WPiK Open Access

24

Eksperymenty myślowe
Jak testujemy rzeczywistość?

Michał Wyrwa

Poznań 2025

24 Eksperymenty myślowe
Jak testujemy rzeczywistość?

WPiK Open Access

24

Eksperymenty myślowe

Jak testujemy rzeczywistość?

Michał Wyrwa

Poznań 2025

Komitet Naukowy
Jerzy Brzeziński, Agnieszka Cybal-Michalska
Zbigniew Drozdowicz (przewodniczący), Rafał Drozdowski
Piotr Orlik, Jacek Sójka

Recenzja
ks. prof. dr hab. Józef Bremer SJ, Uniwersytet Ignatianum w Krakowie

Redakcja naukowa
prof. UAM dr hab. Aleksandra Pilarska

Projekt okładki
Wydział Psychologii i Kognitywistyki UAM

Redakcja i korekta
Klaudia Kulmińska

łamanie
Izabela Baran

Wydanie I

Publikacja finansowana z funduszy
Wydziału Psychologii i Kognitywistyki UAM

© Copyright by Wydawnictwo Nauk Społecznych i Humanistycznych
Uniwersytetu im. Adama Mickiewicza w Poznaniu 2025
© Copyright by Michał Wyrwa 2025

978-83-66983-48-9 ISBN (numer tomu 24)
978-83-66983-22-9 ISBN (numer kolekcji)
https://doi.org/10.14746/9788366983489

Wydawnictwo Nauk Społecznych i Humanistycznych
Uniwersytetu im. Adama Mickiewicza w Poznaniu
60-568 Poznań, ul. Szamarzewskiego 89c
www.wnsh.amu.edu.pl, wnsh@amu.edu.pl, tel. (61) 829 22 54

SPIS TREŚCI

Wprowadzenie 7
1. Eksperyment myślowy jako działanie poznawcze 11
   1.1 Warunki wyobrażania jako eksperymentowania 13
   1.2 Przekonujące działania: przykłady historyczne 15
   1.3 Podsumowanie 22
2. Eksperymentowanie w świecie i w myślach 23
   2.1 Empiryzm bez pomiaru i kreślenie sytuacji wyjątkowej 23
   2.2 Sytuacja wyjątkowa w eksperymencie myślowym ze spadkiem swobodnym Galileusza 25
   2.3 Rola eksperymentów myślowych w nauce XIX i XX wieku 27
   2.4 Współczesne kategoryzacje eksperymentów myślowych 31
   2.5 Światy literackie jako eksperymenty myślowe 34
   2.6 Podsumowanie 36
3. Wyobraźnia: myślenie o tym, co nieobecne 39
   3.1 Wyobraźnia percepcyjna a pojęciowa 40
   3.2 Funkcjonalne ujęcie wyobraźni 42
   3.3 Neuronalne podstawy funkcjonalnie rozumianej wyobraźni 44
      3.3.1 Sieć domyślna i jej podsystemy 44
      3.3.2 Badania przypadków: afantazja, amnezja, demencja 46
      3.3.3 Wartość wyobraźni 47
   3.4 Myślenie warunkowe 49
   3.5 Granice wyobraźni 52
      3.5.1 Spójność jako warunek wyobrażalności 52
      3.5.2 Pojmowalność i jej gradacje 54
   3.6 Podsumowanie 55
4. Mechanizm poznawczy eksperymentu myślowego 57
   4.1 Modele mentalne 59
   4.2 Co wyróżnia eksperymenty myślowe jako modele mentalne? 64
   4.3 Podsumowanie 66

5. Wiarygodność eksperymentów myślowych a intuicja 67
5.1 Kreślenie możliwej sytuacji 70
5.2 Treściowy związek 74
5.3 Struktura wnioskowania w eksperymencie 76
5.4 Wiarygodność intuicji 80
5.5 Podsumowanie 88
6. Studium filozoficznych zombie 91
6.1 Fizykalizm a trudny problem świadomości 92
6.2 Eksperyment myślowy z zombie 95
6.3 Zastrzeżenia co do pojmowalności zombie 99
6.3.1 Słownik i przekonania a częściowy zombie 99
6.3.2 Zombie niczego nie brakuje 102
6.4 Jeśli nawet pojmowalne, to i tak niemożliwe 107
6.5 Przydatność eksperymentu myślowego z zombie 110
Zakończenie 113
Bibliografia 115

WPROWADZENIE

Większość z nas zgodzi się, że przy planowaniu działań warto polegać na solidnych podstawach, takich jak opinie ekspertów, własne doświadczenie czy społecznie wypracowane normy. Szczególnie cenimy sobie wiarygodność postępowania naukowego, testując hipotezy za pomocą eksperymentów. Eksperyment naukowy jest działaniem specyficznym, bo obwarowanym ścisłą metodologią postępowania. Jednak nauka (przyrodoznawcza i ścisła, *science*) to nie tylko eksperymenty empiryczne. Żaden eksperyment nie uzasadnia bezpośrednio kolejnego. Namysł teoretyczny jest niezbędny, aby zadawać właściwe pytania i interpretować wyniki. Bez teorii nie byłoby postępu w nauce.

Obraz filozofii jako czysto fotelowych rozważań jest podobnym uproszczeniem. Czysto teoretyczna filozofia oczywiście dalej istnieje, ale dla celów filozoficznych przeprowadza się również badania empiryczne. W ostatnich latach na popularności zyskał program takiej właśnie filozofii eksperymentalnej i bynajmniej nie jest to pomysł całkiem nowy. Już w pierwszej połowie XX wieku, wychowany w neopozytywistycznej tradycji Koła Wiedeńskiego Arne Næss (1938) „filozofią eksperymentalną" nazywał swoje ankietowe badania semantyki (znaczeń) języka potocznego (zob. Chapman, 2017). Dziś filozofia eksperymentalna to dynamicznie rozwijający się obszar badawczy, budujący mosty między filozofią a psychologią i innymi naukami, obejmujący tysiące artykułów w prestiżowych czasopismach, udowadniających, że można wykorzystać metody empiryczne do analizy tez filozoficznych[1]. Rozróżnienie naukowca od filozofa jest zresztą względnie nowe, wykrystalizowało się dopiero w XIX wieku (Ross, 1962). Współczesną naukę możemy

[1] Rynek publikacyjny filozofii eksperymentalnej do grudnia 2024 roku obejmował ponad 3300 prac w bazie PhilPapers, liczne serie książkowe (jak *Oxford Studies in Experimental Philosophy*), kompendia (Sytsma i Buckwalter, 2016) oraz podręczniki do projektowania badań dla filozofów (Kornmesser i in., 2024; Sytsma i Livengood, 2015). Bibliometryczną analizę kondycji filozofii eksperymentalnej można znaleźć w pracy Jincai Li i Xiaozhen Zhu (2023).

wręcz traktować jako zaawansowaną formę filozofii naturalnej, podlegającej ścisłym standardom metodologicznym.

Książka ta jest próbą opisania pewnego systematycznego sposobu myślenia, charakterystycznego tak dla filozofów, naukowców, jak i ludzi niezwiązanych z akademią – eksperymentowania w myślach. Osoby zainteresowane historią nauki, filozofią czy kognitywistyką znają zapewne co najmniej kilka przykładów, jak zwykł je nazywać Daniel Dennett (2015), dźwigni wyobraźni. Kanoniczny dla nowożytnej filozofii proces wątpienia zawarty w *Medytacjach* Kartezjusza jest typowo filozoficznym eksperymentem myślowym. Wchodząc zaś w świat kognitywistyki, także zaznajamiamy się z eksperymentami myślowymi o filozoficznym rodowodzie: o neurobiolog Mary, która wychowana w czarno-białym środowisku nie doświadczyła nigdy koloru, zombie i chińskim pokoju. Sama praktyka eksperymentowania myślowego znana jest od starożytności i choć ma duże znaczenie dla rozwoju nauk, obecna jest nawet w popkulturze, chociażby w postaci figury kota Schrödingera. Jest działaniem przydatnym wszędzie tam, gdzie chcemy zrozumieć czyjeś argumenty, przekonać innych lub przeanalizować pewne sytuacje bez bezpośredniego w nich uczestnictwa. Stąd też często stosuje się je w szeroko pojętej edukacji (np. Stephens i Clement, 2012).

Praktyka eksperymentów myślowych bazuje na wyobraźni, pozwala wyciągać wnioski z opisu wyjątkowych, zaskakujących przypadków. I choć w ostatnich latach myślenie opierające się na rozważaniu konsekwencji takich przypadków i w ogóle wszelkie czysto fotelowe rozważania są krytykowane tak w dyskusji akademickiej (np. Machery, 2017; Suhler, 2019), jak i w obrębie sfery publicznej przez popularyzatorów nauki i znanych intelektualistów (np. Andersen, 2012; Nagel, 2010; Pigliucci, 2014)[2], to prawidłowo skonstruowany i przeprowadzony eksperyment myślowy jest nie tylko potężnym narzędziem retorycznym. Jak zobaczymy, może mieć też istotną wartość poznawczą.

Współcześnie tematyka ta zajmuje przede wszystkim filozofów. W zasadzie wszystkie wydane w ostatnich dwóch dekadach książki o eksperymentach myślowych są ich autorstwa, przykładowo *The Routledge Companion to Thought Experiments* (Stuart i in., 2018a), *Dźwignie wyobraźni i inne narzędzia do myślenia* (Dennett, 2015), *Epistemic Thought Experiments and Intuitions*

[2] O zniekształceniach obrazu nauki i namysłu teoretycznego wśród popularyzatorów i akademików pisali m.in. Maarten Boudry i Massimo Pigliucci (2018) oraz Jeroen de Ridder i inni (2018).

(Hamdo, 2023), *Thought Experiments* (Miščević, 2022) czy *Thought Experiments, Science, and Theology* (Fehige, 2023). Taki stan rzeczy jest o tyle niefortunny, że eksperymenty myślowe były i są istotne także dla przedstawicieli innych dyscyplin akademickich. W zasadzie my wszyscy możemy – i najpewniej zdarza się tak w codziennym życiu – eksperymentować w myślach. Nie jest to hermetyczne narzędzie dostępne tylko dla wybrańców. Dzisiejsze znaczenie eksperymentów myślowych jest być może dokładnie odwrotne. Wraz z rosnącym zrozumieniem czynników biorących udział w realizowaniu eksperymentów myślowych, mogących skłaniać nas do akceptowania nieuzasadnionych twierdzeń, zarazem możemy nabrać większej rezerwy do ich stosowania, jak i docenić ich elegancję w codziennym przekonywaniu siebie nawzajem.

Kilka wątków dominuje w akademickiej dyskusji. Z uwagi na istotną rolę, jaką przynajmniej niektóre eksperymenty myślowe odegrały w historii nauki, często podnoszoną kwestią jest to, czy taka procedura myślenia prowadzi do uzyskania nieposiadanej wcześniej wiedzy. Przecież nie ruszamy się z przysłowiowego fotela, na ile więc możemy dzięki niej powiedzieć coś nowego o świecie? Wiele prac dotyczy także aspektów historycznych: Kto, kiedy i gdzie praktykował eksperymenty myślowe jako pierwszy? Które słynne argumenty w historii filozofii i nauki faktycznie są eksperymentami myślowymi? Najwięcej zaś napisano o eksperymentach myślowych, odzierając je z narracji i analizując ich poprawność jako układów przesłanek i wniosków. W niniejszej książce odnoszę się również i do tych kwestii.

Oprócz syntezy aktualnego stanu debaty i przedstawienia historii praktyki przyglądam się też krytycznie kilku kluczowym pytaniom:

1. Czym eksperyment myślowy różni się od innych form rozumowania i czy etykieta „eksperymentu" w nazwie jest uzasadniona?
2. Jaki jest mechanizm poznawczy eksperymentowania w myślach?
3. Na czym polega rola wyobraźni i intuicji w eksperymentach myślowych?
4. Czy eksperymenty myślowe mogą być wiarygodnym źródłem wiedzy?

Rzeczone eksperymentowanie odbywa się w umyśle, stąd w trakcie poszukiwania odpowiedzi na te pytania adekwatne wydaje mi się przyjęcie naturalistycznej perspektywy kognitywistycznej. Pokazuję, że eksperyment myślowy to nie tyle abstrakcyjna konstrukcja filozoficzna, ile szczególny przypadek działania poznawczego opartego na wyobraźni, nakierowanego na lepsze zrozumienie wybranego aspektu rzeczywistości. Treść kolejnych rozdziałów napisano tak, aby była jak najbardziej zrozumiała dla studentów i pozostałych czytelników bez dogłębnej znajomości filozofii, psychologii oraz innych

nauk poznawczych. Mam nadzieję, że lektura zachęci takie osoby do sięgnięcia także po bardziej szczegółowe prace, do których liczne odwołania znaleźć można w przypisach. Ci, którzy mają podstawową wiedzę dotyczącą działania ludzkiego umysłu i są otwarci na pojęcia filozoficzne, powinni bez większej trudności odnaleźć się w poruszanej tu tematyce. Jednak niektórych kwestii nie sposób nie wypaczyć bez polegania na technicznym żargonie. Mam nadzieję, że i te treści udało się omówić przejrzyście.

Chociaż odnoszę się w książce do historycznych przykładów eksperymentów myślowych, nie jest ona wyczerpującym studium historycznym. Uwagi te mają natomiast na celu przybliżyć kontekst i umożliwić lepsze zrozumienie specyfiki eksperymentów myślowych i ich znaczenia dla ludzkiego myślenia oraz poznawania świata[3].

Książkę rozpoczyna rozdział zarysowujący ideę eksperymentu myślowego jako działania poznawczego, w tym jego historyczne korzenie w filozofii europejskiej. W rozdziale 2 wyjaśniam, dlaczego eksperyment myślowy jest eksperymentem, śledząc podobieństwa i różnice pomiędzy obiema formami aktywności poznawczej w odwołaniu do zarówno rozwoju metody eksperymentalnej, jak i poglądów naukowców i filozofów na temat ról pełnionych przez eksperymenty myślowe. W rozdziale 3 bronię stanowiska, zgodnie z którym wyobraźnia stanowi bardzo obszerny zbiór aktywności umysłowych. Po jego lekturze powinno być jasne, że nie powinniśmy omijać w analizie eksperymentu jego narracji. Rozdział 4 przybliża najpoważniejszą w momencie pisania tej książki propozycję mechanizmu poznawczego wyjaśniającego eksperymentowanie w myślach: teorię modeli mentalnych. Rozdział 5 dotyczy kryteriów wiarygodności eksperymentów myślowych, roli intuicji i tego, dlaczego i kiedy powinniśmy jej ufać. Ostatni rozdział jest poświęcony jednemu z najsłynniejszych współczesnych eksperymentów myślowych: filozoficznym zombie. Jest to swego rodzaju studium przypadku, którego analiza korzysta z ustaleń wcześniejszych rozdziałów i pozwala lepiej zrozumieć nie tylko złożoność kognitywistycznych problemów naukowych, ale także rolę eksperymentów myślowych na przecięciu myślenia potocznego, naukowego i filozoficznego.

Eksperymenty myślowe nie mają dziś być może dobrej passy, ale gdy je się odczaruje, to jasne jest, że są skutecznym narzędziem poznawczym.

[3] Pełniejsze szkice historyczne czytelnik może znaleźć w pracach Roberto Nitschego (2020), Kateriny Ierodiakonou i Sophie Roux (2011b) oraz w pierwszej części pracy pod redakcją Michaela Stuarta i innych (2018a).

1 EKSPERYMENT MYŚLOWY JAKO DZIAŁANIE POZNAWCZE

Zgodnie z główną ideą eksplorowaną w tej książce eksperymenty myślowe stanowią specyficzny rodzaj aktywności poznawczej. Aby pełniej docenić ich wyjątkowość, warto umieścić je w kontekście innych działań umysłowych. Bez nich niemożliwe byłoby osiąganie podstawowych celów takich jak orientowanie się w aktualnej sytuacji w otoczeniu, unikanie zagrożenia, znajdywanie pokarmu czy budowanie relacji społecznych. Różne sytuacje wymagają różnych działań poznawczych. Przykładowo, wykorzystujemy proces widzenia – percepcji wzrokowej – by ocenić, czy można bezpiecznie przejechać przez przejazd kolejowy. Moglibyśmy polegać na swojej pamięci, znajomości rozkładu jazdy pociągów, ale byłoby to bardziej zawodne: pociąg zawsze może się spóźnić, możemy też znać nieaktualny rozkład. Bezpieczniej jest polegać na wzroku.

Widzenie pozwala bezpiecznie przechodzić przez przejazdy kolejowe, lecz wymaga sprawnych oczu i stosownego sygnału świetlnego padającego na siatkówkę, jest więc ograniczone rodzajem informacji. Są jednak aktywności poznawcze, które takiego ograniczenia nie mają. Wspominanie może obejmować przywoływanie w umyśle obrazów, dźwięków i smaków towarzyszących sytuacjom z naszej przeszłości, może też być tych obrazów pozbawione. Mogę pamiętać, na kogo oddałem głos w ostatnich wyborach, ale nie muszę mieć dostępu do konkretnych wrażeń zmysłowych, które towarzyszyły mi w lokalu wyborczym. Rozumowanie, podobnie jak pamięć, jest wszechobecne w naszym codziennym życiu i kluczowe dla treści tej książki. Powiemy, że nasze myśli są efektami rozumowania, kiedy są uzasadnione przez poprzednie myśli. Na przykład, gdy pozytywnie oceniamy swój wybór koncertów na podstawie wspomnień z dwóch ostatnich, na których byliśmy. Rozumujemy zarówno spontanicznie, jak i celowo, w odniesieniu do realizowanych zadań. Z pierwszym mamy do czynienia, gdy po spojrzeniu w swoje odbicie w lustrze mimowolnie ocenimy, że nasze poranne zniechęcenie ma więcej związku z widocznym na twarzy zmęczeniem niż zielonym kolorem

włożonej bluzki, z drugim zaś, kiedy nauczycielowi pytającemu nas, co wynika z (dość pokracznych) obserwacji, że wszyscy blondyni są ssakami, a my jesteśmy blondynami, odpowiadamy, iż to, że również jesteśmy ssakami.

Uzasadnianie może przybierać różne formy. Możemy zwrócić uwagę na logiczną strukturę rozumowania[4]. Stwierdzimy wtedy, że rozumowanie o koncercie było indukcyjne (myśli o poszczególnych koncertach, na których byliśmy, stanowiły uzasadnienie dla oceny, że dobrze wybieramy dla siebie koncerty, skorzystaliśmy więc z indukcji enumeracyjnej niezupełnej), o senności abdukcyjne (wyjaśniamy to, co już wiadome: zobaczyliśmy swoją twarz i oceniliśmy, że to brak odpoczynku najlepiej wyjaśnia jej stan), a o byciu ssakiem dedukcyjne (niezawodne, jeśli tylko treści naszych myśli o wszystkich blondynach i o kolorze naszych włosów były prawdziwe, to nasza myśl o byciu ssakiem również była zgodna z rzeczywistością). Choć dedukcja jest najbardziej pewna, to na co dzień uznajemy też uprawdopodabniające myślenie indukcyjne i spójne z posiadanymi już przekonaniami hipotezy abdukcyjne.

Jeśli przyjęlibyśmy, że w rozumowaniach liczą się tylko świadome myśli, to zniekształcilibyśmy, jak ludzie faktycznie rozumują. Jeżeli treść tylko takich właśnie myśli ma znaczenie – możemy tę treść traktować jako ten aspekt stanu myślowego, który wyraża sąd, przykładowo myśli o chęci zjedzenia jabłka odpowiadałby sąd „chcę zjeść jabłko" – to nie wyjaśnimy, dlaczego nie zawsze wyciągamy najbardziej uzasadnione wnioski. Uwzględnienie tych dodatkowych zmiennych, zarówno zewnętrznych (środowisko, ludzie, kultura), jak i wewnętrznych (psychika, skłonność do myślenia heurystycznego i inne) pozwala lepiej zrozumieć, które rozumowania uznajemy potocznie za bardziej przekonujące niż inne (Evans, 2017; Pinker, 2021; Tversky i Kahneman, 1983). Wiele z tych wewnętrznych procesów prowadzących do przekonujących wniosków można określić jako intuicje, czyli szybkie, heurystyczne sposoby kierujące nas od pewnych informacji – przesłanek, do których nie mamy pełnego świadomego dostępu – do świadomych wniosków.

Natomiast rolę zewnętrznych czynników sprzyjających rozumowaniu doskonale rozumieją naukowcy: badania nie są nieistotnym, zewnętrznym i pomijalnym elementem rozumowania naukowców, mają one bezpośrednie przełożenie na to, co i jak myślą na temat obszaru ich zainteresowania. Tezy naukowe uzasadniane są poprawnie zanalizowanymi wynikami uzyskiwanymi

[4] Klasyczne wprowadzenie w rodzaje rozumowań znaleźć można w pracy Kazimierza Ajdukiewicza (1985).

w trakcie metodologicznie dobrze zaprojektowanych i realizowanych badań. Przeprowadzanie empirycznego eksperymentu naukowego jest nieodłącznym elementem naukowego postępowania.

