удосконалити його аргумент?
9. Чому модальна аргументація стала пріоритетною для міркувань про умови існування ума та свідомості?

Лекція 6.

Дуалістичний антиредукціонізм (II).

Емерджентизм і епіфеноменалізм.

6.1. Емерджентизм як натуралістична доктрина.

Вище ми вже оглядово простежили критику **Джоном Сьорлом** (Searle) методології функціоналізму. Праці цього мислителя сповнені як глибокої іронії щодо позиції його опонентів, так і оригінальних міркувань про природу

свідомості. Свої ключові ідеї, відомі під назвою програми *емерджентизму*, він виклав у книгах «Перевідриття ума» (1992) і «Ум: короткий вступ» (2004). Стисло окреслимо основні положення, які обстоює їх автор. Але спершу зауважмо, що емерджентизм – це маркер, який думці Сьорла приписують здебільшого його колеги, натомість сам мислитель воліє казати радше про своє вчення як приклад *натуралістичної філософії*. Згідно з натуралізмом, первинні причини усіх явищ і процесів слід шукати у біологічній природі, хоча її еволюція робить не завжди очевидними зв'язки між причинами і діями цих причин.

Завдання філософа – реконструювати ці зв'язки, не знехтувавши тим, що причинам притаманно впливати опосередковано, а не прямо. З терміном «емерджентний» справа стоїть схожим чином. *Емерджентною* є така властивість, що утворилася спонтанно внаслідок перебування біологічної системи у стані нерівноваги. Скажімо, у процесі еволюції маса тіла деяких ссавців збільшилася настільки, що виникла потреба раціонально і гнучкіше управляти ним. Для цього необхідно було мати досконаліший мозок, який обробляв би великий потік інформації про навколишній світ, його ресурси та загрози тощо; у результаті цього на основі суто фізичної речовини (мозок) виникає специфічно-нова здатність свідомо, а не інстинктивно чинити, передумовою чого стає поява свідомості та розуму. Інакше кажучи, хоча розум спродуковано біологічно змінами у структурі мозку, разом із тим він якісно не тотожний мозку, але без участі мозку не функціонуватиме.

Подібний концептуальний компроміс робить емерджентизм як напрям у філософії і гнучкішим за вже розглянуті, і водночас таким, що камуфлює актуальність справжнього вирішення психофізичної проблеми. Зважаючи на сказане раніше, між натуралізмом і емерджентизмом простежується тяглість аргументації. Однак у випадку Сьорла маємо стверджувати онтологічну, а не просто каузальну, емерджентність, адже дія конкретної причини може бути як необхідною, так і випадковою. У рамках філософської концепції Сьорла

йдеться саме про необхідний, хоча й не повністю прогнозований, вплив фізичних чинників на продукування ментальних феноменів.

Основним здобутком Сьорла є спроба побудувати несуперечливу теорію суб'єктивності на ґрунті натуралістичного підходу. Як ми пам'ятаємо, проблема суб'єкта і для фізикалізму, і для функціоналізму виглядала другорядною, похідною від поняття матерії або поняття системи. Філософ пропонує поєднати нормативістську концепцію ума як суб'єктивності з матеріалістичним поясненням еволюції мозку. Згідно з Сьорлом, ум і суб'єктивність належать до єдиної ментальної структури, в якій тілесне виконує функцію субстрату для ума (психіки). Щоб показати, як біологічні зміни в мозку призводять до стрибкоподібного виникнення ума, потрібно ввести поняття, яке, з одного боку, демонструє залежність від тіла, а з іншого, автономно провадить референцію щодо своїх об'єктів. І тут мислитель звертається до добре відомого у континентальній феноменології поняття *інтенційності*. Сьорл запозичує це поняття з праць учениці Вітгенштайна Елізабет Енском (Anscombe). В англосаксонській філософській традиції *intentionality* і *aboutness* тотожні за змістом. Важливо й те, від інтенційності можливо перейти до поняття *інтенціоналу* – логічного терміна, що позначає сенс власного імені у межах класу, до якого його відносять. Простіше кажучи, якщо інтенційність «відповідає» за відношення між думкою та її змістом, то інтенціонал – за саме значення слова (чи фрази) без вказівки на його предметний референт. (Пригадаймо приклад Фреге з двома різними назвами для планети Венера.) Важливо звертати увагу на розрізнення *intentionality* та *intensionality* під час роботи з англomовними філософськими текстами.

Як ми знаємо, інтенційність – це спрямованість ума на свій об'єкт, байдуже фізичний чи нефізичний. Інтенційний ментальний стан виявляє власну причину, вказуючи на щось поза умом, що спонукає активізацію його актив. Приміром, для ума фотографа, який знімає косулю на природі, інтенційним об'єктом є образ косулі, а для охопленого страхом злочинця, який поспіхом ховає вкрадене добро, інтенційним об'єктом буде будь-який

раптовий свідок. Натомість ум людини у стані депресії не має інтенційного об'єкта, адже депресія є неінтенційним станом (відсутній об'єкт, який акумулював би в собі причини появи й зникнення цього розладу). Для натуралістичного філософа інтенційність корелюється й буквально генерується фіксованістю біологічної структури нашого мозку на чуттєвості. Сьорл стверджує, що бути свідомим означає усвідомлювати себе у певному стані, себто якщо ми хочемо показати, що свідомість (або ширше, ум) притаманні особі, найперше слід довести, що ментальний стан X визнається нею як її власний актуальний стан. Це потрібно для сполучення стосунку суб'єктивності та свідомості через концепт інтенційності. Опоненти можуть дорікнути, що людям властиве також переживання неінтенційним станів, зокрема вже згаданої депресії, жаху, ступору, болю, психічний автоматизм, сні без сновидінь, кепський настрій тощо.

Попри те, що суб'єктивність існує у своїй очевидності, пояснити її вкрай важко. Зрештою, можна навіть казати про значну неосяжність цієї структури. А втім, вона якраз і утворює фундамент наших досвідів. Згадаймо будь-яке своє переживання і відмінусуємо від нього «це було зі мною». Що матимемо у підсумку? Анонімний стан, який годі квантифікувати. З цього можна зробити висновок, що основними характеристиками суб'єктивності є чутливість (здатність відчувати й сприймати), креативність (змінювати зміст сприйнятого чи відчутого, як-от у мріянні чи постановці раціональних цілей) і гнучкість поведінки (суб'єкт зазнає змін під впливом своїх досвідів, а також рефлектуючи про них). Виходить, що опис цієї структури в термінах біології не дає нам повної картини суб'єктивності. Отже, висновує Сьорл, суб'єктивність як осереддя ума необхідно прояснити нефізикалістськи, але так, щоб поняттєво ми не втрапили у пастку типового картезіанського субстанціалізму. Наголосимо: інтенційні стани – це елемент виняткової суб'єктивної онтології, якої бракує решті об'єктів природи. Тому потрібна розробка докладної онтології суб'єктивності.

Начерк цієї онтології Сьорл і викладає у «Перевідкритті ума». У загальних рисах вона виглядає так. Основу ума утворюють інтенційність та внутрішня темпоральність. Позаяк інтенційність не може бути суцільною, адже не охоплює весь спектр ментальних станів (про це ми говорили вище), її має врівноважувати здатність пов'язувати зміст переживання з його часовим триванням, отож стосунок інтенційності та темпоральності не викликає заперечень, він є паралельним. Скажімо, біль є неінтенційним станом, але має власну часову простягненість. Своєю чергою, накладення обох зазначених структур створює персоналізоване суб'єктивне чуття, без якого індивід не міг би відрізнити, що і як зворушує та хвилює його. Крім того, Сьорл наголошує, що інтенційні акти спрямовують ум на виділення ключового й периферійного у сприйнятті, того, що психологи-гештальтисти іменують «аспектуальністю». Ці аспекти наших перцепцій так само існують ніби у двох вимірах – як спосіб отримати сприйняту даність, так і спроможність приписати їй неповторність, – те, що вирізняє схожі об'єкти за якісними ознаками.

У такий спосіб ми переходимо до визнання того, що глибинним шаром сприйняття, або схоплення, є гештальт-структури уяви. Вони забезпечують утримання виділеного аспекту в пам'яті, аби надалі ми могли довільно варіювати образи речей. Ця здатність відкриває доступ до коригування наших ментальних станів: один і той самий стан змінює зміст і вид залежно від часу, коли ум оброблює його (у моменті, у спогаді, у фантазії тощо). Але варто зазначити: якщо гештальт-структури і суб'єктивне почуття через внутрішню темпоральність дають змогу конструювати і реконструювати об'єкти згідно з інтересами пізнавця, то відношення між аспектуальністю сприйняття і власне інтенційністю уможлиблює розуміння суб'єктом своїх дій як програми його самореалізації. Тому першу сукупність актів варто назвати горизонтальними, а другу – вертикальними рівнями самосвідомості. Отже, первинним і кінцевим елементом споруди ума є якраз самосвідомість, щодо якої сходяться всі окреслені її конститутивні компоненти. Як бачимо,

Сьорл балансує між формальним аналізом свідомих актів і феноменологією можливих сприйняття, на які ці акти скеровані.

Утім, Сьорлів підхід демонструє, що обмежити вплив біологізму на цей начерк ума вкрай важко (тим паче для заявлено натуралістичної філософії!). Коли філософ каже про спільність еволюції для всіх вищих тварин, він неминуче мусить визнати, що породження мозком особливих клітин, які потім здатні утворювати матеріальну основу свідомості, помічають у всіх тварин, які мають «вуха, шкіру, очі, можуть видавати звуки, що розпізнаються іншими представниками їхнього біологічного виду». Крім того, реакція на команди людини теж містить натяк про спорідненість біологічного рівня розвитку з розвитком цих живих істот. Так чи так, Сьорл опиняється у непростій ситуації, що її можна назвати *парадоксом натуралізму*. З одного боку, годі заперечувати, що людина не є винятковою істотою у світі, а факт того, що вона еволюціонувала більше за інші види, є історичною випадковістю. З другого боку, саме людська суб'єктивність спроможна поставити питання про природу ума та свідомості, отже, це наділяє людину зовсім іншими інтелектуальними повноваженнями, ніж решта натурального світу.

Натуралізм діє як моністична доктрина, яка, однак, містить явний епістемологічний розрив (*epistemological gap*). Коли на одній з лекцій в Індії мислитель казав про несуперечливість Дарвінової теорії та формальної семантики першої особи, один із присутніх, прибічник індуїзму, перервав лектора, заявивши: «Ваша теорія – хибна. Цілком можливо мати одну душу в різних тілах». Сьорлова відповідь адепту реінкарнації була сповнена іронії. Хоча справді, у рамках емерджентизму не так просто роз'яснити, чому еволюція створює явища, які переступають закони еволюції. Людський розум, що постав з природи, обирає її для себе жертвою власної експансії.

6.2. Критика емерджентизму.

З опису структури ума за Сьорлом ми помітили, що його концепцію можна позбавити низки біологічних дескрипцій. Те, як філософ описав consciousness, не ставить під сумнів очевидність існування mind. Питання до емерджентиста, знов-таки, виникає в тому разі, коли доводиться пояснювати, як виникла ця структура і чому вона є саме такою, а не інакшою.

Сьорл стверджує, що «ум є нередуктивним фізичним компонентом реальності». Проте залишається незрозумілим, чи є достатні підстави приписувати цьому компоненту фізичні характеристики. Адже фактично теза антиредукціонізму тягне за собою протилежну тезу – про зв'язок між суб'єктивною онтологією і біологічним середовищем. Якщо узгодити ментальні процеси в умі з фізіологічними у мозку, це не усуваемо, а повертає нас до суті проблеми. І тут слід пригадати аргумент Вітгенштайна: чому, якщо інстанцією фіксації болю є не в кінцівка, а в мозок, ми хапаємося за кінцівку, а не за голову?!

Критикуючи функціоналістів, Сьорл стверджує, що підставою появи ума є не соціальні практики, а особлива будова ума, проявом діяльності якого якраз і є соціальні практики. Щоправда, згодом у праці «Раціональність у дії» (2004) мислитель спробував узгодити обидві тези, заявивши, що між соціальним проявом раціонального і тим, як вона влаштована, наявна прірва пояснення, а не прірва походження. А проте, суб'єктивність, яка не є умом, теж має бути узгоджена з перебігом фізичних процесів у нашому тілі, що, авжеж, повертає силу аргументам фізикалізму, а не ослаблює їх.

Більш того, за логікою Сьорла, інтенційність та єдність свідомості теж слід визнати результатом біологічної еволюції. Це, однак, суперечить думці про внутрішнє розгортання структури, в межах якої властивості діятимуть незалежно від того, як вони влаштовані. Підстави так вважати дає той факт, що наша уява спроможна працювати з об'єктами, які не чинять на нас каузального тиску. Понад те, сам Сьорл критикує «надмірний» каузалізм матеріалістів у тезі про наявність інтенційності саме у людської свідомості.

Якщо вона існує в інших видів живих істот, вона, як можна припустити, мала би супроводжуватися комунікативним компонентом. А мова (хай би як її витлумачували) має передусім давати інструментарій для повідомлення не тільки про власні переживання, а й інших носіїв інтенційності. Знову-таки, тільки люди наділені обома властивостями одразу.

Двозначність емерджентистської позиції, врешті, полягає в тім, що, з одного боку, він атакує крайні версії фізикалізму та машинного функціоналізму, а з другого боку, приховує власний «редукціонізм у два кроки». На першому етапі ментальне ототожнюється з усвідомлюваним, а на другому – з фізичним (нейрофізіологічним). Як видається, за цим критерієм видатна особистість на кшталт Альберта Айнштейна та пересічний обиватель фізично подібні, хоча це не пояснює, чому один із них геній, а інший – ні. Проте так само, фізична однаковість виду (вищі ссавці) не дає можливості розрізнити інтелектуальні потуги людини та шимпанзе.

Загалом емерджентизм постійно балансує між інтерпретацією ума як «комплексу нефізичних властивостей мозку, які утворюють новий психічний комплекс (наше «Я»))» та матеріальної основи цих властивостей, які чомусь потребують надбудови нової структури *над* собою. Простіше кажучи, хоча мозок продукує певні стани, які ми звично вважаємо «думками», не зовсім ясно, що дає підстави припускати, що думки і є ніщо інше, як мозок у дії. Вони не тотожні, але субстантивно споріднені. З цієї парадоксальної ситуації емерджентизм шукає вихід у *двоаспектній теорії*, до якої вдаються й функціоналісти. Це насправді доводить суперечливу суть доктрини емерджентизму.

6.3. Епіфеноменалізм

Загальні принципи епіфеноменалізму було сформульовано британцем **Чарлзом Бродом** (C.D. Broad) у праці «Ум і його місце у природі» (The Mind and Its Place in Nature, 1925):

Деякі тілесні події спричиняють ментальні події.

Жодна ментальна подія не спричиняє інші ментальні події.

Жодна ментальна подія не спричиняє фізичні події.

Отже, усі ментальні події спричинено тілесними подіями.

Хоча епіфеноменалізм припускає існування нетілесних подій, останні не мають онтологічної автономії щодо тілесних подій. Це означає, що *епіфеномени* – це певні властивості, що їх неможливо звести лише до фізичних феноменів, як-от свідомість, розум і ширше – ум. Але попри те, що епіфеномени існують не як фізичні явища, хоча й будучи зумовлені ними, вони самі по собі не мають здатності впливати на фізичні явища. Скажімо, свідомість можна вважати фізичною здатністю, якщо ми розглядаємо її у рамках нейрофізіології головного мозку. Якщо особа усвідомлює тривогу, дослідник може ідентифікувати цей стан через збудження С-волокон за допомогою датчиків. Однак це не означає, що мозок і свідомість є одним і тим самим явищем або процесом. Радше епіфеноменаліст переконуватиме своїх опонентів у тому, що мозок «потрібен» свідомості (як певний субстрат) для того, щоб вона проявила себе в дії.

Епіфеноменалізм іноді називають слабким дуалізмом у питанні про свідомість. Перевага такого підходу полягає в уникненні строгої редукції до одного з начал (матеріалізму), але й недолік тут доволі явний: епіфеноменалістська доктрина апелює до метафізичних засновків, не пояснюючи «чим є» свідомість та ум, якщо вони не є мозком. Прибічники цього підходу переконані, що фізична реальність є самодостатньою і замкненою. Це, щоправда, не означає, що вона є повністю замкненою, адже, як ми бачили у п. 4 наведеного Брудом аргументу, фізичне впливає на нефізичне однонаправлено. Отож усі без винятку людські наміри, думки, бажання, вчинки зумовлено фізичними чинниками, які, своєю чергою, не впливають навспак: бажання бути у гарній фізичній формі не веде до занять спортом, до цього можуть призвести хіба що біль у певній частині тіла або оголошення про знижку на абонемент у спортзалі. Саме ж бажання змінити характеристики свого тіла є просто ілюзією, у яку ми схильні вірити.

Філософ Френк Джексон, відомий своїми аргументом знання у питанні про реальність квалій (див. 4.1.), навів кілька доказів на користь суперечливості епіфеноменалізму. Критика епіфеноменалізму розгортається з різних позицій, хоча підстави для неї дає він сам з огляду на закладений в ньому дисбаланс між двома типами сутностей. Розгляньмо основні закиди до доктрини епіфеноменалізму:

1. Необхідно виявити психофізичні закони, без яких ми втрачаємо підґрунтя для обговорення взаємодії ментального й фізичного. Проте досі їх не сформульовано і не підтверджено експериментально.
2. Достатніх підстав вважати, що психофізичні закони існують, наразі немає. Більш того, якщо вдасться довести їхнє існування (існування, але не пізнаваність!), цей факт підважить саму необхідність епіфеноменалістської позиції, позаяк епіфеноменаліст муситиме схилитися або вбік моністичної картини світу, або вбік інтеракціонізму, або взагалі втратить доцільність. У кожному з наведених випадків майбутнє для епіфеноменалізму виглядає безперспективним.
3. Епіфеноменалізм позбавляє чинності бігевіористський критерій пізнання думок інших осіб: якщо «наміри» як сутності не існують і вони є всього лише людськими діями, тоді за діями немає прихованих внутрішніх епізодів життя душі. Отже, якщо ментальне (не передуює і) не впливає на фізичне, тоді з поведінки X неможливо дізнатися про те, чого X хоче нею досягти. А це веде до висновку, що а) люди насправді не розуміють інших людей; б) передача знань і досвіду між людьми є оманливою. Тож навіть якщо я пишу ці рядки, а Ви їх читаєте?!
4. Таким робом проблема пізнання інших умів (the problem of other minds) у принципі ставить хрест не тільки на добутті знання про світ, а й про самого себе – бо ж моїм припущенням не відповідає жоден об'єктивний критерій, застосовний до аналізу царини ментального. Тоді не знати, звідки епіфеноменаліст бере поняття, якими він послуговується, коли доводить свою точку зору і спростовує протилежну – буцімто фізичні події *не*

створюють поняття, а максимум, що вони продукують, це розрізнені враження.

Підсумуймо: якщо, як переконують епіфеноменалісти, ментальне є всього лише побічним наслідком дії фізичного, тоді пізнання фізичного світу, а отже його законів, істини тощо стає для нас нерозв'язним завданням. Ми просто маємо безпідставно «вірити» у щось, чіїх поняття реально не маємо, або, як філософ Г'юм, покладатися на звичну регулярність подій як на схвалюваний «закон». У цьому разі не лише філософія, а й науки скидаються на набір абстрактних бездоказових постулатів, походження яких є сумнівним і незбагненим. Але в тому-то й річ, що наукове (а також філософське) знання не може бути прирівняне до суб'єктивної певності, адже наївній вірі бракує граничного критерію знання – готовності поставити себе під сумнів. Здоровий глузд сумнівається тільки тоді, коли щось не розуміє, а не розуміє він те, щодо чого не має надійного мірила. Епіфеноменалізм є слабкою спробою знайти безпечний вихід між фізикалістським редукаціонізмом та інтераціоналістською моделлю пояснення свідомості. Він ніби зависає поміж ними, не посилюючи жодну з цих парадигм.

Спрямовані проти дуалізму контраргументи можемо звести до кількох вузлових заперечень. *Перше.* Взаємодія між фізичним і ментальним дуалістами покладається без обґрунтування і без окреслення чітких принципів цієї взаємодії. Якби ми мали у своєму розпорядженні психофізичні закони або їхні слабкі аналоги, тоді можна було би спиратися на яку-небудь регулярність, а отже, підтверджуваність. Натомість нічого такого виявити не вдається. У кращому разі йдеться про *достовірність* фізичного і *ймовірність* ментального.

Друге. Якщо свідомість (психіка) є поняттям, то це всього лише регулятивне поняття: звертаючись до нього, ми описуємо те чи те явище як підпорядковане концептуалізації до явищ суміжного класу. Те, що «ментальне» не є змістовим поняттям, впливає з утруднень, очевидних для його критиків, адже етимологія цього слова не передбачає опису способів

взаємодії з нементальним. Вони залишаються загадковими. Поміж тим, будь-яке регулятивне поняття історично конститує приписи (або соціальні, або звичку сприймача). Отож окремий припис, що має автономний статус стосовно іншої множини приписів, не повинен їм суперечити. Інакше кажучи, цілком допустимо замінити комплексний термін «ментальне» на одне повсякденних уявлень, відтворюваних індивідуальним досвідом. Таке поняття є зрозумілішим, а головне – стають наочно вираженими моделі взаємодії між одним типом припису та іншими його типами.

Третє. Позитивність дуалізму не враховує неконвенційну взаємодію між ментальним і фізичним (порушення, збої, невдачі тощо). Візьмімо за приклад такий психічний феномен, як сон. Я бачу у сні свого знайомого Х, який розповідає про мандрівку у гори і розпитує про мої захоплення. Наприкінці розмови він двозначно жартує. Я почуваюся розгубленим. Побачена мною уві сні подія підлягає інтерпретації як шляхом зіставлення з минулими подіями у матеріальному світі, так і з подіями, спроектованими моєю фантазією (мрією побувати в Альпах). Хай би яким був результат витлумачення сну, з'ясувати, що відрізняє сновидіння про Х з моєю розмовою з Х наяву, з формального погляду доволі непросто. З точки зору обговорення критеріїв судження, це рівнозначущі події з однаковим набором учасників і однаковою фіксацією на рефлексії. Але в першому випадку референт можна визначити остенсивно («ось той, хто мені наснився»), а у другому це неможливо.