W wielu przypadkach, także w nauce, interesuje nas teoretyczne rozważenie konsekwencji pewnych założeń czy twierdzeń. Tu przydatne okazują się eksperymenty myślowe, które w pewnych aspektach przypominają eksperymenty empiryczne. Są aktywnością nakierowaną na poznanie, a dokładniej rozumowaniem opierającym się na wyobraźni. Nie chodzi jednak o ich opis, podobnie jak nie sposób sprowadzić eksperymentu empirycznego do jego opisu. Zwykle analizujemy opis eksperymentu myślowego, ale żeby eksperyment myślowy naprawdę zaistniał, musi być aktywnością umysłową, a nie jedynie jej opisem. Argumentacyjna struktura eksperymentu myślowego nie musi też być jawna dla osoby go przeprowadzającej. I choć eksperymenty myślowe często są redukowane do postaci układu przesłanek i wniosków (Norton, 2004a; zob. Brown, 2004) – co tłumaczy, dlaczego więcej można się o nich dowiedzieć, wpisując w bazy tekstów akademickich „argument z", a nie „eksperyment myślowy z" – to nie każde rozumowanie jest eksperymentem myślowym. Tak jak empiryczny eksperyment naukowy jest aktywnością nakierowaną na poznanie i wykorzystującą procedurę realizowaną w rzeczywistości, tak eksperyment myślowy jest rozumowaniem opartym na procedurze realizowanej w wyobraźni. Polega ona na wyobrażeniu sobie fikcyjnej sytuacji oraz rozważeniu jej konsekwencji, często prowadząc do wniosków na poziomie intuicyjnego wglądu, a więc na nagłym zrozumieniu, uświadomieniu sobie, że zgadzamy się z konkluzją sugerowaną przez autora eksperymentu.

W tym rozdziale przyglądam się bliżej temu, co wyróżnia eksperymenty myślowe na tle innych konstrukcji wyobraźni. Osią moich rozważań jest zaś historia eksperymentów myślowych w europejskiej filozofii.

1.1 WARUNKI WYOBRAŻANIA JAKO EKSPERYMENTOWANIA

Zastanówmy się nad różnicą między eksperymentem myślowym a innym użyciem wyobraźni. Sięgnijmy w tym celu po alegorię jaskini z VII księgi *Państwa* Platona. Często jest ona mylnie identyfikowana jako eksperyment myślowy, tak w tekstach popularnych (np. D'Olimpio, 2019), jak i akademickich (np. Østergaard, 2019). Sokrates opisuje Glaukonowi ludzi przykutych w jaskini, którzy mogą widzieć jedynie cienie przedmiotów rzucane na ścianę przez niewidziane przez nich ognisko. Cienie te wydają się im prawdziwymi

przedmiotami. Po przedstawieniu tej fikcyjnej sytuacji Sokrates wyjaśnia, dlaczego poznanie prawdy, uwolnienie się z kajdanów i jaskini jest bolesne, ale też wartościowe:

> […] ci ludzie tam [w łańcuchach] nie co innego braliby za prawdę, jak tylko cienie pewnych wytworów. […] ile razy by kto został wyzwolony i musiałby zaraz wstać i […] patrzeć w światło, cierpiałby. […] Dopiero na końcu, myślę, mógłby patrzeć w słońce […]. Potem by sobie wymiarkował o nim, że […] ono rządzi wszystkim w świecie widzialnym, i że jest w pewnym sposobie przyczyną także i tego wszystkiego, co oni tam poprzednio widzieli (Platon, 2010, 515c–516g).

By zrozumieć, dlaczego fragmenty o jaskini nie tworzą eksperymentu myślowego, odwołajmy się do kryteriów Alexandra Beckera (2018, s. 48–51). Jego zdaniem eksperyment myślowy musi spełniać trzy warunki (zob. Nitsche, 2020, s. 106–108):

1. Musi pełnić funkcję epistemiczną – fikcyjna sytuacja powinna mówić coś o rzeczywistym świecie.
2. Narracja eksperymentu myślowego musi być argumentacyjnie niezbędna: nie może być jedynie ilustracją, musi być potrzebna, aby wykazać konkluzję eksperymentu.
3. Eksperyment musi być rozstrzygalny, to znaczy zawarte w nim informacje muszą wystarczyć, żeby czytelnik był w stanie ocenić, czy zgadza się z jego konkluzją.

Alegoria jaskini symbolicznie przedstawia proces poznawania prawdy, relację między światem zmysłowym a światem idei, a nawet edukacyjno-polityczny wymiar filozofii platońskiej (osoby już wyzwolone winny pomagać innym uwalniać się z kajdan). Czytelnik ma poszczególne jej aspekty odnosić do świata rzeczywistego. Platon sam pisze, że jeśli „wychodzenie pod górę i oglądanie tego, co jest tam wyżej […], weźmiesz za wznoszenie się duszy do świata myśli, to nie zbłądzisz" (Platon, 2010, 517b). Spełnione jest więc pierwsze kryterium i, co najmniej częściowo, trzecie. Narracja jest bowiem wystarczająco realistyczna, aby czytelnik mógł ocenić, czy wyzwolenie się z jaskini jest faktycznie bolesne, czy rzucane na ścianę cienie są jedynie niedoskonałymi odbiciami innych przedmiotów itd.

Jednak opowieść o jaskini nie spełnia drugiego kryterium – nie jest argumentacyjnie niezbędna. Stanowi jedynie ilustrację filozofii Platona, a nie argument na rzecz jej prawdziwości. Platon nie przekonuje nas samą historią do istnienia świata idei ani do tego, że od idei Dobra „w świecie widzialnym

pochodzi […] światło" oraz że rodzi ona „prawdę i rozum" (Platon, 2010, 517b). Musimy tę filozofię poznać niezależnie od opisu jaskinii, aby tę drugą odczytać jako alegorię (Becker, 2018, s. 50–51).

1.2 PRZEKONUJĄCE DZIAŁANIA: PRZYKŁADY HISTORYCZNE

Historia eksperymentów myślowych oferuje wiele przykładów przekonujących działań umysłowych. Przeanalizujmy je, aby lepiej zrozumieć ich funkcję poznawczą oraz powiązanie narracji z argumentacyjnością eksperymentów.

Alegoria jaskini eksperymentem myślowym nie jest, niemniej znajdziemy u Platona przykłady tej praktyki, także w *Państwie*, np. na początku Księgi II. Glaukon referuje w niej potoczne przekonanie, zgodnie z którym sprawiedliwość nie jest sama w sobie dobra, jest pożądana jedynie z uwagi na jej konsekwencje: zmniejsza prawdopodobieństwo doświadczania przez nas krzywd za cenę utrudnienia wyrządzania przez nas krzywd innym osobom. Aby pokazać, że ludzie taki układ społeczny stworzyli niechętnie, Glaukon proponuje rzecz następującą:

> Że ci, którzy ją [sprawiedliwość] praktykują, robią to niechętnie, dlatego tylko, że nie potrafią krzywd wyrządzać, to może najlepiej potrafimy dojrzeć, jeżeli coś takiego potrafimy zrobić w myśli: damy jednemu i drugiemu wolność – niech każdy z nich robi, co chce, i ten sprawiedliwy, i niesprawiedliwy. A potem pójdziemy w trop za nimi, zobaczyć, dokąd każdego z nich żądza zaprowadzi. Wtedy byśmy ich na gorącym uczynku schwytali, bo na jedno i to samo wychodziłby sprawiedliwy i niesprawiedliwy – aby więcej mieć – do tego każda żywa istota z natury dąży, bo to jest dobro, a tylko prawo gwałtem ją ściąga do tego, żeby cenić równość mienia (Platon, 2010, 359b–359c).

Potrzebną wolność w wyobrażanym scenariuszu zapewnić mają pierścienie działające na wzór pierścienia Gygesa. Jak tłumaczy dalej Glaukon, pasterz Gyges z Lidii znalazł pierścień, który zapewniał mu niewidzialność. Dzięki temu udało mu się uwieść żonę panującego króla, dokonać zamachu i samemu przejąć tron. Dysponując tego rodzaju pierścieniami, moglibyśmy sprawdzić zachowanie wolnych osobników, sprawiedliwego i niesprawiedliwego:

> Więc gdyby istniały dwa takie pierścienie i jeden by sobie na palec włożył człowiek sprawiedliwy, a jeden niesprawiedliwy, to nie znalazłby się chyba żaden człowiek taki kryształowy, żeby wytrwał w sprawiedliwości i nie śmiałby wyciągać ręki po cudze ani go tykać, chociażby mu wolno było, i z rynku bez obawy brać, co by tylko chciał, i do domów wchodzić, i obcować z kim

> by mu się podobało, i zabijać, i z więzów uwalniać, kogo by tylko zechciał. […] Postępując tak, nie robiłby nic innego niż drugi – na jedno i to samo wychodziliby obaj. Przecież to wielkie świadectwo […], że nikt nie jest sprawiedliwy z dobrej woli, tylko z musu, bo sprawiedliwość to nie jest dobro dla jednostki, bo każdy, gdzie tylko myśli, że będzie mógł krzywdy wyrządzać bezkarnie, tam je wyrządza. Każdy człowiek uważa, że o wiele większy pożytek przynosi jednostce niesprawiedliwość niż sprawiedliwość (Platon, 2010, 360c–360d).

Czy pierścień Gygesa, bo tak zwana jest powyższa argumentacja, spełnia kryteria Beckera? Zawiera fikcyjny scenariusz, który ma nam coś powiedzieć na temat świata realnego, a mianowicie czy ludzie chcą sprawiedliwości dla niej samej. Opisana sytuacja nie jest też trudna do oceny: opierając się na własnych intuicjach, możemy stwierdzić, jak zachowają się osoby sprawiedliwe i niesprawiedliwe, dysponujące pierścieniami dającymi niewidzialność. Intuicja odgrywa tu kluczową rolę – to ona pozwala nam ocenić spójność wyobrażanych konsekwencji z naszymi przekonaniami o rzeczywistości. Nad pełniejszym zrozumieniem intuicji pochylimy się w kolejnych rozdziałach, ale już teraz warto zauważyć, że jest ona nieodłącznym elementem udanego eksperymentu myślowego. Glaukon zdaje się po cichu zakładać kilka dodatkowych prawd na temat natury ludzkiej, np. to, że ludzie zawsze chcą jak najlepiej dla siebie i nie są specjalnie sami z siebie altruistyczni. Możemy się z nim w tej sprawie nie zgodzić, lecz jest jasne, że środkiem naszej oceny będą nasze intuicje na temat nieskrępowanego zachowania innych ludzi. Pierwsze dwa kryteria są więc spełnione.

Podobnie jest z trzecim kryterium (Becker, 2018, s. 49). Na co dzień nasze działania podlegają społecznym i prawnym konsekwencjom: możemy się spotkać z karą prawną, jak i z dezaprobatą społeczną, jeśli złamiemy przyjęte reguły. Normy te wpływają na nasze motywacje i utrudniają ocenę tego, czy pożądamy sprawiedliwości dla niej samej. Glaukon dostrzega ten czynnik i niejako wyłącza go z równania. Mamy osobników niewidzialnych, a więc cokolwiek zrobią, pozostaną bezkarni. Zgodnie z powszechnym poglądem rekonstruowanym przez Glaukona (a więc dla niego intuicyjnym) obaj osobnicy wyrządzaliby krzywdę innym. Pierścień Gygesa pełni więc ważną funkcję argumentacyjną. Ułatwia ocenę prawdziwości potocznego obrazu sprawiedliwości (pożądanej jedynie z uwagi na konsekwencje) poprzez przedstawienie skrajnej sytuacji, w której na działania osób nie wpływa czynnik zakłócający (kara). Ocena ta polega na odwołaniu się do intuicji tego, kto słucha opowieści Glaukona. Cokolwiek byśmy nie sądzili o słuszności oceny dokonywanej

przez samego Glaukona, odwołanie się do intuicji jest tu niezbędne: oceniany pogląd na sprawiedliwość ma przecież być domyślny dla myślenia potocznego.

Pierścień Gygesa jest często przytaczanym w literaturze przykładem starożytnego eksperymentu myślowego (np. Gendler, 2000, s. 19, 28; Mišćević, 2022, s. 5; Rescher, 2017, s. 75; Roux, 2011, s. 2). Jednak nie ma zgody co do tego, który eksperyment myślowy można uznać za najstarszy. Sytuacja ta przypomina określanie granicy rozproszonego cienia: gdzie kończy się cień, a gdzie jest już obszar nasłoneczniony? Nicholas Rescher (1991, 2017) twierdzi, że trzeba cofnąć się do czasów presokratycznych, przed IV wiek p.n.e. Przywołuje przy tym liczne przykłady wyjaśnień Talesa i Anaksymandra z Miletu, Pitagorasa z Samos, Ksenofanesa z Kolofonu i Heraklita z Efezu. Przytoczmy przykład z pierwszego filozofa. Dobrze obrazuje on kolejną cechę eksperymentów myślowych:

> Według niego [Talesa] Ziemia utrzymuje się w spoczynku, bo jest w stanie pływać po wodzie jak drewno lub jaka inna materia tego rodzaju (żadna z nich nie spoczywa naturalnie w miejscu w powietrzu, lecz na powierzchni wody) (Arystoteles, 1980, B294a).

Wyobrażalnym założeniem, czyniącym przytaczane we fragmencie zjawisko – według Reschera – eksperymentem myślowym, jest analogiczne do drewna unoszenie się Ziemi na wodzie. Jeśli to założenie przyjmiemy, możemy wyjaśnić, dlaczego Ziemia utrzymuje się w spoczynku (w powietrzu by spadała, ale nie spada, jeśli unosi się na wodzie). Powód, dla którego część badaczy (np. Brown i Fehige, 2022; Ierodiakonou i Roux, 2011b; Irvine, 1991) nie uznaje przykładów Reschera za eksperymenty myślowe, jest natomiast następujący: jeśli tak ma wyglądać eksperyment myślowy, to są też nimi wszystkie rozumowania, w których czynione są wyobrażalne przy pomocy obrazów umysłowych założenia[5]. Andrew Irvine (1991, s. 154–156) wskazuje, że od rozumowania-eksperymentu myślowego powinniśmy oczekiwać bogatszego kontekstu teoretycznego i obserwacyjnego na temat świata, a także testowania hipotezy z tego kontekstu wyrosłej. Katerina Ierodiakonou (2011b, s. 48) twierdzi natomiast, że charakterystyczną cechą starożytnych eksperymentów myślowych – czy też *paradeigmata* (przykładów), jak jej zdaniem (2018, s. 33) określali je starożytni Grecy – jest to, że wyobrażana w nich sytuacja może się rozwiązać jedynie na dwa sposoby i w momencie startu „eksperymentowania"

[5] Ściślejsze rozumienie wyobrażalności przedstawione jest w dalszej części tekstu.

nie wiadomo, który sposób jest poprawny. U Talesa tak nie jest: naszym problemem jest to, jak Ziemia może utrzymywać się w spoczynku, a nie pozostawać w ruchu (spadającym). Unosi się na wodzie – brzmi treść wyobrażenia. Nie jest to jednak początek eksperymentowania, tylko w istocie rozwiązanie zagadki[6].

Takie uzasadnianie założeniem *ad hoc*, mimo że jest przypadkiem hipotetycznego myślenia, nie jest drogą do przekonującego eksperymentu myślowego. Podobnie nie mamy do czynienia z przekonującym empirycznym eksperymentem naukowym, jeśli jako badacze sztucznie, *ad hoc* tworzymy dane obserwacyjne tak, aby pasowały do naszej hipotezy.

Wyobraźnia należycie wykorzystana w eksperymencie myślowym polega na stworzeniu w myślach sytuacji, której konsekwencje są następnie przez przeprowadzającego eksperyment rozważane. Widzieliśmy to już w dość rozbudowanym przykładzie pierścienia Gygesa, ale rozpatrzmy następujący:

> Gdybym stanął na krańcu nieba gwiazd stałych, czy wyciągnąłbym rękę lub laskę na zewnątrz, czy nie? Nie móc wyciągnąć, to niedorzeczne; jeśli zaś ją wyciągnę, czy ciało, czy miejsce będzie na zewnątrz […]. Zawsze bowiem dąży się w ten sam sposób do wciąż osiąganego kresu i pyta o to samo; i jeśli będzie ciągle czymś innym to […] oczywiste, że jest bezkres (Eudemos, *Fizyka*, 467, tłum. M. Wesoły, za: Rodzeń, 2008, s. 242–243)[7].

Ten pochodzący z pierwszej połowy IV wieku p.n.e. eksperyment myślowy również bywa uznawany za najstarszy (np. Ierodiakonou, 2018), a przypisuje się go Archytasowi z Tarentu. Eksperyment ten stał się pierwowzorem tzw. eksperymentu z oszczepem z *O naturze wszechrzeczy* Lukrecjusza, a także innych rozumowań na rzecz nieskończoności świata, tak starożytnych, jak i średniowiecznych oraz nowożytnych (Ierodiakonou i Roux, 2011b, s. 37–42). Archytas również chce nas przekonać, że świat jest nieskończony. Zwróćmy uwagę, że tym razem wyobrażana sytuacja nie opisuje wprost konkluzji, jedynie zarysowuje samą sytuację problemową (próba wyciągnięcia ręki bądź laski poza świat). Pozostała część eksperymentu polega na rozważeniu konsekwencji tej sytuacji. Co się stanie, gdy spróbuję wyciągnąć rękę

[6] Wyjaśniliśmy już, dlaczego alegoria jaskini nie jest eksperymentem myślowym, ale zwróćmy uwagę, że nawet gdyby ją za taki uznać, to nie spełnia ona warunku Ierodiakonou: alegoria nie przedstawia sytuacji, której możliwe rozwiązania są początkowo nierozstrzygnięte. Zamiast tego ilustruje tezy Platona, które albo przyjmujemy, albo odrzucamy, niezależnie od samej alegorii.

[7] Pełne tłumaczenie z języka starogreckiego na współczesny angielski, wraz z komentarzem, znaleźć można w pracy Carla Huffmana (2005, s. 240–250).

poza świat? Są dwie możliwości: albo się nie da (coś mnie powstrzyma), albo będę mógł ją wyciągnąć. Zdaniem Archytasa niedorzeczne byłoby uznanie, że się nie da. Pozostaje druga możliwość. Jeśli tak i znowu stanę na krańcu świata, to ponownie będę mógł wyciągnąć rękę. A zatem świat jest bezkresny.

Nie musimy się zgadzać z Archytasem, że świat jest bezkresny. Zwróćmy jednak uwagę na podobieństwo w strukturze obu przywołanych eksperymentów myślowych. Działanie umysłowe w obydwu przypadkach zaczynało się od wyobrażenia pewnej sytuacji – przyznania pierścieni niewidzialności oraz stanięcia na krawędzi świata. W obu przypadkach rozważane były następnie konsekwencje tych sytuacji. Według Platona sprawiedliwy, pozostający bezkarnym, zachowywałby się tak samo jak niesprawiedliwy. Jeśli uznamy ten rozwój wydarzeń za intuicyjny, to tym samym zgodzimy się z wnioskiem Platona, że sprawiedliwość cenimy za ochronę przed krzywdą, a nie ze względu na nią samą. Zdaniem Archytasa natomiast zawsze można wyciągnąć rękę za krawędź świata, a więc ten jest bezkresny. Tok myślowy kończy się w momencie, gdy uznajemy dany bieg wydarzeń – zgodnie z tezami obu filozofów – za najbardziej intuicyjny lub gdy odmawiamy takiego uznania.

W przeciwieństwie do powyższej struktury w alegorii jaskini wyobrażana sytuacja nie była punktem wyjścia toku myślenia, tezy filozofii Platona zaś nie były punktem dotarcia. Historia rozgrywająca się w jaskini była równoległa do filozoficznych tez Platona, które należało przyjąć, aby móc tę historię zgodnie z duchem autora zinterpretować. Wyobrażany scenariusz nie pełnił więc funkcji argumentacyjnej. Nie trzeba przekonywać, bo żeby zrozumieć symboliczny wymiar alegorii, założyliśmy już to, do czego dopiero powinien nas przekonać eksperyment myślowy. Z kolei choć w przytaczanym toku myślenia Talesa mieliśmy do czynienia z przekonywaniem, to było ono *ad hoc*. Wyobraziliśmy sobie taki scenariusz, który od początku nam (Talesowi) pasował. Dlaczego Ziemia utrzymuje się w spoczynku? Bo podobnie jak drewno – używamy wyobraźni – unosi się na wodzie.

Jak wskazałem, ocena, które z starożytnych rozumowań są eksperymentami myślowymi, jest i pewnie pozostanie przedmiotem kontrowersji. Takich przykładów filozoficznych w starożytności znajdziemy więcej, chociażby u Arystotelesa (Corcilius, 2018) czy wśród starożytnych matematyków (Lakatos, 1976)[8]. Nieprzypadkowo traktuję tu wyłącznie o kontekście

[8] Imre Lakatos (1976, s. 9) eksperymentami myślowy nazwał przykłady *deíknūmi*, czyli „pokazywać" (*show*, *point out*), „wyjaśniać" (*explain*), „wydobywać na światło dzienne" (*bring*

europejskim. Prace nad historią omawianej praktyki – podobnie zresztą jak w przypadku całej historii filozofii – skupione są właśnie na tej perspektywie. Skoro jednak trudno u starożytnych Greków doszukać się osobnego pojęcia eksperymentu myślowego, a mimo to z nich korzystali, to nie ma powodu nie przypuszczać, że korzystano z nich również w innych kulturach[9].

W zasadzie całą historię filozofii możemy napisać, wymieniając kolejne eksperymenty myślowe. Takiego zadania podjęła się Peg Tittle, która w swojej książce pt. *What If… Collected Thought Experiments in Philosophy* (2005) opisuje ponad 110 eksperymentów myślowych znaczących dla historii filozofii. Do dzisiaj jest to zresztą stały element wachlarza metod filozoficznych: ziemia bliźniacza Hilary'ego Putnama (1975, 1998) i żuk w pudełku Ludwiga Wittgensteina (1953, 2000) okazały się kluczowe dla debaty na temat teorii znaczenia języka w filozofii; w etyce zawrotną karierę zrobiły dylemat wagonika Philippy Foot (1967; Thomson, 1976) i skrzypek Judith Jarves Thomson (1971); w filozofii tożsamości paradoks teleportacji Dereka Parfita (1984); w epistemologii problem Edmunda Gettiera (1963); w filozofii społecznej zaś zasłona niewiedzy Johna Rawlsa (1971)[10].