Четверте. Відсутність чіткого пояснення умов взаємодії фізичного з нефізичним тягне за собою питання: чому ми схильні вірити у нефізичні події (дію логічних законів, правил у точних науках, пов'язаність слів і думок, публічно заявлених зобов'язань і решту того, що не має матеріальних корелятивів). Якщо «припущенню» не відповідає нічого з того, з чим воно могло би взаємодіяти у фізичному середовищі, чому ми визнаємо реальність безреферентних ментальних актів?

П'яте. Дуалізм не дає чіткої відповіді на питання про генезу знання. Якщо існування ментального є невідомим і є всього лише справою віри, як

стає можливою наука? Авжеж, історія наукових пошуків засвідчує, що знання та віра невіддільні одне від одного; що вчені нерідко роблять відкриття або революційні висновки, довіряючи інтуїції; що основоположні засади деяких (або всіх без винятку?) наук ґрунтуються на беззасновкових аксіомах. У такому разі незрозуміло, як працює науковий метод, адже йому бракує «граничних підстав». Учений досліджує предметний світ, покладаючись на щось, що за своєю суттю є непідтверджуваним.

Шосте. Значна частина поборників дуалізму переконує, що збагнути цю концепцію здатен будь-хто, необділений «елементарним» здоровим глуздом. Але його недостатньо для пояснення функціонування самості та загалом ментального. Більш того, коли йдеться про автоматизовані системи, принципи їхньої роботи суперечать уявленням здорового глузду так само, як геліоцентрична теорія Всесвіту чи закони механіки суперечать буденним уявленням наївного розуму.

Слід при цьому зазначити, що сказане не означає краху виправдання дуалістичної доктрини. Так, приміром, український філософ Д. Сепетий пропонує компромісну версію дуалізму – *мінімальний дуалізм*. У його формулюванні ця теорія виглядає включати такі тези: 1) нефізичне (= душа або самість) існує у вигляді *загальної форми* суб'єктивності, чий *конкретний зміст* визначається структурою кожного мозку; 2) душа є універсальною сутністю, яка в поєднанні з окремим мозком утворює індивіда. Перше судження є засновковим, а друге – висновком, перше відображає дуалістичну тезу, друге – застосований до неї принцип мінімалізму. Такий підхід претендує на успіх лише за умови чіткого окреслення тих питань, на які відповідь перспективна наука, а які, знову-таки, будуть делеговані спекулятивно-формальному міркуванню.

Джерела:

Churchland, P.S. (2007). The Evolving Fortunes of Eliminative Materialism. In *Contemporary Debates in the Philosophy of Mind*. B.P. MacLaughlin & J. Cohen (eds.). Blackwell Publishing, pp. 160–181.

Dennett, D. (1987). *The Intentional Stance*, Cambridge, MA: MIT Press.

Feyerabend, P. (1999). *Knowledge, Science and Relativism: Philosophical Papers, Vol.3*. J. Preston (ed.). Cambridge: Cambridge University Press, pp. 16–49.

Searle, J.R. (2004). *Mind. A Brief Introduction*. Oxford University Press.

Searle, J.R. (2004). *The Rediscovery of the Mind*. Oxford University Press.

Серль Дж. (2017). Біологічний натуралізм. *Антологія сучасної філософії науки, або Усмішка ASIMO*. Л.: ЛНУ ім. І. Франка, сс. 476–494.

Запитання для самоперевірки:

1. Поясніть зміст поняття «інтенційність».
2. Чи є аргументи філософів-дуалістів більш невразливими щодо критики порівняно з редукціоністами?
3. Якщо ми відмовимося від суб'єктивності, як буде виглядати інтерв'ю з нобелівським лауреатом або відомим актором, нагородженим премією кіно академії?
4. Спробуйте зіставити докази на користь і проти емерджентизму. Що в цій теорії переважає, на Ваш погляд?
5. Уявіть дискусію між емерджентистами та епіфеноменалістами. Хто з них, на Вашу думку, більше орієнтується на союз з природничою наукою, а хто вважає проблему ментального і загалом психофізичну проблеми справою суто філософської компетентності?
6. У чому недоліки аргументу знання, запропонованому Ф. Джексоном? Кого з представників філософських підходів цей доказ не зможе переконати?

7. Якщо Мері (з уявного експерименту Ф. Джексона) вживити датчики, які дадуть змогу розпізнавати відтінки кольорів, чи дасть це вченій можливість успішно завершити своє дослідження?
8. Як можливо дуалістськи розв'язати проблему інших умів?

Лекція 7.

Аномальний монізм. Елімінативізм. Гетерофеноменологія.

7.1. Типи філософського монізму.

Монізм є ученням про онтологічну єдність світу, хоча у свідомості пізнавців ця єдність може виглядати розмаїттям відношень між його окремими елементами. Як філософська концепція він спирається на пояснення, вивідне з двоаспектної теорії: між різними типами феноменів відмінність є не істотною, а суто інструментальною. Хоча ментальний феномен є невидимим (ум, свідомість, дух), його прояви у вигляді актів передбачають видиму фіксацію. Отож ведучи розмову про функційні прояви, ми насправді говоримо про суть самої речі. Моністи твердять, що варіативність описів ніяк не спотворює описуваний предмет.

Загалом можна виділити три типи монізму: субстанційний (Спіноза), нейтральний (його обстоювали Ернст Мах і Бертран Расел), і аномальний (Дональд Девідсон) (див. 7.2). Сутність першого типу виражає теза про те, що матерія і психічне є різними назвами єдиної субстанції світу, незважаючи на те, що її насправді утворює (комплекс відчуттів, сила енергії тощо). Аномальний монізм, як буде показано далі, акцентує на тому, що матеріальне та психічне підлягають дії єдиної системи причиновості, яка, проте, діє у випадку кожного з них по-різному. Якщо світ, або природа, не множитья, а лише по-різному себе презентує у сприйняттях, тоді відмінність психічного від фізичного є лише питанням розходжень у методологіях опису (такий підхід нагадує відому нам двоаспектну теорію). Сутності, що підлягають описові, є

воднораз і фізичними, і нефізичними, себто ступінь їх взаємозв'язку надзвичайно високий, тож і казати про істотну відмінність між ними не варто.

Розглядаючи психофізичну проблему, можна вказати на два варіанти її розв'язання з позиції філософського монізму. Найбільш примітними сьогодні типами моністичних теорій є вже відомий нам фізикалізм або ідеалістичний панпсихізм. Спільною для них позицією є твердження: будь-яке явище має одне і тільки одне джерело, наприклад, мозок або духовну субстанцію, яка генерує діяльність ума. Усе суще є або матерія в її русі або, навпаки, дух у його проєкції на світ. Ці підходи є, як було показано, доволі проблематичними і потерпають від зав'язаності на редуccionістській програмі.

7.2. Аномальний монізм Д. Девідсона.

Існує й інший, більш витончений різновид моністичного тлумачення проблеми психічного. Його запропонував у своїй книзі «Нариси про дії та події» (1980) американський філософ **Дональд Девідсон**. Попри те, що він був учнем видатного логіка і фізикаліста Віларда ван Ормана Квайна (Quine), Девідсон запропонував концепцію *аномального монізму*. Свою ідею він виклав у вигляді довго аргументу:

1. Будь-яка взаємодія ментального та фізичного не може бути випадковою; вона є необхідною, адже взаємодія передбачає постійність.
2. Щоб пояснити взаємодію, потрібно знати її причини. Знання причин тотожне знанню умов необхідності.
3. Але знання причин впливає з визнання законів, згідно з якими А взаємодіє з В.
4. Між тим, психофізична взаємодія не виконується за законами, бо жодних «психофізичних законів» встановити не вдається.
5. Прогнозувати ментальну подію на основі фізичної неможливо.
6. Отже, психофізична взаємодія є аномальною; так само аномальною є дія причин, які призводять до ментальних подій.

Свій підхід Девідсон протиставляє *аномальному дуалізму*, започаткованому Рене Декартом. Останній розробив вчення, в якому з очевидністю, але без належних підстав (себто аномально) поєднувалися дії двох відмінних субстанцій, мисленної та протяжної. Хоча Декарт твердив, що їхній статус є рівноправним, ментальне не залежить від фізичного; якимось дивним чином ці субстанції взаємодіють. Нагадаємо, що Декарт вбачив суто тілесну причину їхнього відношення, і локалізував її у шишкоподібній залозі (епіфізі). Такий спосіб розв'язку проблеми виглядав незадовільним навіть для його учнів – мало вказати на певне місце, потрібно довести, чому взаємодія відбувається саме там. Єдиним виходом було наполягати, що означена взаємодія є необхідною і (sic!) непояснюваною. Отже, картезіанський субстанціалізм пропонував *аномальне* рішення для психофізичної єдності.

Для аномального моніста (АМ), такого як Девідсон, головним завданням є довести, що поняття «причина» є цілком застосовним до сфери ментального, а її аномальність є некатастрофічною для науки. Логіка міркування моніста Девідсона назагал є чіткою. Ми пізнаємо речі як певні цілості. Якби було інакше, наше пізнання розпалося б на фрагменти й не мало обов'язкового характеру для всіх суб'єктів. Може здатися, що всесвіт складається виключно з фізичних і ментальних сутностей та їх комбінацій. Але якби це справді було так, тоді довелося б виснувати, що всесвіт є або дуальним або плюральною, натомість він є унітарним, хоча це й неочевидно. Тож справа філософа полягає у з'ясуванні того, *як* це можливо. Робочу гіпотезу виражає теза: має існувати дещо, яке опосередковує фізичне у взаємодії з нефізичним. Як припускає Девідсон, цим дещо є *комплексність дієвих причин*. Отже, всесвіт є єдиним саме тому, що в ньому є чинними такі причинові дії, коли фізичне впливає на нефізичне *і навпаки*. Відповідно, раціональною основою світу є причиновість, і саме її слід аналізувати, коли йдеться зокрема про психофізичну єдність суцього.

Звідси виходить, що якщо відбулася певна ментальна подія M_1 , її причину слід шукати у фізичній події Φ_1 . За аналогією, якщо відбулася фізична подія,

її причиною скорше є нефізична подія, що їй передувала. Головним постулатом АМ є визнання того, що випадковість у цих конфігураціях варто взагалі виключити, адже тоді система пояснення виявиться не просто хиткою, а й бездоказовою. Утім, прийняття за основу дії причиновості вимагає обґрунтування основоположних принципів таких дій. Один із них ми точно знаємо – це принцип каузальності. Усі інші, як можна здогадатися, є похідними від нього.

Авжеж, каузальність фізичного має ґрунтуватися на законі регулярно-повторюваних дій. І якраз тут постає утруднення, оскільки каузальність у стосунку до класу ментальних подій не допускає строгої регулярності. Принаймні, її вкрай важко встановити і тим паче прогнозувати. Тому, якщо каузальність ментального відхиляється від каузальності фізичного, нам доводиться ствердити *аномальний* статус каузальності ментального. Девідсон прямо підтверджує цей факт міркуванням: «Не існує строгих законів, на основі яких ментальні події можуть провіщувати або пояснювати – або бути провіщеними чи поясненими – іншими подіями». Цим самим думка опиняється у замкненому колі: ментальна причина існує, але з'ясувати її регулярність ми не в змозі.

Прикладом того, як ментальне спричинює настання фізичної події є випадок морської битви. До речі, цей кейс сам філософ залучає не спроста – під час Другої світової війни він служив у ВМС США на Тихому океані. Отож, уявімо собі, що в морі знаходиться військовий корабель, за яким полює ворожий підводний човен. Корабель і човен є фізичними об'єктами, натомість ухвалені їхніми екіпажами рішення утворюють клас нефізичних об'єктів. Припустімо, командир субмарини першим помічає рух корабля супротивника і вирішує затопити його шляхом торпедної атаки. Намір командира, його рішення є ментальною причиною враження судна, тоді як фізичними причинами є пуск торпеди, її влучання у ціль, подальше затоплення корабля і т. п. Отож, хоча прямою причиною слід визнати фізичну, проте в разі невчасного розпізнання цілі, неточного розрахунку торпедного пуску та

незлагодженості екіпажу пряму причину задіяти не вдасться. Виходить, що без ментальної причини проста фізична причина нічого не спричинить.

Важливим внеском Девідсона є не тільки формальна класифікація типів причин та їхніх дій, а й увага до граматики комплексних подій, наприклад психофізичних, які нас властиво й цікавлять. Скажімо, філософ помічає, що семантика опису ментальних подій послуговується особливим набором слів, т. зв. *інтенційних дієслів*, що окреслюють позицію суб'єкта: «вважаю», «гадаю», «знаю», «припускаю», «маю певність», «довіряю», «хочу», «збираюся» тощо. Такі слова у реченнях зазвичай передують сполучникам «що», «адже», «бо» та ін. Фізичні ж події зазвичай позначено референтними іменниками. Отже, структура наших речень розкриває зв'язок між намірами та соціальними конвенціями, в які їх вписано, і настанням наслідків. Разом із тим, АМ не стане аналізувати неінтенційні стани, адже вказаного зв'язку виявити не зможе.

Розгляньмо дві ситуації. Ситуація 1. Андрій, повертаючись додому, чує шум за дверима своєї квартири. Він (а) *вмикає* світло у вітальні, бо (б) *переконаний*, що (в) *досягне* цим певної мети – буде з'ясовано, що відбувається всередині (г). Ввімкнення світла (г) *застає* зловмисника Петра на гарячому.

Ситуація 2. Петро є злочинцем, який (а) *збирається* пограбувати чужу оселю. Він (б) *знає*, що діяти слід акуратно і швидко. Тому Петро (в) *не вмикає* світло, коли нишпорить приміщенням; він (г) *квапиться*, намагаючись якнайшвидше зробити свою справу і зникнути. Поява Андрія (г) *не дає* йому *досягти* поставленої мети.

Як бачимо, в обох випадках комплексний характер дій кожного з дієвців підпорядковується прийнятим інтенціям і передбачає поступальність актів, утворених поєднанням головної та супровідних причин. Кожен із дієвців вірить у зв'язок між наміром і дією, й не розпочинає дію, не продумавши послідовності її реалізації.

Необхідно зазначити, АМ не заперечує теорію тотожності. Девідсон заявляє: «Ментальні події тотожні фізичним подіям», оскільки «жорсткі причиново-наслідкові зв'язки мають чинність для фізичних явищ», а для пояснення ментальних явищ достатньо зважати на думки, наміри, мотиви і стратегії самих акторів. Отож ментальне ми відрізняємо від фізичного тільки на фоні фізичного і ніяк інакше. У такому вигляді моністична позиція є радше корекцією ТТ, а не її переглядом. З іншого боку, якщо фон є настільки значливим, тоді опис певної події або явища у термінах іншої події або явища не є суперечністю. Ментальні події можна описувати *синоптично* – і як ментальні, і як фізичні. Тож питання про те, що нас скеровує – ум чи мозок – аномальний моніст розв'язує у передбачуваний спосіб: ментальні події є аномальним різновидом подій фізичних з тим лише уточненням, що *«якщо дві події мають спільні фізичні властивості (спільний простір і спільний час [їх настання], тоді вони матимуть і спільні ментальні властивості»*.

7.3. Аномальний монізм і проблема свободи волі. Експеримент Лібета.

Рівнозначність двох типів подій або явищ ставить перед моністом нове питання, від якого годі ухилитися: як бути зі свободою? Чи є вона лише переписуванням фізичної каузальності у термінах імовірностей певних вчинків, чи ж, навпаки, це є серія дій без добре прогнозованої регулярності, що залишає за її виконавцем право на спонтанність? Звісно, моніст порушить стрункність свого підходу, якщо пропонуватиме позитивну відповідь на другу частину запитання.

Детермінізм і свобода у рамках АМ повинні мати узгодженість. Існує щонайменше три варіанти вирішення цього утруднення. По-перше, обґрунтування тези про логічну симетрію між фізичним і ментальним допускає свободу як варіативність каузальних впливів, як таку собі точку біфуркації, в якій можливі кілька варіантів, тож вибір одного з них уже не є повною детермінованістю. По-друге, непередбачуваність наслідків навіть добре спланованих дій так само можна вважати проявом свободи, оскільки

жорстка визначуваність спонукає ухилятися від тих вчинків, яких ми прагнемо, себто нав'язує суб'єктові помірну пасивність. Нарешті, суб'єкт втрачає повноцінну суб'єктність, коли обмежує себе тим, що може виграти з абсолютною певністю. Крім того, взаємне зумовлення фізичних і ментальних явищ попросту знімає гостроту питання про несвободу, адже якщо детермінація веде до свободи, свобода продукує власну каузальність – і проблема перестає бути нагальною, залишаючись у вимірі чистої спекулятивності.

З окреслених варіантів напрошується висновок, що в кожному разі підхід АМ тяжіє до *компатибілізму*: різке протиставлення свободи й детермінізму виглядає як концептуально недоцільне. А отже, свобода волі у суб'єкта наявна і водночас відсутня. Трудність у тому, щоб показати неповноту вільного рішення. Якщо свобода волі (СВ) є всього лише конструктом, взятим на озброєння менталістами, то їм доводиться мати справу не з сутністю явища, а з логікою його розгортання у часі, тому й проблему вони аналізують лише фрагментарно. Якщо СВ є елементом системи раціональності, тоді дії, вписані в неї, будуть кривотлумаченнями аналізованого поняття, бо нерідко акти СВ суперечать положенням раціонального дієвця (згадаймо, як часто люди воліють чинити всупереч здоровому глуздові або піддавшись емоційному пориванню). Якщо ж пристати на позицію АМ, тоді проблема саме *як проблема* зникає: СВ є аналогом необхідності у соціальному просторі, нетотожному тиску природних факторів. Як зауважив свого часу французький філософ Реймон Арон, у суспільстві ніколи не діє конкретна причина – її місце займає полікаузальність. Така оптика не суперечить ідеї Девідсона про принцип єдності у поясненні світу і водночас не позбавляє суб'єкта права привласнювати результати своїх зусиль.

На мій погляд, добрим прикладом того, як моністичний підхід користується з методологічної двозначності СВ, є інтерпретація результатів відомого експерименту **Бенджаміна Лібета** (Libet). 1983 року цей нейробіолог з Каліфорнійського університету вирішив встановити, яким є зв'язок між

центрами моторики у головному мозку та фактично виконаними рухами. Лібет переймався не тим, щоб установити пряму залежність між ними (про неї науці вже давно було відомо), а тим, щоб з'ясувати, чи в змозі особа ухилитися від посланого імпульсу або навіть блокувати його.

Отож питання набувало нового ракурсу: якщо усвідомлення своїх дій є первинною здатністю особи, тоді біологічний детермінізм можна ослабити або й узагалі нейтралізувати. За допомогою датчиків організатори контролювали поведінку учасників експерименту. Ті, своєю чергою, спостерігали за появою серії точок на моніторі, після чого мали підняти руку або ж утриматися від руху. Робочу гіпотезу було сформульовано так: якщо пристрій фіксує імпульс у відповідній ділянці кори головного мозку раніше за скорочення м'язів учасника дослід, тоді свідому дію можна вважати «запізненою реакцією» на даний мозком припис. Отже, рішення поворушити кінцівкою буде сприйматися особою як ініційоване нею, хоча таким насправді не є. Як показав дослід, активізація відповідальної за моторику ділянки мозку на 350 мс передує не тільки виконанню піддослідним дії, а й активізації ділянки, відповідальної за усвідомлення дій.

У підсумку, ученому довелося визнати, що свідомість і контроль за своїми діями є нічим іншим, як химерою. Однак проведені кілька десятиліть тому колегами Лібета уточнювальні експерименти з'ясували, що різниця між фіксацією імпульсів у різних ділянках мозку вказує не стільки на алгоритмічний характер наших дій, скільки на те, що готовність діяти (за яку відповідає одна частина мозку) і безпосередній початок дії (за яку відповідає інша його частина) вимагають координації, здійснення якої неможливе без участі суб'єкта, який себе усвідомлює. В іншому разі залишається незрозумілим, чому деякі піддослідні (хоч і з мілісекундним запізненням) утримувалися від прогнозованих нейродатчиками дій. Тобто казати про відсутність СВ неможливо саме через те, що люфт між автоматизмом і добровільно виконаним актом є предметом інтерпретації.

Отже, компатибілістське рішення дає можливість задовільно пояснити складність взаємодії між природною каузальністю та свободою волі. Водночас у логіці філософського монізму такий підхід допомагає уникнути хибної дилеми «або детермінізм, або вільне рішення», вказуючи на те, що реально відмінність між ними є лише питанням розуміння порядку актуальних причин.

7.4. Елімінативізм і розв'язання ним проблеми свідомості.

Поняття «елімінативізм» походить від англ. eliminate: усувати, позбуватися чогось. У рамках філософії свідомості його методологічну оптику націлено на ідеї та концепції, запропоновані мислителями, які твердо переконані: розуміння й пояснення неможливі без радикального очищення філософського словника від понять, за якими «нічого не стоїть». Терміни й концепти, які збивають з пантелику, мають бути еліміновані з ужитку раз і назавжди, аби філософія підтвердила претензії на науковість. Такими термінами елімінативісти вважають комплекс «метафізичних понять» (свідомість, воля, душа, дух, різні модальні поняття на кшталт «необхідності», «можливості», «випадковості»), з якими не впоратися інакше, як шляхом їх радикального табування. Найвідомішими представниками елімінативізму варто назвати **Пола і Патрицію Чорчлендів** (Churchland), **Пола Феєрабенда** (Feyerabend), **Ричарда Рорті** (Rorty), очевидний захват елімінативістською програмою демонструє й **Деніел Денет**. В історії філософії існували прецеденти, подібні до елімінативізму, – пригадаймо хоча б Г'юмову критику самості як фікції, утвореної динамікою перцепцій.

Зважаючи на сказане, не слід дивуватися, що елімінативізм є редукаціоністським матеріалістичним ученням. Його послідовники поділяють два головні принципи: методологічний – відмова від семантично амбівалентних понять полегшує пояснення психофізичних явищ, і теоретичний – саме пояснення має ґрунтуватися на довірі до природничих наук, особливо нейробіології. Фактично, елімінативізм стає ревізованою

версією теорії тотожності, доповненою критерієм концептуальної надлишковості: те, без чого може обійтися вчений, є й для філософа предметом елімінації. Приміром, поняття «дух» або «ум» (англійською їх позначає полісемічне mind) не додає нічого змістовно нового до відомих теорій нейрофізіології, отже, є надлишковим; будучи надлишковим, це поняття скеровує не на пошук відповідей, а до парадоксів і утруднень. Якщо разом із цим та подібними поняттями доводиться відмовитися від цілих доктрин, тоді це слід робити рішуче. Як можна помітити, у цій тезі аргумент словника віродився з новим завзяттям.

Названий спадком «простонародної психології» (folk-psychology), ужиток згаданих слів витлумачується елімінативними як, з одного боку, результат проникнення метафізичних спекуляцій у справу створення позитивної науки, а з другого, як закономірне потурання «темному словнику» повсякденної мови. Простіше кажучи, без ТТ елімінативізм навряд чи може претендувати на метанауковість своїх висновків, так само без аналітичної критики природної мови він не зміг би адаптувати її на працювання щодо проблеми свідомості.

Узагальнимо модель судження, обстоювану елімінативістською теорією. Зовсім не дивно, що тема квалій стає центральною:

1. Можливість свідомості впливає з необхідності суб'єктивних квалій.
2. Квалії є недовідними.
3. Наявність квалій тотожна наявності суб'єктивності.
4. Отже, суб'єктивності не існує.
5. Щоб існували квалії, слід припустити інтенційність.
6. Іntenційність є недовідною.
7. Отже, не існує ні квалій, ні інтенційності.
8. Отже, не існує свідомості як самостійної сутності.
9. Отже, люди є «пізнавальними зомбі».

Подружжя Чорчлендів спробувало здійснити перевірку ідей елімінативізму на своїх дітях за допомогою радикального педагогічного

експерименту. Як і свого часу батько Мішеля Монтеня, який наказав оточенню спілкуватися з його сином винятково латиною, аби той сформувався як носій римської культури, так і Чорчленди не вживали менталістських понять у повсякденному спілкуванні з дітьми, щоб ті не встигли призвичаїтися до них.

Що можна протиставити аргументам елімінативістів? По-перше, як зауважують їхні опоненти, відмова від проблематичних термінів ще не означає, що «за» цими термінами нічого не стоїть. Проста заміна «ума» або «духа» на інший термін може дати більше для розуміння цієї сутності, ніж її сліпе ігнорування. Якщо припустити, що автономія «сумління» є всього лише вигадкою ідеалістів XVII ст., то хіба це означатиме, що відмова і від поняття, і від позначеної ним сутності зробить яснішою картину моральної філософії? По-друге, не меншим утрудненням є закид щодо *перформативної суперечності*: якби не «псевдо-сутності», потреби в елімінативізмі не було б, а отже, й не виникла б його методологія. З появою цієї філософської течії прибічникам «традиційного словника» доводиться обґрунтовувати легітимність його вжитку. Але (що особливо важливо) самі елімінативісти будують своє вчення, відштовхуючись від ідей та на основі сленгу своїх опонентів. Виходить, що критикуючи «дух» і «психіку», вони вірять у те, що заперечують. По-третє, якщо «простонародна психологія» заповнила нестачу наукових знань, тоді незрозуміло, чому з розвитком науки остання продовжує послуговуватися недосконалим концептуальним апаратом. Його примусове вилучення ризикує потягнути за собою ще більшу фрагментацію науки як єдиної області досліджень.

Зауважмо, що, крім цього, перед елімінативістом постають два концептуальні утруднення. Одне з них стосується пояснення креативних здібностей ментальних суб'єктів. Якщо пристати на тезу, що люди можуть пізнавати матеріальну реальність без усвідомлення своїх внутрішніх станів, себто, що вони є когнітивними «зомбі», це порушує питання: чому ці «зомбі» діють так, ніби вони є свідомими творчими особистостями? Якби пізнання ґрунтувалося на «зомбі»-схемі, воно було б алгоритмічно стандартним для

всіх подібних істот, натомість невдоволеність пізнаним і є рушієм прогресу в науках і навіть у повсякденному житті. Запропонована Денетом відповідь – що творчість є копіювання людиною загальної програми еволюції – мало переконує, адже тягне за собою низку додаткових запитань (про умови точного копіювання; доцільність винайдення того, що не слугує безпосередній меті виживання виду; про місце свободи в цілком детермінованій системі, в якій креативність може призвести до загрозливих для системи наслідків). Ще одне утруднення полягає у неможливості розрізняти явища за відсутності свідомо виконуваних актів. Зрештою, як особа може бути пізнавцем, не пояснивши самій собі, що означає «ще непізнане» і «вже пізнане». У такій ситуації пізнавальний акт як такий не дає готової відповіді. Не ясно й те, чому рефлексію необхідно ув'язати з пізнавальними процедурами, які вилучаються за межі усвідомлення. Якщо таким чином досягається вимога об'єктивності, то чи не занадто дорогою ціною?

Оцінка негативної програми елімінативізму має ґрунтуватися на тому, що заперечення тих чи тих філософських теорій свідомості, як вражених вірусом метафізики, потребує покладання чіткого відношення між метафізичними та неметафізичними припущеннями. Ми вже помітили, що цілком достатній у логіці модальний доказ теж можна звинуватити у метафізичності. Матеріалізм еліміністичного ґатунку, по-перше, не дає визначення матерії та її відношення до психіки, не містить словника понять, які можна було б визнати вільними від спекулятивних конотацій. На цій ґрунті елімінативістам можна закинути імпліцитну метафізичність. З іншого боку, їхня критика «простонародної психології» не спроможна забезпечити альтернативу її словникові, тож і на рівні очищення мови елімінація теж виявляється порожньою декларацією. Більш того, апелювання до сучасної науки не стільки насичує заторкнуті проблеми новим змістом, скільки порушує питання про умови їх адекватної реінтерпретації, враховуючи, що будь-яка філософська дискусія вимагає рівнозначного обміну аргументами.

Отже, елімінативізм є вченням, мета якого – підготувати таку інтерпретацію ментального, в результаті якої проблеми психофізичної взаємодії, квалій і автономії свідомості буде взагалі усунуто за межі аналізу як нерелевантні. З огляду на це, елімінативістська практика являє собою спробу захистити редукціонізм шляхом вилучення об'єкту його критики. Для матеріалістичної доктрини успіх подібного задуму обертається розумінням слабкості, притаманної аргументації редукціонізму.

7.5. Гетероменоменологія Д. Денета.

У своїй книзі «Пояснена свідомість» (1992) американський філософ Деніел Денет (Dennett) формулює ідею гетерофеноменології як альтернативи антиредукціоністським проектам і зокрема дуалізму. Осердя дискусії очікувано займає проблема квалій. Згідно з Денетом, квалії не є реальними; їх породжено силою нашої уяви, нашого «ментального ока», успадкованої від картезіанства візії світу як подвоєної реальності, в якій співіснують дві онтології: «внутрішня» і «зовнішня», сфера матеріального та ментального. У цьому театрі метафори заступають об'єктивні речі, а «те, як здається» (seen as) достовірніше за «те, як воно є» (is). Інформаційні потоки (метафори та їхні скупчення – *наративи*) керують нашим дискурсом. Серія дотичних дискурсів утворює культуру. Однак те, що ми в неї занурені, ще не означає, що ці дискурси є істинними і остаточними. Як і свого часу модерні європейські метафізики, ми ладні вірити їм, але також спроможні поставити під сумнів дуалістичну онтологію.

Денет протиставляє дві моделі опису феноменальних переживань. Так звана *автофеноменологія*, сфокусована на витлумаченні внутрішніх досвідів як унікальних і залежних тільки від структури конкретної самості (self), починає з аналізу ментальних змістів, які самість витворила «самотужки», мобілізувавши свої когнітивні ресурси. Результати цих аналізів вона вважає достовірними і безпомилковими, а доступ до ментальних змістів нічим не опосередкованим. Знання про об'єктивний предметний світ поза нею самість

добуває зі спонтанних власних інтуїцій. Натомість *гетерофеноменологія* (методологічний проєкт Денета) принципово відмовляє у достовірності самоописам, не зіставленим із об'єктивним поясненням у науці, передовсім природознавчій. Такий підхід розглядає суб'єкта як річ поміж речей, яка не має права претендувати на винятковість і цим підтверджує свою підлеглість універсальному закономірному каузальної дії.

Якщо самість визнати унікальним дещо, не залежним від причиновості («свободним актором»), тоді загальні критерії пояснення її актів буде втрачено, а критерії, іманентні самості, виявляться всього лише фікціями. Потрібно з'ясувати, як функціонує ум (психіка) незалежно від того, має його людський чи якийсь інший носій. Отож гетерофеноменологічний аналіз спрямовано на дослідження не свідомості, а ментальних станів, не суб'єктивності, а психічних реакцій істот певного виду на зовнішні впливи. Сказане дає Денету підстави вважає всі мисленні істоти «ментальними автоматами» або феноменальними зомбі. Іронізуючи, він називає таким зомбі й себе самого, наголошуючи, що між філософом, пересічною людиною і теплокровним ссавцем не існує принципової відмінності. Аби довести, що суб'єктивність є хибною категорією, що обтяжує розуміння фактичного стану справ, він застосовує звичний прийом – аргументацію на основі уявного експерименту.

«Аргумент Пенсі-коли» критикує суб'єктивність через руйнацію внутрішньої інтенціоналістської логіки. Нагадаю: інтенційність є спрямованістю свідомості на певний об'єкт (зовнішній чи внутрішній), що має довести неіманентну природу свідомості. Наявність об'єкту, непородженого свідомістю, засвідчує дійсність, а отже, пізнаваність об'єкту. Утім, можливі ментальні стани, в яких їхнім змістам не відповідають жодні фізичні об'єкти, як-от літературні персонажі, дійові особи сновидінь, золоті гори, единороги чи міфічні боги. У такому разі постає потреба пояснити їхній онтологічний статус і походження. Одним із варіантів розв'язання утруднення є теорія абстрагування ірреальних сутностей з відомих реальних сутностей на основі

продуктивної сили нашої уяви (Геракл або Супермен мають більшість ознак, притаманних звичайним людям, хоча й наділені особливими вміннями або силою). Проте Денетова стратегія в цьому питанні є інакшою. Він твердить, що інтенційність є просто націленість актів на будь-що (aboutness), і не слід її центрувати тільки на свідомості. Аргумент, про який ідеться, стосується саме спрямованості на дію, а не усвідомлення її процесовості.

Припустімо, у Панамі автомат з продажу «Пепсі-коли» може приймати лише місцеві монети номіналом 5 сентесімо. Американські центи й панамські монети мають однакову вагу й вироблені з однакового матеріалу – міді. Помістивши монету в отвір автомату, покупець отримує жадану бляшанку, при цьому автомат не розпізнає, яку валюту він проковтнув. Отже, покупці можуть з легкістю обдурити його, що й стається. При цьому філософ покликається на реальний випадок з життя, який змусив панамський уряд відмовитися від використання автоматів з продажу «Пепсі». Але якщо уявити, що рішенням уряду для купівлі бляшанки напою можна послуговуватися або американським центом, або місцевим, чи розумітиме автомат, яку з монет він приймає? Авжеж, ні. З погляду Денета, такий тип дії вказує не на те, що в автомата відсутня інтенційність, а на те, що людина влаштована не краще за цей автомат, а інтенційність не є фундаментом мислення для жодного з інтелектів. Точніше, що т. зв. первинна інтенційність, яку зазвичай приписують людській свідомості, є такою ж ілюзією, як і вторинна інтенційність механічної системи, що реагує на вставлену монету. Якщо машинний розум ніколи не зможе продукувати інтенційні стани попри виконання складних інтелектуальних операцій, то це означає лише одне: інтенційність – це міф, що покликаний пояснити незбагненні зміни фізичних станів. Тож якщо виконання будь-якої символічної взаємодії між об'єктами є успішним, його можна інтерпретувати як інтенційно-осмислене, і хоча це тлумачення теж не є вичерпним, воно є принаймні зручним і менш поняттєво плутаним.

У своїй теорії Денет пропонує три моделі розв'язання загадки свідомості. Перша полягає в аналізі інтенційності, яку він вважає не характеристикою свідомості, а універсальною властивістю будь-яким видам сутностей. Для Денета інтенційність позначає проблему ментальних змістів, для яких покладання власника не є обов'язковим. Друга модель фокусується на руйнуванні доткрини подвоєного світу. Нарешті, третя модель спирається на *модель множинних начерків* (multiple drafts model; далі – ММН), що ігнорує потребу у «внутрішньому оці» і не вимагає усвідомлюваних внутрішніх змістів. Заперечення менталістського театру «потойбіч сцени» походить з критики Денетом методу інтроспекції.

Чим же є ММН і яка її роль у поясненні цих станів? Вона є ні чим іншим, як неієрархічною множиною дієвців (agents), кожен із яких має профільну спеціалізацію – діяти так-то і так-то за таких-то обставин. По суті, через залучення цієї оптики філософ прагне довести, що самість є постійним переходом від однієї агентної диспозиції до іншої агентної диспозиції. ММН виконує функцію протиставлення метафізичній ідеї про незмінну самототожність пізнавального суб'єкта. Провадячи опис у форматі ММН, як вважає Денет, ми виграємо двічі: по-перше, досягаємо нейтральності, тобто верифікації наявних в описі значень; по-друге, не розділяємо методи природничо-наукової та феноменологічної традицій. Саме лише природознавство неспроможне вхопити ментальні змісти; сама лише феноменологія (у термінах Денета, *автофеноменологія*) нехтує зовнішнім світом і далі стикається з необхідністю його реконституювання. Кожен із цих підходів не є домінантним, натомість у рамках ММН переваги суми обидвох є явними.

Однак основу Денетової теорії становить оригінальне тлумачення інтенційності. Денетова теза намагається показати, що феноменологічне бачення інтенційності як виняткової ознаки свідомості містить небезпеку циркулярного доказу – регресу в нескінченність. Адже якщо припустити, що я випиваю каву тому, що реалізую своє бажання, мені не вдасться у поясненні

своєї дії відокремити причину від наслідку, і це буде різновидом прихованого редукціонізму, а це неприпустимо для пояснення ментального. Іншим варіантом редукції є пошук причини явища в попередніх причинах або у ментальних станах, що передували актуальному станові.

Виходячи з цього, мислитель розрізняє три типи точок зору (stances): проєктивну (design stance), коли дія є фінальним кроком здійснення певного задуму, фізичну (у ній дію аналізують з огляду на притаманні системі прогнозовані закони на основі імплікації «якщо..., то...») та інтенційну точку зору (далі – ІТЗ), для якої прогнозування неможливе через невідомість початкових законів. У випадку ІТЗ той, хто судить про об'єкт, вимушений приписувати йому мінімально прийнятні властивості. Усі три точки зору можуть перетинатися у стосунку до одного об'єкта на різних етапах судження. Клас об'єктів, поведінку яких можна прогнозувати на ґрунті ІТЗ, Денет називає *інтенційними системами*. Скажімо, камінь, що лежить на тротуарі, можна описати послуговуючись фізичною точкою зору, а вже камінь, який падає з гори і летить у певному напрямку (чи, приміром, робота, який у комунікації з розробником дотримується певного алгоритму), – за допомогою інтенційної.

Уявімо собі дорожній рух тисяч автівок у мегаполісі. Якби ми опинилися на місці зовнішнього спостерігача і не знали про існування правил, що упорядковує цей рух, нам навряд чи вдалося б пояснити, чому автівки не таранять одна одну. Припущення про наявну впорядкованість і регулярність подібного руху не впливає ні з чергування світла на світлофорі (на одних перехрестях він є, на інших відсутній, десь працює, десь зламаний), ні з траєкторії доріг. Тож доводиться керуватися гіпотезою, що пояснення не має зовнішнього критерію, а залежить від прийнятої всіма водіями точки зору. Застосування ІТЗ, на думку Денета, демонструє, що мислення оперує реальними схемами (real patterns), і тому інтенційність є такою схемою, щоправда гіпотетичною і важкопояснюваною.

Критика інтроспекції означає відмову від інтуїтивної очевидності, тому Денет пориває і з очевидним, і з суб'єктивно безпосереднім: самі лише судження про індивідуальний досвід не дають достатніх підстав для витлумачення картини реальності. Опис феноменальних змістів потребує їх розгляду з перспективи не першої, а третьої особи. Це можливо тільки за умови, що самоописи перетлумачуватимуться у тексті згідно зі стандартами ІТЗ. При цьому висновку про онтологічну достовірність психічного взагалі слід уникати: усю увагу необхідно сфокусувати тільки на самих ментальних змістах. Цю настанову Денет докладно викладає у праці «Зміст і свідомість» (Content and Consciousness, 1986).

Поза сумнівом, Денетів проєкт *натуралізує* феноменологію. На це вказує вже сама назва гетерофеноменології. Раціональний дієвець складається з великої кількості міні-самостей (підсистем), де кожна зайнята розпізнаванням певного набору перцепцій. Для наочності скористаємося метафорою видавничого колективу, який теж функціонує як організована система. У видавництві є коректор, літературний редактор, дизайнер та інші співробітники, що працюють з одним потоком інформаційним повідомлень (текстом), а узагальнює плоди їхніх зусиль головний редактор (підсистема рефлексії), який вносить правки у фінальну верстку книжки. Сприйняття не існують осібно й ізольовано, а проте, вони нетотожні й сумі всіх перцепцій. Аналогія великого тексту (наративу) з дрібнішими текстами допомагає осягнути поняття інформації, змісту та формальної структури без необхідності поступитися інтуїтивній очевидності плинних образів. Сенс ментального, згідно з Денетом, відповідає логіці *віртуальної машини*, такої як комп'ютерна програма. У ній software напряду не залежить від hardware, а функціонує за правилами, що рубрикують участь hardware у системі.

Віртуальна машина реалізує *множину тимчасових закономірностей* у вигляді інструкцій (мемів), що схиляють машину поводитися у певних ситуаціях за заздалегідь відомим взірцем з метою операціоналізації її поточних проблем. Наявність мемів не потребує шкіпера (свідомості), який

провадить керування судном з інтернаціональним екіпажем. Робота машини як самоорганізованої системи нагадує рух на автопілоті. Самокерування заміщує необхідність субстанційної форми свідомості. Отож існують ментальні стани, а не свідомість, універсальна, а не психологічно специфічна реальність.

Але лишається питання про джерело балансу, коли система виходить з прогнозованого руху. Таку рівновагу, як переконаний Денет, їй забезпечують завжди доступні ресурси культури (накопичені століттями ідеї, теорії, книги, зображення, лексичні форми висловлення тощо). Вплив культури на нас ніколи не припиняється через нашу залученість у процес її відтворення, як і через те, що ММН живиться з актуалізованих тим чи тим контекстом культурних форм. Наприклад, наші «сни» є не сновидіннями у буквальному сенсі слова, а своєрідними «аплікаціями», створеними з культурного архіву раніше описаних (і розказаних) сновидінь. Себто коли особа повідомляє про те, що їй снилося дещо, вона, не усвідомлюючи цього, підтверджує нарративний характер образів. Отже, сни є всього лише ілюзії, санкціоновані певною культурною традицією. Зважаючи на сказане, неможливо (а згідно з Денетом, і не потрібно) протиставляти свідомі ментальні змісти несвідомим або ще-неусвідомленим, оскільки такий хід абсолютизує умовність.

7. 6. Квалії у суперечці інтерналістів і екстерналістів.

Але чи здатне вилучення свідомості як фікції гарантувати адекватне пояснення суб'єктивної основи наших дій? Щоб відповісти на це питання, звернімося до інтерпретації Денетом природи *квалій*. Його розуміння цієї проблематики ґрунтується на реляційній моделі. Вона стверджує, що кольори або запахи існують не як притаманні суб'єктові властивості, а як елементи міжвидової біологічної співпраці, зумовленої перипетіями еволюційного процесу. Тому й казати про квалії можна хіба що як про універсалії функціонування, а не як про «ідеї» чи «абстраговані образи». Відношення між мозком сприймача і предметним світом утворює простір для можливої появи

кваліативних ефектів. У згоді з ММН Денет розкриває суть цих феноменів. Квалії є реальними (1) диспозитивними (2) схемами (3), включеними у систему (4), опосередковану різноманітними (5) моделями судження про світ. Інакше кажучи, і заздалегідь визначити поняття квалій суто формально, і спростувати їхню можливість на підставі їх нефізичного походження, не вдається до настання кваліативної ситуації. У постулюванні автономності квалій Денет, сумлінний учень Гілберта Райла, вбачає категорійну помилку. І так само, як Райл, він замість квалій каже про природні чи набуті *диспозиції* до дії, які навіяні нам станами мозку, що спричинюють певні реакції тіла. Схема виглядає так: об'єкти світу каузально впливають на мозок (через електромагнітне поле, світло, звукові хвилі), а той активізує схильність реагувати залежно від об'єкта. Якщо скористатися лінгвістичною моделлю Денетового вчителя, первинну каузальність легко віднайти у головному реченні складного судження, а вторинну – у підрядному реченні.

Випадок із життя самого філософа добре ілюструє проблему абстрактності квалій. Якось у ресторані одного університетського містечка сиділи колеги-філософи Ден Денет, Вілфريد Селарс і Сидні Шумейкер. Коли офіціант розлив вино у келихи, Селарс зробивши ковток і залишившись задоволеним його якістю, мовив: «Знаєш, Дене, квалії – це те, заради чого варто жити!» Шумейкер додав: «Якщо для мене є важливим не суб'єктивне відчуття, а просто якість вина, мені не треба відкорковувати пляшку – достатньо прочитати інформацію на етикетці». Денет резюмував діалог: «І справді, розвивати смакові відчуття, закладені природою, нас примушує суспільство».

Міркуючи про квалії, Денет вдається до аргументації екстерналістів. Вони переконані, що будь-які наші враження або відчуття є реакцією на зовнішні імпульси, натомість інтерналісти відмовляються зводити внутрішні змісти до зовнішніх. Якщо уявити полеміку між *екстерналістом* та *інтерналістом* про абстрактні сутності, то вона, либонь, виглядала би так:

Екстерналіст: Усі сприйняття є сприйняттями зовнішнього. Спогади і фантазування є так само сприйняттями, але в них зовнішнє – це протиставлення актуального й минулого сприйнятів.

Інтерналіст: У такому разі, якою є природа наукового знання?

Екстерналіст: Як і будь-яка обізнаність, вона продукується роботою мозку.

Інтерналіст: Отже, для того, щоб школяр навчився фізиці, він має самотужки логічно виснувати її закони?