Współczesne eksperymenty myślowe w filozofii nie odbiegają też formą od tych starożytnych. Dla ilustracji rozważmy uwagi Wittgensteina na temat możliwości prywatnego języka, mającego opisywać nasze subiektywne doznania:

> Jeżeli mówię o sobie, że jedynie sam po sobie wiem, co znaczy słowo „ból", – to czy nie muszę powiedzieć tego również o innych? I jak mogę jeden przypadek uogólniać w tak nieodpowiedzialny sposób? Cóż, każdy mówi mi o sobie, że tylko sam po sobie wie, czym jest ból! Przypuśćmy, że każdy miałby pudełko zawierające coś, co nazywamy „żukiem". Nikt nigdy nie może zajrzeć do cudzego pudełka; i każdy mówi, że tylko z widoku swego żuka wie, co to jest

to light), „przedstawiać" (także fizycznie; *exhibit*, *display*) (Lidell i in., 1968). Była to praktyka starożytnych matematyków polegająca na „sprawianiu by prawdziwość lub fałszywość wyrażenia matematycznego stawała się w jakiś sposób widoczna" (Szabó, 1978, s. 189). Zastrzec trzeba, że nie każde *deíknūmi* było eksperymentem myślowym, bo nie zawsze polegało na *myślowym* działaniu – można twierdzenie uczynić widocznym np. wizualnie.

[9] Przykładowo, zidentyfikowano je w średniowiecznej myśli muzułmańskiej (McGinnis, 2018) i tradycji buddyjskiej (Hertogh, 2018; Williamson, 2019, rozdz. 4). Póki co jednak nie dysponujemy wieloma próbami wyłonienia nieeuropejskich przykładów starożytnych eksperymentów myślowych (zob. Kerr, 2019). Także antropologicznie nie wiemy zbyt wiele na temat tego, na ile eksperymentowanie w myślach jest *skłonnością* przejawianą przez przedstawicieli naszego gatunku (zob. Zawadzki, 2017, s. 124).

[10] Przytoczone zostały takie przykłady, o których z perspektywy czasu można z pewnością powiedzieć, że odbiły się szerokim echem wśród filozofów.

> żuk. – Mogłoby się zdarzyć, że każdy miałby w swym pudełku co innego. Można by sobie nawet wyobrazić, że rzecz ta stale by się zmieniała. – Gdyby jednak mimo to słowo „żuk" miało dla tych ludzi jakieś zastosowanie? – Wtedy byłoby ono zastosowaniem nazwy pewnej rzeczy. Rzecz w pudełku nie należy w ogóle do gry językowej; nawet nie jako coś, gdyż pudełko mogłoby być też puste. – Nie, przez ową rzecz w pudełku można „upraszczać"; ona się znosi, czymkolwiek by była. To znaczy: Jeżeli gramatykę wyrazu doznania konstruujemy według wzorca „przedmiot – oznaczenie", to przedmiot wypada z rozważań jako nieistotny (Wittgenstein, 2000, § 293).

Powyższy cytat ilustruje kluczowy element późnej myśli Wittgensteina, który przekonywał, że znaczenia słów wyznacza sposób ich używania w grach językowych[11]. W konsekwencji słowa mające być nazwami dla wewnętrznie, pierwszoosobowo przeżywanych stanów umysłowych *de facto* nie mogą się do nich odnosić – ani w takim sensie, że wyrażenia rodzaju „czuję ból w lewej łydce" w sposób zrozumiały informują o doznaniach drugiej osoby, ani w takim, że są zrozumiałe dla samej osoby doznającej. O bólu, rozumianym jako doznanie, nie sposób nic sensownego powiedzieć. Eksperyment myślowy z żukiem przekonuje do takiej konkluzji, pokazując niepożądane konsekwencje przyjęcia założenia o istnieniu języka prywatnego: z perspektywy praktyki językowej to, co kryje się w pudełku – cokolwiek jest doznawane – jest obojętne dla poprawności tej praktyki. Przy czym filozof nie odmawia doznaniom istnienia, a jedynie znaczeniotwórczej roli[12].

Czy eksperyment myślowy z żukiem spełnia kryteria Beckera? Po pierwsze, w przytoczonych fragmentach *Dociekań filozoficznych* Wittgenstein korzysta z fikcji, by powiedzieć coś na temat świata realnego, a dokładniej na temat nieprywatnego charakteru języka. Przedstawiona sytuacja jest spójna i możliwa do wyobrażenia, odbiorca może sam ocenić ciąg wyobrażanych zdarzeń, tj. czy faktycznie z perspektywy praktyki językowej zawartość pudełka nie jest

[11] „Znaczeniem słowa jest sposób użycia go w języku" (Wittgenstein, 2000, § 43). Według standardowego ujęcia nie sposób ściśle zdefiniować gry językowej. Niemniej przywołajmy wstępną charakterystykę: „«Grą językową» nazywać też będę całość złożoną z języka i z czynności, w które jest on wpleciony" (Wittgenstein, 2000, § 7). Przykładowymi grami językowymi są więc witanie się, modlitwa w zgromadzeniu, dialog przy ladzie w sklepie, dziękowanie, prowadzenie wykładu, odpowiadanie na pytania itd.

[12] Przytoczone uwagi są bardzo fragmentaryczne względem całości rozumowania Wittgensteina. Więcej na temat interpretacji eksperymentu myślowego z żukiem znaleźć można w pracy Józefa Bremera (2013). Dodać też trzeba, że filozofia Wittgensteina jest trudna do zrekonstruowania do tego stopnia, że same interpretacje różnych wątków z prac filozofa żyją własnym życiem, stając się istotnym wkładem w filozofię języka (np. Kripke, 2007).

istotna. Omawiane fragmenty są też argumentacyjnie niezbędne: bez rozpatrzenia, co się stanie, jeśli zawartość pudełek nie będzie publicznie dostępna, nie możemy przyjąć konkluzji, czyli tego, że zawartość ta nie ma znaczenia dla poprawności praktyki językowej.

1.3 PODSUMOWANIE

- Eksperyment myślowy jest umysłowym działaniem poznawczym zorientowanym na przekonanie osoby przeprowadzającej go do określonej konkluzji. Jest więc narzędziem argumentacyjnym, a nie tylko ilustracją poglądów autora.
- Niewiele wiemy na temat historii używania tego narzędzia poza kontynentem europejskim; w Europie sięga natomiast co najmniej czasów presokratejskich (Tales z Miletu, Archytas z Tarentu).
- Procedura eksperymentowania w umyśle polega na wyobrażeniu sobie fikcyjnej sytuacji oraz rozważeniu jej konsekwencji. Struktura wnioskowania warunkowego zawartego w eksperymencie może pozostać niejawna dla osoby przeprowadzającej eksperyment, swój udział w przekonywaniu mają zaś kwestie związane z samym aktem wyobrażania. To, czy osoba przeprowadzająca eksperyment zgodzi się z proponowanymi przez autora konsekwencjami rozważanej sytuacji, często opiera się na jej intuicji.
- Wymienia się (za Beckerem) następujące cechy odróżniające eksperyment myślowy od innego wykorzystania wyobraźni (np. rozumienia alegorii): (1) jego wniosek mówi coś na temat realnego świata, (2) sytuacja eksperymentu musi być opisana w rozstrzygalny dla osoby go przeprowadzającej sposób, (3) wyobrażana sytuacja jest argumentacyjnie niezbędna dla uznania wniosku autora eksperymentu.
- Dysponując bardzo wstępnym wyróżnieniem eksperymentów myślowych jako działania poznawczego, możemy postawić bardziej szczegółowe pytania. Zaczniemy od następujących: Skoro omawiana tu praktyka jest swojego rodzaju rozumowaniem, to dlaczego nazywana jest „eksperymentowaniem"? Co dokładnie wyróżnia ją na tle innych rozumowań?

2 EKSPERYMENTOWANIE W ŚWIECIE I W MYŚLACH

Do tej pory, charakteryzując eksperymenty myślowe, odwoływaliśmy się przede wszystkim do kontekstu filozoficznego. Jest to o tyle uzasadnione, że – jak powiedzieliśmy – omawiana metoda rozumowania jest często utożsamiana z filozofowaniem właśnie. Niektórzy jednak uznają, że pojawienie się eksperymentów myślowych jest ściśle związane z wykształceniem praktyki eksperymentowania naukowego. W kontekście tej ostatniej podkreśla się zwykle okres nowożytności (Butterfield, 1958; Shapin, 1996). Rozważmy teraz, jaka jest relacja między „eksperymentem" jako empiryczną procedurą naukową a będącą w centrum naszego zainteresowania metodą myślenia.

2.1 EMPIRYZM BEZ POMIARU I KREŚLENIE SYTUACJI WYJĄTKOWEJ

Motywacja stojąca za wiązaniem historii eksperymentów myślowych z historią empirycznego eksperymentu naukowego jest dość banalna: aby praktykować te pierwsze, musimy przecież rozumieć, czym jest eksperyment (w domyśle ten drugi: naukowy, empiryczny). I tak przykładowo James McAllister (2018) jako pioniera eksperymentów myślowych wskazuje Galileusza. Jest to podejście radykalne o tyle, że dociekanie natury świata, także empiryczne, ewoluowało stopniowo i upraszczające jest utożsamianie powstania metody naukowej z jedną osobą. Wczesne wizje nauki eksperymentalnej kreślono przed czasami Galileusza i Francisa Bacona, chociażby w XIII-wiecznych pismach Roberta Grosseteste'a i Rogera Bacona (Hackett, 1997; Oliver, 2004). Idea obserwacji dającej bazę doświadczeniu (*empeiríā*), które to można następnie rozumowo porządkować, jest natomiast co najmniej arystotelesowska (Hepburn i Andersen, 2021). Warto zaznaczyć, że choć Arystoteles traktował obserwację zasadniczo jako czynność pasywną, a nie – jak dziś kojarzymy eksperymentowanie naukowe – jako działanie kontrolujące i wpływające na elementy obserwowanej sytuacji (Corcilius, 2018, s. 57–58), empiryzm Arystotelesa miał zasadniczy wpływ na rozwój

myślenia naukowego[13], także poprzez rolę, jaką odegrał w średniowiecznej filozofii przyrody. Tę ostatnią kojarzy się często jedynie ze spekulacjami religijnymi, ale przyrodę wyjaśniano, odwołując się również do specyficznie rozumianego doświadczenia. Historyk średniowiecza Edward Grant (2001, s. 179) nazywał ten okres „empiryzmem bez obserwacji i pomiaru". Teorie rzadko kiedy testowano empirycznie, natomiast często sięgano po „eksperymentalne argumenty", polegające na odwoływaniu się do naocznych obserwacji autorów, obserwacji zapisanych w greckich, rzymskich lub arabskich traktatach czy do powszechnie podzielanych przekonań („wszyscy wiedzą, że…"). Jako odwołanie do doświadczenia traktowano także nierzadko przybierające postać eksperymentów myślowych obserwacje domniemane: rozważanie konsekwencji wyobrażanych sytuacji hipotetycznych (Grant, 2001, s. 160–182, 2004; Grellard, 2011; King, 1991).

Wspomnieliśmy już, że konstrukcja empirycznego eksperymentu naukowego przypomina konstrukcję eksperymentu myślowego. By przetestować swoje hipotezy, badacze nie mierzą dowolnie, czyli czegokolwiek w jakikolwiek sposób. Przedmiot obserwacji wynika ze stawianych pytań badawczych, tak jak sposób tej obserwacji wynika ze standardów metodologicznych danej dyscypliny. W zależności od celu badacz musi zadbać o wystarczającą wielkość (albo i nawet reprezentatywność) próbki badawczej, minimalizację wpływu potencjalnych zmiennych zakłócających, powtarzalność procedury pomiędzy osobami badanymi, ich komfort czy choćby o uwzględnienie warunków kontrolnych w eksperymencie. To wszystko często czyni sytuację eksperymentalną sztuczną, mało realistyczną. W badaniach behawioralnych mówimy w tym kontekście o niskiej ekologiczności, niskim stopniu odwzorowania warunków, w których dane zachowanie ma miejsce w rzeczywistości pozaeksperymentalnej[14]. W przypadku eksperymentów myślowych osiągamy wyizolowanie istotnych czynników oraz eliminację tych nieistotnych i potencjalnie zakłócających poprzez abstrakcję i idealizację. Wyobrażamy sobie przecież sztuczną sytuację. Dzięki abstrakcji jest ona ograniczona, przynajmniej pozornie, jedynie do tych czynników, które autor eksperymentu uznał za istotne dla jego rozwiązania. Idealizujemy te czynniki, w określonym stopniu abstrahując

[13] Szczegółowe opracowanie arystotelesowskiego rozumienia doświadczenia znajduje się w pracach Petera Kinga (2003) i Travisa Butlera (2003).

[14] Im więcej aspektów sytuacji badawczej kontrolujemy, strukturyzujemy, tym trudniej o jej wysoką ekologiczność (Shamay-Tsoory i Mendelsohn, 2019).

od ich wpływu na siebie nawzajem bądź też przypisując temu wpływowi stałą wartość (Stuart, 2016b, s. 460–461; Zielińska, 1981).

Ponadto sytuacje badawcze najcenniejsze są wtedy, kiedy dzieje się coś dotąd niewyjaśnionego. Podobnie jest w przypadku eksperymentów myślowych: nie wyobrażamy sobie dowolnych sytuacji, tylko takie, które – posługując się terminem Tamary S. Gendler (2000, s. 2–18) – są wyjątkowe (*exceptional case*). Nie zostały one dotąd wyjaśnione przy pomocy naszego aparatu teoretycznego, są zaskakujące, nietypowe. Pozwalają przetestować zakres przydatności, siłę danej teorii czy zrewidować spojrzenie na daną klasę zjawisk – wartościowy eksperyment myślowy jest więc przydatny. Wyjątkowość opisywanej sytuacji jest jedną z przyczyn, dla której przeprowadzanie eksperymentu myślowego jest w stanie spełnić swoją funkcję, czyli przekonać nas do tezy jego twórcy. Teza ma nam pomóc uporać się z tą wyjątkową sytuacją, wyjaśnić, co się w niej dzieje. Podobną rolę odgrywa interpretacja wyników empirycznego eksperymentu naukowego.

2.2 SYTUACJA WYJĄTKOWA W EKSPERYMENCIE MYŚLOWYM ZE SPADKIEM SWOBODNYM GALILEUSZA

Rozważmy przykład nowożytnego eksperymentu myślowego, aby dobrze zilustrować kreślenie w wyobraźni sytuacji wyjątkowej na podobną modłę projektowania empirycznego eksperymentu naukowego. Naturalnie, w epoce stopniowego wypracowywania empirycznej metody eksperymentalnej wielu badaczy korzystało z eksperymentów myślowych: Simon Stevin (Brown, 2011, s. 3–5), Isaac Newton czy chociażby Gottfried Leibniz (Arthur, 2018). Jednym z najsłynniejszych jednak przykładów jest argumentacja Galileusza odnośnie do spadku swobodnego. Wywód z *Rozmów i dowodzenia matematycznego w zakresie dwóch nowych umiejętności dotyczących mechaniki i ruchów miejscowych* (1930) prowadzony jest w formie dialogu między trzema postaciami: Salviatim (alter ego Galileusza), Simplicio i Sagredo:

> Salviati: Bez żadnych doświadczeń, możemy krótkim a ścisłym wywodem wykazać jasno, że jest niemożliwe, aby większy ciężar poruszał się prędzej niż mniejszy, jeżeli oba są z jednej materji i wreszcie takie, o jakich mówi Arystoteles. Więc powiedz, panie Simplicio, czy godzisz się na to, że każde ciało spadające ma prędkość przez naturę oznaczoną, tak że, aby się ta powiększała lub zmniejszała, potrzeba użyć siły przyspieszającej lub opóźniającej. […] [Simplicio się zgadza – przyp. M.W.]. Salviati: Gdybyśmy więc mieli dwa ciała, których naturalne prędkości są różne, to jasnem jest, że gdy

> powolniejsze złączymy z prędszem, to wtedy to ostatnie będzie opóźniane przez tamto, a powolniejsze będzie przyspieszone przez prędsze. […] Jeżeli więc jest prawdą, że wielki kamień porusza się z prędkością ośmiu stopni a mniejszy z prędkością czterech stopni, to te dwa kamienie złączone razem będą miały prędkość mniejszą od ośmiu stopni. Ale połączone dwa kamienie są większe od wielkiego kamienia, który miał prędkość ośmiu stopni, więc w tym razie ciężar większy poruszałby się wolniej od mniejszego, co sprzeciwiałoby się Waszemu założeniu. Widzicie więc jak z założenia, że większe ciało porusza się prędzej od mniejszego, dochodzę do wniosku, że większy ciężar porusza się wolniej. […] Sagredo: […] dowiedliście jasno, że ciała nierównego ciężaru w tym samym ośrodku nie poruszają się z prędkościami proporcjonalnemi do ich ciężkości, lecz z jednakowemi, przyjmując przytem, że ciała są z tegoż samego materjału albo tegoż samego ciężaru gatunkowego (Galilei, 1930, s. 53, 57).

Galileusz jasno nakreśla wyabstrahowaną z nieistotnych elementów wyobrażalną sytuację: dwa kamienie połączone ze sobą spadają z pewnej wysokości bez działania dodatkowych sił. Jaka jest prędkość tego spadku? Przyjmując, że Arystoteles ma rację, ciała mają prędkość spadku proporcjonalną do swojej ciężkości. Wobec tego połączenie powolniejszego (lżejszego) z szybszym (cięższym) powinno spowolnić prędkość spadku względem prędkości spadku samego szybszego kamienia. Ale jednocześnie połączenie mniejszego z większym sprawia, że spadający przedmiot jest jeszcze większy, więc powinien spadać szybciej niż niepołączony większy kamień. Nasze założenie prowadzi zatem do sprzeczności. Ciała nie spadają z prędkością proporcjonalną do swojej ciężkości, tylko z jednakową. Wyobrażana sytuacja jest wyjątkowa: oto dwa kamienie zostały połączone i zrzucone z wysokości. Samo to w sobie nie jest może niespotykane (często spadają nam przecież przedmioty złożone z kilku elementów jak kluczyki z brelokiem), lecz mamy też pominąć działanie innych sił na kamienie oraz, nade wszystko, sytuacja ta wskazuje na problematyczność teorii Arystotelesa na temat prędkości spadku: do jednej i tej samej sytuacji fizycznej dała ona sprzeczne prognozy.

Dodajmy, iż proponowane przez Galileusza wyjaśnienie prędkości spadku swobodnego ciał, że jest ona w tych idealnych warunkach jednakowa, jest być może dla nas oczywiste, jednak nie wynika wprost z zanegowania tez Arystotelesa. Jak wskazuje Paolo Palmieri (2018, s. 102–106), znajdziemy u Galileusza argumenty uzasadniające jego pozytywne wyjaśnienie, ale są one niezależne od samego sformułowania przytoczonego tu eksperymentu myślowego. Dlaczego ten ostatni został więc tak, a nie inaczej sformułowany?

Zdaniem Palmierego Galileuszowi bliskie było powszechne w średniowieczu, jak i XVI wieku[15] myślenie o ludzkim intelekcie jako władzy biernej poznawczo. Uważano, że faktyczne poznanie prawdy na temat świata – nie tylko dostrzeżenie sprzeczności w dotychczasowych teoriach – wymaga udziału Boga, będącego intelektem czynnym.

Zostawmy kwestię ominięcia działania innych sił i zastanówmy się, czy opisany eksperyment rzeczywiście miał miejsce. Salviati podkreśla w dialogu, że obalenie tezy Arystotelesa nie wymaga faktycznego doświadczenia empirycznego, ale pewną kością niezgody jest to, czy Galileusz oprócz opisywania eksperymentu realnie go przeprowadził. Z jednej strony, eksperyment ze spadkiem swobodnym uznaje się za jeden z najpiękniejszych empirycznych eksperymentów naukowych – tak przynajmniej wynika z ankiety zrealizowanej na początku XXI wieku wśród fizyków (Crease, 2003). Zgodne jest to z uwagami ucznia i autora biografii Galileusza, Vinenzo Vivianiego, który wspomina, że Galileusz zrzucał kule o różnej wielkości z Krzywej Wieży w Pizie (Viviani, 2019, s. 19), w obecności zresztą innych uczonych i swoich studentów. Z drugiej strony, ze względu na brak dodatkowych źródeł potwierdzających tę relację, większość badaczy Galileusza (przynajmniej od XIX wieku) jest zdania, że opisywana sytuacja nie miała miejsca, a nawet jeśli miała, to nie była testowaniem teorii Arystotelesa, lecz demonstracją zależności, którą Galileusz udowodnił wcześniej przy pomocy rozumu (np. Palmieri, 2018; Segre, 1989; zob. Drake, 2003). Taką samą opinię znajdziemy pośród badaczy eksperymentów myślowych (np. Brown, 2004, s. 24). Pośrednie stanowisko w tej kwestii wysuwa Palmieri (2018), stwierdzając, że w przypadku Galileusza nie sposób odseparować eksperymentów myślowych od empirycznych w tym sensie, że zarówno doświadczenia rozumowe, jak i empiryczne były elementem tego samego procesu dociekania.