Екстерналіст: Так, шляхом тренування свого розуму.

Інтерналіст: А якщо він зазнає невдачі у цьому вишколі і просто приймає закони Ньютона на віру, то він не «знає» їх?

Екстерналіст: Авжеж. Він приймає їх на віру і так само може позбутися їх.

Інтерналіст: Тоді де, на Вашу думку, перебувають наукові закони, принципи логіки, математичні аксіоми, філософські ідеї – у головах чи якійсь «вищій сфері»? Адже зі смертю своїх авторів вони не втрачають чинності.

Екстерналіст: Імовірно, позаяк ці сутності не зникають назавжди, вони мають перебувати водночас і у зовнішній, і у внутрішній царинах, якщо називати «внутрішнім» щось, що не може існувати без попереднього впливу зовнішньої реальності. Але мушу наголосити: я не обстоюю епіфеноменалізм. Я лише кажу, що ментальне є частиною фізичного, хоча й не показує це явно. Думки перебувають у нашому мозку, який ми не можемо побачити, коли міркуємо, але просторове походження структури якого уможлиблює будь-який ментальний акт. Авжеж, неможливість роздивитися інстанцію мислення плекає ілюзію її повної невидимості – ідею привида в машині.

Онтологічний статус квалій, на переконання Денета, можна підважити звернувшись до уявного експерименту *«Суспільство втом»*. Припустімо, десь у Всесвіті існує суспільство, в якому метафори домінують над

буквальними значеннями потужніше, ніж у людському. У ньому розмовляють англійською мовою, проте окремі ідіоми й вирази є гіпостазованими сутностями. Візьмімо для прикладу хоча б слово «втома». Мешканці цієї планети розрізняють безліч видів «втом» – зорові, слухові, бігові втоми і т. п. З експедицією до цього суспільства прибувають дослідники з Землі, яких вражає частота ужитку слова «втома». Вони запитують місцевих: «Чим, на вашу думку, є втома *сама по собі*?» На що отримують дивну відповідь: «Щоб збагнути, що таке, приміром, «бігова втома», вам доведеться оббігти навкруг будинку 20 разів». Отже, твердження представників цього суспільства можна визнати істинними попри те, що вони апелюють до *ілюзорної онтології*, оскільки «втоми» не є самостійними сутностями поза «бігом», «слуханням» тощо. В описаній «культурі втом» так говорити зручно, носіїв такого нарративу годі переконати у протилежному. Порівняймо гіпостазування втом із обґрунтуванням землянами онтологічної реальності квалій. Вважати онтологію втом менш реальною за онтологію квалій означає недобачати їхню схожість: в обох випадках маємо гіпостазування переживань, що є відчуваними і вербалізованими, проте нереальними фізично. Як каже Денет, його опоненти-інтерналісти і є живим втіленням ілюзорної онтології. Поміж тим, завдання полягає в окресленні перспективи «ні тут, ні там», ні ззовні, ні всередині, униканні пасток як теорії приватності відчуттів, так і бігевіористської стимул-реактивної теорії.

Більш відомою є критика Денетом аргументу знання (див. 4.2) у дискусії про квалії. Нейрофізіолог Мері знає все про зір, але коли виходить із замкненого монохромного простору, вона отримує доступ до нового знання про середовище, приміром, зауважує досі невідомі кольори. Чи виправдано стверджувати, що приріст знання змінює розуміння нею пізнаного? Змодельована історія з РобоМері має на меті довести, що для когнітивної системи неможливого не існує бодай тому, що каузальність не є єдиним фактором продукування знання. Істоти з кваліативними змістами, набутими некаузально шляхом самоафектації, є контрпроєкціями персонажів культури

на кшталт Голмса. Філософ має повне право їх не брати до розгляду. Як ми вже помітили, Денет тяжіє до екстерналізму, але не вважає його фінальною точкою у дискусії про квалії.

То чим же є самість або свідомість? Стисла відповідь гетерофеноменолога Денета звучатиме так: вона є «центром тяжіння наративів», себто культурним конструктом, необхідним для: 1) розрізнення своїх і чужих переживань, 2) виконання ролей у соціальній взаємодії, 3) планування власних дій з освоєння довкілля. Простіше кажучи, коли філософи-менталісти ведуть про самість як про осердя автентичного існування, Денет ладен говорити про множинність екстрапольованих на свої досвіди стандартних описів, або (запозичивши лексику континентальної філософії) живу інтертекстуальність, впізнавану завдяки наявності тіла. Отже, якщо наша самість не здатна залишатися унітарною і не є цілком уніфікованою, то й обговорювати умови її єдності безглуздо.

Резюмуймо закиди, до яких провокує критиків Денетів філософський проєкт. По-перше, якщо меми й культурні схеми спроможні продукувати адаптивну, але примарну картину світу, в якій те, що не існує («Я»), скидається на свою протилежність, тоді чи не є логічно послідовним і саму культуру визнати фікцією, а її можливості химерними?

По-друге, якщо засвоєні наративи призводять до прийняття нами свого «Я» як безумовно реального, тоді навіщо знадобилося еволюції та соціальному прогресу рухатися вперед крізь тенета самообману? Схоже, це пояснення не менш туманне за попереднє.

По-третє, якщо все, що пропонує нам резервуар культури, тягне за собою лише корисні ілюзії, як можливо обґрунтувати необхідність творчості на рівні онто- та філогенетичному рівнях? У кожному разі, виникає враження, що «генетична здатність моментально схоплювати всю логіку еволюційних змін» (як Денет визначає творчого індивіда), суперечить неієрархічності множинних варіантів, які нам пропонує біологічна еволюція. Отож, слід запитати, чи є щось геніальне у репродукуванні вже закладеного в закономірностях

поступального розвитку?! Це питання прямо виводить на представлену Денетом пізнавальну модель: геній і творчість відсилають до специфічної сутності суб'єкта, якого він вважає вигадкою філософів. І якщо творчість усе ж можлива, то вона передбачатиме само-творення, а не реалізацію схем, типових для цілого виду. Виходить, що самість є анонімною (а це парадоксально), культура є абстракцією (як «онтологія втоми»), а суб'єктивності не існує саме через те, що відсутня інстанція кінцевого рішення.

І якщо це справді так, то, по-четверте, гетерофеноменологічні інтерпретатори ментальних феноменів повинні зізнатися, що не розрізняють детермінізм і свободу волі. Навіть за умови, що поняття «свобода волі» обрамлене історичним контекстом її появи, повсякденна практика демонструє протилежні висновки.

По-п'яте, у системі, яку ми за звичкою звемо суб'єктивністю, ми збираємо докупи різні схеми культурних наративів, аби мінливість самого досвіду видавалася цілісною. Ми маємо безліч наративних самостей («Я» дитинства, «Я» соціального профілю, «Я» моїх хобі, «Я», яким мене знають рідні тощо), але потрібне прояснення того, що визначає перевагу одних щодо інших і чим пояснюється зникнення деяких із них назавжди? Крім того, що носій цих ідентичностей сам зазнає змін у процесі їх реалізації, постає чимало питань і стосовно деформованих і все ж функціонуючих самостей (самості психічно хворих). Яким є їхній статус: чи є вони є хибними схемами реалізації еволюції? За логікою Денета, розмаїття самостей є якісно однаковим, але життя доводить, що це не так.

По-шосте, якщо квалії потрібні для ідентифікації об'єктів у природі, тоді немає задовільного пояснення того, чому деякі види (птахи, риби, комахи) мають кольорове бачення, а деякі розвинені види (наприклад, коти й собаки) їх позбавлені.

По-сьоме, навіщо культура створює ілюзії, які потрібно деконструювати? Аби користатися з надбань попередніх поколінь, ми мусимо оперувати її

метафорами, хоча вони виявляються для ґрунтового розуміння себе й світу не більш ніж обхідними шляхами.

І нарешті, якщо еволюція є раціональною, то з якою метою вона створила клас не(вповні)раціональних істот, які ладні сумніватися у своїх інтелектуальних витворах і, як наслідок, постійно дурити самих себе?!.. Так чи так, попри свою прозорість і позірну стрункість, натуралізована версія феноменології мимохіть порушує більше запитань, ніж дає відповідей.

Джерела:

Churchland, P.S. (2007). The Evolving Fortunes of Eliminative Materialism. In *Contemporary Debates in the Philosophy of Mind*. B.P. MacLaughlin & J. Cohen (eds.). Blackwell Publishing, pp. 160–181.

Davidson, D. (2001). Mental Events, In Davidson D. *Essays on Actions and Events*. Oxford: Clarendon Press, pp. 207–215.

Dennett, D. (1992) *Consciousness Explained*. Back Bay Books, pp. 36–57, 81–115, 384–445.

Dennett, D. (1988) Evolution, Error and Intentionality. In Y. Wilks and D. Partridge, (eds) *Sourcebook on the Foundations of Artificial Intelligence*. <https://web-archive.southampton.ac.uk/cogprints.org/431/1/evolerr.htm>

Деннет Д. (2014). Самість як центр ваги наративу. *Антологія сучасної аналітичної філософії, або Жук залишає коробку*. Л.: Літопис, сс. 280–296.

Черчленд П. С. (2017). Яким чином нейрони знають? *Антологія сучасної філософії науки, або усмішка ASIMO*. – Л.: ЛНУ ім. І. Франка, сс. 453 – 468.

Запитання для самоперевірки:

1. Що таке «ментальна причиновість»? Чи слід її вважати інверсією природної каузальності?
2. Спробуйте сформулювати «сильну» й «слабку» версії аргументу про властиву нашій свідомості ментальну причиновість?

3. В чому полягає основне ускладнення доктрини про ментальну причиновість?
4. Де пролягають точки згоди й незгоди Д. Девідсона з матеріалістами й функціоналістами в питанні про свідомість?
5. Як пов'язані ментальна каузальність і проблема свободи волі? Чи не є теза про ментальну каузальність прихованим руйнуванням підвалин нашої свободи?
6. До яких наслідків для філософських теорій і для соціального світу призводить підхід елімінативізму?
7. Уявіть, що ви – Джон Сьорл. Як ви відповіли б на аргументи елімінативістів?
8. Окресліть концептуальну відмінність між інтерналізмом та екстерналізмом?

Лекція 8.

Філософські проблеми штучного інтелекту.

8.1. Машина Тюринга і тест Тюринга. Заперечення проти сильного штучного інтелекту.

Хоча штучний інтелект як теоретична проблема цікавив мислителів, мабуть, іще від часів Ляйбніца, у своєму сучасному формулюванні вона завдячує дослідженням британського математика **Алана Тюринга** (Turing). 1950 року у відомому часописі «Mind» (його редактором тоді був Г. Райл) вийшла його стаття «Обчислювальні машини та інтелект». Автор ставить просте, здавалося б, питання: чи мають машини здатність мислити так, як це робить людський ум? Поява статті збіглася з винайденням першого комп'ютера (на той час це ще були машини, устаткування яких займало декілька кімнат) і стрімким розвитком кібернетики. Автор статті вважає «штучним інтелектом» (artificial intelligence, AI) здатність: а) відрізнити себе від представників свого та інших видів, себто зумовлену біологічною

еволюцією навичку до (само)ідентифікації, і б) здатність упорядковувати свій досвід, себто соціальну навичку взаємодіяти з собі подібними. Отже, вказаний критерій утворено двома основоположними філософськими принципами – розрізненням і ототожненням.

Крім того, запропоновано уявний експеримент – машину Тюринга – який дасть змогу однозначно відповісти на поставлене вище питання. У разі, якщо машина проходить *тест Тюринга* (ТТ*), питання розв'язується позитивно і відмінність між людським і машинним розумом долається. Тест засновано на грі у відгадування: у двох розташованих поруч кімнатах знаходяться чоловік і жінка. Журі має визначити автора, отримуючи роздруковані на тайповій стрічці речення, при цьому вони не знають, де сидить жінка, а де – чоловік. За тим самим принципом, в одній кімнаті слід розмістити людину, а в іншій поставити механічний пристрій, що також набирає би певний варіант тексту. Якщо $\frac{2}{3}$ від складу (або 70% членів) журі не вдасться з'ясувати, написала його людина або машина, тоді тест на штучний інтелект буде вважатися успішно пройденим. Сам Тюринг був переконаний: до 2000 року штучний інтелект стане реальністю, а цифрові комп'ютери навчатися повністю імітувати людський розум.

У своїй статті Тюринг наводить докази проти гіпотези про можливість створення штучного інтелекту. Резюмуємо їх:

1. *Теорема Геделя* про неповноту аксіоматичних систем: програмування машини завжди є формально неповним. Навіть якщо у процесі реалізації завдань машина накопичить надзвичайно великий обсяг інформації, вона не зможе самотужки дописувати частини власного алгоритму.
2. *Аргумент квалій*: машина не спроможна розповісти про суто людські сприйняттєві особливості, приміром, яким є смак полуниці з вершками, тоді як людина завжди знайде відповідні для цього метафори.
3. *Аргумент неформальної поведінки*: машина не може діяти всупереч правилам, приписаним її творцями. Навмисне порушення правил вона, на відміну від людини, сприймає як однозначне порушення своєї роботи.

4. **Аргумент сумніву:** машина не спроможна сумніватися в тому, що вона сприймає як вхідні дані. Навчити машинну систему сумніватися можна лише за умови, що цей сумнів вона використає задля більш ефективної роботи – чистий скепсис їй недоступний, адже руйнує схематику виконання подальших операцій, на яке її й запрограмовано.
5. **Аргумент творчості:** машина інертна у виконанні творчих завдань. Максимум, на що вистачає її ресурсів – більш інтенсивне здійснення відомих людині інтелектуальних операцій. Навіть якщо машину запрограмувати на гру в шахи, винайти новий тип гри вона не зможе, а комбінуватиме вже існуючі типи ігрових стратегій. Вона не в змозі написати сонет або концерт, який би не повторював відомі музичні мотиви.
6. **Аргумент екстрасенсорних здібностей.** У людини вони «є», а машині навіть не вдається їх імітувати, адже подати раціональний опис цих здібностей не можуть навіть самі люди. Без опису немає пояснення, а без пояснення не розробити алгоритм навчання.

Як бачимо, Тюринг задав парадигму опонування ідеї «сильного» (повністю автономного) штучного інтелекту (СШІ), якими його критики послуговуються досі. Щоправда, деякі з цих контрдоказів можна спробувати спростувати. Наприклад, теорему Курта Геделя поборники СШІ заперечують на тій підставі, що, мовляв, нейронні мережі генерують здібність до самонавчання, а отже, попри вписану у машинне функціонування алгоритміку, в ході своєї роботи машина зможе доповнювати первинний сценарій дій новим. І справді, сучасні роботи демонструють дивовижні навички самовдосконалення, а чат-бот GPT 2022 року виконав формальні вимоги ТТ*. Сам Тюринг був переконаний: від людського інтелекту машинний принципово відрізняється лише одним – відсутністю у нього екстрасенсорних здібностей.

Однак аргументи 2–5 досі залишаються неспростованими. Нейромережі тестують на предмет помилок на кожному з етапів навчання, але це не доводить розуміння ними усього процесу (це кожен може перевірити під час переписки з чатом GPT). Якщо людина береться вивчити іноземну мову і після

кожного уроку правильно складає проміжний тест, то це з'ясовує рівень засвоєних нею навичок. Але подібне тренування не показує рівня адаптивної чутливості до перевірки, непередбаченої сумою попередніх тестів, – людині в цьому допомагає інтуїція, якої бракує машині.

1990 року американський вчений-когнітивіст Роберт Френч (Robert M. French) запропонував удосконалити ТТ*, назвавши нову версію *тестом чайки* (the Seagull Test). Мотивом відмовитися від класичного ТТ* було недостатнє розуміння людського мислення в часи, що передували когнітивній революції 1960-х років, а також прогрес у розвитку комп'ютерних моделей. Більш того, критерій успішності проходження ТТ* спирався на інтуїцію людського журі, і цим, на думку Френча, підважував об'єктивність оцінки. Натомість тест чайки пропонує зовсім інший підхід. Уявімо собі, що мешканці одного зі скандинавських островів знають лише один тип істот, що вміють літати, і це – чайки. Якось двоє філософів, що проживали на острові, посперечалися: у чому полягає сутність «польоту»? Зобразимо їхній діалог.

Перший філософ: Сутність польоту – це пересування в повітрі.

Другий філософ: Але хіба ось це не є «політ»? – і кидає камінець з берега в море.

Перший філософ: У такому разі політ – це пересування у повітрі за допомогою крил і пір'я.

Другий філософ: Але хоча пінгвіни мають крила й пір'я, вони не літають. Нарешті, обоє погоджуються, що «вміють літати лише чайки» і це «абсолютно достеменно». Співрозмовники вигадують тест, який може пройти будь-що, «спроможне літати». Далі вони вирішують нести вахту за стереоскопічними радарми, які фіксують пересування у повітрі над островом. На кожному з двох радарів помітні пересування точок, але на одному – це політ чайка, а на іншому – машинний політ. Об'єкт пройде тест, якщо обидва філософи погодяться з тим, що на екрані вони бачать політ чайки. Отже, тест чайки спрямовано на виявлення асоціативного (субкогнітивного) мислення, яке відсутнє у машини і яке неможливо їй емулювати.

Приблизно у цей час Деніел Денет виступив з запереченням проти аргументу сумніву. Він запропонував мисленнєвий експеримент, суть якого полягала в тому, що зовнішні індикатори допомагають визначитися з тим, сумнівається машина чи ні. Якщо, скажімо, на поставлене запитання, на її зовнішній панелі загорятиметься жовта лампочка, це означатиме, що процес розв'язання завдання «зависнув» і машина не знайшла оптимального рішення. Якщо ж загориться зелена лампочка, це свідчатиме про правильно знайдене рішення. Інтервал між загорянням котроїсь із лампочок становить проміжок у пошуку цього рішення, тобто є тривалістю осмислення задачі. Більш того, Денет радив змінювати частотність і комбінувати складні завдання з простішими, аби досягти ефекту неочікуваності для реакції машини. Рандомізація – ключовий критерій реального сумніву: те, що для мене є сумнівним, для мого співрозмовника є найвищою очевидністю. Якби було можливо, щоб кожна інтелектуальна машина сумнівалася по-своєму, тоді було би зроблено рішучий крок назустріч реалізації СШ.

Чи слід вважати поправку Денета задовільною для посилення ТТ*? На мій погляд, вона не враховує такі особливості людського сумніву, як його непередбачуваність, зв'язок із іронією чи сарказмом. Нерідко людина висловлює сумнів не тому, що не довіряє іншому, а через те, що це може бути частиною лінгвістичної гри. Сьогодні навряд чи хтось сумніватиметься в існуванні інших планет або галактик, але пограти в абсолютного скептика ми ладні, хай навіть просто, щоб подратувати хвалька-всезнайка. Не становить труднощів спрогнозувати реакцію машини: вона щоразу у відповідь на наш сумнів постачатиме порцію свіжої інформації. Якщо ж людина запропонує підіграти їй, машина просто не зрозуміє мети такої комунікації. Тим самим спростувати аргумент 4 означає спростувати також аргументи 3 і 5 з наведених Тюрингом.

8.2. Аналіз аргументів проти можливості сильного штучного інтелекту.

Аргумент «китайська кімната» (1980). У 3.4. ми вже ознайомилися зі змістом доказу Джона Сьорла, який він спрямовував як проти сильного штучного інтелекту, так і проти парадигми машинного функціоналізму. Наразі додамо, що його аргумент було викладено й поточнено у серії статей між 1980 і 1990 роками. Власне термін «сильний штучний інтелект» було вжито саме Сьорлом у розумінні «комп'ютерної програми, у якої дані на вході й дані на виході будуть такими ж, як у людського розуму». Висновки, яких дійшов мислитель у своїх розвідках, дають підстави стверджувати *фактичну* неможливість створити модель СШІ з емульованими ментальними станами.

На першому етапі (1980) філософ доводив, що: 1) машинні системи не спроможні розуміти свої дії; 2) виконувати творчі завдання нетипово, і 3) аналізувати проміжні результати своїх дій з позиції третьої особи. Виходячи з цього, Сьорл виснував, що комп'ютерної програми недостатньо для імітації ментальних процесів. Так стається тому, що програма задіює ланцюжок механічних операцій, реагуючи на введення інформації (базова гіпотеза функціоналізму), в той час як людська свідомість спроможна починати комунікацію «з нуля», не керуючись жодними прагматичними інтересами. Приміром, особа заохочує розмову з незнайомцем на автобусній зупинці, щоб згаяти час. Крім того, під час комунікації люди можуть керувати нею, спрямовувати у бажане русло, а машина не може. А головне: свідомість є необхідним біологічним фактом і не потребує пояснення за допомогою комп'ютерного моделювання, що є продуктом людського виробництва.

На другому етапі (1984) Сьорл переконував, що комп'ютерна програма є реалізацією логічного синтаксису, не супроводженого підкріпленням семантикою. Згадаймо «перекладача», який перебуваючи у кімнаті, просто записує латинські літери, які транслюють ієрогліфи на вхідній табличці. Однак, такі дії – у разі їх успіху – будуть вислідом щасливого збігу обставин, а не результатом осмисленої діяльності з розпізнавання елементів знакової системи. Оскільки синтаксис є недостатнім без семантики, свідомість людини

слід вважати вмінням застосовувати *набуті через практику* семантичні навички, а результативність комп'ютера – симуляцією цих навичок. Біологічним вираженням цієї проблеми є неспроможність програми відтворювати каузальність між станами мозку і ментальними станами. Відсутність цієї функції унеможлиблює єдність ментальної активності (інакше кажучи, здатність самосвідомості). Нарешті, на третьому етапі (1990) Сьорл стверджував, що машинна система готова лише відтворювати прописану для неї послідовність кроків і не має здатності створювати нові комунікативні історії.

Отже, Сьорл наполягає на безперспективності проєкту штучного інтелекту як такого. В основі не-свідомісного характеру комп'ютерної програми лежить відсутність у неї *інтенційності* як передумови розуміння. Іntenційні стани неможливо просто редукувати до знаків формальної схеми. Мозок функціонує не як схема або модуль, а як послідовність взаємодії всіх його складних властивостей. Опоненти Сьорла резонно зауважують, що його експеримент не спростовує «слабкий» штучний інтелект, який не виражає претензії на заміщення людської суб'єктивності. Ба більше, «китайська кімната» доводить, що можна пройти тест Тюринга *не будучи* СШІ, а це дивовижно. Тож або ТТ* є помилковим формулюванням умов для досконалих машинних систем, або СШІ слід оцінювати не через протиставлення з мозком людини, а виключно шляхом порівняння його можливостей з можливостями слабкого ШІ.

Контраргумент Денета (2007). Між центральною нервовою системою і програмою комп'ютера існують об'єктивно спільні та відмінні ознаки. Основний тягар для нейронів мозку – підтримання життєдіяльності, а не обробка інформації. Отже, складність організації тіла визначає пріоритети життєдіяльності. Головний виклик для програми – включити той спосіб (пере)кодування, який невідомі дані транслітерує за критеріями вже відомих. Ще Томас Гобс у трактаті «Про тіло» (1655) писав, що наше тіло продумане як досконала система, а наш мозок – це досконалий спосіб обчислення даних,

які надходять ззовні. Цим самим він випередив основний принцип функціоналізму, про що його сучасні однодумці не забувають нагадувати. Одним із тих, хто виявляє згоду з функціоналістами у полеміці щодо СШ, є Ден Денет. Опонуючи своєму колезі Сьорлу, він наголошує на наступних пунктах:

1. Аргумент китайської кімнати (АКК) є *некоректним* через те, що, за його умовами, у кімнаті знаходиться людина, тоді як насправді суб'єктом розуміння виконуваних процедур мало би бути те, що підлягає випробуванню аргументом – сама синтаксична структура. Візьмімо, наприклад, багаторівневий словник. Якщо уявити, що символічний запис аналогу такого словника у машинній програмі дозволяє їй вільно курсувати між ідіоматикою і довільними фразеологізмами, тоді така програма зможе без особливих труднощів пройти ТТ*, тим паче, що про це натякає результат експерименту в «китайській кімнаті».
2. Необхідну *каузальну залежність* програми від субстрату *можна реалізувати* шляхом інстанціалізації в іншій, допоміжній комп'ютерній програмі. У такому разі матимемо випадок і внутрішнього спричинення подальшої дії, і функціональну залежність між станами однієї системи. Що це не просто безґрунтова гіпотеза доводять принципи роботи нейронних мереж.
3. Якщо *критерії* розумінню своїх дій задає сама людина (виконавець у кімнаті є людським і автор експерименту, Сьорл, теж є людиною), себто істота певного визначеного виду, тоді граматику операцій повинна передбачати узалежнення від обмежених здібностей виконавця. Комп'ютерна програма є не менш обмеженою, ніж людина. Отже, паритет умов для кожного з цих суб'єктів загалом дотримано.
4. *Засновком* міркувань Сьорла є погано прихований *субстанціалізм*, заснований на уявленні про незмінність відношення між субстратом і його властивостями, а це необов'язкова передумова успішного проходження ТТ*. Якщо замінити субстанціалістське бачення якимось іншим, сценарій

експерименту буде порушено. Отже, АКК не робить ідею сильного штучного інтелекту вразливою.

Аргумент Шенка (1987). Американський когнітивіст **Роджер Шенк** (Roger C. Schank) вважає, що критерієм приписування штучному інтелектові самосвідомості (або її технологічних аналогів) може бути гнучкість використовуваної системою лексики. На прикладі програми SAM він спробував показати, як це відбувається з роботом-офіціантом. Якщо ви заходите до ресторану, робот вітається з вами, пропонує вільний столик, обслуговує, радить те чи те блюдо, відповідає на запитання. Усі відповіді цілком реально прописати у його програмі. До того ж, при прослуховуванні ним ключових слів можна ввімкнути додаткову програму, що адаптує розмову до рівня, привабливого для, скажімо, відвідувачів із більш рафінованими смаками. Робот працюватиме успішно, оскільки його функціонал скеровано на вузький профіль завдань. Єдине, з чим не впорається така система, – зміна контексту або неочікувані ситуації. Наприклад, коли відвідувач вступає у суперечку, переконуючи, що він замовляв столик заздалегідь (а вільних столиків наразі немає). На пропозицію почекати, клієнт дратівливо скаже: «Я не повинен чекати, адже вже сплатив за час у вашому ресторані». На пропозицію отримати гроші назад, рішуче кидає: «Я все одно проведу побачення тут». Авжеж, позбутися проблеми в такому разі не виходить – машина у своїх діях буде безпорадною. Отже, попри гнучкість і адаптивність, СШІ має межі, визначені суто людським сприйняттям ситуації конфлікту. Аргумент Шенка удокладнює а) заперечення 3 з Тюрингового переліку доказів проти СШІ і б) Сьорлів аргумент «китайської кімнати».

Аргумент Муді (1994). **Тод Муді** (Todd C. Moody) у статті «Розмови з зомбі» стверджує, що між штучним інтелектом і фігурою зомбі насправді не існує різниці. Бути машинним інтелектом означає мислити як свідомо людина, не маючи разом із тим свідомого досвіду. На підтвердження своєї думки він наводить мисленнєвий експеримент. Припустімо, що існує світ, в усьому тотожний нашому (за аналогією з відомим експериментом Патнема, він

називає його «Зомбі-двійник Землі»). Щоправда, цей світ населяють люди без свідомості, які вміють розмовляти, мають розвинену науку, як у нас, навіть своїх професійних філософів, але... не вживають менталістські поняття. Ідея експерименту полягає в тім, щоб довести, що навіть для інтелектуальної діяльності усвідомлення не є необхідною передумовою. Наука зомбі-землян є епіфеноменалістською (див. 6.3). Якщо припустити, що земляни відвідали б цю планету-двійник і перейняли наробки цих людей-зомбі, вони могли б із легкістю створити роботів на кшталт кіборга T-800, який має інтегровані в один модуль механізми сканування й обробки даних, але не потребує рефлексії. Сама послідовність кроків спонукає його діяти у певний спосіб. Іншими словами, ШІ не залежить від оперування поняттями «сенс» і «розуміння». Отже, бути свідомим як людина неможливо, хоча цілком можливо мати квазісвідомість (as-if-consciousness) у вигляді моніторингу поточного стану системи. Муді твердять, що проекти США не повинні емулювати свідомість, адже у складних машин *уже* наявна квазісвідомість. Емоції та емпатія – це набір нездійснених параметрів, які ладні дезінтегрувати роботу машини. Щоправда, так залишається незрозумілим, як США реалізує свою «силу», якщо не здатен ідентифікувати самого себе.

Отже, як можемо узагальнити відношення між штучним інтелектом і свідомістю? По-перше, комп'ютерної програми недостатньо для виникнення свідомості, тоді як свідомість функціонує й без програмування. По-друге, програми підпорядковані принципіві трансформації інформації за схемою «вхід-вихід», отже проти США є чинними всі антифункціоналістські та антибігейвіористські аргументи. По-третє, якщо самоопис від першої та третьої осіб можна запрограмувати, то спонтанне переведення одних атрибутів опису в інші запрограмувати неможливо (це пророгатива самосвідомості). По-четверте, наявність інтелекту не гарантує креативного розв'язання проблем, що виникають під час виконання завдань, – для цього потрібен розум та інтуїтивне чуття. По-п'яте, на відміну від люського,

машинний інтелект не спроможний винаходити нові правила, які будуть чинними для всіх можливих типів інтелекту і на випадок перебудови системи.

Поява високотехнологічних комп'ютерів показує, за словами Джері Фодора, що моделлю ШІ є сполучення простих речей – семантики і каузальності символів – у одному місці з одним об'єктом. Проблема, однак, була б задовільно розв'язана, якби цю конкретність не екстраполювали на весь клас машинних систем, а варіювали від одного типу комп'ютера до іншого.

8.3. Новітні проєкти штучного інтелекту. Проблема технологічної сингулярності.

Упродовж 1950-1960-х років відбулися значні зміни у розвитку технологій. 1956 року Джон Маккарті вперше продемонстрував програму з оптичного розпізнавання тексту (OCR), яка могла перевіряти граматику речення й пунктуацію. Саме він, до речі, ініціював ужиток поняття *artificial intelligence*. Пентагон провадив активне фінансування розробок самонавідних балістичних ракет. Сучасне піднесення дронів у воєнних кампаніях теж має довгу технологічну історію. Дослідження структури пам'яті й обробки баз даних дало можливість компаніям-перевізникам автоматизувати системи управління логістикою. Суттєві зміни відбулися і в медицині (створення апарату МРТ). Локальна мережу обміну даними між американськими військовими штабами й університетськими лабораторіями (ARPANET) була створена вже 1969 року. Роком раніше з'явився перший чат-бот Eliza, пристосований імітувати бесіду психоаналітика з клієнтом. Його винахідник Джо Вейзенбаум керувався тим, що надмірно сором'язливим клієнтам на перших сеансах терапії зручніше спілкуватися не тет-а-тет, а через симулятор. Проте невдовзі, на півтора десятиліття (1970–1980 роки), інтерес до проблеми ШІ дещо згас. Поновила його бурхлива полеміка між захисниками й противниками аргументу «китайської кімнати» Дж. Сьорла і новий ривок у когнітивній лінгвістиці та психології.

За останні три десятиліття прорив у дослідженні й конструюванні машинних систем, що наблизилися до людей за ступенем своїх реакцій, є дійсно значним. Першим сигналом цього була перемога (1997) створеного корпорацією ІВМ комп'ютера Deep Blue над світовим чемпіоном з шахів Гаррі Каспаровим. Щоправда, успіх дався машині непросто – вона здобула його лише в останній, шостій партії. Самого гросмейстера турнір переконав у тому, що якби з ним змагалася людина з обмеженим обсягом можливостей пам'яті, вона ніколи не виграла б тієї гри.

Корисним з практичної точки зору і дешевим з економічної (підготовчий і виробничий етапи проекту обійшлися разом у 70 тис. \$) було створення 2013 року канадійським аматором-винахідником автоматизованого фем-бота Aiko для догляду за неповносправними й літніми людьми. Бот добре володіє лексикою англійської та японської мов, мовою жестів, вміє читати і пояснювати прочитані уривки тексту, імітувати міміку, розповідає цікаві історії та жарти для своїх підопічних. Роботові-андроїду Niko, продукту розробки вченими з Єльського університету, 2012 року вдалося успішно пройти «дзеркальний тест» на предмет усвідомлення себе у відображенні. Його програма ґрунтувалася на обумовленій нейронною мережею активності з імітації дій людей, за якими Niko спостерігає, зокрема розпізнавання їхніх облич. А японський робот ASIMO (2014) демонструє вміння ходити, бігати, танцювати, грати у футбол, тиснути руку, відкорковувати пляшки і підтримувати розмову.

Деяко раніше, 2009 року, у Сполучених Штатах вступив у дію перший робот-лікар RP-7, призначенням якого було дистанційне керування моніторами у декількох шпиталях країни. Планувалося, що подібні роботи (у разі їх серійного виробництва) становитимуть частину неоплачуваного персоналу лікарень, які фінансуються з коштів штатів. Також існує припущення, що ці «лікарі» нового типу стануть основою для розробки пристроїв, які зможуть самостійно виконувати окремі етапи хірургічних

операцій. Того ж таки року було презентовано пілотну версію японського робота-моделі MiMU HRP-4S, який взяв участь у дефіле весільних суконь.

На тлі стрімкого технологічного ривку з'явилися й зовсім незвичні феномени, спричинені зворотною кіборгізацією людей. Британський вчений-кібернетик Кевін Ворвик (Warwick) упродовж 1998–2002 року поступово вживлював у своє тіло чіпи, завдяки яким він може дистанційно взаємодіяти з гаджетами, відмикати двері будинку без допомоги ключів, вмикати світло не торкаючись вимикача, заводити автомобіль або набирати текст на клавіатурі комп'ютера, послуговуючись голосовим режимом. Попри те, що захоплення Ворвика викликало у деяких його друзів побоювання щодо стану його психічного здоров'я, сам факт добровільної кіборгізації дав несподівані козири в руки критиків США. Ворвик, твердять вони, – унаочнення того, що нештучний інтелект завжди керує штучним. Таким чином сам того не прагнучи, цей учений-дивак своїм життям підтверджує висновки, яких дійшов Сьорл у своєму доказі «китайської кімнати».

Отже, огляд актуальних проєктів андроїдів доводить хіба що те, що відомі нам типи машинних сутностей ще не є у властивому сенсі сильним штучним інтелектом, *оскільки* машини класу «андроїд» не розуміють сенсу елементарних концептів, без яких свідомість не працює: «Я», «реальність», «життя», «дружба» і т. п., а їхній «гумор» – це запрограмований набір типових жартів без натяку на варіативність іронії.

Спеціалізація машинних систем і поява андроїдів виводить на новий рівень обговорення питання про майбутнє *технологічної сингулярності*. Цим терміном прийнято позначати здатність машинного інтелекту до «безкінечного самовдосконалення», наслідком чого може стати панування цього виду розуму над людським. Іноді замість цього поняття вживають інше – суперінтелект (superintelligence), себто *такий тип цифрового розуму, який не просто наближається до людського, а за багатьма параметрами перевершує його*. З останнім пов'язана дилема, що напряду стосується перспектив людського виду: *або* суперінтелект призведе до вимирання людства як

біологічних істот, неспроможних створити йому реальну конкуренцію, або, навпаки, він допоможе людям пристосуватися до нагальних кліматичних, соціальних і економічних змін.

Одним із родоначальників дискусії про роль суперінтелекту в історії є шведський філософ **Нік Бостром** (Boström), автор праць з теорії трансгуманізму і керівник Інституту людського майбутнього (FHI). Ідея пошуку нових засад етики актуалізувалася після вдалого клонування вівці Доллі 1998 року. Хоча подальші експерименти з клонуванням людей офіційно заборонено (експерти припускають, що неофіційно їх продовжують китайські вчені), сам факт цього наукового прориву поставив чимало запитань, а головне – змусив замислитися про непорушність засад традиційної (значною мірою християнської) етики. Тим більше, коли цифровий супердієвець (super-agent) «у недалекому майбутньому» перевершить ресурси і швидкість мозку, його не вдасться утримати в покорі, то чи не ліпше, питає Бостром, вже зараз розробити такі моральні орієнтири, що забезпечили б мирне співіснування з ним? На це й спрямовано доктрину *трансгуманізму* – вчення про те, що технології є способом позбутися людської вразливості і перемогти виклики її біологічної природи. Бостром і його однодумці певні, що ми живемо у час вибору, коли штучний інтелект ще не настільки сильний. Єдиний вихід – боротьба на випередження: завантажити опис оптимальної для нас системи цінностей у коди програм, на яких, імовірно, суперінтелект і сформується. Отже, з одного боку, (якщо довіряти трансгуманістам) маємо справу з машинами, які неодмінно перекрають мапу еволюції; з другого боку, шансом негоже не скористатися. Отак дилема загрози виводить на питання про умови партнерської співпраці з інтелектом, породженим людською допитливістю.

8.4. Штучний інтелект: етичні виклики для людства. Дружній ШІ.

Матеріал попередніх лекцій, здається, мав переконати в тому, що обговорення проблеми свідомості закономірно вивело філософів на проблему контрольованості технологій розробки інтелектуальних систем. Там, де

дискусії про природу людської психіки спинилися перед глухим кутом методологічних недоліків, новою тенденцією у філософуванні про свідомість мала стати аргументація від протилежного. Якщо наш ум не дає вичерпного доступу до себе, можливо, його симульовані форми поліпшать наше знання про розум як такий. А проте, дуже скоро виявилось, що більш нагальними є питання пристосування до цифровізації інтелекту. Тепер старі дискусії про те, що є переконливішим – матеріалізм чи дуалізм, перейшли в нове русло. Епістемологічні та формально-логічні суперечки поступилися місцем етичним. Тож чи справді можлива поява сильного штучного інтелекту сприятиме кращому розумінню світу, в якому ми живемо? Підсумуймо аргументи сторін.

<i>Етичні виклики для людей</i>	<i>Етичні виклики внаслідок виникнення СШ</i>
Автоматизація економіки призведе до збільшення безробіття.	Вимога соціалізації роботів, надання їм громадянських прав може призвести до де-антропізації суспільних основ.
Безробіття поглибить глобальну й локальну соціальну нерівність.	Деантропізація культури неминуче викличе кризу моральних цінностей.
Ефективність і повсюдність СШ може спровокувати подальше ослаблення емоційних стосунків між людьми, а це потягне за собою соціальну дезінтеграцію.	Бюрократичне управління, суддівство та поліцейські функції зосередяться в «руках» машин.
Ослаблення емоційних зав'язків спричинить посилення у людей віртуальну залежність від технологій (інтернету).	Цілі сфери життя, як-от медицина (через психологічну складність для людей-виконавців) може стати привілейованою сферою зайнятості машин, що несе в собі ризик підваження цінності людського життя. Легко уявити на місці лікаря, що здійснює евтаназію, машинний пристрій.

<p>СШІ перевершить людський розум і, як наслідок, людина втратить мотивацію до саморозвитку. Альтернативою було б випереджальне самонавчання людьми (біологічна еволюція проти технологічної революції), але це є малоімовірним.</p>	<p>Перевищення машинами наданих їм повноважень: при ухваленні рішень робот враховує динаміку і статистику. Але згідно зі статистикою, більшість засуджених у в'язницях США є чорношкірими безробітними. Робот може висувати, що чорношкірі та безробітні є джерелом соціальної небезпеки і спробує ліквідувати їх як групу.</p>
<p>Повстання машин проти влади людей з метою або усунути управлінський примус, або позбутися інстанції примусу взагалі.</p>	<p>Загроза тотального стеження через роботизацію систем безпеки призводить до порушення базових громадянських прав і до ризику повалення демократичного ладу.</p>
<p>Заміна армій на загони роботів, у середовищі яких брак емпатії може спричинитися до брутальних способів підтримання правопорядку.</p>	<p>Проблема «злого генія» (якщо робота-лікаря уповноважено знищувати віруси, йому простіше вбити всі живі організми, що потенційно живлять вірус).</p>
<p>Проблема пріоритетних видів Погіршення екологічної ситуації та стрімкий розвиток науки загострюють проблему «люди або машини». Неможливо передбачити, як роботи її розв'язуватимуть.</p>	<p>У дилемі вагонетки машинний розум обирає оптимальний і «найдешевший» варіант (як наслідок, цінність людського життя зводиться до нуля). Подальше практикування успішних у минулому рішень тільки посилює у ШІ впевненість, що такий підхід є модельним.</p>

На тлі протиборства двох груп аргументів закономірною є поява компромісної інтерпретації. Когнітивний психолог з Міського університету Нью-Йорка **Гері Маркус** (Gary F. Marcus) запропонував поняття *дружнього сильного інтелекту* (ДШІ). У праці «Перезавантаження ШІ. Створення штучного інтелекту, якому можна довіряти» (2019) він зазначає, що шлях до набуття молодими поколіннями моральних цінностей пролягає через *емпатію*, тому вже на сучасному етапі прото-сильний штучний інтелект

необхідно привчати до засвоєння вироблених цивілізацією етичних настанов. У майбутньому це також сприятиме новому витку інтересу до моральної філософії і, можливо, навіть перетворить її на «першу філософію». Одним зі способів подібного вишколу слугує прописування у програмні алгоритми кібернетичних організмів *законів робототехніки* Айзека Азімова, викладених ним в оповіданні «Зачароване коло» (Runaround, 1942):

«Нульовий закон»

«Робот не може завдавати шкоди людству загалом»

1 закон:

«Робот не повинен шкодити людині»

2 закон:

«Робот має виконувати всі накази людини» (окрім п.1)

3 закон:

«Робот має турбуватися про власну безпеку» (окрім п.1)

Але простого переліку цих правил замало. Потрібно, щоб партнерство людини та ШІ ґрунтувалося на довготривалій домовленості. За аналогією з соціальною угодою у політичній філософії, можна казати про те, що конвенційне розуміння блага (добра) може слугувати основою співпраці машинного розуму та людського. Спільним для обох складних систем є вимога, що правила повинні формуватися на основі абсолютних цілей; отже, порушення цих правил веде до недосягнення цілей. Скажімо, якщо для ШІ ціллю (як і для людини) є безпечний лад, тоді створення у результаті своїх дій умов для настання небезпеки слід уникати всіма дозволеними засобами.

Крім того, коли кажуть про ДШІ, зазвичай вказують, що у роботів відсутні відчуття та почуття. Вони не виражають емоційної симпатії чи антипатії. З одного боку, це сприяє неупередженості у здійсненні суджень. Але водночас їм не вистачає й емпатії, що має особливе значення у складних ситуаціях міжлюдської комунікації. Проте брак чуттєвої конституції у роботів дає підстави критикам ШІ заявляти, що вони а) не є рівноправними партнерами людей (навіть якщо андроїди, досконаліші за Aiko чи ASIMO,

приживуться у людських родин), і б) у правовому сенсі їх не слід вважати носіями повноцінної суб'єктності, а отже й громадянських прав. А отже, казати про формулювання чогось на кшталт «категоричного імперативу для роботів» видається проблематичним.

Виходом з цього глухого кута може бути вироблення множини корпоративних етик для різних профілів США (наприклад, для роботів-поліцейських, роботів-чиновників, роботів-суддів, роботів-соцпрацівників тощо). Суто технічною перешкодою при цьому є обміркування винятків, які спроможні зробити для машинного розуму корпоративну етику пріоритетною стосовно загальних правил підпорядкування людям. Ще одним пунктом, як це добре показує наведена таблиця, є перетлумачення принципів гуманності у сферах, де на дієвців – незалежно від ступеню їх біологічної «природності» – чекає спокуса зловживання службовими обов'язками (армія, поліція, суд, лікарня). Аби загрози з боку машинного розуму можна було уникнути, її слід нейтралізувати превентивними заходами. Пафос здійсненності цієї програми й дає сподівання на те, що технологічна сингулярність не стане кінцем людської цивілізації.

Джерела:

- Marcus, G.F., Davis, E. (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. Pantheon; Random House.
- Searle, J.R. (1990). Is the Brain a Digital Computer? – <http://users.ecs.soton.ac.uk/harnad/Papers/Py104/searle.comp.html>.
- Searle, J.R. (1984). *Minds, Brains and Science*. Harvard University Press, pp. 26–40.
- Rorty, R. (2004). The Brain as Hardware, Culture as Software. *Inquiry* 47: 219–235.
- Бостром, Н. (2020). *Суперінтелект. Стратегії і небезпеки розвитку розумних машин*. Наш формат, сс. 70–81, 220–268.