2.3 ROLA EKSPERYMENTÓW MYŚLOWYCH W NAUCE XIX I XX WIEKU

Oczywiście, w nowożytności – podobnie jak w starożytności – nie posługiwano się terminem *eksperyment myślowy*. Pojęcie to pojawiło się dopiero w XIX wieku, wtedy też zaczęto systematycznie teoretyzować samą metodę, którą rozumiano już nie tylko jako narzędzie wyjaśniające konkretne

[15] Sięgające, a jakże, Arystotelesa.

fenomeny fizyczne. Mówiąc ściślej, termin wprowadził duński chemik i fizyk Hans Ch. Ørsted, który w swoim rodzimym języku pisał o *Tankeexperiment* oraz po niemiecku – *das Gedankenexperiment* i *das Gedankenversuch* (Buzzoni, 2018b; Ierodiakonou i Roux, 2011a; Ørsted, 1811). W swym pierwotnym użyciu pojęcia te odnosiły się do metody dbania o spójność wyjaśniania przyrody. *Tankeexperiment* testować miał zakres stosowalności teorii fizycznych, poprzez stosowanie wcześniej ustalonych praw przyrody do sytuacji hipotetycznych i sytuacji zdających się anomaliami. Duński uczony twierdził, że dobrze określone danymi empirycznymi teorie fizyczne będą wpływać na potoczny sposób widzenia świata. Wyjaśnienia zgodne z daną teorią fizyczną powinny zatem w eksperymentach myślowych narzucać się jako intuicyjne. Udany *Gedankenexperiment* zwiększać miał zaufanie do naukowego wyjaśniania świata, nieudany zaś zmniejszać.

Duński źródłosłów terminu jest dziś względnie znany, jednak przez długi czas za jego twórcę uznawano filozofa i fizyka Ernsta Macha. Terminologię Ørsteda przekładano na *mental experiment*, *imaginary experiment* i *experimental idea* (Piotrowski, 2011, s. 10), po raz pierwszy anglojęzyczny zwrot *thought experiment* pojawił się zaś właśnie w książce Macha *Erkenntnis und Irrtum* (*Knowledge and Error*) w tłumaczeniu Thomasa J. McCormacka, wydanej osiem lat po czasopismowej premierze tekstu w 1897 roku[16]. Nie wiadomo natomiast, czy Mach znał terminologię Duńczyka. Czym jest eksperyment myślowy zdaniem Macha? Jest specyficznym sposobem myślenia, wynikającym z wrodzonej ludzkiej skłonności do eksperymentowania:

> […] szanowany kupiec tak samo jak badacz z pasją dokonuje tej samej rzeczy. Każdy z nich wyobraża sobie [*conceives*] okoliczności i wiąże z nimi ideę, oczekiwania bądź założenia dotyczące pewnych wyników; tworzą oni eksperyment myślowy (Mach, 1973, s. 451).

Takie myślenie nie jest oderwane od świata empirycznego. Sprawne teoretyzowanie według Macha opiera się na doświadczeniu, bo każda wiedza ma dla niego źródło w empirii. Eksperymentowanie w myślach, o ile ma być poprawne, musi więc pozostać związane z doświadczeniem. U Macha dostrzegalna jest też rola intuicyjności – wiedza instynktowna, nieuświadamiana porcja informacji pochodząca z doznań, koryguje proces eksperymentowania.

[16] Z tłumaczeniem tym zapoznać się można też we wznowieniach książki (zob. Mach, 1976).

W konsekwencji jedne scenariusze wydają się bardziej prawdopodobne niż inne (Sobczyńska, 2001, s. 144)[17].

Warunkiem przeprowadzenia empirycznego eksperymentu jest wcześniejsze przeprowadzenie eksperymentu myślowego (Mach, 1973, s. 452). Badacz-empirysta, stawiając hipotezę, domniemuje rezultatu swojego (empirycznego) działania eksperymentalnego. W umyśle symuluje zatem scenariusz zachodzenia interesującego go zjawiska, co pomaga mu też skonstruować adekwatny względem badanego zjawiska plan eksperymentu empirycznego (zob. Roux, 2011, s. 6–11). Oba momenty (generowania hipotezy i dobierania procedury) nie muszą być zresztą przedmiotem świadomego namysłu. Wprawny badacz doskonale wie, które plany eksperymentalne i metody analizy danych są właściwe dla interesujących go problemów badawczych. Im bardziej rutynowy proces badawczy, tym mniejsza szansa, że badacz będzie świadomie myślał o tym dwuetapowym procesie przygotowawczym.

Jeszcze silniej o związku między eksperymentem naukowym a myślowym wypowiadał się Albert Einstein. Podobnie jak Galileusz, był on autorem jednych z najsłynniejszych eksperymentów myślowych w historii fizyki, później potwierdzanych empirycznymi badaniami (np. Aspect i in., 1982; zob. Norton, 1991, 2013): eksperymentu z windą (dotyczącego zasady równoważności masy grawitacyjnej i bezwładnej, Einstein i Infeld, 1967, s. 226–234), ściganiem fali światła (Einstein, 1951) czy choćby eksperymentu EPR (rozważającego niekompletność mechaniki kwantowej, jego nazwa pochodzi od nazwisk autorów, Einstein i in., 1935). Eksperymenty myślowe były dla Einsteina tworzonymi myślą wyidealizowanymi eksperymentami naukowymi (Einstein i Infeld, 1967, s. 9). Im dalej jednak od rzeczywistości fizycznej i uznanych teorii naukowych z ich prawami, tym bliżej „igraszkom umysłu" (jak zwykł nazywać eksperymenty myślowe) do absurdu. Krytykę takich niepoprawnych eksperymentów dobrze wyraża następujące zdanie Alexandra Moszkowskiego, przyjaciela Einsteina i jednego z pierwszych popularyzatorów teorii

[17] Wiedza instynktowna może budzić skojarzenia z procesami intuicji. Jest to częściowo uzasadnione, bo w przypadku obu traktujemy o procesach korzystających z niedostępnych świadomości informacji i docieraniu dzięki nim do świadomych wniosków (na tym właśnie, jak zobaczymy w rozdziale 5, polega intuicja). Nie powinniśmy jednak jednego z drugim utożsamiać, jako że dla Macha wiedza instynktowna jest silniej związana ze zmysłowymi informacjami niż szersza kategoria intuicji, dla której nie ma znaczenia, czy nieświadoma informacja jest zmysłowa, czy jakaś inna. O godzeniu roli wyobraźni z sensualizmem Macha można szerzej przeczytać w pracy Char Brecevic (2021; zob. Buzzoni, 2018a).

względności: „To nie eksperyment wyobraźni: to farsa, albo nawet śmielej, to tylko szwindel”[18] (1921, s. 116).

Wspomnijmy jeszcze o kocie Schrödingera – eksperymencie myślowym zaprojektowanym w celu skrytykowania kopenhaskiej interpretacji mechaniki kwantowej. Erwin Schrödinger, jeden z twórców tej dziedziny, jest także w kulturze masowej znany z tego właśnie eksperymentu (Schrödinger, 1935). Oto kot zostaje zamknięty wraz z urządzeniem uwalniającym lotną truciznę w momencie, gdy rozpadowi ulegnie jeden atom radioaktywnej substancji, również znajdującej się w urządzeniu. Substancji tej jest tak niewiele, iż to, że w ciągu godziny jej jeden atom się rozpadnie, jest równie prawdopodobne jak to, że tak się nie stanie. Intuicyjnie rzecz biorąc, uznalibyśmy, że po godzinie kot jest albo martwy, albo żywy. Jednak zgodnie z kopenhaską interpretacją mechaniki kwantowej atomy, fotony i cząsteczki subatomowe do momentu obserwacji mogą znajdować się w dwóch stanach jednocześnie, w tzw. superpozycji. Obrazując to w makroskali, trzeba by powiedzieć, że po godzinie, w momencie zajrzenia do pokoju, kot naprawdę jest zarazem żywy i umarły. To zaś, eufemistycznie mówiąc, wydaje się niemożliwe. Wbrew pozorom i popularnonaukowym opisom kot Schrödingera nie miał być ilustracją mechaniki kwantowej jako takiej. Eksperyment ten jest krytyczny, gdyż Schrödinger uważał interpretację kopenhaską za absurdalną. Eksperyment myślowy miał tę absurdalność wydobyć na światło dzienne i skłonić odbiorcę do odrzucenia interpretacji[19].

[18] Zdanie wypowiedziane przy omawianiu idei Lumena, postaci poruszającej się szybciej od światła, która dzięki temu mogłaby obejrzeć bitwę pod Waterloo od końca do jej początku, a zatem z prędkością niemożliwą w praktyce do osiągnięcia zgodnie z teorią względności.

[19] Jeszcze innym znanym naukowym, ale myślowym eksperymentem jest demon Maxwella opisany przez Jamesa C. Maxwella, pokazujący paradoksalność drugiej zasady termodynamiki. Zgodnie z nią w termodynamicznie izolowanych układach entropia się nie zmniejsza. Prawo to bywa często dyskutowane także w publikacjach popularnych, m.in. w kontekście strzałki czasu (jako uzasadnienie dla upływu czasu od przeszłości do przyszłości) i śmierci termicznej Wszechświata (zob. Piotrowski, 2011). Szerszą rekonstrukcję poglądów na temat eksperymentów myślowych w XX-wiecznej filozofii nauki można znaleźć w pracy Nitschego (2020, s. 111–118).

2.4 WSPÓŁCZESNE KATEGORYZACJE EKSPERYMENTÓW MYŚLOWYCH

O ile w XIX i pierwszej połowie XX wieku naukowcy tacy jak Mach czy Einstein pogłębili rozumienie roli eksperymentów myślowych w kontekście empirycznej nauki, o tyle współczesne analizy filozoficzne usiłują sklasyfikować ich różnorodność i funkcje w bardziej zniuansowany sposób.

Zdaniem XX-wiecznego filozofa nauki Karla Poppera poprawnym użytkiem z eksperymentowania w myślach jest wykazanie przy jego pomocy sprzeczności bądź też zilustrowanie teorii naukowej (eksperymenty, odpowiednio, krytyczne i heurystyczne). Złym sposobem natomiast byłaby próba obrony teorii odpowiednio skonstruowanym *Gedankenexperiment* (eksperymentem apologetycznym). Popper wskazał także na dodatkowy warunek dla eksperymentów krytycznych: powinniśmy wyobrażać sobie taką sytuację, która będzie zgodna z poglądami tego, kogo atakujemy (Popper, 2002, s. 459). Jeśli warunku tego nie spełnimy, to grozi nam arbitralność (niemerytoryczna zmiana założeń celem uratowania stawianych tez) oraz niepełność wyjaśniania (Popper, 2002, s. 460–464). Poprzez taki niepoprawnie skonstruowany eksperyment myślowy nieadekwatnie odniesiemy się do czyjejś teorii lub jej krytyki.

Kiedy najłatwiej konstruuje się eksperymenty myślowe, a więc i identyfikuje sytuacje wyjątkowe? Według Thomasa Kuhna, autora *Struktur rewolucji naukowych* (2011), eksperymenty myślowe są symptomami kryzysu naukowego. W duchu kuhnowskiego modelu rozwoju nauki, nasilające się obserwacyjne anomalie, czyli obserwacje niepasujące do aktualnego paradygmatu prowadzenia badań i wyjaśniania zjawisk, wskazują na nieadekwatność obowiązującego systemu pojęciowego względem „subtelnej struktury świata w całości" (Kuhn, 1985, s. 360). Choć nazywa je krytycznymi, eksperymenty myślowe nie muszą, jego zdaniem, nawet ujawniać sprzeczności teorii, bo te są już widoczne na etapie kryzysu paradygmatu. Dobre *Gedankenexperimente* mogą pokazać, że niektóre anomalie wcale nimi nie były. Innymi słowy, według Kuhna eksperymenty myślowe nie tylko dostarczają wiedzy na temat stosowania pojęć, ale i ugruntowują nową wiedzę o świecie. Co czyni zaś eksperyment dobrym? Kuhn wyróżnił dwa kryteria. Po pierwsze, wskazał na posługiwanie się terminologią dotychczasowego paradygmatu. Badacz chce przecież ocenić, na ile aktualny słownik nie przystaje do obserwowanej przyrody. Po drugie, choć w praktyce wyobrażana sytuacja nie musi być

możliwa do realizacji, to z samą prezentowaną w niej anomalią badacz musiał się już zetknąć (Kuhn, 1985, s. 369).

Oprócz tych rozróżnień w literaturze funkcjonują też inne klasyfikacje. John Norton (1991) dzieli eksperymenty myślowe ze względu na ich strukturę logiczną na dedukcyjne („Typ 1") i indukcyjne („Typ 2"). Podział w oparciu o strukturę logiczną eksperymentów proponuje także Roy Sorensen (1998, s. 132–166). W jego ujęciu eksperymenty myślowe obalają twierdzenia poprzez negowanie ich konsekwencji modalnych, to jest tego, co dozwolone/zabronione, wiedzowe/przekonaniowe, możliwe/konieczne itd. Może się to dziać na co najmniej dwa sposoby. Po pierwsze, eksperyment może pokazać, że to, co wydaje się niemożliwe według założeń narracji, jest jednak możliwe. Sorensen nazywa takie eksperymenty podważaczami konieczności (*necessity refuters*). Po drugie, zachodzi prawdopodobieństwo zaistnienia sytuacji, w której uznanie treści narracji wiąże się z potrzebą przyjęcia kłopotliwych konsekwencji. Przykładem może być narracja kreśląca obraz wszechpotężnego Boga, która prowadzi do tzw. paradoksu omnipotencji. Jedno z jej lepiej znanych sformułowań brzmi: czy Bóg może stworzyć kamień tak ciężki, że sam nie będzie w stanie go podnieść? Jeśli może, to znaczy, że istnieje coś, czego nie może podnieść, zatem nie jest wszechmocny. Jeżeli nie może, to również nie jest wszechmocny. Paradoks ten pokazuje potencjalną sprzeczność w pojęciu wszechmocy[20]. Sorensen nazywa eksperymenty myślowe oparte na takich narracjach, ujawniających sprzeczności pojęciowe, podważaczami możliwości (*possibility refuters*).

Najbogatszy podział *Gedankenexperimente* podaje James R. Brown (2011; Brown i Fehige, 2022)[21]. Rozróżnia on dwa naczelne typy – eksperymenty destruktywne (negatywne) i konstruktywne (pozytywne). Do tych pierwszych należą takie, które wykazują sprzeczności w danej teorii, analogicznie do eksperymentów krytycznych w sensie Poppera. Brown nazywa je eksperymentami *reductio ad absurdum*. Do destruktywnych należą też eksperymenty kontrintuicyjne – wykazują one sprzeczność teorii skonfrontowanej z naszymi posiadanymi już przekonaniami, w szczególności tymi zdroworozsądkowymi. Trzecim rodzajem eksperymentów destruktywnych są eksperymenty

[20] Potencjalną, bo nie brakuje badaczy uważających, że paradoks omnipotencji wynika z błędnych założeń na temat pojęcia wszechmocy (zob. Mordarski, 2015).

[21] Przyjmuję tu translację rodzajów eksperymentów za Robertem Piotrowskim (2011, s. 28, 50). Przykłady poszczególnych rodzajów eksperymentów można znaleźć w pracy Browna (2011, rozdz. 2 i 3).

podkopujące centralne założenie/przesłankę eksperymentu, np. poprzez analizę związków pomiędzy pojęciami stosowanymi w narracji. Ostatnim rodzajem są kontreksperymenty – negujące ustalenia wcześniejszych eksperymentów myślowych.

Eksperymenty konstruktywne Brown dzieli na bezpośrednie, postulatywne i pośredniczące. Pierwsze kończą się teorią wyjaśniającą opisywane w narracji zjawisko. W eksperymentach postulatywnych przedstawiane są sytuacje niekoniecznie możliwe w naszym świecie – zaliczymy do nich, oprócz kontrfaktycznych, eksperymenty z narracją o wysokim stopniu idealizacji. Trzecie natomiast przypominają heurystyczne w rozumieniu Poppera – wyjaśniają zjawiska poprzez ilustrację związków pomiędzy wyobrażaną w narracji sytuacją a eksplanacją w ramach danej teorii, której treść zakładana jest w toku eksperymentu[22].

Dodajmy jeszcze, że niektóre eksperymenty myślowe mają komponenty zarówno destruktywne, jak i konstruktywne. Brown nazywa takowe eksperymenty platońskimi, gdyż, w jego rozumieniu, dostarczają one nowej wiedzy o świecie mimo bycia czysto rozumowymi operacjami. Przykładem jest analizowany wcześniej eksperyment myślowy ze spadkiem swobodnym – Galileusz jednocześnie obala w nim ujęcie Arystotelesa i proponuje dla niego alternatywę. Jest ona „nowym" dodatkiem, gdyż nie wynika wprost z zanegowania tez Arystotelesa. Czasem też przyporządkowanie eksperymentu do danego rodzaju zależy od kontekstu, na którym się skupiamy w trakcie jego przeprowadzania. Kontreksperyment, dla Browna destruktywny, może być krytyczny w sensie Poppera, jeśli myślimy o jego kontekście użycia, czyli ataku na konkluzję wcześniejszego eksperymentu myślowego. Jednocześnie może być traktowany jako eksperyment apologetyczny, bo broni teorii krytykowanej przez ten wcześniejszy eksperyment. Podział Nortona na indukcyjne i dedukcyjne jest natomiast przydatny w ocenie ich niezawodności, nie powinien być jednak utożsamiany z kryterium ich wartości. W praktyce naukowej, podobnie jak w codziennych rozumowaniach, korzystamy zarówno z wnioskowań niezawodnych, jak i zawodnych.

[22] Przykładami eksperymentów bezpośrednich są wspomniane wcześniej eksperymenty z windą Einsteina oraz EPR. Eksperymentem postulatywnym jest natomiast rozumowanie znane jako wiadro Newtona, które miało uzasadniać istnienie absolutnej przestrzeni poprzez analizę ruchu obrotowego ciał. Z kolei przykładem eksperymentu pośredniczącego jest demon Maxwella, który został już krótko omówiony w jednym z wcześniejszych przypisów (Brown, 2011, s. 35–41).

Relacja między nauką a eksperymentami myślowymi jest, jak zobaczyliśmy, ujmowana rozmaicie: *Gedankenexperimente* ułatwiają przejścia do nowych paradygmatów, ilustrują i wskazują braki teoretyczne, służą jako metoda uprawiania nauki przez wyidealizowane eksperymentowanie, ale też umożliwiają przeprowadzanie eksperymentów empirycznych. Podkreśla się również ich rolę w pogłębianiu zrozumienia teorii naukowych (np. Camilleri, 2014; Stuart, 2016a)[23]. Praktyka – istotna dla emancypującej się historycznie nauki – nie ogranicza się zresztą do fizyki, co mógł sugerować dobór przywołanych przykładów. Eksperymenty myślowe stosuje się także w biologii (Schlaepfer i Weber, 2018) czy chociażby matematyce (Starikova i Giaquinto, 2018)[24].

2.5 ŚWIATY LITERACKIE JAKO EKSPERYMENTY MYŚLOWE

Naukowe eksperymenty myślowe rzadko spotykają się z krytyką, w przeciwieństwie do tych filozoficznych. Nawet w głosach krytycznych można jednak dostrzec uznanie dla przynajmniej niektórych naukowych przykładów tej praktyki (np. Thagard, 2014; zob. Hacking, 1992). Wystarczy zajrzeć do tekstów Nortona czy Poppera, aby zrozumieć, że punktem ich zainteresowania były właśnie naukowe eksperymenty myślowe[25]. Niemniej przywoływane wcześniej kategoryzacje są też odnoszone do kontekstu filozoficznego (chociażby przez Browna i Sorensena). Nauka i filozofia są jednak przedsięwzięciami specyficznymi, dobrze pasującymi do treści wyobrażanych w eksperymentach myślowych: swój cel lepszego poznania lub zrozumienia rzeczywistości często osiągają poprzez obserwację bądź rozważanie przypadków skrajnych, nietypowych, czy tworzenie sytuacji sztucznych. Jak jest w innych obszarach działalności intelektualnej człowieka, takich jak literatura?

W kontakcie z literaturą intensywnie korzystamy z wyobraźni, ale czy możemy mówić o niej w kategoriach eksperymentów myślowych? Filozofka

[23] Ujęć przydatności eksperymentów myślowych jest naturalnie więcej, zob. przykładowo rekonstrukcję poglądów na temat eksperymentów myślowych Paula Feyerabenda w pracy Stuarta (2021b).

[24] Najprawdopodobniej nie w każdej dyscyplinie naukowej równie łatwo o eksperymenty myślowe. Często przywoływaną dziedziną, w której odgrywały i odgrywają one niewielką rolę, jest chemia (zob. Snooks, 2006).

[25] Może zrodzić to podejrzenie, że naukowe eksperymenty myślowe różnią się czymś zasadniczym od innych, np. filozoficznych. W niniejszej pracy skupiam się na wspólnym rdzeniu omawianej praktyki, niezależnym od zakresu jej aplikacji: eksperymentowaniu w myślach jako działaniu umysłowym.

Catherine Elgin przekonuje, że tak. Dzieła literackie „funkcjonują jako eksperymenty myślowe. Zarówno fikcje literackie jak i naukowe eksperymenty myślowe są konstrukcjami wyobraźni pogłębiającymi rozumienie poprzez egzemplifikację cech i rozpatrywanie ich konsekwencji" (Elgin, 2007, s. 43). Egzemplifikacja oznacza tu relację pomiędzy próbką a tym, czego jest próbką. Słowo „kot domowy" w sposób dosłowny egzemplifikuje kota mieszkającego w domu, a przenośnie – przyjaciela czy członka rodziny. Obraz lecącego samolotu egzemplifikuje sam samolot, jak i latanie. Egzemplifikacja X-a nie jest odwołaniem się do dowolnego przykładu należącego do kategorii X, tylko do takiego, który manifestuje, podkreśla wybrane – dla autora najważniejsze – cechy X-a i marginalizuje te nieistotne (zob. Goodman, 1965). W ujęciu kognitywistycznym powinniśmy oczywiście poprawić Elgin i powiedzieć, że literatura może inicjować praktykę eksperymentów myślowych. Możemy rozważać wpływ na tę praktykę całych dzieł lub poszczególnych scen. Elgin w cytacie sugeruje wprawdzie pierwszą opcję, jednak niektóre jej przykłady sugerują drugą[26]. W przypadku najprostszych fabularnie dzieł literackich, takich jak baśnie, możemy najpewniej myśleć o nich całych jako inicjujących eksperymentowanie w myślach, ale powieści oraz opowiadania często są obszerne i zbyt interpretacyjnie złożone.