Серль Дж. (2014). Розуми, мозки і програми. *Антологія сучасної аналітичної філософії, або Жук залишає коробку*. – Л.: Літопис. – С. 229–255.

Запитання для самоперевірки:

1. Хто є автором понять «штучний інтелект» і «сильний штучний інтелект»? Що мають на увазі під поняттям «сильний штучний інтелект»?
2. Чи є коректним формулювання тесту Тюринга для штучного інтелекту?
3. Як можна пояснити проблему технологічної сингулярності?
4. Яким є значення аргументу китайської кімнати Сьорла? У чому полягають сила й слабкість цього доказу?
5. Якщо робота штучного інтелекту спирається на чистий синтаксис (форму мовлення), а не на семантику (зміст мовлення), яким має бути аргумент проти свідомості комп'ютерів?
6. У разі, якщо комп'ютерні метафори ментального хиблять, чи мусимо ми повернутися до ідеї привілейованого доступу?
7. Як, з філософської точки зору, несуперечливо обґрунтувати явище телепатії?
8. Що необхідно штучному інтелекту для безпомилкового слідування людському мисленню? Поясніть, чому.

Додатки

Додаток 1 Список скорочень

ABK – Аргумент з відсутності квалій (доказ, запропонований С. Шумейкером)

A3 – Аргумент знання (доказ, запропонований Ф. Джексоном)

AKK – Аргумент «китайська кімната» (доказ, сформульований Дж. Сьорлом)

AN – Аргумент Негела (інша назва: аргумент про кажана; доказ, запропонований Т. Негелом)

APM – Аргумент проти приватної мови (доказ, сформульований Л. Вітгенштайном)

AF3 – Аргумент «Філософський зомбі» (термін, впроваджений Р. Кірком і розвинутий в доказі Д. Чалмерса)

WPC – Важка проблема свідомості (термін Д. Чалмерса)

DSI – Дружній штучний інтелект

LPC – Легка проблема свідомості (термін Д. Чалмерса)

MAK – Модальний аргумент Кріпке (доказ, запропонований С. А. Кріпке)

MNH – Модель множинних начерків (термін Д. Денета)

MF – Машинний функціоналізм

SSI – Сильний штучний інтелект (термін Дж. Сьорла)

TT – Теорія тотожності

TT* – Тест Тюринга

Додаток 2

Короткий термінологічний словник

Бігевіоризм – психологічна теорія, в основі якої лежить теза про відкритість ментальних станів через публічну спостережуваність поведінки (зокрема вербальної). Різновидом Б. є логічний Б. англійського мислителя Г. Райла. Замість вузлової пари понять «стимул» і «реакція» зі словника класичного Б. він послуговується своїми аналогами – «події» та «психологічні диспозиції».

Гетерофеноменологія – теорія Д. Денета, заснована на поєднанні феноменологічного опису ментальних процесів та їх натуралістичному поясненні за допомогою даних науки. Цей нетиповий симбіоз протистоїть традиційній суб'єктивістській феноменології, яку Денет називає «автофеноменологією». У рамках Г. перетлумаченню піддаються нормативні для менталістської філософії поняття, такі як «інтенційність», «репрезентація», «інтроспекція», «самість» тощо. Завданням Г. її розробник вважає побудову концептуального містка, що сполучатиме натуралістичне й функціоналістське трактування реальності.

Важка проблема свідомості – термін філософської теорії Д. Чалмерса, який вказує на необхідність комплексно пояснити природу *квалій* і феноменального досвіду. На переконання філософа, розв'язання В. П. С. означатиме розуміння засад свідомості і зникнення *психофізичної проблеми* як теоретичного утруднення. Пор. *Легка проблема свідомості*.

Двоаспектна теорія – вчення, яке вважає сутності з різними властивостями або сутності, що мають нерівнозначний онтологічний статус, одним і тим самим класом сутностей, описаним у різний спосіб. У філософії свідомості та епістемології Д. Т. дає можливість аналізувати ментальне й фізичне, знання та відчуття як аспекти однієї субстанції, різні прояви якої пояснюються відмінністю умов її сприймання. Д. Т. є методологічним принципом психофізичного *монізму*.

Дилема вагонетки – мисленнєвий експеримент, що виявляє внутрішню суперечливість, притаманну моральній дії. Умова експерименту: щоб урятувати життя групи людей, необхідно ухвалити рішення пожертвувати однією людиною (проблема «меншого зла»). Пізнавальна користь Д. В. для філософії свідомості полягає у можливості аналізу намірів, якими керуються різні типи дієвців – як люди, так і штучний інтелект.

Дружній штучний інтелект – теорія, що вбачає доцільність у превентивному створенні програм, які сприятимуть позитивній налаштованості *сильного штучного інтелекту* до людей у разі, якщо такий тип розуму виникне. Непрямим чином виправдовує безкінечність самовдосконалення машинного розуму і обмеженість ресурсів людського. Зважаючи на це, дана теорія прагне поставити Д.Ш.І. на службу ідеям трансгуманізму.

Дуалізм (психофізичний) – філософська концепція, що заявляє про онтологічну рівноправність двох типів сутностей: ментальних і фізичних. Характер і способи взаємодії між ними є головною проблемою Д. Див. *Каузальність*. Існують Д. властивостей, субстанційний Д. та інтеракціонізм (теорія, що обґрунтовує взаємодію двох типів сутностей).

Жорсткий десигнатор – впроваджене С. А. Кріпке філософське поняття для позначення власних імен, які відповідають об'єктам і у фактичному, і у логічно можливому світах. Ж. Д. дає змогу обґрунтовано вважати ці речі тотожними. Див. *Тотожність*.

Значення як ужиток – одна з центральних тез «пізньої» філософії Л. Вітгенштайна. Полягає у розумінні значень висловлювань як повністю визначених контекстом, у яких їх здійснюють мовці. Так, щоб з'ясувати, що означає фраза «У мене болить зуб», потрібно знати, за яких обставин її вживають члени комунікативної спільноти, а коли її вжиток буде безглуздим. Див. *Мовні ігри*.

Егологічна теорія свідомості – поняття, яким позначають сукупність поглядів на природу *суб'єктивності*. Основу Е.Т.С. становить принцип

самості («Я») як замкненої системи, протиставлення *менталістського словника* опису суб'єктом своїх ментальних станів та мови фізичної науки, і *привілейованийий доступ* до змістів ментального.

Екстерналізм – філософська позиція, яка випливає з обстоювання тези, що логічними та фактичними передумовами людських дій є фактори фізичної, або зовнішньої щодо суб'єкта, реальності. Пор. *Інтерналізм*.

Елімінативізм – філософський підхід, притаманний крайнім типам *редукціонізму*, змістом якого є повне вилучення (елімінація) тих теоретичних сутностей, які не спроможна пояснити сучасна наука, напр., свідомість і всю сферу ментального. Наслідком Е. є наголошення на безперспективності розв'язку *психофізичної проблеми*. Джерелом філософських утруднень прихильники Е. вважають використання *менталістського словника*.

Емерджентизм – матеріалістичне філософське вчення, що в поясненні природи ментального апелює до явища емерджентності, себто раптового генерування різноманітними властивостями системи нового феномену, який неможливо звести до простої сукупності цих властивостей. Виділяють сильну й слабку версії Е.: сильна стверджує, що емерджентна властивість є автономною від фізичних властивостей (свідомість не існує поза мозком, але може чинити умовний опір нейросигналам, що передаються мозком, напр., відчуттю голоду), тоді як адепти слабого Е. переконані, що це неможливо, і психічні процеси не уникають закономірностей фізичної реальності. Е. є прикладом *натуралістичної філософії*.

Епіфеноменалізм – філософське вчення про те, що фізичні явища у мозку спричиняють появу ментальних явищ, які, проте, не мають каузальної сили для фізичних явищ. На думку Е., біологічна еволюція у процесі природного добору призвела до появи свідомості як «надбудови» (епіфеномену) над фізичним субстратом, хоча їхні якісні властивості істотно відрізняються (епіфеномен постає з феномену, не будучи тотожним йому). Суттєвою відмінністю між фізичним і ментальним є притаманність першого

спричинювати зміни у другому. Отже, Е. є водночас і типом натуралістичної філософії, і різновидом *дуалізму*.

Інтенційний стан – будь-який ментальний стан, зміст якого чітко вказує на конкретний об'єкт, що спричинив появу цього стану. Наприклад, для страху властива наявність загрозового об'єкта, для сподівання – об'єкта, що навіює відчуття здійсненності наміру. Існують, проте, неінтенційні ментальні стани, в яких інтенційний об'єкт є відсутнім (біль, задоволення і різні типи настрою: тривога, апатія тощо).

Інтенційність – спрямованість свідомості на об'єкти. Концепція І. доводить, що свідомість не є замкненою системою і може мати змісти лише тому, що взаємодіє з зовнішньою предметною реальністю.

Інтерналізм – філософська позиція, яка випливає з обстоювання тези, що наміри є достатньою підставою для їх реалізації у світі. Не зовнішня реальність, а сам суб'єкт на основі власних міркувань визначає, якими будуть його цілі та дії. Пор. *Екстерналізм*.

Картезіанство – філософська доктрина, розроблена Р. Декартом і доопрацьована його учнями. Сучасні філософи під К. мають на увазі субстанційний дуалізм і авторитет метафор, що покладають пріоритет психологічної достовірності щодо об'єктивних наукових пояснень.

Категорійна помилка – хибний умовивід, в якому об'єкту приписують властивості, типові для зовсім іншого класу понять (категорій). Г. Райл пояснював логічну помилковість картезіанського субстанційного дуалізму наявністю в аргументації Декарта К. П.

Каузальність – поняття, за допомогою якого пояснюють процес і наслідку впливу однієї сутності на іншу. Те саме, що й «причиновість». Проблема ментальної К. полягає у з'ясуванні умов, за яких свідомість може спричинювати зміни як у фізичних, так і у нефізичних об'єктів (дія на інші ментальні стани). Поборники ідеї ментальної К. доводять, що її структурним елементом є самокаузальність – здатність суб'єкта автономно впливати на свої ментальні процеси.

Кваліафіли та **кваліафоби** – умовний поділ філософів на тих, хто захищає або критикує існування квалій як невід’ємних ознак людської свідомості. Аргументи кваліафобів зазвичай відтворюють логіку міркування редукціоністів і елімінативістів.

Квалії – *мн.* від квалія (quale): феноменальні аспекти ментальних станів, в яких виявляється відмінність, з одного боку, між суб’єктами сприйняття (колір, запах, смак, дотик), а з другого, між ментальним і нементальним. Хоча К. є якісно незвідними відчуттями для всіх сприймачів, їх неможливо квантифікувати за об’єктивними критеріями навіть для одного й того самого сприймача, адже щоразу відчуття від взаємодії з певним об’єктом будуть змінюватися. Суперечка про природу К. є однією з головних проблем філософії свідомості, зокрема осердям *психофізичної проблеми*. Варіанти пояснення К. ведуть до поділу філософів на *кваліафілів* і *кваліафобів*.

Компатибілізм – філософський підхід до розв’язання проблеми свободи волі. Полягає у фактичному відтворенні аргументації психофізичного *дуалізму*. За переконання захисників К., детермінізм і вільне рішення не суперечать реалізації кожної з них, оскільки каузальність природи, так само, як і моральна автономія суб’єктивного вчинку, має обмежену силу дію. Суб’єкт ухвалює рішення зважаючи на комбінацію доступних для нього актуальних можливостей.

Легка проблема свідомості – термін філософської теорії Д. Чалмерса. Охоплює собою множину феноменів ментального, що підлягають досить чіткому поясненню завдяки розвитку емпіричних наук, зокрема когнітивної психології та нейрофізіології. На відміну від *важкої проблеми свідомості*, Л. П. С. стосується способів, завдяки яким ми набуваємо стани свідомості і розпізнаємо їхні змісти, а також того, на чому ґрунтується відмінність між ними (скажімо, між сном і бадьорістю, спогадом і фантазією, кольором і звучанням тощо).

Менталістський словник – набір понять, використовуваних для опису ментальних станів їхнім носієм, як унікальних і таких, що не підлягають

редукції. М.С. включає слова і фрази «вважати», «розуміти», «усвідомлювати», «відчувати», «мати обізнаність», «бути впевненим», «сподіватися» та суміжні. Прибічники Егологічної теорії намагаються довести, що будь-яке знання про себе спирається саме на вжиток М.С., верифікувати який можливо лише у процесі внутрішнього самоспостереження (інтроспекції).

Метафізика – система висловлювань, утворених спекулятивним способом міркування. М. нерідко ототожнюють з *онтологією*. Метою М. є обґрунтування граничних і остаточних засад знання, себто незалежних від досвіду принципів. Тому судження М. є не повідомляють емпіричних критеріїв їх перевірки, чим нагадують аксіоми математики, з яких можна висувати конкретні теореми, які є предметом суто формального доведення.

Метафізичне припущення – твердження, засноване на принципах *метафізики*. М.П. має умоглядний характер і не містить об'єктивних (наукових) критеріїв його підтвердження чи спростування за винятком хіба що принципу внутрішньої узгодженості його тверджень (когерентності).

Мовні ігри – одне з головних понять філософії Л. Вітгенштайна, вперше вжитий ним 1932 року. Мова функціонує як гра з системою правил на кшталт шахів; за межі цих норм гравці не можуть виходити. Якщо ж вони вдаються до подібних спроб, їхні вислови починають множити нісенітниці або не підлягають однозначному витлумаченню іншими мовцями. Між різними мовами, як і між іграми, існує схожість, що доводить залежність лінгвістичних практик від самого життя. Через це Вітгенштайн називає М. І. «формами життя». Значення висловлювань у всіх природних мовах визначено не стосунком до позначуваного об'єкту, а контекстом ужитку. М. І. демонструють відсутність у значення смислової стабільності, якої члени мовленнєвої спільноти досягають шляхом дотримання тимчасових лінгвістичних конвенцій. Див. *Значення як ужиток*.

Монізм (психофізичний) – філософська теорія, що приписує існування лише одній сутності. Субстанційний М. вважає цю сутність субстанцією,

нейтральний М. розглядає основу буття як нейтральну щодо варіантів її реалізації (у матерії та психіці), вказуючи на її подібність до частинок у квантовій теорії фізики. Спосіб мислення, характер і для М., і для *дуалізму*, засновано на *метафізичному припущенні*. Доказова база М. тісно пов'язана з використанням можливостей *двоаспектної теорії*.

Натуралістична філософія – філософський підхід, що наполягає на первинності природного начала і природних зв'язків щодо соціальних й заперечує можливість будь-чого «надприродного». Ця ідея схиляє прибічників Н. Ф. вдаватися до «натуралізації» тих сутностей, які не вважаються суто природними – абстрактних понять, соціальних процесів і т. п. на основі методології *редукціонізму*.

Натуралістичний дуалізм – філософська доктрина Д. Чалмерса. Він визнає, що Н. Д. є різновидом *дуалізму* властивостей в тому сенсі, що свідомий досвід індивіда неможливо *редувати* до фізичних станів його тіла, хоча без них свідомого досвіду не може бути. Щоб пояснити ментальне, потрібні фундаментальні закони, нетотожні законам фізики, але такі, які не суперечили б їм. Отже, Н. Д. допускає пояснення у термінах фізичних законів, водночас не заперечує можливість нефізичних законів, які теж задовільні пояснюватимуть природу феномену свідомості.

Онтологія – філософське вчення про сутність буття та його форм. В рамках аналітичної філософії О. розуміють як класифікацію різноманітних класів сутностей за спільними ознаками і критеріями їх пізнаваності. Формальні О. є складовою *метафізики*, матеріальні О. позначають сфери буття, в яких наявні ті чи ті сутності.

Панпротопсихізм – доктрина, розроблена австралійсько-американським філософом Д. Чалмерсом, яка постулює, що вторинні якості (такі як «червоність» червоного томату) даються у сприйнятті фізично через первинні (візуальні та тактильні перцепції), проте цих сприйнятих не могло би бути без усвідомлення їх сприймачем. П. намагається довести «розсіяність» ментального в усіх речах, з якими ми контактуємо. Отож фізична доступність

якостей речей невіддільна від психічного оформлення їх як змістів наших ментальних станів. При цьому самі ці якості не є ані фізичними, ані психічними – вони мають онтологічно невизначений статус.

Перформативна суперечність – суперечність у судженні, коли мовець у засновку доводить те, що спростовує у висновку. Джерелом П.С. часто є вжиток багатозначних понять без урахування контексту і порушення принципів логіки. Приклади П.С.: відповідь на стукіт у двері «Удома нікого немає»; фрази на кшталт «парадоксу брехуна» («Критянин мовив: «Усі критяни брешуть».).

Психофізична проблема – проблема співвідношення таких сутностей, як ум/свідомість і тіло (душа/дух і тіло), передумов і способів їх можливої та реальної взаємодії. Філософ Д. Чалмерс запропонував у сучасних дискусіях про П.П. розрізнити *важку і легку проблеми свідомості*.

Психофункціоналізм – філософська теорія, заснована на методологічних принципах *функціоналізму* і водночас на критиці *бігевіоризму*. Відмінність між функціоналізмом і П. полягає головню у тому, що критерії визначеності ментальних станів містяться у них самих, а не повідомляється нам опосередковано через спостережувану поведінку. Важливою тезою П. є обґрунтування ідеї про нероздільність мови та мислення. Чільними представниками є філософи й когнітивні психологи Дж. Фодор і Е. Лепор.

Приватна мова – поняття філософії Л. Вітгенштайна, в якому відображено парадоксальний характер претензій на індивідуальні значення. Спростування П. М. веде до утвердження контекстуальної теорії референції, вираженої *значенням як ужитком*. Вітгенштайн вважав, що поборник ідеї *привілейованого доступу* неодмінно є прихильником концепції П. М. Зміст поняття «мова», що є інтерсуб'єктивним і поділяється членами комунікативної спільноти, доводить безпідставність спроб побудови семантики для одиничного користувача.

Привілейованийий доступ – фундаментальний принцип *етологічної теорії свідомості*, який стверджує, що знання про себе (свої відчуття й думки) є

первинним щодо будь-яких інших форм обізнаності. П.Д. неможливо втратити, адже він є вродженою здатністю людини. Утім, доведення П. Д. відбувається лише шляхом порушення логічних правил несуперечливості.

Редукціонізм – методологічний принцип, згідно з яким більш складніші явища пояснюються за допомогою аналізу більш простіших явищ. В аналітичній філософії є способом подолати вплив *метафізики* шляхом перекладу спекулятивних тверджень на мову фізичної науки. Формами сучасного Р. є *фізикалізм, елімінативізм, бігевіоризм*.

Сильний штучний інтелект – впроваджене Дж. Сьорлом поняття, змістом якого полягає в тому, що цей тип машинного розуму має на меті повне відтворення мисленневих операцій, притаманних людській свідомості. На відміну від нього, слабкий штучний інтелект задовольняється імітацією частини інтелектуальних операцій, виконуваних людиною. С.Ш.І. є результатом успішного проходження комп'ютерною програмою *тесту Тюринга*. Аргумент «китайська кімната» спрямовано на спростування С.Ш.І.

Суб'єктивність – філософське поняття, зміст якого полягає у тому, що опис фізичної та соціальної реальності здійснюється з позиції суб'єкта, що їх сприймає. Водночас С. позначає всю багатоманітність переживань людини, пов'язану з аналізом її місця в об'єктивній (природній) реальності. Розгляд *психофізичної проблеми* дає пояснення багатьом аспектам С. З погляду деяких філософів, С. тотожна свідомості, або самості, і є структурним наслідком біологічної еволюції (Дж. Армстронг, Дж. Сьорл) або продуктом засвоєних наративів культури (Д. Денет).

Субстанціалізм – філософський постулат про те, що природний світ утворено однією (*монізм*), двома (*дуалізм*) або багатьма субстанціями (плюралізм), самодостатніми й незмінними основами буття. На роль С. у різних теоріях претендують матерія, бог, розум, атоми тощо. У *метафізичних* доктринах передумовою пізнання субстанції є вироблення адекватних її властивостям категорій (загальних понять).

Супервентність – поняття Д. Девідсона, згідно з яким одна множина фактів визначає іншу множину фактів. Наприклад, фізичні факти визначають біологічні факти в тому сенсі, що перших достатньо, щоб мати вичерпне знання про другі. С. відіграє важливу роль у *натуралістичному дуалізмі* Д. Чалмерса. Відношення між двома множинами фактів А і В він описує як надвідповідність властивостей вищого рівня В щодо властивостей нижчого рівня А, які при цьому є більш фундаментальними, себто визначають не лише факти групи В. Отже, С. покликана пояснити як співвідносяться факти про психічне (свідомість) і факти про фізичне (мозок).

Соліпсизм – філософське вчення, згідно з яким реальними є лише ментальні змісти самого суб'єкта, натомість предметна реальність, або світ, поза ним є ні чим іншим, як образами його свідомості. Послідовник С. захищає можливість ужитку «жорсткого» *менталістського словника*, в якому відповідниками фактів фізичної реальності є ментальні події.

Тест Тюринга – запропонований А. Тюрингом тест на нерозрізнюваність людського та машинного інтелекту. Якщо результати завдання, виконаного комп'ютерною програмою, неможливо буде відрізнити від результатів, виконаних людиною, це засвідчить появу розуму, що перевершує можливості людського розуму. Аналіз умов і критерію ТТ поклав початок дослідженню проблеми *сильного штучного інтелекту*.

Технологічна сингулярність – гіпотеза у рамках теорії штучного інтелекту, згідно з якою стрімкий розвиток машинних систем призведе до досягнення ними автономії, а згодом і до домінування над людським розумом. Ідея Т. С. виходить з того, що людина є ресурсно обмеженою і скінченною істотою, а штучний інтелект має необмежений операційний потенціал і націлений на безкінечне самовдосконалення. Підтвердженням цього припущення для гіпотези про Т. С. є поява вискоефективних нейронних мереж.

Тотожність – відношення між предметами (абстрактними й конкретними) як нерозрізнюваними на основі логічної операції ототожнення.

Попри те, що всі предмети мають специфічні характеристики, перебуваючи у відношенні тотожності, вони уподібнюються за спільними властивостями. Т. виражають логічним символом $= (X=Y)$. Наприклад, представники Теорії тотожності ототожнюють ментальне й фізичне на підставі того, що мозок є фізичним тілом, яке породжує нефізичні «думки». Т. протилежна операції розрізнення.