Aby uznać interpretowanie literatury jako podstawę eksperymentów myślowych, musimy przyjąć, że za dziełem (scenami) stoją tezy na temat świata realnego, do których przekonują nas autorzy. Musimy przyjąć kognitywizm co do literatury, czyli traktować ją jako źródło wiedzy (zob. Carroll, 2002; Egan, 2016). Dla Elgin (1993) dzieła literackie są „laboratoriami umysłu" takimi samymi jak w nauce i filozofii. Wyobrażamy sobie odpowiednie sytuacje wyjątkowe, a następnie rozpatrujemy ich konsekwencje zgodnie z hipotezami autorów i oceniamy, na ile naszym zdaniem są one prawdopodobne. W odróżnieniu od eksperymentów w nauce, w literaturze mogą być bardziej szczegółowe, choć pozbawione osadzenia w teoriach naukowych (Elgin, 2007, 2014)[27].

Jednak nie każda literacka egzemplifikacja opisuje sytuację wyjątkową, której konsekwencje moglibyśmy ocenić, aby lepiej zrozumieć świat. Ujęcie

[26] Przykładowo, rozważa scenę procesu z *Zabić drozda* Harper Lee jako eksperyment myślowy właśnie (Elgin, 2014, s. 221).

[27] Więcej na temat eksperymentów myślowych w literaturze i sporu o kognitywizm znaleźć można w pracach Davida Daviesa (2018) oraz Falka Bornmüllera i Mathisa Lessau (2019).

Elgin jest inkluzywne, ale jeśli w dziele brakuje opisu takiej sytuacji, to trudno mówić o przekonującym eksperymencie myślowym. Czytanie o kolejnym zmęczonym życiem detektywie może i jest dla nas olbrzymią przyjemnością, lecz niekoniecznie zmieni nasze poglądy na temat prawa i porządku – nie przekonały nas przecież wcześniej czytane, podobne opowieści[28].

W literaturze nie brakuje oczywiście dzieł skłaniających czytelników do nowego spojrzenia na świat. Przykładowo, Katja Hettich przeanalizowała jako podstawę eksperymentu myślowego *Teresę Raquin* Émile'a Zoli (2019), Giulia Agostini dzieła Samuela Becketta i Julio Cortázara (2019), a Alexander Fischer *Dżumę* Alberta Camusa (2019). Ujmowanie literatury jako podstawy eksperymentowania w myślach jest relatywnie nowe, sięga początku lat 90. XX wieku (Elgin, 1993). Literatura jednak nie ogranicza się do funkcji epistemicznej i może odkrywać świat w sposób nieprzypominający eksperymentowania. Czytelnicy fikcji mają większą swobodę interpretacyjną niż filozofowie i naukowcy. Jak przykładowo zauważa David Egan (2016, s. 143–144), możemy na *Śmierć Iwana Iljicza* Lwa Tołstoja spojrzeć jako na jeden eksperyment myślowy przekonujący, że wiedza na temat swojej śmiertelności nie zawsze sprawia, iż własna śmierć nas nie zaskakuje, ale nie jest to jedyna możliwa interpretacja. Utwór może być odczytany jako prosta historia sędziego, fragmentaryczna biografia nieistniejącej osoby. Możemy też docenić i analizować *Śmierć…* jako dzieło artystyczne. Widać więc, że eksperymenty myślowe w literaturze nie są tak jednoznaczne jak w filozofii czy nauce, w których dany tok myśli albo jest *Gedankenexperiment*, albo nie jest. W kontakcie z literaturą można eksperymentować w myślach, lecz nie trzeba.

2.6 PODSUMOWANIE

Eksperymentowanie myślowe i empiryczne różnią się – jedno przeprowadzamy w umyśle, drugie w świecie. Mają jednak ten sam cel: zrozumienie i wyjaśnienie rzeczywistości. Obie metody opierają się na badaniu sytuacji wyjątkowych – takich, które zaskakują, prowokując do rewizji teorii, testowania granic stosowalności określonych pojęć, lub też ujawniają nieoczywiste konsekwencje przyjmowanych założeń. Historycznie eksperymenty myślowe i empiryczne były uznawane za wiarygodne źródła wiedzy, a ich zastosowania ewoluowały wraz z rozwojem nauki i filozofii. W nauce oraz filozofii

[28] Chyba że to nasz pierwszy kryminał w życiu.

Gedankenexperimente pozwalają testować teorie, identyfikować i rozwiązywać anomalie, podczas gdy w literaturze mogą służyć za poszerzanie perspektywy i refleksji nad światem.

Idąc zarówno za przyrodnikami, jak i humanistami, możemy poczynić jeszcze kilka kluczowych wniosków:

1. Z uwagi na cel konstrukcji eksperymentów myślowych można wyróżnić ich dwa podstawowe rodzaje:
   - pozytywne eksperymenty myślowe – ich celem jest obrona istniejących teorii i wyjaśnień (eksperymenty apologetyczne), rozjaśnienie ich niuansów (eksperymenty heurystyczne/pośredniczące) lub prowadzenie do powstania nowych teorii i wyjaśnień (eksperymenty bezpośrednie);
   - negatywne eksperymenty myślowe – ich celem jest pokazanie niepoprawności istniejących już teorii i wyjaśnień (eksperymenty destruktywne/krytyczne), poprzez wykazanie ich sprzeczności wewnętrznej (eksperymenty *reductio ad absurdum*) albo sprzeczności w zestawieniu z posiadanymi już przekonaniami (eksperymenty kontrintuicyjne), a także ujawnienie niezauważonych wcześniej związków pojęciowych (eksperymenty pojęciowe). Eksperymenty negatywne mogą także wykazywać błędność ustaleń wcześniejszych eksperymentów myślowych (kontreksperymenty).
2. Spoglądając na eksperymenty myślowe od strony struktury wnioskowania, możemy rozróżnić:
   - indukcyjne („Typ 1" u Nortona) i dedukcyjne eksperymenty myślowe („Typ 2" u Nortona, podział z racji na typ wnioskowania);
   - podważacze konieczności i możliwości (podział Sorensena, dyktowany różnicą w kontekście modalnym eksperymentów).
3. Nie każde użycie wyobraźni stanowi równie silną podstawę eksperymentu myślowego, podobnie jak nie każda sytuacja zewnętrzna nadaje się do pomiaru naukowego. Zarówno empiryczne eksperymenty naukowe, jak i myślowe obierają swój przedmiot zgodnie z pytaniami badawczymi autorów i charakteryzują się swoistą sztucznością: izolowaniem czynników istotnych i eliminacją tych nieistotnych. Wspólną cechą obu form eksperymentowania jest więc ich zainteresowanie sytuacjami wyjątkowymi.

Eksperymenty myślowe, mimo swojej subiektywności, pozostają skutecznym narzędziem w nauce, filozofii i humanistyce. Ich argumentacyjna siła nie wynika jednak wyłącznie z logicznej struktury. Jak to się dzieje, że niektóre z nich potrafią nas przekonać do swoich konkluzji, nawet jeśli nie

wyprowadzają ich dedukcyjnie? Niektórzy, jak Brown, widzą w tym platoński wymiar – dostarczanie nowej wiedzy o świecie przez czysto rozumowe operacje. Nie musimy jednak popadać w skrajnie filozoficzną debatę o pochodzeniu wiedzy. Kognitywistyczna perspektywa sugeruje bardziej pragmatyczne pytanie: jaka jest rola wyobraźni i intuicji w eksperymentowaniu w myślach? Zrozumienie tych procesów pozwoli lepiej uchwycić, w jaki sposób eksperymenty myślowe mogą być przekonujące, niezależnie od (nie)zawodności ich formalnej struktury logicznej.

Co to więc znaczy: „wyobrażać sobie"?

3 WYOBRAŹNIA: MYŚLENIE O TYM, CO NIEOBECNE

Spróbujmy zwizualizować sobie w umyśle liść klonu porwany przez wiatr. Nie będziemy mieć większych trudności w ocenie, czy przeżywany przez nas obraz przedstawia jesienne odrywanie się od drzewa złoto-czerwonego liścia wskutek mocnego, świszczącego wiatru. Być może wyobraziliśmy sobie opadający liść bez drzewa, a może też liść był zielony. Być może wyobraziliśmy sobie działanie wiatru, ale nie jego dźwięk. Możliwości jest wiele, lecz istota procesu pozostaje niezmienna. Świadomie przeżywamy pewne treści zmysłowe – zarówno eksterocepcyjne (wzrokowe, słuchowe, dotykowe itd.), jak i interocepcyjne (czucie stanu ciała towarzyszące chociażby wyobrażaniu sobie bólu czy radości)[29] – mimo braku bodźców, które by te treści mogły zapewnić. Używamy wyobraźni.

Wyobraźnia jest obecna nawet w podejściu Nortona (1991, 2004b), uznającego eksperymenty myślowe za układy przesłanek i wniosków ubrane w „obrazkową" formę. Każde ujęcie eksperymentów myślowych traktuje ją jako element niezbędny, nierzadko też jej rola jest wręcz podkreślana (Stuart i in., 2018b). Wyobraźnia jest jednak specyficznym rodzajem myślenia, dlatego aby móc zrozumieć jej znaczenie dla przebiegu eksperymentu myślowego, musimy być bardziej przejrzyści co do jej natury. Nie jest to łatwe, osoba pragnąca wyłuskać konsensus na temat istoty wyobraźni i zapoznająca się chociażby z jedną z przekrojowych prac zbiorowych poświęconych tej tematyce – jak *The Routledge Handbook of Philosophy of Imagination* (2016) albo *The Cambridge Handbook of the Imagination* (2020) – znajdzie się w nie lada tarapatach.

[29] Panuje pewna kontrowersja co do tego, czy można wyobrazić sobie odczuwanie emocji, ale idea ta znajduje obecnie poparcie zarówno eksperymentalne, jak i filozoficzne (Cocquyt i Palombo, 2023; Todd, 2023).

3.1 WYOBRAŹNIA PERCEPCYJNA A POJĘCIOWA

Nie każdy przypadek myśli o treści percepcyjnej, jak rozważane przed chwilą spadanie liścia klonu, będzie niekontrowersyjnym przykładem wyobraźni. Umysłowe obrazowanie (*mental imagery*) jest odmiennym pojęciem niż wyobraźnia, nawet jeśli uznamy, że wyobrażać możemy sobie jedynie treści zmysłowe. Nagle powracające do nas wspomnienia nieprzyjemnych sytuacji często mają określone treści zmysłowe (kolor koszuli naszego szefa podczas ostatniej oceny okresowej, soczystość zerwanego z drzewa jabłka w ogrodzie z dzieciństwa), ale nie pojawiają się one, bo tego pragniemy. Są przypadkami obrazowania umysłowego, natomiast nie wszyscy nazwaliby je wyobraźnią. Niektórzy tę ostatnią kategorię rezerwują wyłącznie dla celowego, wolicjonalnego przywoływania sobie pewnych treści w umyśle (Nanay, 2018; zob. Williamson, 2016, s. 114–115).

Ponadto wyobraźnia nie zawsze jest tak wyraźnie percepcyjna jak umysłowe obrazowanie[30]. Również myślenie pozbawione jednoznacznej treści zmysłowej wielu badaczy nazywa wyobraźnią: postawową, propozycjonalną czy też pojęciową (Kind i Kung, 2016; Langland-Hassan, 2020; van Leeuwen, 2013; Williamson, 2016, s. 117). Rozważmy graniczny przypadek, możliwy do wyobrażenia zarówno z obrazem, jak i bez niego. Czytając cykl o Geralcie Andrzeja Sapkowskiego, zdajemy sobie w pewnym momencie sprawę, że wiedźmini starzeją się wolniej. Możemy wyobrazić sobie tę treść wizualnie jako film puszczony w szybkim tempie, ukazujący twarze starzejących się zwykłych ludzi i wiedźmina. Twarze tych pierwszych szybko pokrywają zmarszczki i zmęczenie, twarz wiedźmina pozostanie zaś względnie niezmieniona. Taki akt wyobraźni jest czasochłonny (musimy przecież „obejrzeć film") i wymaga szczegółowego odwzorowania (zmian na skórze twarzy). Nie uchwycimy też takim „filmem" istoty starzenia – to zdecydowanie bardziej złożone zjawisko niż tylko postępujące zmiany skórne. Obejmuje zarówno zmiany biologiczne, jak i psychologiczne. Alternatywnie możemy po prostu przyjąć, że stwierdzenie o wolniejszym starzeniu się wiedźminów jest prawdziwe w kontekście fikcyjnego świata Sapkowskiego. Taki akt będzie zgoła prostszy, wymaga od nas jedynie przyjęcia stosownej postawy (powiedzmy roboczo: wyobrażeniowej) względem długowieczności wiedźminów.

[30] Być może trafniej byłoby nazywać wyobrażanie sobie liścia klonu wyobraźnią „quasi-percepcyjną" albo „quasi-zmysłową". Wybieram „percepcyjną" dla wygody.

Rozpatrzmy drugi przykład: wiedźmini zwykle są sierotami. Jak jednoznacznie i bez szczegółów, percepcyjnie, wyobrazić sobie bycie sierotą? Moglibyśmy wyobrazić sobie nieco dłuższy film, obejmujący narodziny przyszłego wiedźmina, śmierć rodziców bądź porzucenie i odnalezienie przez starszych wiedźminów. Taki film byłby jednak jeszcze bardziej złożony, a puszczenie go jeszcze bardziej czasochłonne niż w przypadku poprzedniego. Tymczasem w praktyce dopuszczanie myśli o sierocym pochodzeniu wiedźminów nie wiąże się wcale z wiele większym trudem niż dopuszczenie myśli o wolniejszym starzeniu się wiedźminów. Wystarczy, że przyjmiemy, iż tak właśnie jest w świecie wiedźminów.

Trzeci przypadek jest najtrudniejszy do wyobrażenia przy pomocy obrazów umysłowych: wyobraźmy sobie, że żaden wiedźmin nie marzy o posiadaniu konia. Zadaniem karkołomnym byłaby próba wyobrażenia sobie tej sytuacji przy pomocy treści zmysłowych. Jaka treść percepcyjna reprezentuje brak marzeń? Kiedy wyobrażamy sobie tę sytuację, w naszych umysłach mimowolnie mogą pojawić się pewne obrazy, reprezentujące konia, idylliczne krajobrazy czy wiedźminów, ale nie są to obrazy odnoszące się do braku marzeń. Co najwyżej moglibyśmy powiedzieć, że przy okazji wyobrażania sobie wiedźmińskiego braku marzeń o koniach dodatkowo wyobrażamy sobie wiedźmina i konia[31].

Niektórzy przekonują, że myśli pozbawione treści percepcyjnych wyobrażeniami być nie mogą (Weatherson, 2004). Podążają w tym sensie za tradycją Kartezjusza, który pisał następująco[32]:

> […] zwracam najpierw uwagę na różnicę pomiędzy wyobraźnią a czystym ujęciem intelektualnym czy pojmowaniem. […] Gdy na przykład wyobrażam sobie trójkąt, to nie tylko pojmuję, że jest to figura składająca się z trzech linii, lecz zarazem napotykam te trzy linie jako coś obecnego dzięki sile i wewnętrznej pracy mego umysłu. I właśnie to nazywam wyobrażeniem sobie. Jednakże gdy chcę pomyśleć o tysiącboku, to pojmuję prawdziwie, że jest to figura złożona z tysiąca boków […], ale nie mogę wyobrazić sobie tysiąca boków […]. I chociaż z nawyku, jaki mam, aby posługiwać się wyobraźnią, mogę

[31] Więcej o różnicach między obrazowaniem umysłowym a wyobraźnią znaleźć można w pracach Margerity Arcangeli (2020) i Petera Langlanda-Hassana (2015).

[32] Kartezjusz pisał *Medytacje* po łacinie, a przytaczany cytat był tłumaczony w różny sposób. Alternatywne, starsze tłumaczenie jego najważniejszego fragmentu było następujące: „zbadam najpierw różnicę, jaka istnieje między wyobrażaniem a czystym pojmowaniem (*intellectio*). […] wyraźnie spostrzegam, że dla wyobrażenia sobie czegoś potrzeba szczególnego wysiłku ducha, którego nie potrzebuję przy pojmowaniu czegoś" (Descartes, 1948).

> przedstawić sobie niejasno jakąś figurę, to jednak jest całkiem oczywiste to, że figura owa wcale nie jest tysiącbokiem, bo i nie różni się ona w niczym od tego, co przedstawiłbym sobie, myśląc o dziesięciotysiącoboku […] jak również i to, że w żadnej mierze nie przyda się ona do wykrycia własności, które odróżniają tysiącobok od innych wieloboków. […] pojmuję jasno, że potrzeba mi znacznego wysiłku umysłu, aby coś sobie wyobrazić, podczas gdy pojęcie czy zrozumienie czegoś wcale go nie wymaga (Descartes, 2005, s. 81–82).

Zwolennicy takiego rozumienia wyobraźni podkreślają, że wyobrażenia percepcyjne mają specyficzną fenomenologię: są one pierwszoosobowymi przeżyciami odzwierciedlającymi to, czego naszym zdaniem doznawałaby osoba znajdująca się w wyobrażanej przez nas sytuacji[33]. Wyobrażając sobie picie coca-coli, doświadczamy również w pewnym stopniu jej smaku. Wyobrażając sobie widok ze szczytu Mont Blanc, rekonstruujemy świadome przeżycie wzrokowe, które w naszej opinii miałaby osoba znajdująca się na szczycie. Magdalena Balcerak Jackson (2016) nazywa tę cechę wyobraźni doświadczeniową perspektywicznością (*experiential perspectivity*).

3.2 FUNKCJONALNE UJĘCIE WYOBRAŹNI

Zdaniem zwolenników wąskiego, percepcyjnego rozumienia wyobraźni, kiedy myślimy o wiedźmińskim braku marzeń o koniach, przypuszczamy (*suppose*), że wiedźmini nie marzą o koniach. Przypuszczanie Balcerak Jackson rozumie jako akceptowanie, w technicznym sensie wprowadzonym przez Roberta Stalnakera (1984): przypuszczać, że jest tak, jak głosi stwierdzenie X, oznacza traktować X jako prawdziwe w danym kontekście i ignorować, że mogłoby być fałszywe. Badaczka (Balcerak Jackson, 2016, s. 46–48) wskazuje też na szereg innych różnic – obok fenomenologii – między wyobrażeniami a przypuszczeniami. Po pierwsze, wyobrażanie może się nie powieść, przypuszczenia zaś zwykle się powodzą. Jeśli nigdy nie jeździłem na koniu, mogę nieadekwatnie zobrazować przeżycia doznawane przez konnego jeźdźca. Mimo tego nie mam problemu, by przypuszczać, że pewien ktoś jeździ na koniu. Po drugie, łatwiej przypuszczać stany rzeczy stwierdzające nieposiadanie pewnej cechy (np. wiedźmini nie marzą o koniach), niż je sobie wyobrazić. Po trzecie, jesteśmy często niechętni, by wyobrazić sobie pewne sytuacje (np. zachowania traumatyzujące) – nazywamy to oporem wyobrażeniowym

[33] Przystępne wprowadzenie do współczesnej, zorientowanej naukowo myśli fenomenologicznej czytelnik znajdzie w pracy Shauna Gallaghera i Dana Zahaviego (2015).

(*imaginative resistance*) (Gendler i Liao, 2016) – ale jednocześnie nie mamy problemu z przypuszczaniem, że takie sytuacje mogą mieć miejsce. Po czwarte, wyobraźnia może w przeciwieństwie do przypuszczania być celem samym w sobie – wyobrażamy sobie pewne treści jedynie po to, aby je odczuć (np. by odczuć przyjemność, którą jeszcze mocniej czulibyśmy, gdyby wyobrażana sytuacja się ziściła). Po piąte, wyobrażenie może zmotywować do działania (np. wyobraźmy sobie radość po ukończeniu wyjątkowo trudnego projektu, co skłania nas do szybszego uporania się z nim), przypuszczenie natomiast motywuje najwyżej do rozmyślań.

Jeśli Balcerak Jackson ma rację odnośnie do obserwowanych różnic, to czy dowodzą one, że przypuszczanie nie jest formą wyobraźni? Wydaje się, że niepotrzebnie dzielilibyśmy włos na czworo. Różnice między przypadkami są normalne, nawet jeżeli należą one do jednej ogólniejszej klasy. Modraszka zwyczajna ma charakterystyczne żółte upierzenie na brzuchu i zwykle żyje w grupie, a jastrząb ma ciemno-jasny, prążkowany brzuch i żyje głównie samotnie. I jedno, i drugie to ptak. Podobnie jest z percepcją. Wzrokowo postrzegane obiekty są przede wszystkim rozciągłe, zajmują określoną przestrzeń w otoczeniu, elementy odbierane słuchowo są zaś przede wszystkim czasowe, trwają określony czas. Aczkolwiek oba procesy dostarczają nam tego samego: przydatnego obrazu otoczenia. Podobnie, pomimo różnic w sposobie reprezentowania treści w umyśle, treści te służą temu samemu – myśleniu o nieobecnym i przyjmowaniu na chwilę perspektywy, której nie przyjmujemy w kontakcie z aktualnym otoczeniem.

Wyobraźnia w funkcjonalnym, szerokim ujęciu obejmuje każde myślenie na temat „możliwego, udawanego, nierealnego […] fikcyjnego i fantastycznego" (Langland-Hassan, 2020, s. 58). Do jej aktywności zaliczymy wspominanie, prospekcję (myślenie o przyszłości w rozmaitych jego formach, zob. Schacter i in., 2012; Szpunar i in., 2014), myślenie hipotetyczne i kontrfaktyczne (Kulakova i in., 2013)[34], snucie marzeń, myślenie towarzyszące nam w czasie podejmowania działań udawanych, przypisywanie innym stanów

[34] Myślenie kontrfaktyczne i myślenie hipotetyczne są przykładem myślenia warunkowego, ale są od siebie odróżnialne. W tym pierwszym zakładamy, że rzeczy potoczyły się inaczej niż faktycznie lub że pewna nasza wiedza jest fałszywa (np. „Załóżmy, że Bolesław Chrobry nie był pierwszym królem Polski", „Załóżmy, że Układ Słoneczny nie powstał"). W myśleniu hipotetycznym natomiast zakładamy, że mają miejsce rzeczy (jeszcze) niewiadome, ale możliwe (np. „Załóżmy, że człowiek skolonizował Marsa", „Załóżmy, że świadomość jest iluzją").

umysłowych, proces odbioru dzieła sztuki czy chociażby reprezentowanie w umyśle fikcyjnego świata w trakcie lektury książki (Abraham, 2016; Kushnir, 2022). Aktywności te nie są wzajemnie się wykluczające – elementem odbioru dzieła sztuki może być proces myślenia warunkowego, a tego zaś elementem snucie marzeń, myślenie o stanach umysłowych innych osób itd.

Funkcjonalne rozumienie wyobraźni nie pozbawia nas możliwości badania i mówienia na temat szczegółowych aspektów i mechanizmów zaangażowanych w akty wyobraźni. Jako że badania nad różnymi formami wyobraźni są w zasadzie prowadzone osobno, to łatwo zapomnieć o ich pokrewieństwie[35]. Możemy niemniej pokusić się o wskazanie kilku zależności. Rozważanie alternatywnych strategii, które mogła przyjąć Kartagina w wojnach punickich, przypuszczalnie bardziej niż wyobrażanie sobie jazdy na rowerze angażuje myślenie analityczne. Wspomnienia osobiste mogą wywołać silne emocje w przeciwieństwie do niezwiązanego z naszą biografią wyobrażania sobie jaskini platońskiej[36].

3.3 NEURONALNE PODSTAWY FUNKCJONALNIE ROZUMIANEJ WYOBRAŹNI

3.3.1 Sieć domyślna i jej podsystemy

W badaniach eksperymentów myślowych rzadko zwraca się uwagę na wyniki neuronauki. Również neuronauka nie zainteresowała się póki co *Gedankenexperimente*. Na przykład w *The Routledge Companion to Thought Experiments* jedynym odwołaniem do neuronalnych podstaw procesów poznawczych jest wzmianka na temat neuronów lustrzanych (Stuart i in., 2018b, s. 317)[37]. Jeśli weźmiemy pod uwagę popularność eksperymentów myślowych w filozofii i sięgniemy po *The Oxford Handbook of Philosophical Methodology*, to znajdziemy po jednym tekście o roli neuronauki i psychologii

[35] Przykładowo, badania nad mentalizacją prowadzone są przede wszystkim w kontekście rozwojowym, psychopatologii i poznania społecznego (Spaulding, 2019), z rzadka natomiast są łączone z badaniem nad snuciem marzeń (Poerio i Smallwood, 2016).

[36] Ujęcie funkcjonalne nie determinuje, czy stany wyobrażeniowe (*imaginings*) są stanami *sui generis*, swoistego rodzaju, niesprowadzalne do konfiguracji bardziej podstawowych stanów umysłowych (jak przekonania czy pragnienia). Nad tą kwestią aktualnie intensywnie się debatuje (zob. Abraham, 2016, 2020; Langland-Hassan, 2015, 2020).

[37] Warto zastrzec, że po początkowym zachłyśnięciu się ideą neuronów lustrzanych świat naukowy i popularnonaukowy przeżywa dziś swego rodzaju otrzeźwienie (zob. Hickok, 2016).

w filozofii (Cappelen i in., 2016, s. 554–606). Sytuacja wygląda nieco lepiej w przypadku badań wyobraźni (zerknąć warto chociażby do *The Cambridge Handbook of the Imagination*, Abraham, 2020).

Zgodnie z jedną z pierwszych propozycji integrujących – wyjaśniających pracę wyobraźni w jej różnych formach (zarówno percepcyjnej, jak i abstrakcyjnej, pojęciowej) – z zakresu neuronauki poznawczej wyobraźnia realizowana jest przy udziale mózgowej sieci domyślnej (*default mode network*, Andrews-Hanna i Grilli, 2021). Sama sieć domyślna jest w ostatnich latach intensywnie badana przez neuronaukę. Jej aktywność jest przede wszystkim zorientowana do wewnątrz systemu poznawczego – jest zaangażowana przy skupianiu się na swoim środowisku wewnętrznym, umysłowym (Davey i in., 2016). Wspomaga przetwarzanie emocjonalne, samozwrotną aktywność umysłową (a więc taką, która korzysta z informacji o nas samych), swobodne – nieograniczone dodatkowymi poleceniami – myślenie, proces wspominania i myślenia o przyszłości, percepcji społecznej, błądzenie w myślach, a także uczestniczy w przygotowywaniu i realizacji zachowań refleksyjnych (Andrews-Hanna i in., 2014; Raichle, 2015; Uddin i in., 2019).

Wyobraźnia realizowana ma być przy udziale dwóch podsystemów sieci domyślnej, nazywanych okiem umysłu i umysłem umysłu (Andrews-Hanna i Grilli, 2021). Oko umysłu odpowiada za wyobraźnię percepcyjną, sprecyzowane kontekstowo myślenie o tym, co aktualnie nieobecne w środowisku: wyobraźnię towarzyszącą śledzeniu sytuacji opisywanych w książkach, umysłowemu nawigowaniu znajomej sobie przestrzeni czy wspominaniu przeszłych wydarzeń z życia. Umysł umysłu natomiast odpowiada za wyobrażanie sobie abstrakcji, gdy wnioskujemy, wykorzystując zależności między poszczególnymi znanymi nam pojęciami (przykładowo, kiedy ze stwierdzenia „Każdy pies jest ssakiem" wyciągamy wniosek „Każdy jamnik jest ssakiem"), ale także gdy myślimy o cudzych i własnych stanach umysłowych, próbujemy zrozumieć sens całych konstrukcji narracyjnych (w przeciwieństwie do wyobrażania sobie poszczególnych, opisywanych w tych narracjach, scen) czy też wydobywamy abstrakcyjną wiedzę ze swojej pamięci semantycznej. Różnice w funkcjach mają także swoje odbicie w różnicy aktywacji sieci domyślnej. Oko umysłu to przede wszystkim aktywność przyśrodkowo-skroniowej części sieci domyślnej: kory retrosplenialnej, formacji hipokampa i kory przyhipokampowej, tylnej części dolnego płata ciemieniowego i brzusznej części przyśrodkowej kory przedczołowej, jak również większa aktywność centralnej dla sieci domyślnej kory obręczy, podczas gdy umysł umysłu to aktywność

grzbietowo-przyśrodkowej kory przedczołowej, styku skroniowo-ciemieniowego, a także przednich i bocznych części płata skroniowego (Andrews-Hanna i in., 2010, 2014; Andrews-Hanna i Grilli, 2021; Raffaelli i in., 2020).

3.3.2 Badania przypadków: afantazja, amnezja, demencja

Przynależność różnych aktywności umysłowych do wspólnej klasy form wyobraźni sugerują też zaobserwowane zależności neuronalno-behawioralne. Afantazja to przypadłość objawiająca się trudnościami w obrazowaniu umysłowym, dotycząca około 4% populacji (Dance i in., 2022). Osoby cierpiące na afantazję w znacznie mniejszym stopniu doznają (quasi)percepcyjnych treści oderwanych od aktualnej sytuacji w ich środowisku. W skrajnych przypadkach ich myślenie ma charakter czysto pojęciowy (Zeman i in., 2015). Afantazja nie ogranicza się do widzenia, może obejmować także inne zmysły, np. słuch (Hinwar i Lambert, 2021)[38]. Ma również swoje reperkusje na poziomie przetwarzania emocjonalnego: zaobserwowano chociażby, że osoby z afantazją mają mniej intensywną skórno-galwaniczną reakcję – ich dłonie mniej się pocą – w trakcie lektury opowieści z dreszczykiem niż osoby bez afantazji. Okazuje się ponadto, że afantazja przekłada się na procesy pamięciowe i prospekcyjne. Afantastycy przywołują mniej szczegółów na temat przyszłych (domniemanych) i przeszłych (ze swojego życia) wydarzeń niż osoby neurotypowe (Dawes i in., 2022).

Jeszcze inną zależność zaobserwowano u cierpiących na amnezję w wyniku uszkodzeń hipokampa oraz wśród osób z demencją. Ich trudności nie ograniczają się do problemów z przywoływaniem faktów ze swojego życia, obejmują też prospekcję, wyobraźnię percepcyjną, a nawet korzystanie z mocno sugestywnego percepcyjnie języka (Hassabis i in., 2007; Irish i Piolino, 2016). Demencja semantyczna zdaje się mieć efekt behawioralny częściowo odwrotny. Na poziomie neuronalnym objawia się degeneracją m.in. skroniowych części podsystemu umysł umysłu, na poziomie behawioralnym zaś stopniowym pogarszaniem się zdolności do przypasowywania słów do znaczeń, niemniej osoby sprawniej posługują się sugestywnymi percepcyjnie słowami niż czysto abstrakcyjnymi (Jefferies i in., 2009).

[38] Afantazja nie oznacza automatycznie posiadania bogatszej niż u osoby neurotypowej mowy wewnętrznej. Wyniki Risha Hinwara i Anthony'ego Lamberta (2021) sugerują zależność dokładnie odwrotną, ale bardziej systematyczne badania nad związkiem między mową wewnętrzną a afantazją dopiero są prowadzone.

Ograniczenia w wyobraźni percepcyjnej mogą wpływać na zdolność do pełnego angażowania się w narracje eksperymentów myślowych, szczególnie tych opartych na quasi-zmysłowych treściach. Nie oznacza to jednak, że osoba z afantazją nie jest zdolna do przeprowadzania eksperymentów myślowych. Może jednak oceniać ich konkluzje inaczej od osoby z pełniejszą wyobraźnią percepcyjną, ponieważ w większym stopniu polega na operacjach czysto pojęciowych.

3.3.3 Wartość wyobraźni

Życie bez wyobraźni byłoby bardzo trudne, by nie rzec – niemożliwe. Osoby z zaawansowaną demencją, amnezją i innymi podobnymi zaburzeniami pamięci i prospekcji nie są w stanie samodzielnie funkcjonować. To samo dotyczy innych form wyobraźni. Zanik spontanicznego myślenia kontrfaktycznego, konsekwencja uszkodzenia kory przedczołowej, przekłada się na ogromną trudność w uczeniu się na błędach, odczuwaniu żalu i dostrzeganiu konsekwencji decyzji, jak i samym ich podejmowaniu (Camille i in., 2004)[39]. Również problemy w rozpoznawaniu stanów umysłowych utrudniają życie, odnajdywanie się w sytuacjach społecznych, tak w kontekście rozwojowym, jak i u osób dorosłych (Schaafsma i in., 2015; zob. Crespi i in., 2016).

Wartość wyobraźni jest też widoczna na ogólniejszej płaszczyźnie, nie tylko w odniesieniu do przypadków nietypowych. Brak wyobraźni to brak przewidywania, planowania, przygotowywania się na to, czego jeszcze w środowisku nie ma, uczenia się na podstawie tego, co w nim już przeminęło, czy nawet rozumienia komunikatów językowych. Bez wyobraźni nie stalibyśmy się – posługując się nomenklaturą Daniela Dennetta – istotami gregoriańskimi, pozostawiającymi w środowisku gotowe rozwiązania problemów poznawczych w postaci fizycznych i kulturowych narzędzi, ani nawet popperowskimi, mogącymi dokonywać preselekcji behawioru dzięki umiejętności odwzorowywania środowiska zewnętrznego w swoim środowisku wewnętrznym (Dennett, 1997, rozdz. 4). Istotnie, w antropologiczno-kognitywistycznych historiach naszego gatunku podkreśla się wartość przystosowawczą, jaką dla *homo sapiens* miało oderwanie od tego, co bezpośrednio dane w kontakcie z otaczającym środowiskiem. Dodajmy tu, że historia ewolucji wyobraźni jest tak naprawdę dłuższa. Jej różne postaci są często traktowane przez badaczy

[39] Nie ma przy tym znaczenia rodzaj uszkodzenia, czy jest ono wynikiem wypadku, choroby neurodegeneracyjnej (McNamara i in., 2003) czy schizofrenii (Roese i in., 2008).

jako kryterium poznania złożonego, którego przejawy zaobserwowano w często odległych od siebie gałęziach drzewa życia: u krukowatych, naczelnych (szczególnie u człekokształtnych), głowonogów (zob. Emery i Clayton, 2004; Krupenye i Call, 2019; Mather i Dickel, 2017)[40]. W porównaniu z innymi gatunkami ludzie wykazują wyjątkowo rozwiniętą dyspozycję do umysłowego odrywania się od aktualnych sytuacji w środowisku (Fuentes, 2018). Dyspozycja ta manifestuje się w różnorodny sposób, w tym w umiejętności długofalowej współpracy (Tomasello, 2015), językowej komunikacji i myśleniu symbolicznym (Fuchs, 2016) czy choćby umysłowych podróżach w czasie (Suddendorf i in., 2018).

Dalsze badania pozwolą uszczegółowić nasze rozumienie mechanizmów neuronalnych zaangażowanych w różne formy wyobraźni i ich wzajemne zależności. Prowadzone są obecnie badania nad rolą form wyobraźni w kontekście procesów pamięciowych i prospekcyjnych (Martin-Ordas, 2020; Schacter i in., 2012), relacją między śnieniem, percepcją a wyobraźnią (Windt, 2020), znaczeniem wyobraźni dla poznania społecznego w dzieciństwie (Kushnir, 2022) czy neuronalnym podłożem warunkowego rozumowania o alternatywnych i przyszłych scenariuszach działań (Miyamoto i in., 2023).

Neuronalne splecenie różnych form wyobraźni wskazuje, jak należy patrzeć na jej rolę w kontekście eksperymentów myślowych. Jako że są one rozumowaniami o fikcyjnych sytuacjach, to zawsze wykorzystywaną w ich przypadku formą wyobraźni jest myślenie warunkowe. Jednakże na ocenę powodzenia eksperymentu – czyli ewentualnego zgodzenia się z autorem – wpływ mogą mieć potencjalnie wszystkie inne formy aktywności: sięganie do biograficznych wspomnień i wiedzy pojęciowej (chociażby celem oszacowania prawdopodobieństwa kreślonej w eksperymencie fikcyjnej sytuacji) czy też refleksja nad własnymi i cudzymi poglądami, przyszłością lub kwestiami abstrakcyjnymi. Te aktywności przynależą do tej samej klasy wyobrażeniowej, w której są percepcyjne treści *Gedankenexperimente*, jak w przypadku pierścienia Gygesa, gdzie możemy wyobrazić sobie osoby znikające po włożeniu pierścienia, czy rozważań Archytasa o nieskończoności świata, gdzie towarzyszyć może nam wizja osoby wyciągającej rękę w kosmosie. Przeprowadzając

[40] Lista ta nie jest wyczerpująca, wymieniane często są też chociażby gryzonie. Trudności i kontrowersje w badaniu umysłowych podróży w czasie u zwierząt innych niż ludzie zostały szerzej opisane w pracach Lucy Cheke i Nicoli Clayton (2010), a także Ekrema Derego i innych (2018).

rozumowanie Wittgensteina z żukiem, możemy zobaczyć w wyobraźni barwę i inne cechy wyglądu żuka ukrytego w naszym pudełku (od zapałek?). Śledząc, w końcu, tok myślenia Galileusza, możemy percepcyjnie wyobrażać sobie kamienie (o kulistym kształcie?) spadające z wysokości (zrzucone z Krzywej Wieży?). Niektóre z tych wyobrażanych treści uzupełniają to, o czym mówi tekst przeprowadzanych eksperymentów myślowych: nie było wszak informacji, że pudełko, w którym znajduje się żuk, jest pudełkiem od zapałek. Podobnie w dialogu Galileusza nie było mowy o Krzywej Wieży. Zwykle jednak nie jesteśmy w stanie wyobrazić sobie percepcyjnie wszystkich aspektów eksperymentów. Sprawiedliwość czy kraniec świata pozostaje poza możliwością wiernego obrazowego oddania[41].

Choć zapewne przesadą jest wyświechtane powiedzenie, że obraz wart jest tysiąca słów, to udział wyobraźni percepcyjnej ma fundamentalne znaczenie dla siły oddziaływania narracji eksperymentów myślowych. W naturalnym, codziennym nastawieniu do świata jesteśmy naiwnymi realistami – automatycznie i mimowolnie akceptujemy to, co spostrzegamy, jako istniejące (Martin, 2002). I trudno, aby było inaczej, percepcja umożliwia nam szybkie zrozumienie tego, co dzieje się dookoła i w naszych ciałach. Gdy widzimy samochód niezwalniający mimo naszego wejścia na pasy, korzystniej jest odskoczyć niż zastanawiać się nad istnieniem nadjeżdżającego bolidu. Jako że neuronalnie wyobraźnia percepcyjna angażuje przetwarzanie percepcyjne (Ganis i in., 2004), to nic dziwnego, iż często łatwiej przekonać drugą osobę, kiedy oprócz abstrakcyjnego argumentu odwołamy się do czytelnej ilustracji (Birdsell i Groarke, 2007).

3.4 MYŚLENIE WARUNKOWE

Skupmy się jednak na samym myśleniu warunkowym, bo bez niego eksperyment myślowy nie może istnieć. I tak, wyobrażając sobie, jak mogłaby przebiec inaczej niepomyślna sytuacja (której doświadczyliśmy bądź o której usłyszeliśmy), jesteśmy skłonni skupić się na tych jej elementach, które uznamy za wyjątkowe, nietypowe (Kahneman i Tversky, 1982). Jeżeli jeździmy

[41] Możliwe, że istnieją czysto pojęciowe eksperymenty myślowe, przeprowadzaniu których nie towarzyszą jakiekolwiek wyobrażane treści zmysłowe – musiałyby one jednak dotyczyć czysto formalnych rzeczy. Być może w niektórych postaciach matematyki i logiki przeprowadza się tego rodzaju operacje (zob. Arana, 2016; Poincaré, 1956).

z przyznaną ulgą tramwajem, a zapomnimy dokumentów i otrzymamy mandat, prawdopodobnie będziemy żałować tego, że nie wzięliśmy dokumentów, a nie tego, że nie wybraliśmy innego środka transportu. Nietypowy był brak dokumentów, nie tramwaj.

Umysł naturalnie dąży do ekonomiczności w myśleniu warunkowym, nie tylko realizując zasadę wyjątkowości, ale też ograniczając rozważane scenariusze do tych uznawanych za najbardziej prawdopodobne. Jeśli dopuszczamy tylko jeden rozwój danego scenariusza, to postrzegamy go wręcz jako „prawdziwy" (Byrne, 2005, rozdz. 2). Przyjmując warunek „Jeśli Adam przybył na czas, to zjadł obiad" bez dodatkowych informacji, zignorujemy inne możliwości, skupiając się na najbardziej oczywistym rozwiązaniu, że Adam faktycznie przybył i zjadł obiad. Pozostałe kombinacje, mimo że są logicznie możliwe, wymagają od nas większego wysiłku umysłowego i dlatego, dopóki nie zostaniemy wprost poproszeni, ich nie rozważymy (Espino i in., 2009).

Czasami rozważamy od razu kilka możliwości, czego koronnym przykładem jest rozumowanie kontrfaktyczne (zob. Byrne, 2018). Aby stwierdzić, że mamy do czynienia z treścią kontrfaktyczną, musimy wiedzieć, że choć wyrażaną przez tę treść możliwość traktujemy jako wydarzoną, to w istocie się nie wydarzyła. Wymóg utrzymywania więcej niż jednej możliwości w umyśle naraz, a więc i większą złożoność myślenia kontrfaktycznego niż czysto hipotetycznego, potwierdzają dane neuroobrazowe (Kulakova i in., 2013). Jednak nawet tu umysł stosuje oszczędne strategie. Na przykład w sytuacji z tramwajem i mandatem, jeśli nie wspomnimy o braku dokumentów („Wracając tramwajem, dostaliśmy mandat"), trudniej jest nam wyobrazić sobie konieczność ich posiadania, niż kiedy informacja o mandacie jest już reprezentowana w naszym umyśle (Johnson-Laird i Byrne, 1991). Takie zbyt intensywne korzystanie z heurystyki dostępności jest podstawą dla błędów poznawczych, takich jak błąd koniunkcji (Tversky i Kahneman, 1983), gdy wydaje nam się bardziej prawdopodobne, że Linda – singielka, inteligentna absolwentka filozofii, zainteresowana sprawami społecznymi i klimatycznymi – jest bankierką i aktywistką feministyczną, a nie tylko bankierką.