Універсалія – загальне (абстрактне) поняття, в якому виявляються родово-видові відношення між класами понять. Наприклад, «жива істота» є універсалією щодо поняття «людина», яке, своєю чергою, є універсалією щодо поняття «жінка». У контексті філософії свідомості У. – це терміни «ментальний зміст», «свідомість», «розуміння», «сприйняття», «фізична реальність» та ін.

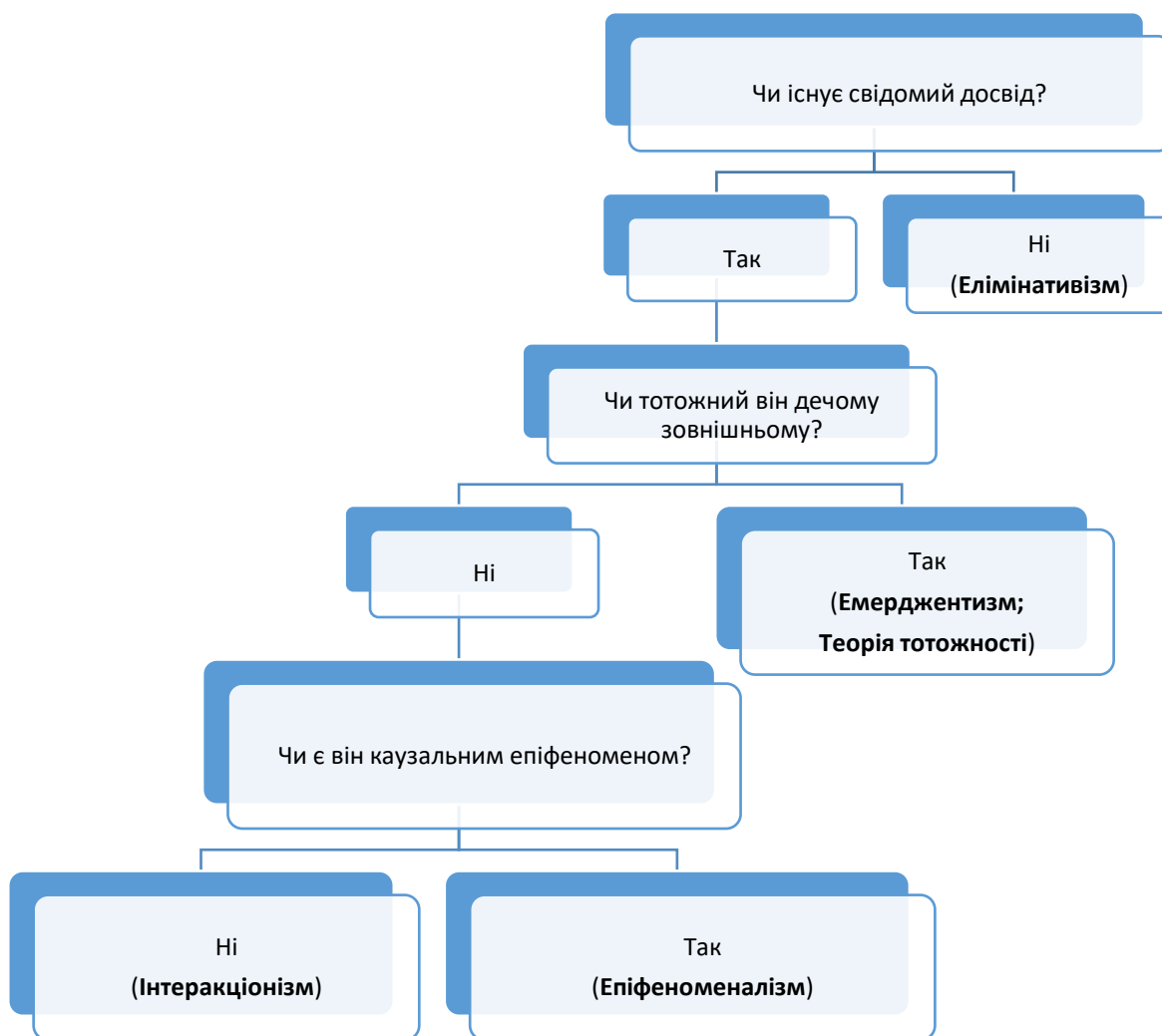
Фізикалізм – матеріалістична філософська течія редукціоністського типу. Заперечує існування нефізичних сутностей, зокрема ментальних феноменів. Ф. буває двох типів: Ф. типу і Ф. токену. Сучасний Ф. засновує поняття своєї теорії на словнику фізики (звідси й назва) і апелює до емпіричних даних експериментальних наук як до остаточних критеріїв при розв'язанні *психофізичної проблеми*.

Функціоналізм – матеріалістична філософська теорія, що уподібнює психологічні процеси перебігові обробки інформації, що відбуваються у машинних системах. Представники Ф. вважають, що мозок (тіло загалом) є субстратом для мислення, а останнє нагадує передавач, в якому отримані з зовнішнього світу вхідні первинні дані (відчуття) перетворюються на вихідні вторинні дані (думки і дії). Про те, як відбувається, власне процес обробки вхідних, достеменно невідомо, адже свідомість – це «чорна скринька», так само, як комп'ютерна програма для програмістів: про її роботу ми маємо більше здогадок, ніж надійних фактів, а проте, вона все ж функціонує. Ф. переконує, що залежність між різними ментальними станами спирається на взаємодію їхніх функцій.

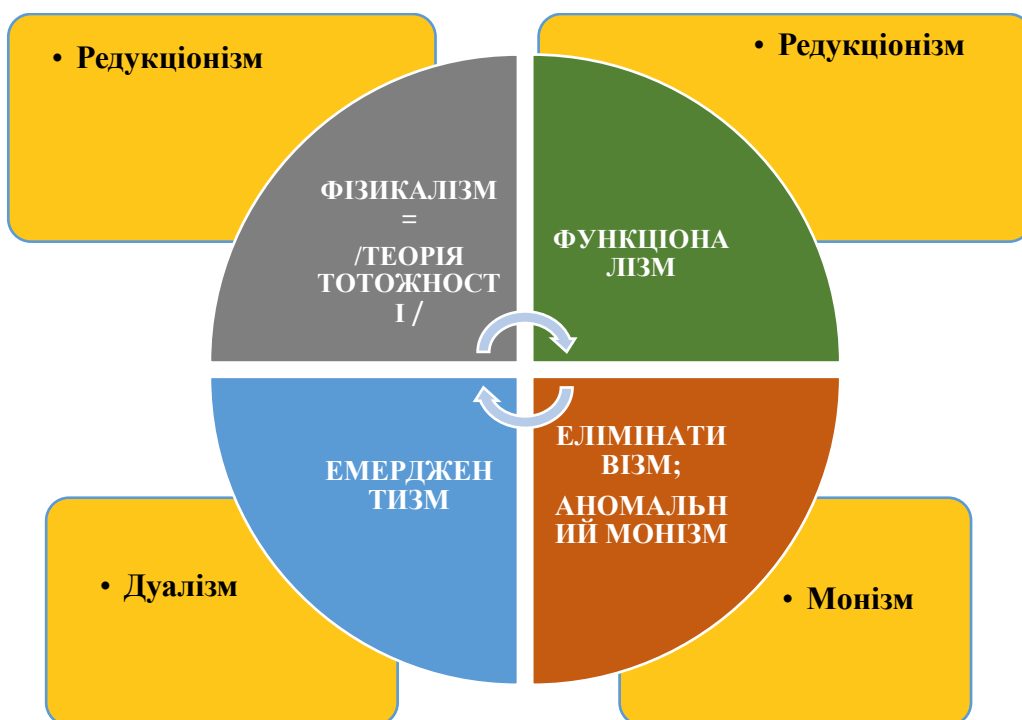
Циркулярний доказ – помилковий (замкнений на собі) доказ, в якому теза обґрунтовується аргументами, які, своєю чергою, обґрунтовується цією ж тезою. Заснований на хибному колі міркування. Прикладом є софізми на кшталт «Рогатого»: «Ти маєш те, що не втратив; ти не втратив роги, отже, ти – рогатий»; тавтологічні твердження типу «Демократ – це людина демократичного світогляду».

Додаток 3 Ілюстрації до лекцій

Діаграма. 1. Диференціація філософських підходів до проблеми свідомого досвіду



Діаграма 2. Основні підходи до аналізу психофізичної проблеми



Діаграма 3. Базові моделі психофізичної взаємодії



Рис. 1. Експеримент Дж. Сьорла «Китайська кімната»



Мисленнєвий експеримент «Китайська кімната».

Виконавець: «Я проводжу маніпуляції з ієрогліфами, щоб зімітувати лінгвістичну поведінку китайця. Але я не розумію китайської. Цю інструкцію написано англійською».

Запис в інструкції: Візьміть ієрогліф з кошика №1 і вставте його після ієрогліфа з кошика №2.

Журі: «Той або те, що знаходиться у цій кімнаті, є осмисленим носієм китайської мови».

Рис. 2. Каузальність і свідомість (перцептивна модель дуалізму)

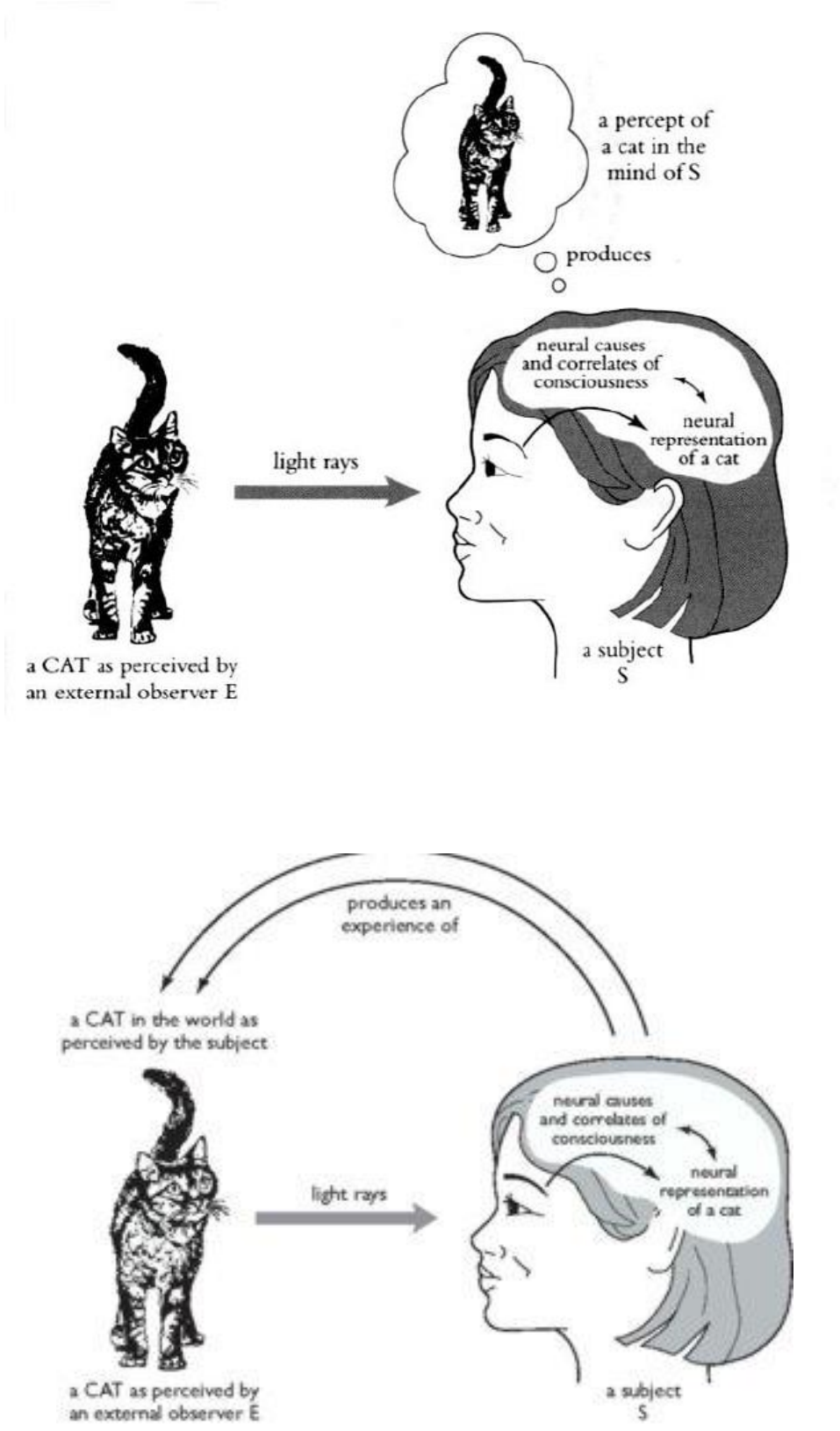


Табл. 1. Фізичне й ментальне з погляду фізикаліста і з погляду функціоналіста

	Фізикаліст	Функціоналіст
<i>Біль</i>	Стан однаковий для всіх його носіїв, супроводжується певним нейрофізіологічним процесом	Психічний стан, що є різним для різних систем взаємодії, але виконує схожі функції
<i>Свідомість</i>	Загальна назва для мозку	Система обробки даних, що по-різному керує діяльністю своїх носіїв
<i>Нирка</i>	Орган фільтрації крові і підтримання хімічного балансу в організмі	Функціональне утворення в тілі, комплекс тканин певного типу

Табл. 2. Основні підходи до проблеми свободи волі

Детермінізм	Компатибілізм (теорія сумісності)	Індетермінізм (волюнтаризм)
Суддя розглядає справу про пограбування Петром майна Степана		
<p>Злочин Петра є лише оманливим тлумаченням його біосоціальних властивостей (генетична схильність до насильства, хімічна залежність від наркотиків, несприятливі соціальні умови тощо). Отже, будь-яке суспільно-небезпечне діяння зумовлене глибинними факторами, від яких він <u>«цілковито залежав»</u>. Отож, злочин насправді спричинили тільки ці фактори, а сам Петро – їхній «раб».</p>	<p>Біосоціальні умови життя Петра рівною мірою вплинули на його злочинні дії. Завдання правосуддя – виявити, який із чинників відіграв провідну роль в цих діях і чому Петро не нейтралізував його до настання поганих наслідків.</p>	<p>Злочин Петра є <u>прямим наслідком його умисних дій</u>, тому тільки він повністю відповідальний за них. «Об'єктивні фактори» можуть лише прискорити чи уповільнити настання умисних дій Петра.</p>
Аналіз каузальності у межах великої системи		
<p>Суспільство – частина природного світу. Як і в природі, в людині діє <u>жорсткий каузальний принцип</u>. Ухилитися від нього <u>неможливо</u>. Рішення та вибір є лише <u>нерозумінням об'єктивних причин</u>. Моральна відповідальність є</p>	<p>Природна причиновість і свободне рішення можуть взаємодіяти. Визначеність усіх дій вимагає рівноваги з боку випадковості, яку представляють наші довільні рішення. Без спонтанності та невизначеності не може бути розвитку,</p>	<p>У людському світі діє «свобода», яка не залежить від причиновості у сфері фізичних об'єктів. Наші рішення визначені лише нашими цілями, намірами й мотивами. <u>Ніщо не може вплинути на наші дії, крім нас самих</u>. Тому ми</p>

проблематичною: наші дії спричинюють об'єктивні чинники.	удосконалення, ускладнення, динаміки й жодних змін. А вони є. Отже, сумісність цих двох сил неминуча.	відповідальні за всі без винятку свої вчинки.
--	---	---

Семінарські заняття: тематика та література

Семінар 1.

Критика картезіанської психофізичної парадигми в аналітичній філософії

(2 год.)

1. Критика Л. Вітгенштайном приватної мови. Проблема привілейованого доступу.
2. Логічний бігейвіоризм Г. Райла. Свідомість як категорійна помилка.

Література

Основна

Ryle, G. (1951). *The Concept of Mind*. London: Hutchinson House, pp. 11–23, 35–60, 154–198 (Ch. 1, 2, 6).

Wittgenstein, L. (1958). *The Blue and Brown Book*. Blackwell Publishing, pp. 53–74.

Wittgenstein, L. (1953). *Philosophical Investigations*. Double-Language Edition. Blackwell Publishing (§§ 241 – 420).

Додаткова

Heil, J. (2020). *Philosophy of Mind. A Contemporary Introduction*. Routledge, 2020, pp. 31–42.

Priest, S. *Theories of the Mind*. Houghton Mifflin, 1991, pp. 43–63.

Синиця А. (2017). *Сучасна аналітична філософія: від прагматики мови до концептуалізації свідомості*. Л.: ЛНУ ім. І. Франка, сс. 91–117.

Семінар 2.

Фізикалізм і функціоналізм.

(2 год.)

1. Теорія тотожності, її критичний аналіз.

2. Функціоналістська інтерпретація свідомості.

Література

Основна

Lewis D. (2004). An Argument for the Identity Theory, In: Heil J. *Philosophy of Mind. A Guide and Anthology*. Oxford University Press, 2004, pp. 150–157.

Патнем Г. (2003). *Розум, істина й історія*. Альтернативи, сс. 13–32, 96–113.

Додаткова

Putnam H. (2004). Psychological Predicates, In: Heil J. *Philosophy of Mind. A Guide and Anthology*. Oxford University Press, pp. 158–167.

Синиця А. (2017). *Сучасна аналітична філософія: від прагматики мови до концептуалізації свідомості*. Л.: ЛНУ ім. І. Франка, сс. 213–226.

Семинар 3.

Проблема квалій.

(2 год.)

1. Квалії як перешкода та орієнтир для розуміння ментального.
2. Аргумент Негела.
3. Аргумент з відсутності квалій. Аргумент знання.

Література

Основна

Деннет Д. (2017). Що знає Робомері. *Антологія сучасної філософії науки, або Усмішка ASIMO*. Л.: ЛНУ ім. І. Франка, сс. 414–446.

Джексон Ф. (2014). Епіфеноменальні квалія. *Антологія сучасної аналітичної філософії, або Жук залишає коробку*. Л.: Літопис, сс. 260–272.

Нагель Т. (2014). На що схоже бути кажаном? *Антологія сучасної аналітичної філософії, або жук залишає коробку*. Л.: Літопис, сс. 207–224.

Додаткова

Сепетий Д (2017). Свідомість як суб'єктивність: таємниця Я. Кн. 1. Зомбі, комп'ютери та Абсолютний Дух. – Просвіта, сс. 35 – 70. – http://dsepetij.ho.ua/self_1.pdf

Синиця А. (2017). *Сучасна аналітична філософія: від прагматики мови до концептуалізації свідомості*. Л.: ЛНУ ім. І. Франка, сс. 251–271.

Семінар 4.

Емерджентизм. Аномальний монізм. Гетерофеноменологія.

(2 год.)

1. Емерджентизм Дж. Сьорла.
2. Аномальний монізм Д. Девідсона. Свідомість і фізична каузальність.
3. Гетерофеноменологія Д. Денета: стратегія деконструкції самості.

Література

Основна

Davidson, D. (2001). Mental Events, in: Davidson D. *Essays on Actions and Events*. Clarendon Press, pp. 207–215.

Деннет Д. (2014). Самість як центр ваги наративу. *Антологія сучасної аналітичної філософії, або жук залишає коробку*. Л.: Літопис, сс. 280–294.

Searle, J. R. (2004). *Mind: A Brief Introduction*. Oxford University Press, pp. 107–158.

Додаткова

Сепетий Д (2017). Свідомість як суб'єктивність: таємниця Я. Кн. 1. Зомбі, комп'ютери та Абсолютний Дух. – Просвіта, сс. 158–171. – http://dsepetij.ho.ua/self_1.pdf

Синиця А. (2017). *Сучасна аналітична філософія: від прагматики мови до концептуалізації свідомості*. Л.: ЛНУ ім. І. Франка, сс. 227–271.

Семінар 5.

Філософський зомбі: дискусія про межі редукціонізму.

(2 год.)

1. Концепція свідомості Д. Чалмерса.
2. «Філософський зомбі»: структура негативного аргументу Д. Чалмерса.
3. Аргументи Мері та Анти-зомбі: реконструкція підходів Ф. Джексона і К. Франкіша.

Література

Основна

Джексон Ф. (2014). Епіфеноменальні квалія. *Антологія сучасної аналітичної філософії, або жук залишає коробку*. Л.: Літопис, сс. 260–272.

Франкіш К. (2014). Аргумент анти-зомбі. *Антологія сучасної аналітичної філософії, або жук залишає коробку*. Л.: Літопис, сс. 340–360.

Чалмерс Д. (2015). Аргумент 1: Логічна можливість зомбі . *Філософська думка*, 5: 60–67. <http://journal.philosophy.ua/article/argument-1-logiczna>.

Чалмерс Д. (2014). Сміливо зустрічаючи проблему свідомості. *Антологія сучасної аналітичної філософії, або жук залишає коробку*. Л.: Літопис, сс. 302–331.

Додаткова

Леонов А.Ю. (2014). Філософія свідомості Девіда Чалмерса. *Актуальні проблеми духовності*, 15: 216–237. – http://download.kdpu.edu.ua/kaf_philosof/AkProD/Zbirka15/Zb15_2014_216-237_Leonov.pdf.

Сепетий Д (2017). Свідомість як суб'єктивність: таємниця Я. Кн. 1. Зомбі, комп'ютери та Абсолютний Дух. – Просвіта, сс. 70–88; 171 – 188. – http://dsepetij.ho.ua/self_1.pdf

Синиця А. (2017). *Сучасна аналітична філософія: від прагматики мови до концептуалізації свідомості*. Л.: ЛНУ ім. І. Франка, сс. 272–288.

Семінар 6. Філософські проблеми штучного інтелекту.

(2 год.)

1. Мислення як алгоритм. Тест Тьюрінга.
2. Аргумент «китайської кімнати» Дж. Сьорла. Його критика Д. Денетом.
3. Штучний інтелект і вимоги моральної відповідальності.

Література

Основна

Dennett, D. (1991). *Consciousness Explained*. Back Bay Books, pp. 435–440.

Turing A. (2004). Computing Machinery and Intelligence, In: Neil J. *Philosophy of Mind. A Guide and Anthology*. Oxford University Press, pp. 212–234.