Wydaje się też, że w kontekście skłonności do spontanicznego kontrfaktycznego myślenia rolę odgrywa to, na ile jesteśmy aktywnymi bohaterami wyobrażanej treści. Łatwiej wyobrażamy sobie alternatywne możliwości dla działań niż dla braku działań (Byrne, 2005, rozdz. 3). Mamy także większą skłonność do wyobrażania sobie sytuacji kontrfaktycznych, w których czujemy kontrolę nad wydarzeniami, niż tych, które są od nas niezależne

(Girotto i in., 1991). Nasze myślenie zależy również od tego, czy widzimy siebie jako bohaterów kontrfaktycznej sytuacji. Jeśli negatywne zdarzenie dotyczy innej osoby, to najpewniej zasugerujemy, że mogła zachować się inaczej, aby uniknąć nieprzyjemności. Natomiast jeżeli to my jesteśmy bohaterami historii, to raczej stwierdzimy, że do niczego negatywnego by nie doszło, gdyby zewnętrzne okoliczności były inne. Jeśli ktoś źle opłacił rachunki i teraz musi spłacić odsetki, to prędzej stwierdzimy, że powinien się bardziej skupić przy płaceniu rachunków, ale kiedy pomyślimy, że sytuacja dotyczy nas, to szybciej przyjdzie nam do głowy kontrfaktyczny scenariusz, w którym to rachunek był prostszy do odczytania, a my zachowaliśmy się tak samo (Girotto i in., 2007). W myśleniu o własnych przyszłych działaniach – prospekcja jest formą myślenia hipotetycznego – skupiamy się na ich kontrolowalnych aspektach, czyli na tym, co możemy zrobić, aby działanie nam się powiodło (Ferrante i in., 2013). Chętniej też wyobrażamy sobie pozytywne alternatywy dla rzeczywistych wydarzeń odległych w czasie, a negatywne dla bliskich (Rim i Summerville, 2014). Chętniej kontrfaktycznie myślą osoby przekonane o istnieniu wolnej woli i skłonne do snucia marzeń na jawie (Alquist i in., 2015; Bacon i in., 2013). Ponadto osoby intensywnie odczuwające zazdrość mają więcej kontrfaktycznych myśli niż osoby aktualnie zazdrości nieprzeżywające (van de Ven i Zeelenberg, 2015). Chorzy na depresję częściej niż osoby zdrowe mają negatywną ocenę bieżących sytuacji, bo częściej wyobrażają sobie ich pozytywne alternatywy (Broomhall i in., 2017).

Przetwarzanie emocjonalne jest elementem większości – jeśli nie wszystkich – procesów poznawczych: pamięci, uwagi, percepcji i każdego rodzaju rozumowania (Dolan, 2002; Pessoa, 2018), w tym myślenia warunkowego. Doświadczanie emocji wiąże się także z tym, na ile akceptujemy faktyczny stan rzeczy. Takie emocje nazywane są wprost kontrfaktycznymi (Beck i in., 2014). Przeżywamy wstyd i żal, odczuwając pragnienie cofnięcia się w czasie i podjęcia innych decyzji, niż faktycznie podjęliśmy – pewne kontrfaktyczne scenariusze oceniamy więc jako bardziej preferowane niż rzeczywisty przebieg wydarzeń (Niedenthal i in., 1994). Antycypujemy przyszłość, bo konkretny jej scenariusz – a więc hipotetyczny – wydaje się nam niemal pewien. Użycie emocjonalnego języka do prezentacji przesłanek pogarsza naszą zdolność do wyciągania z nich poprawnych wniosków (Blanchette i Richards, 2004), chyba że nasze własne doświadczenie pokrywa się z tematyką rozumowania: weterani wojenni lepiej radzą sobie z oceną wnioskowań o tematyce militarnej niż strukturalnie identycznymi wnioskowaniami o innej tematyce,

podobnie jak ofiary nadużyć seksualnych z wnioskowaniami na temat nadużyć (Blanchette i Caparos, 2013).

Przytoczone wyżej zależności nie są w żadnej mierze wyczerpującym opisem wkładu badań nauk poznawczych w zrozumienie ludzkich rozumowań. Celem było pokazanie, dlaczego nie należy oczekiwać od czysto formalnej analizy, abstrahującej od psychologicznych i sytuacyjnych czynników, wyczerpującego wyjaśnienia mocy przekonywania argumentów realizowanych za pomocą eksperymentów myślowych. Ich siła nie polega na byciu przekonanym, że wiemy, iż ich konkluzja jest prawdziwa. Eksperymenty myślowe nierzadko przedstawiają sytuacje odległe od naszego utartego, codziennego doświadczenia tego, jak działa świat. Siła eksperymentu leży w wywołaniu w nas poczucia zgody, nawet jeśli początkowo odczuwaliśmy opór względem konkluzji. Warto przy tym pamiętać, że w tym procesie kluczową rolę odgrywa specyfika działania wyobraźni i mechanizmów myślenia warunkowego, uwidaczniających różnorodne tendencje naszego myślenia. Dopiero późniejsza pieczołowita rekonstrukcja pozwala ocenić, czy nasza intuicyjna zgoda była uzasadniona w sensie poprawności wnioskowania zawartego w eksperymencie.

3.5 GRANICE WYOBRAŹNI

3.5.1 Spójność jako warunek wyobrażalności

Na koniec tego rozdziału zastanówmy się nad granicami wyobraźni. Aby przeprowadzić eksperyment myślowy, musimy móc wyobrazić sobie jego treść. Skoro realizujemy go w umyśle, to skąd możemy wiedzieć, że nie mylimy się, czy możemy go sobie wyobrazić? Na pierwszy rzut oka takie pytanie może być niezrozumiałe. Czy nie możemy sobie wyobrazić wszystkiego, nawet najbardziej abstrakcyjnego i niezrozumiałego scenariusza? Spróbujmy: kwadratowe koło wielkości żółwia je łyżką planetę wielkości Marsa. Uwaga, to koło nie ma rąk ani ust, nie unosi telepatycznie łyżki ani żadna inna siła jej nie unosi. Jeśli poświęcimy choć chwilę na próbę wyobrażenia sobie tej sytuacji, staje się jasne, że jest to zadanie niemożliwe. Nie tylko percepcyjnie, ale też pojęciowo.

Skoro zatem istnieje granica wyobraźni, to gdzie leży? Powyższy przykład jest wewnętrznie sprzeczny, więc może spójność jest granicą? Wyprzedzając nieco tok wywodu, odpowiedź brzmi – tak, lecz jest też tutaj więcej niuansów dotyczących tej modalności. Rozważmy następujący, klasyczny dla

epistemologii modalności, cytat: „Cokolwiek umysł pojmuje jasno, to zawiera implicite ideę możliwego istnienia; albo, innymi słowy, że cokolwiek sobie możemy wyobrazić, to nie jest abolutnie niemożliwe" (Hume, 2005, s. 51). Wyobrażalność jest przez Hume'a wyrażana poprzez dwa inne pojęcia modalne: pojmowalność i możliwość. Pojmowalność ma być równoważna z wyobrażalnością („cokolwiek umysł pojmuje jasno […] innymi słowy, że cokolwiek sobie możemy wyobrazić"), wyobrażalność sytuacji oznaczać ma zaś możliwość jej faktycznego zajścia[42]. Co zatem oznacza pojmowalność sytuacji i możliwość jest zajścia? Sięgnijmy po ujęcie Christophera S. Hilla:

> […] pojmować coś oznacza uformować reprezentację tego czegoś, złożoną z pojęć jako budulca i strukturalnych relacji gramatyki logicznej jako cementu. […] Jeśli mamy dysponować rzetelnym testem możliwości, musimy natomiast polegać na ograniczonej pojmowalności – pojmowaniu kompatybilnym z prawami logiki, zasadami składowymi dla pojęć i wszelkich innych zdań zakładanych w danym kontekście za konieczne (2016, s. 327–328).

Pojmujemy X, jeśli mamy jego reprezentację umysłową i nie dostrzegamy sprzeczności ani w pojęciu X-a, ani w innych pojęciach i informacjach relewantnych dla kontekstu występowania X-a, czyli w naszych zasobach epistemicznych. Choć to sformułowanie jest mało zgrabne językowo, wyraża intuicyjną ideę: pojąć można to, co oceniliśmy jako w danym kontekście spójne. Nie da się natomiast pojąć sprzeczności.

Moglibyśmy się upierać, że pojmowalne jest wszystko, co da się wyrazić za pomocą poprawnie sformułowanych konstrukcji logicznych (przyjmijmy, że zdania języka naturalnego da się zawsze sformalizować systemem logicznym), i uznać, że stwierdzenie „Urodziłem się w Kioto i nie urodziłem się w Kioto" jest jak najbardziej pojmowalne. O ile jednak taka wypowiedź nie jest traktowana jako metaforyczna albo odnosząca się do różnych rozumień „Kioto" albo „urodzić się", o tyle ogranicza się jedynie do stwierdzenia sprzeczności i nie dostarcza żadnej specyficznej treści. Takie „pojmowanie" wypowiedzi nie czyni jej przekonującą. Jeśli natomiast pojmujemy X w proponowanym tu sensie, to oznacza, że jesteśmy w stanie przedstawić sobie X w umyśle w sposób niekłócący się z innymi naszymi reprezentacjami dotyczącymi kontekstu dotyczącego X-a. Tylko jeśli X się nie kłóci, a racje na jego

[42] W przeciwieństwie do Kartezjusza Hume nie separuje wyobraźni od pojmowania. Ma to głęboki sens w kontekście empiryzmu Hume'a i pasuje do wprowadzonego wcześniej funkcjonalnego rozumienia wyobraźni.

rzecz są dla nas przekonujące, może X zostać włączone w zbiór naszych przekonań. Jednoczesne rodzenie się i nie w Kioto łamie chociażby logiczną zasadę niesprzeczności, trudno też uznać biologiczny proces narodzin za coś, co może odbywać się kilkukrotnie dla konkretnego organizmu. Stąd problematyczne jest uznanie takiej Kioto-sytuacji za pojmowalną.

3.5.2 Pojmowalność i jej gradacje

Dodajmy nieco gradacji (za: Chalmers, 2002). Niekiedy możemy pochopnie uznać coś za pojmowalne, gdy sumienne zastanowienie wykazałoby niepojmowalność. Tak *prima facie* pojmowalne może być stwierdzenie „Gwiazda Wieczorna nie jest Gwiazdą Poranną", gdy zapomnimy (bądź nie wiemy), że obie nazwy odnoszą się do planety Wenus. Gdybyśmy zawsze idealnie byli w stanie ocenić zachodzenie sprzeczności, w oderwaniu od naszych ograniczeń i tendencji poznawczych, takiego błędu byśmy nie popełnili. Treści mogą też być negatywnie bądź pozytywnie pojmowalne. W tym pierwszym przypadku, po dowolnie długim zastanawianiu się, stwierdzamy, że nie jesteśmy w stanie odrzucić X-a jako sprzecznego z prawami logiki i innymi ustaleniami na temat relewantnego kontekstu. Możemy nie znać tradycji kulinarnych różnych nacji, ale nawet wtedy negatywnie pojmowalne będzie dla nas stwierdzenie, że jest różnica pomiędzy kuchnią włoską a polską, choćby z uwagi na domniemane stosowanie nieco innych składników, przepisów oraz różną historię kultury włoskiej i polskiej. Pozytywna pojmowalność X-a wymaga czegoś więcej niż tylko nieodrzucenie X-a. Jeśli X jest pozytywnie pojmowalny, to potrafimy podać pozytywną charakterystykę stanu rzeczy, kiedy X zachodzi. Nie tylko stwierdzimy różnice między kuchniami, ale także wyobrazimy sobie pierogi i kotlety schabowe z jednej strony, a z drugiej rozmaite pasty, wskazując na unikalne cechy obu tradycji[43].

Warunek ograniczonej pojmowalności wyznacza granice wyobraźni, a więc i eksperymentowania w myślach. Nie jest on szczególnie obciążający, ale gradacja pomaga. *Prima facie* pojmowalność jest mniej pewna niż idealna, pozytywna zaś bardziej niż negatywna. Idealna pozytywna pojmowalność jest najtrudniejsza do osiągnięcia, łatwiej jest bowiem podać negatywną charakterystykę – przez wskazywanie na brak określonych cech albo niezachodzenie jakiejś zależności – niż pozytywną.

[43] Niech znawcy kuchni wybaczą uproszczoną charakterystykę. Pozytywną i negatywną pojmowalność nazywa się też *silną* i *słabą* pojmowalnością (zob. van Cleve, 1983).

3.6 PODSUMOWANIE

- Choć niektórzy badacze do aktywności wyobrażeniowych zaliczają jedynie stany (quasi)percepcyjne, to uprawnione jest też ujmowanie wyobrażeń jako bardziej szerokiej klasy stanów, również czysto pojęciowych. Kluczowym warunkiem wyobraźni jest jej odniesienie do przedmiotów i treści niedostępnych w aktualnym, bezpośrednim kontakcie z otoczeniem.
- Na korzyść takiego funkcjonalnego rozumienia wyobraźni przemawiają bogate dane: neuronaukowe, wskazujące na wspólne podłoże neuronalne różnych form wyobraźni, tak percepcyjnej, jak i pojęciowej, zaobserwowane trudności osób z afantazją nie tylko w zakresie wyobraźni percepcyjnej, ale także procesów pamięciowych (wyobrażania sobie przeszłości) oraz prospekcyjnych (wyobrażania sobie przyszłości), czy chociażby widoczne pogorszenie wyobraźni percepcyjnej i poziomu języka opisującego doświadczenia zmysłowe u osób z amnezją i demencją.
- W praktyce eksperymentów myślowych współuczestniczą zarówno percepcyjne, jak i pojęciowe formy aktywności wyobrażniowej. Udział tych pierwszych tłumaczy, dlaczego nie powinno się pomijać roli narracji w analizie eksperymentów myślowych – pozwala nam do pewnego stopnia doświadczyć sytuacji eksperymentu, odczuć ją. Im bardziej sugestywna dla nas narracja, tym łatwiej zaakceptować konkluzję eksperymentu. Z kolei udział wyobraźni pojęciowej pozwala nam zrozumieć i ocenić tkwiący w eksperymencie argument, jego strukturę i zależności pomiędzy poszczególnymi abstrakcyjnymi pojęciami w nim wykorzystywanymi.
- Aby móc eksperyment myślowy uznać za poprawnie skonstruowany, jego treść musi być wyobrażalna, to jest pojmowalna. Pojmowalność treści oznacza jej wewnętrzną spójność. W zależności od charakteru oceny tej spójności – *prima facie*, idealnej, pozytywnej lub negatywnej – możemy mówić z mniejszym lub większym stopniem pewności o pojmowalności eksperymentu myślowego.

4 MECHANIZM POZNAWCZY EKSPERYMENTU MYŚLOWEGO

Eksperymenty myślowe są przejawem naszych codziennych zdolności myślowych, takich jak wyobraźnia czy wnioskowanie. Korzystamy z nich najczęściej bezwiednie. Ale jak rozumieć sam mechanizm poznawczy tej praktyki? Na samym początku książki korzystaliśmy z kryteriów Beckera wyróżniających eksperymenty myślowe na tle innych rozumowań. Nie będą jednak one specjalnie pomocne dla dociekania mechanizmu poznawczego. Pewnym punktem wyjścia może być dla nas stanowisko Nortona, którego zdaniem eksperymenty myślowe są po prostu wnioskowaniami przedstawianymi fabularnie. Rekonstrukcja eksperymentu jako wnioskowania dedukcyjnego bądź indukcyjnego kończy pracę badawczą, ocena przekonującej mocy eksperymentu sprowadza się bowiem do oceny prawdziwości jego przesłanek i poprawności wyprowadzonej z nich konkluzji (Norton, 1991, 2004a, 2004b).

Podejście Nortona zakłada, że czynniki związane z osobą przeprowadzającą eksperyment myślowy są pomijalne, bo kluczowa jest poprawność logiczna samego rozumowania. Jeśli rekonstrukcja jego struktury wskazuje na błędność rozumowania, to i cały eksperyment jest niepoprawny. Taka ocena zyskałaby na jednoznacznym przekładzie eksperymentów myślowych na układ przesłanek i wniosków. Metody takiej nie mamy. Wiele eksperymentów rekonstruowanych jest na różne sposoby, chociażby eksperyment Galileusza o spadku swobodnym, który jest inaczej formalizowany przez samego Nortona (1996) niż przez innych badaczy (np. Gendler, 1998). Choć dla niektórych eksperymentów istnieje względny konsensus co do ich rekonstrukcji, to nie jest to reguła.

Co więcej, rozróżnienie między poprawnymi i niepoprawnymi eksperymentami w duchu Nortona nie zawsze pokrywa się z tym, które eksperymenty są przekonujące. Dla naszych umysłów „ubiór" rozumowań ma jednak znaczenie, często oceniamy je jak książki, po okładce. Widzieliśmy już, że o niektórych treściach łatwiej nam myśleć warunkowo niż o innych. Ale rozważmy

dwa przykłady, historyczny i współczesny: eksperyment z krańcem świata oraz eksperyment o odpowiedzialności i wolnej woli Joshuy Greene'a i Jonathana Cohena (2004). Rozpatrzmy wpierw historyczny *Gedankenexperiment*:

1. Jeśli świat jest skończony, to ma granice.
2. Możemy umieścić osobę na granicy świata.
3. [Z uwagi na stwierdzenie 1.] Osoba ta nie może wyciągnąć ręki poza granicę świata.
4. Jeśli nie można wyciągnąć ręki, to oznacza, że napotyka się na przeszkodę.
5. [Z uwagi na stwierdzenie 2. i 4.] Osoba na granicy świata napotyka na barierę leżącą poza granicą świata.

   Konkluzja: [Z uwagi na stwierdzenie 5.] Świat jest nieskończony.

Niekoniecznie jest to najwierniejsza rekonstrukcja eksperymentu myślowego Archytasa z Tarentu, lecz wystarczy na nasze potrzeby. W tradycyjnej geometrii euklidesowej źródłem kłopotu jest przesłanka 2. Jeśli traktujemy „granicę świata" jak granicę między państwami, umowną linię podziału, to choć całkiem możliwe jest stanięcie dokładnie w jej miejscu, to raczej nie zmieści się w niej całe ciało – ta umowna linia nie ma bowiem grubości. Ale dlaczego kraniec rzeczywistości miałby przypominać linie oddzielające kraje? Dlaczego też na granicy świata nie mogłoby już być jakiegoś innego ciała, pozbawionego rąk, które mogłyby zostać wyciągnięte? Strukturalnie niczym nie manipulujemy, jednak nagle zmiana w znaczeniu sformułowania „granica świata" sprawia, że trudniej uznać 2. przesłankę za prawdziwą.

Rozważmy teraz eksperyment dotyczący odpowiedzialności i posiadania wolnej woli. Greene i Cohen argumentują, że przekonanie o wolnej woli jest potrzebne, abyśmy czuli się odpowiedzialni za swoje czyny. Proponują wyobrazić sobie stworzonego przez naukowców Pana Marionetkę, który popełnia morderstwo podczas nieudanej akcji narkotykowej. Przekonania, pragnienia, motywy Pana Marionetki są naprawdę jego, ale powstały w wyniku manipulacji niecnych naukowców. Pan Marionetka, zdaniem autorów, jest godny raczej naszej litości niż potępienia. My również mamy zaś być niczym Pan Marionetka, będąc produktem zewnętrznych wpływów, takich jak kultura czy społeczeństwo (Greene i Cohen, 2004, s. 1780).

Dennett argumentuje, że narracja Greene'a i Cohena wprowadza w błąd, i proponuje następującą jej modyfikację:

> Kapitan Autonomia skończył ekonomię na Harvardzie, a po ukończeniu studiów poszedł do pracy w Lehman Brothers, gdzie wszyscy wokół niego oszukiwali i zbijali na tym fortunę. Zakochał się w uwodzicielce, której zależało

> tylko na pieniądzach i która zagroziła mu, że zostawi go, jeśli szybko nie stanie się bogaty. Dostrzegł swoją szansę, praktycznie niewykrywalną malwersację, prawie na pewno niedostrzegalne przestępstwo, i z pełną świadomością podjął ryzyko, lecz niestety, na przekór wszystkiemu, pojawił się świadek, który popełnił błąd polegający na tym, że stał zbyt blisko balustrady w ogrodzie luksusowego apartamentu na ostatnim piętrze wieżowca, … szybkie „potknięcie" – Ups! – i świadek został zepchnięty na znajdującą się poniżej ulicę, gdzie zginął. Wzbudziło to podejrzenia, a Kapitan Autonomia wkrótce został aresztowany (Dennett, 2015, rozdz. 8).

Czy Kapitan Autonomia też jest bardziej wart litości niż potępienia? Jego działania także ukształtowały zewnętrzne czynniki, niemniej oba przypadki się różnią mimo użycia takiego samego wnioskowania. Dennett nie rozstrzyga, jak na odpowiedzialność wpływa rodzaj determinujących nas czynników, natomiast pokazuje swoją modyfikacją, jak łatwo „ubiór" eksperymentu myślowego wpływa na nasz jego odbiór[44].

4.1 MODELE MENTALNE

Zgodnie z alternatywnym podejściem eksperymenty myślowe funkcjonują jako modele mentalne (Miščević, 1992, 2007, 2022; Nersessian, 1991, 1992, 2007, 2008, 2018)[45]. W tych ujęciach rozumowanie polega na przekształcaniu reprezentacji w naszej pamięci roboczej[46]. Modele te są obrazami spostrzeganych elementów otoczenia, a także pamiętanych lub fikcyjnych sytuacji i procesów.

Teoria modeli mentalnych wpisuje się w kognitywistyczne podejście do umysłu, zgodnie z którym jest on systemem obliczeniowo modelującym zjawiska zewnętrzne i wewnętrzne oraz generującym reakcje organizmu na bazie tych reprezentacji[47]. Zdaniem zwolenników logiki mentalnej (do których

[44] Podobną argumentację można znaleźć w pracy Przemysława Zawadzkiego (2017), a bardziej szczegółową krytykę stanowiska Nortona u Stuarta (2016b).

[45] Idea modelu mentalnego sięga samych początków rewolucji poznawczej w nauce (zob. Craik, 1943; Johnson-Laird, 2006a).

[46] Pamięć robocza to mechanizm, który utrzymuje aktywne informacje i jednocześnie uczestniczy w ich przetwarzaniu; ze względu na jej ograniczoną pojemność aktywnych informacji w danym momencie jest niewiele (zob. Baddeley, 2003; Johnson-Laird, 1983).