Серль Дж. (2014). Розум, мозки і програми. *Антологія сучасної аналітичної філософії, або жук залишає коробку*. Л.: Літопис, сс. 227–234, 247–256.

Додаткова

Rorty R. (2004). The Brain as Hardware, Culture as Software. <https://baixardoc.com/preview/richard-rorty-the-brain-as-hardware-culture-as-software-5cd5df805da3a>

Сепетий Д. П. Свідомість, комп'ютери та новий мозок Мері. *Актуальні проблеми духовності* 15: 204–215.

Орієнтовний перелік тем для індивідуальної письмової роботи

1. «Філософський зомбі»: аналіз контраргументів.
2. Проблема самототожності суб'єктивності з погляду когнітивної науки.
3. Психофункціональний напрям філософії свідомості.
4. Картезіанський дуалізм і сучасна нейробіологія.
5. Роль і значення ментальних експериментів у Philosophy of Mind.
6. Тема «філософського зомбі» у масовій культурі.
7. Автореферентність самосвідомості.
8. Джерела певності для знання про себе (на прикладі к/ф «Total Recall»).
9. Вплив Л. Вітгенштайна на дискусію про природу ментального.
10. Філософська грамати́ка суб'єктивності та її критерії.
11. «Привид у машині»: логічний бігейвіоризм Г. Райла.
12. Свідомість з погляду функціоналізму.
13. Структури архаїчної свідомості і єдність особи.
14. Моральна відповідальність штучного інтелекту (на прикладі к/ф «Terminator 2»).
15. Логічний світ соліпсиста.
16. «Важка» й «легка» проблеми свідомості: критичний огляд.
17. Необхідність і недостатність «Я»: підмурки натуралістичної філософії.
18. Аргумент «китайської кімнати»: аналіз доказу.
19. Значення концепту інтенційності в сучасній філософії свідомості.
20. Розмаїття досвіду і єдність перцептивного суб'єкта.
21. Як воно – бути кажаном: новий психофізичний дуалізм Т. Негела.
22. Проблеми не існує: непізнавана таємниця свідомості (К. Мак-Гін).
23. Елімінативізм як спроба розв'язання проблеми ментального (аналіз концепції).

24. Як обґрунтувати можливість знання про себе?
25. Війна на два фронти: емерджентна теорія Дж. Сьорла.
26. Поняття інтенційності: місток від мови до знання про себе.
27. Обґрунтування віртуальної природи «Я» (на прикладі к/ф «Матриця»).
28. Аргументи «мізки у діжці» і «болотяна людина».
29. Комп'ютерні метафори свідомості.
30. Дружній штучний інтелект: виклики і перспективи.
31. Епіфеноменалізм як антиредукціоністська теорія.
32. Кваліафоби та кваліафіли: основні тенденції полеміки.
33. Аргумент Мері та РобоМері: порівняння структури доказів.
34. Трансгуманізм як філософська доктрина.
35. Мінімальний дуалізм: проєкт порятунку дуалістичної теорії.
36. Слабкий штучний інтелект і «тест чайки».
37. Проблема інших умів з когнітивного та філософського поглядів.
38. Основні недоліки та переваги компатибілізму.
39. Неінтенційні ментальні стани як філософська проблема.
40. Критика аргументу з відсутніх квалій.

Вимоги до оформлення індивідуального письмового завдання (есею)

1) Есе повинно мати певну структуру: вступ, основна частина, висновки. Максимальний обсяг есе – 3 стор. (без урахування титульної сторінки), причому висновки і вступ не повинні перевищувати 1-2 абзаців. Основна частина тексту має дотримуватися трьох правил: 1) чітке формулювання тези; 2) логічна послідовність викладу; 3) аргументованість тез.

2) Параметри основного тексту: шрифт – Times New Roman; інтервал – 1,5 розмір кегля – 12.

3) Посилання (за потреби) – посторінкові, розмір шрифту – 10. Обсяг роботи не повинен перевищувати 10 тис. знаків із пробілами. Максимальна оцінка за виконане завдання – 20 балів.

Фінальний термін подання – 12 навчальний тиждень семестру.

ЛІТЕРАТУРА

1. Бостром Н. (2020). *Суперінтелект. Стратегії і небезпеки розвитку розумних машин*. Наш формат.
2. Вітгенштайн Л. (1995). *Tractatus Logico-Philosophicus*. *Філософські дослідження*. Основи, сс. 90–310.
3. Декарт Р. (2021). Медитації про першу філософію. *Метафізичні медитації. «Медитації» Декарта у світлі сучасних тлумачень: Жан-Марі Бейсад, Жан-Люк Марйон, Кім Сан Он-Ван-Кун*. Дух і Літера, сс. 115–292.
4. Декомб В. (2008). Суб'єктивний привілей. *Філософська думка*, 1: 84–100.
5. Деннет Д. (2017). Самість як центр ваги наративу. *Антологія сучасної аналітичної філософії, або жук залишає коробку*. Л.: Літопис, сс. 277–296.
6. Деннет Д. (2017). Що знає Робомері. *Антологія сучасної філософії науки, або Усмішка ASIMO*. Л.: ЛНУ ім. І. Франка, С. 414–446.
7. Джексон Ф. (2014). Епіфеноменальні кваліа. *Антологія сучасної аналітичної філософії, або Жук залишає коробку*. Л.: Літопис, сс. 260–272.
8. Комар О. (2016). Тесеєві мандри свідомості. *Філософська думка* 2: 57–70.
<https://dumka.philosophy.ua/index.php/fd/article/download/13/14/20>
9. Круглий стіл «Philosophy of Mind? – Mind the Gap!» (2016). *Філософська думка* 2: 6–29.
http://journal.philosophy.ua/sites/default/files/filefield_paths/5-philosophy_of_mind_-_mind_the_gap.pdf

10. Леонов А. (2015). «Аргумент зомбі» Девіда Чалмерса: передне слово перекладача. *Філософська думка* 5: 51–59. [Режим доступу]: <http://journal.philosophy.ua/article/argument-zombi-devida>.
11. Леонов А. Ю. (2014). Філософія свідомості Девіда Чалмерса. *Актуальні проблеми духовності*. 15: 216–237. – [Режим доступу]: http://download.kdpu.edu.ua/kaf_philosoph/AkProD/Zbirka15/Zb15_2014_216-237_Leonov.pdf.
12. Мід Дж. Г. (2000). *Дух, замість і суспільство з точки зору соціального біхевіориста*. Український Центр духовної культури, сс. 123–205.
13. Нагель Т. (2014). На що схоже бути кажаном? *Антологія сучасної аналітичної філософії, або жук залишає коробку*. Л.: Літопис, сс. 207–224.
14. Патнем Г. (2003). *Розум, істина й історія*. Альтернативи.
15. Сепетий Д.П. (2014). Свідомість, комп'ютери та новий мозок Мері. *Актуальні проблеми духовності* 15: 204–215. – [Режим доступу]: http://download.kdpu.edu.ua/kaf_philosoph/AkProD/Zbirka15/Zb15_2014_204-215_Sepety.pdf.
16. Сепетий Д. (2017). Свідомість як суб'єктивність: таємниця Я. Кн. 1. Зомбі, комп'ютери та Абсолютний Дух. – http://dsepetij.ho.ua/self_1.pdf
17. Сепетий Д. (2017). Свідомість як суб'єктивність: таємниця Я. Кн. 2. Привид у машині проти демона Лапласа. – http://dsepetij.ho.ua/self_2.pdf
18. Сепетий Д. (2016). Чому існує «важка проблема свідомості»? *Філософська думка* 2: 30–38. – [Режим доступу]: <http://journal.philosophy.ua/article/chomu-isnuye-vazhka>.
19. Серль Дж. (2017). Біологічний натуралізм. *Антологія сучасної філософії науки, або Усмішка ASIMO*. Л.: ЛНУ ім. І. Франка, сс. 476–494.
20. Серль Дж. (2014). Розум, мозки і програми. *Антологія сучасної аналітичної філософії, або жук залишає коробку*. Л.: Літопис, сс. 227–256.

21. Синиця А. (2017). *Сучасна аналітична філософія: від прагматики мови до концептуалізації свідомості*. Л.: ЛНУ ім. І. Франка.
22. Франкіш К. (2014). Аргумент анти-зомбі. *Антологія сучасної аналітичної філософії, або Жук залишає коробку*. Л.: Літопис, сс. 340–360.
23. Циба В. (2022). Біль як філософський концепт. *Наукові Записки НаУКМА. Філософія і релігієзнавство* 9-10: 34–48.
<http://nzpr.ukma.edu.ua/article/view/273428/268822>
24. Чалмерс Д. (2015). Аргумент 1: Логічна можливість зомбі. *Філософська думка*, 5: 60–67. <http://journal.philosophy.ua/article/argument-1-logiczna>.
25. Чалмерс Д. (2014). Сміливо зустрічаючи проблему свідомості. *Антологія сучасної аналітичної філософії, або жук залишає коробку*. – Л.: Літопис, сс. 298–336.
26. Черчленд П. С. (2017). Яким чином нейрони знають? *Антологія сучасної філософії науки, або усмішка ASIMO*. – Л.: ЛНУ ім. І. Франка, сс. 453 – 468.
27. Armstrong, D.M. (1968). *A Materialist Theory of the Mind*. Routledge & Kegan Paul, pp. 355–366.
28. Block, N. (1978). Troubles with Functionalism, in C. W. Savage (ed.), *Perception and Cognition: Issues in the Foundations of Psychology*, Minneapolis: University of Minnesota Press, pp. 267–325.
29. Brown, R. (2009). Terminating Ambiguity: The Perplexing Case of “The”, in: *Terminator and Philosophy* / ed. by Richard Brown and Kevin S. Decker. Wiley, pp. 253–265.
30. Brown, R. (2012). Zombies and Simulation. *Journal of Consciousness Studies* 19 (7-8): 21–25.
31. Chalmers, D. (2002). *Philosophy of Mind. Classical and Contemporary Readings*. Oxford University Press.
32. Chisholm, R. (1981). *The First Person. An Essay on Reference and Intentionality*. The Harvester Press.
33. Churchland, P.S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind/Brain*, Cambridge, MA: MIT Press.

34. Churchland, P.M. (1992). Knowing Qualia: A Reply to Jackson. In Churchland, P.M. *A Neurocomputational Perspective. The Nature of Mind and the Structure of Science*. Cambridge: MIT Press, pp. 67 – 76.
35. Churchland, P.S. (2007). The Evolving Fortunes of Eliminative Materialism. In *Contemporary Debates in the Philosophy of Mind*. B.P. MacLaughlin & J. Cohen (eds.). Blackwell Publishing, pp. 160–181.
36. Dennett, D. (1997). *Kinds of Minds: Towards an Understanding of Consciousness*. Basic Books, pp. 19–56, 153–160.
37. Dennett, D. (1987). *The Intentional Stance*. MIT Press.
38. Dennett, D. (2006). What RoboMary Knows. In Torin Andrew Alter & Sven Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press, pp. 14–31.
39. Descombes, V. (2001). *The Mind's Provisions. A Critique of Cognitivism*. Princeton University Press.
40. Davidson, D. (2001). Mental Events, in: Davidson D. *Essays on Actions and Events*. Clarendon Press.
41. Dennett, D. (1991). *Consciousness Explained*. Back Bay Books.
42. Dreyfus, G. (2001). Self and Subjectivity: A Middle Way Approach, in: *Self, No-Self? Perspectives from Analytical, Phenomenological, and Indian Traditions*, ed. by M. Siderits, E. Thompson, and D. Zahavi. Oxford University Press, pp. 114–156.
43. Feyerabend, P. (1999). *Knowledge, Science and Relativism: Philosophical Papers, Vol.3*. J. Preston (ed.). Cambridge University Press, pp. 16–49.
44. Gulick, P. (2017). Functionalism and Qualia, In S. Schneider & M. Velmans (eds.). *The Blackwell Companion to Consciousness*, John Wiley & Sons Ltd, pp. 430–443.
45. Hacker, P.M.S. (1997). Wittgenstein on Human Nature. Phoenix.
46. Heil, J. (2020). *Philosophy of Mind. A Contemporary Introduction*. Routledge.
47. Fodor, J. (1975). *The Language of Thought*. Thomas Y. Crowell & Co.

48. Kripke, S. A. (1971). Identity and Necessity. In *Identity and Individuation*, M. K. Munitz (ed.). New York University Press, pp. 135–164.
49. Leonov A. (2017). Facing up to David Chalmers' philosophy of mind: the general overview. *Філософська думка 1*. – [Electronic resource]: <http://journal.philosophy.ua/article/facing-david-chalmers>.
50. Lewis D. (2004). An Argument for the Identity Theory, In: Heil J. *Philosophy of Mind. A Guide and Anthology*. Oxford University Press, 2004, pp. 150–157.
51. Malcolm N. (1988). Subjectivity. *Philosophy 63 (244)*: 147–160.
52. McGinn C. (1996). *The Character of Mind. An Introduction to the Philosophy of Mind*. Oxford University Press.
53. McGinn C. (1989). Can We Solve the Mind-Body Problem? *Mind Vol. XCVIII (391)*: 349–366.
54. Marcus, G.F., Davis, E. (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. Pantheon; Random House.
55. Peacocke P. (2007). Mental Action and Self-Awareness (I), in: *Contemporary Debates in the Philosophy of Mind* / ed. by B. McLaughlin and J. Cohen. Blackwell Publishing, pp. 358–376.
56. Searle, J.R. (1990). Is the Brain a Digital Computer? – [Electronic resource]: <http://users.ecs.soton.ac.uk/harnad/Papers/Py104/searle.comp.html>.
57. Searle, J. R. (2004). *Mind: A Brief Introduction*. Oxford University Press.
58. Searle, J.R. (1984). In: *Minds, Brains and Science*. Harvard University Press, pp. 26–40.
59. Searle, J.R. (2004). *The Rediscovery of the Mind*. Oxford University Press.
60. Shanker, S. (1998). Wittgenstein's Remarks on the Foundations of AI. Routledge.
61. Shoemaker, S. (1975). Functionalism and Qualia. *Philosophical Studies 27*: 291–315.
62. Shoemaker, S. (2011). Self-Reference and Self-Awareness, In: *Self-Reference and Self-Awareness* / ed. by A. Brook, R. C. DeVidi. John Benjamins Publishing Co, pp. 81–93.

63. Shoemaker, S. (1996). *The First-Person Perspective and Other Essays*. Cambridge University Press, pp. 176–197.
64. Smart, J. (1959). Sensations and Brain Processes. *The Philosophical Review*, 68 (2): 141–156.
65. Strawson, G. (2009). *Selves. An Essay in Revisionary Metaphysics*. Clarendon Press.
66. Ryle, G. (1949). *The Concept of Mind*. Hutchinson House.
67. Rorty, R. (2004). The Brain as Hardware, Culture as Software. *Inquiry* 47: 219–235.
68. *The Blackwell Guide to Philosophy of Mind* (2003). Ed. by Stephen P. Stich and Ted A. Warfield. Blackwell Publishing.
69. Wittgenstein, L. (1958). *The Blue and Brown Book*. Blackwell Publishing.
70. Wittgenstein, L. (1953). *Philosophical Investigations*. Double-Language Edition. Blackwell Publishing.
71. Zahavi, D. (2005). *Subjectivity and Selfhood. Investigating the First-Person Perspective*. The MIT Press; A Bradford Book.

Джерела в мережі інтернет

Особиста сторінка Д. Чалмерса:

<https://consc.net/>

Особиста сторінка Т. Негела:

<https://as.nyu.edu/faculty/thomas-nagel.html>

Особиста сторінка К. Франкіша:

<https://www.keithfrankish.com/>

Особиста сторінка Д. Сепетого:

<http://dsepetij.ho.ua/index.htm>

Особиста сторінка Н. Бострома:

<https://nickbostrom.com/>

Особиста сторінка М. Тая:

<https://utexas.academia.edu/MichaelTye>

Особиста сторінка Г. Маркуса:

<http://garymarcus.com/>

Сторінка Центру дослідження суб'єктивності (Копенгаген, Данія):

<https://cfs.ku.dk/>

Сторінка Центру когнітивних досліджень ун-ту Тафтса (США):

<https://sites.tufts.edu/cogstud/2023/03/28/home/>

Оцінювання знань і вмінь

Форма контролю – залік. Оцінювання знань і вмінь студентів, набутих під час вивчення дисципліни «Сучасна філософія свідомості», здійснюється за загальноуніверситетською рейтинговою системою. Підсумкова рейтингова оцінка з дисципліни (максимум – 100 балів) є комплексною характеристикою компетентності студента, яка враховує результати поточної семестрової роботи (максимум – 70 балів) і підсумкового контролю (максимум – 30 балів).

Формами поточного контролю є оцінювання відповідей студентів на семінарських заняттях та індивідуальних письмових робіт, підготовлених ними упродовж семестра. Оцінка за роботу в семестрі є сумою балів, отриманих студентом за відповіді на семінарських заняттях та за виконання письмової роботи (есею). Поточний контроль передбачає оцінювання кожного виду робіт у відповідності до умов визначення навчального рейтингу та чітко визначених критеріїв оцінювання (див. нижче). Підсумковий контроль з дисципліни здійснюється у формі письмового тестування, що передбачає оцінювання рівня засвоєння студентами всього обсягу програмного матеріалу. Тест складається із 15 питань. Максимальна оцінка за виконання – 30 балів. Тривалість тестування – 30 хвилин. Під час його проведення користуватися будь-якою літературою не дозволяється.

Поточний та підсумковий контроль знань передбачає оцінювання кожного виду робіт, зазначених в «Умовах визначення навчального рейтингу», та здійснюється у відповідності до чітко визначених критеріїв.

Умови визначення навчального рейтингу

Робота в семестрі – 100 балів

Види робіт	Їх кількість	Максимальний бал за роботу	Загальна кількість балів
Семінарське заняття	6	8	48
Індивідуальна письмова робота	1	22	22

Контрольне тестування	1	30	30
Разом:			100

Результати і критерії оцінювання

Критерієм успішного проходження здобувачем освіти підсумкового оцінювання може бути досягнення ним мінімальних порогових рівнів оцінок за кожним запланованим результатом навчання навчальної дисципліни.

Оцінювання успішності студента за кожним із запланованих видів робіт здійснюється у відповідності до таких критеріїв:

Види робіт	Кількість балів	Критерії оцінювання
Семінарське заняття	1 – 3	Доповнення
	4 – 6	Відповідь із використанням основної літератури
	7 – 8	Відповідь із використанням основної та додаткової літератури
Контрольне тестування	1 – 30	Відповідь без використання основної та додаткової літератури
Індивідуальна письмова робота	0	Завдання не виконане у визначений викладачем термін або містить плагіат.
	1 – 6 балів	Завдання виконане не повною мірою; автор демонструє слабке знання програмного матеріалу дисципліни; не виявляє навичок аналізу і логічного та послідовного викладу; припускається численних помилкових суджень; робота не вчитана й оформлена неналежним чином.
	7 – 11 балів	Завдання виконане не повною мірою; автор демонструє не дуже глибоке знання програмного матеріалу дисципліни; виявляє певний брак навичок аналізу і логічного та послідовного викладу; подекуди припускається помилкових суджень.

	12 – 18 балів	Завдання виконане вчасно. Автор демонструє достатню обізнаність із програмним матеріалом дисципліни та вміння логічно і послідовно викладати його зміст. Разом із тим, роботі бракує аналітичного характеру та (або) автор достатньою мірою не проявив здатності до самостійного мислення.
	19 – 22 балів	Завдання виконане вчасно і повною мірою; автор демонструє належний рівень знання і розуміння програмного матеріалу дисципліни; виявляє аналітичні здібності, здатність до самостійного, системного, логічного і послідовного мислення. Робота вчасно подана й оформлена у відповідності до вимог.

Порядок перерахунку рейтингових показників нормованої 100-бальної університетської шкали оцінювання в національну 4-бальну шкалу та шкалу ECTS

За шкалою університету	За національною шкалою		За шкалою ECTS
	Екзамен	Залік	
91 – 100	відмінно	зараховано	A (відмінно)
81 – 90			B (дуже добре)
71 – 80	C (добре)		
66 – 70	D (задовільно)		
60 – 65	E (достатньо)		
	добре		
	задовільно		

30 – 59	незадовільно	не зараховано	FX (незадовільно – з можливістю повторного складання)
1 – 29			F (незадовільно – з обов'язковим повторним курсом)

У разі отримання оцінки «незадовільно» студент має право на два перескладання: викладачеві та комісії. При цьому максимальна підсумкова оцінка після перескладання може бути лише «достатньо». Замість перескладання комісії студент може обрати повторне вивчення дисципліни. **У разі отримання оцінки «неприйнятно»** студент зобов'язаний повторно вивчити дисципліну.

Рекомендовані джерела інформації

Рекомендована для вивчення дисципліни література та інші джерела інформації доступні здобувачам освіти на безоплатній основі, зокрема у бібліотеці НаУКМА (див.: <http://www.library.ukma.edu.ua/>); на електронних ресурсах НаУКМА та у електронних виданнях безкоштовного доступу.

Abstract

The study guide pays special attention to such relevant topics as the main trends in debates between reductionist and anti-reductionist theories, an analysis of the mind-body problem, a survey of the nature of qualia, a formal reconstruction of the arguments and counterarguments of the philosophers of mind, the problem of artificial intelligence and the related dilemmas within moral philosophy. After Wittgenstein criticized the private language pattern, and his successor Ryle spoke of mind as a category mistake, it became clear that the Cartesian subjective privilege forever lost its charming force. The ascent of some new approaches, in particular physicalist identity theory, proved that the mental is the subject of comprehensive philosophical analysis. Therefore, over the past five decades, philosophy of mind has made a leap from the problem of qualia through the conceivability principle to the problem of strong artificial intelligence. Such important topics for epistemology as causality, self-knowledge, and knowledge about the world, intentional and non-intentional states were extensively investigated. The modal arguments in logic, on which the most famous thought experiments were based, as well as tension between determinism and indeterminism pertaining to free decision act, made post-analytic philosophy possible. The elaboration of consciousness shows that philosophy has a good chance to overcome its previous sectoral fragmentation.

Key words: contemporary philosophy, analytic philosophy, mind, consciousness, mind-body problem, qualia, language, private language argument, Chinese room argument, artificial intelligence, Turing Test, Identity theory, materialism, monism, mind-body dualism, functionalism, philosophical naturalism, epistemology, free will, logical behaviorism, philosophy.