[47] Tak zwany paradygmat kognitywizmu w naukach poznawczych nie jest, oczywiście, jedynym. Dobrze funkcjonują także chociażby podejścia predykcyjne i ucieleśnione. Niektórzy bronią też stanowiska, że umysł w ogóle nie działa na reprezentacjach mentalnych. Niemniej do dziś jest to domyślny sposób myślenia w psychologii i neuronauce poznawczej,

zaliczylibyśmy Nortona) zasady przekształcania aktywnych w rozumowaniu reprezentacji umysłowych odzwierciedlają logiczne reguły syntaktyczne, składniowe (np. Rips, 1994). Modele mentalne mają być w przeciwieństwie do tego generowane znaczeniowo, semantycznie. Są odbiciem cech tego, co modelują, takich jak struktura przedmiotów, zachowanie czy funkcja, dlatego mówi się o nich, że są modelami ikonicznymi (Nersessian, 2008, s. 93). Nie zawsze towarzyszą im przeżycia wyobraźni percepcyjnej, ale mogą się one pojawić (Johnson-Laird, 2010, s. 18244; Nersessian, 2018). Przekształcanie modeli mentalnych wykracza poza formalne operacje logiczne i operowanie na jawnej (przyjmijmy: dostępnej dla świadomości), pojęciowej wiedzy. Korzysta ono z wiedzy dziedzinowej o czasowo-przestrzennych parametrach obiektów, ich fizyce, motoryce, zachowaniach oraz związkach przyczynowo-skutkowych. Struktura modeli konstruowanych w ramach eksperymentów myślowych bazuje na narracji, nazywane zaś są dyskursywnymi lub sytuacyjnymi. Obrazują bowiem „nie strukturę zdań, tylko wyobrażanej lub percypowanej sytuacji" (Johnson-Laird, 1989, s. 471).

Choć tworzenie modeli mentalnych odbywa się w pamięci roboczej, to ich treść zależy od informacji przechowywanych w pamięci zarówno krótkotrwałej, jak i długotrwałej (Nersessian, 2018, s. 317). Oznacza to, że na zawartość modeli wpływają nie tylko bieżące dane z otoczenia, ale także nasze przekonania, wiedza czy stany emocjonalne[48].

Wspomnieliśmy wcześniej o ekonomiczności myślenia warunkowego. Podobnie sytuacyjny model mentalny jest ekonomiczny, gdyż pozwala na szybsze niż wnioskowanie w czysto językowej formie ustalenie rozwinięć modelowanych sytuacji, np. jak ruch jednego obiektu wpłynie na wzajemne relacje przestrzenne wszystkich postrzeganych obiektów. Zamiast przekształcania symbolicznych, językowych przedstawień dynamiki sytuacji model symulować ma bowiem sytuację poprzez przetwarzanie przestrzenne (Ragni i Knauff, 2013; Vandierendonck i in., 2004).

językoznawstwie kognitywnym, filozofii umysłu i sztucznej inteligencji (Bermúdez, 2020; von Eckardt, 1995; zob. Hutto i Myin, 2013).

[48] Znaczenie emocji dla modeli mentalnych bywa kontestowane (np. Zawadzki, 2017, s. 129), ale zależność ta jest bez wątpienia silna. Dla zwolenników idei modeli mentalnych są one obiektami, do których odnoszą się i na które oddziałują stany emocjonalne (zob. Oatley i Johnson-Laird, 1996).

Aby lepiej zrozumieć różnicę między myśleniem opartym na modelach mentalnych a myśleniem rozumianym jako przekształcanie przesłanek, wyobraźmy sobie następującą sytuację[49]:

1. A jest na lewo od B.
2. B jest na lewo od C.
3. D jest przed A.
4. E jest przed C.

Pytanie brzmi: jaka jest relacja między D i E? Jeśli rozumujemy poprzez przekształcanie przesłanek, to poprawną odpowiedź znajdziemy w czterech krokach (w tym dwóch założeniach o relacjach przestrzennych):

5. [Przechodniość bycia po lewej stronie] Jeśli obiekt znajduje się na lewo od drugiego obiektu, a trzeci obiekt jest na lewo od pierwszego obiektu, to trzeci obiekt jest na lewo od drugiego obiektu.
6. [Relacja między byciem przed a byciem po lewej i prawej stronie] Bycie przed obiektem oznacza, że jest się po lewej i po prawej stronie tych obiektów, po których lewej i prawej stronie jest obiekt, przed którym się jest.
7. [Z uwagi na stwierdzenie 1., 2. oraz 5.] A jest na lewo od C.
   Konkluzja: [Z uwagi na stwierdzenie 6. i 7.] D jest na lewo od E.

Zmieńmy teraz jeden element:

1. A jest na lewo od B.
2. B jest na lewo od C.
3. D jest przed A.
4. E jest przed B.

Tym razem udzielenie odpowiedzi na pytanie o relację D i E wymaga dwóch kroków:

5. [Relacja między byciem przed a byciem po lewej i prawej stronie] Bycie przed obiektem oznacza, że jest się po lewej i po prawej stronie tych obiektów, po których lewej i prawej stronie jest obiekt, przed którym się jest.
   Konkluzja: [Z uwagi na stwierdzenie 1., 3., 4. i 5.] D jest na lewo od E.

Drugi wariant jest prostszy obliczeniowo, wymaga mniejszej liczby kroków i założeń, aby dotrzeć do odpowiedzi na pytanie. Rozumując poprzez przekształcanie układu przesłanek, spodziewalibyśmy się więc, że pierwszy wariant okaże się trudniejszy dla badanych. Tak jednak nie jest (Byrne i Johnson-Laird, 1989, s. 570). Zwolennicy modeli mentalnych interpretują te wyniki

[49] Jest to uproszczona wersja rozumowań użytych w badaniu Ruth Byrne i Philipa Johnsona-Lairda (1989; zob. także Johnson-Laird, 2010).

jako świadczące o korzystaniu przez ludzi z ikonicznych modeli sytuacyjnych, kolejno uzupełnianych przez te elementy, o których mówią przesłanki. Odpowiedź na pytanie polega więc na analizie pojedynczego modelu, a nie przekształcaniu reprezentowanych osobno w umyśle przesłanek.

Zgodnie z teorią modeli mentalnych trudność rozumowań zależy od liczby modeli koniecznych do pełnego przedstawienia sobie sytuacji. Jeśli opis jest jednoznaczny, to istnieje tylko jeden model. Sytuacja opisana niejednoznacznie wymaga wielu modeli. Dla ilustracji rozważmy taki przypadek:

1. B jest na prawo od A.
2. C jest na lewo od B.
3. D jest przed C.
4. E jest przed A.

Sytuację tę opisują dwa modele. Możemy je sobie zobrazować następująco:

C A B	A C B
D E	E D

W obu przypadkach spełnione są warunki 1–4, ale w pierwszym D jest na lewo od E, a w drugim D jest na prawo od E. Nie możemy jednoznacznie określić, jaka jest relacja między D i E. Ale żeby móc stwierdzić, że nie da się udzielić odpowiedzi, należy przeanalizować oba modele. Rozumowania dotyczące takich niejednoznacznych sytuacji sprawiają nam więcej trudności niż te, które wymagają przeanalizowania wyłącznie jednego modelu mentalnego (Byrne i Johnson-Laird, 1989; Johnson-Laird, 2010).

Modele mentalne mają reprezentować sytuacje, uwzględniając jedynie to, co jest bezpośrednio podane w opisie lub spostrzeżeniu sytuacji (Johnson-Laird i Savary, 1999). Wyjaśnia to, dlaczego często dostrzegamy zależności tam, gdzie ich nie ma, czyli tzw. iluzoryczne wnioskowania. Rozważmy klasyczny przykład: naprzeciw nas siedzi gracz z kartami i wiemy, że jedna z poniższych wskazówek dotyczących go jest prawdziwa, a jedna fałszywa:

1. Jeśli gracz ma w ręce króla, to ma też asa.
2. Jeśli gracz nie ma króla w ręce, to ma asa.

Czy gracz ma asa? Dla wielu osób intuicyjne wydaje się, że tak lub że jest to możliwe. Niemniej poprawna odpowiedź brzmi: gracz nie ma asa.

Kluczowe jest tu rozważenie fałszywości jednej ze wskazówek, przy czym nie ma znaczenia, którą z nich uznamy za fałszywą, bo musimy sprawdzić, co się dzieje przy fałszywości każdej z nich. Standardowo myślimy o okresach

warunkowych, że są niespełnione (fałszywe), o ile ich poprzednik jest prawdziwy, a następnik fałszywy. Rozważmy analogię: „Jeśli mam 2 banknoty po 50 złotych i nie mam więcej pieniędzy, to mam 100 złotych". Zasada ta nie mówi nic o sytuacji, gdy nie mam 2 banknotów po 50 złotych. Jeżeli mam zamiast tego 5 banknotów po 20 złotych, to też będę mieć 100 złotych, bez złamania reguły o 2 pięćdziesiątkach. Podobnie, gdybym miał 3 banknoty po 10 złotych, to miałbym 30 złotych – taki przypadek również nie łamałby naszej reguły. Reguła wypowiada się tylko na temat posiadania 2 pięćdziesiątek.

Podobnie, jeśli druga wskazówka o kartach jest fałszywa, to jej poprzednik jest prawdziwy, a następnik fałszywy, a więc gracz nie ma ani króla, ani asa. Analogicznie, kiedy uznamy pierwszą wskazówkę za fałszywą, to gracz ma w ręce króla (prawdziwy poprzednik), ale nie ma asa (fałszywy następnik).

Wedle teorii modeli mentalnych kłopotliwość tego typu rozumowań wynika z faktu, że treść modelu mentalnego jest mimowolnie postrzegana jako prawdziwa. Ocena przykładu z kartami i wskazówkami wymaga rozważenia fałszywości sytuacji zgodnej ze wskazówkami i dodatkowo wymaga rozważenia większej liczby modeli – implicytnych – takich, których zazwyczaj nie bierzemy pod uwagę (zob. Johnson-Laird i Savary, 1999, s. 196–201). W przeciwieństwie do pierwszego zawsze konstruowanego dla rozumowania modelu, korzystanie z tych dodatkowych modeli nie jest automatyczne (Johnson-Laird, 2010, s. 18245).

Teoria modeli mentalnych bywa krytykowana, a dla jej wyników eksperymentalnych proponowane są alternatywne wyjaśnienia (np. Manktelow, 2012, rozdz. 5 i 6; Stenning i van Lambalgen, 2008). Pozostaje jednak wpływowym i nieustannie rozwijanym podejściem w psychologii rozumowań (np. Johnson-Laird, 2023). Jest szeroko stosowana do wyjaśniania różnego typu rozumowań: dedukcyjnych (Johnson-Laird, 2008), przestrzennych (Ragni i Knauff, 2013), przyczynowych (Khemlani i in., 2014), temporalnych (związanych z myśleniem o czasie) (Schaeken i in., 1996), probabilistycznych (Johnson-Laird, 1994) czy właśnie eksperymentów myślowych (wcześniej cytowane). W momencie pisania tej książki jest tak naprawdę jedynym podejściem kognitywistycznym wprost wykorzystywanym do opisu mechanizmu poznawczego stojącego za eksperymentami myślowymi.

Zgodnie z teorią Johnsona-Lairda zrozumienie przesłanek – obojętnie czy dostarczanych jako zdania czy percepty – rozpoczyna proces rozumowania, dyktując treść modelu mentalnego (Johnson-Laird i Byrne, 1991;

Manktelow, 2012, s. 46–48). Manipulacja modelem zaś prowadzi nas do wniosków. W przypadku eksperymentów myślowych manipulacje te oddają przyczynowy rozwój sytuacji opisanej w narracji. Dodajmy, że nawet zwolennicy alternatywnych podejść do rozumowań, argumentujący na rzecz możliwości szerokiej formalizacji ich we współczesnych logikach, podkreślają, aby nie zapominać o tym pierwszym etapie. Przykładowo, Keith Stenning i Michiel van Lambalgen (2008) wskazują, że najpierw rozumujemy do interpretacji, dookreślając znaczenia słów i spostrzeganych treści percepcyjnych, zadań stawianych przed nami przez środowisko. Myślenie problemowe nie przebiega w próżni, zawsze pracujemy już na jakimś odczytaniu sytuacji, aby móc rozumować od interpretacji.

4.2 CO WYRÓŻNIA EKSPERYMENTY MYŚLOWE JAKO MODELE MENTALNE?

Aby modelowanie mentalne było przydatnym opisem mechanizmu eksperymentowania w myślach (nazwijmy je modelami$_e$), musimy także wskazać cechy odróżniające je od innych modeli mentalnych. Nancy Nersessian (2018) wskazuje, że w przeciwieństwie do codziennego modelowania, pomocnego chociażby przy ustalaniu orientacji mebli przy przechodzeniu przez framugę drzwi, naukowe eksperymenty myślowe informują nas nie tylko o konkretnej sytuacji, ale także całej ich klasie. Abstrahujemy od nieistotnych czynników, zakładając, że niezależnie od nich rozwój sytuacji pozostanie niezmienny. Obojętnie, jakie związane ciała Galileusz zrzuci z wysokości – o ile tylko jedno będzie cięższe, a drugie lżejsze, rezultat będzie taki sam: „Ogólność eksperymentów myślowych leży w zdolności umysłu do uchwycenia, że wniosek z konkretnej sytuacji ma zastosowanie do [każdej] sytuacji tego typu" (Nersessian, 2018, s. 321).

Modele mentalne$_e$, w odróżnieniu od rozważanych przed chwilą sytuacji z obiektami A, B, C, D i E, korzystają też z narracji. Zapoznanie się z narracją rozpoczyna zarówno interpretację prowadzącą do powstania modelu mentalnego$_e$, jak i proces oceny zgodności rozwoju sytuacji w modelu$_e$ z końcowym stanem rzeczy opisywanym w narracji eksperymentu. Taki sposób prezentowania i uruchamiania eksperymentu myślowego od razu w pewnym stopniu uwiarygadnia jego treść dla przeprowadzającej eksperyment osoby, bo – jak wspomnieliśmy – automatycznie powstający model mentalny traktowany jest mimowolnie przez nas jako prawdziwy.

W przypadku *Gedankenexperimente* z zakresu nauki najczęściej modelujemy$_e$ w umyśle wyidealizowaną sytuację fizyczną, postrzeganą jak gdyby zmysłowo. Możemy percepcyjnie wyobrazić sobie upadek ciał z wysokości czy nawet sięganie za granicę świata ręką. Wiele narracji jednak nie odwołuje się do różnic obserwowanych zmysłowo. Motywacje i zachowania niewidzialnych osób z pierścieniami Gygesa są całkowicie domniemane. Skrajnym przykładem takiego eksperymentu, jak zobaczymy, jest eksperyment z zombie, którego narracja skłania nas wprost do dostrzeżenia różnicy niedostrzegalnej zmysłowo.

Na jakich zasobach opiera się modelowanie mentalne w takich przypadkach? W kontekście eksperymentów myślowych Nenad Miščević (2022, rozdz. 4) rozróżnia dwie odmiany modelowania$_e$: budowanie modelu i symulację mentalną. Eksperymenty myślowe pierwszego typu mają narrację wyobrażalną zmysłowo, konstrukcja modelu$_e$ polega w nich na dopasowaniu treści narracji do naszych pojęć i poziomu fizycznego zrozumienia świata. Można to nazwać modelowaniem trzecioosobowym[50]. Symulacja mentalna natomiast opiera się na perspektywie pierwszoosobowej, na wczuciu się, jak gdyby było się uczestnikiem owej modelowanej sytuacji, i na rozpoznaniu stanów umysłowych, które naszym zdaniem są adekwatne.

Odnieśmy się do eksperymentu myślowego z żukiem Wittgensteina. Choć skupiamy się w nim na zachowaniu językowym – postrzegalnym, bo możemy „usłyszeć", co mówią posiadacze pudełek z żukiem – to nie jest on przypadkiem czystego „budowania modelu", mówiąc językiem Miščevicia. Nasze przypuszczenia zależą od tego, co sami myślimy o ukrytych w pudełku żukach. Podobnie jest z pierścieniami Gygesa – oceniamy motywacje niewidzialnych osób na podstawie własnych przeczuć, jak zachowalibyśmy się, będąc bezkarni. Modelowanie w obu eksperymentach jest zatem przykładem symulacji mentalnych.

[50] Miščević nie ma tutaj, oczywiście, na myśli procedury modelowania i zasad tworzenia teorii naukowych w takim sensie, w jakim traktuje się o tej materii w filozofii nauki, uporządkowanych zbiorów twierdzeń, które poprzez uproszczenie reprezentują określony wycinek rzeczywistości, zezwalając na przewidywanie i wyjaśnianie obserwacji empirycznych. Modele naukowe mają charakter formalny, bywają matematyczne lub fizyczne. W kontekście modeli mentalnych natomiast „budowanie modelu" nadal oznacza mechanizm poznawczy przebiegający w umyśle osoby przeprowadzającej eksperyment myślowy.

4.3 PODSUMOWANIE

- Teoria modeli mentalnych jest eksperymentalnie ugruntowanym wyjaśnieniem mechanizmu poznawczego leżącego u podstaw naszej zdolności do rozumowania. Zgodnie z nim rozumowanie oparte jest na ikonicznym reprezentowaniu sytuacji problemowej w pamięci roboczej. Ten proces modelowania dynamicznego odzwierciedlania strukturalnych, funkcjonalnych i innych cech zadanej sytuacji może przybrać formę trzecioosobowego budowania modelu bądź pierwszoosobowego symulowania mentalnego. Pierwszy sposób polega na wyznaczeniu prawdopodobnych konsekwencji sytuacji przez wzgląd na nasze zrozumienie świata fizycznego i posiadane pojęcia, drugi zaś na wczuciu się w agentów zadanych w sytuacji i ocenie adekwatnych naszym zdaniem ich stanów umysłowych.
- O ile w przypadku codziennego rozwiązywania problemów treść modelu mentalnego wyznacza kontekst otoczenia, w którym się aktualnie znajdujemy, o tyle treść modelu mentalnego eksperymentu myślowego wyznacza narracja eksperymentu.
- Modele mentalne eksperymentów myślowych reprezentują każdorazowo całe klasy sytuacji, abstrahuje się w nich od nieistotnych szczegółów.
- Modelowanie mentalne jako propozycja teoretyczna mechanizmu poznawczego leżącego u podłoża eksperymentowania w myślach ma przewagę nad traktowaniem tej praktyki jedynie jako formalnie poprawnego lub nie układu przesłanek i wniosków, bo: (1) wyjaśnia, w jaki sposób wywodzimy z eksperymentów nowe konkluzje bez nowych obserwacji empirycznych (poprzez analizę rozwoju zbudowanego modelu), (2) tłumaczy, dlaczego indywidualne zasoby (wiedza dziedzinowa, przekonania, emocje, pamięć) osoby przeprowadzającej eksperyment mogą wpływać na jej uznanie konkluzji eksperymentu.

5 WIARYGODNOŚĆ EKSPERYMENTÓW MYŚLOWYCH A INTUICJA

Czy modele mentalne tworzone podczas eksperymentowania w myślach mogą stanowić źródło *uzasadnienia*? Powiedzieliśmy, że modele mentalne są automatyczne, często wykraczające poza operowanie na wyłącznie jawnych, dostępnych nam świadomie informacjach. Fundamentalne znaczenie ma więc dla nich proces intuicji, która jest nieodłącznym elementem naszego codziennego myślenia. Czy więc intuicja może dostarczać wiarygodnych wniosków, czy też zawsze zasługuje na sceptycyzm?

Pionier teorii modeli mentalnych, Johnson-Laird, nazywa intuicję „nieświadomymi procesami heurystycznymi, w których od nieświadomych przesłanek przechodzimy do świadomego wniosku" (2006b, s. 427). Mieć intuicję to więc posiadać utarty sposób myślenia na jakiś temat, dyspozycję do generowania o nim konkretnych modeli mentalnych. Może ona uczynić eksperyment myślowy przekonującym dla osoby przeprowadzającej, gdy pierwszy automatycznie utworzony model potwierdza konkluzję autora eksperymentu.

Tego typu rozumienie intuicji jest dość standardowe zarówno w naukach poznawczych, jak i w filozofii. Podobnie jest z szerszym rozróżnieniem między szybkim, intuicyjnym myśleniem („System 1") a wolnym, nieintuicyjnym („System 2") (Kahneman, 2012). Myślenie intuicyjne jest automatyczne, łatwe, niejawne, heurystyczne i ewolucyjnie starsze, podczas gdy nieintuicyjne jest świadomie kontrolowane, celowe i analityczne. Choć kwestia niezależności tych dwóch systemów myślenia pozostaje dyskusyjna (zob. De Neys, 2023), takie dwuprocesowe wyjaśnienia są powszechne w naukach poznawczych. Wykorzystuje się je do opisu zachowań zdrowotnych i ekonomicznych, teorii umysłu i percepcji społecznej, moralności, kreatywności, wyobraźni, błędów poznawczych i ogólnej dyspozycji do rozumowania oraz podejmowania decyzji (De Neys, 2018; De Neys i Pennycook, 2019; Stuart, 2021a).

Intuicję często postrzegamy pozytywnie. Dobrze przyswojona wiedza staje się intuicyjna. Kierowcy nie monitorują świadomie zmian biegów i intuicyjnie

je redukują, wiedząc z doświadczenia, że na niższym biegu łatwiej przyspieszyć i wyprzedzić inny pojazd. W codziennych zadaniach intuicja często podpowiada trafne rozwiązania (np. Gigerenzer, 2008; Mousavi i Gigerenzer, 2014). Niektóre sprawy oceniamy też jako tak intuicyjnie oczywiste, że nie warto ich dodatkowo uzasadniać: słońce wschodzi rano i zachodzi wieczorem, a wieloletnie opychanie się słodyczami kończy się problemami zdrowotnymi. Jednocześnie intuicja może być problematyczna. Trudniej zmienić nam te przekonania, które uznajemy za szczególnie intuicyjne. Intuicyjne, szybkie i mimowolne decyzje, jak przyspieszanie przy zbliżaniu się do skrzyżowania, aby zdążyć przed czerwonym światłem, bywają niezgodne z przepisami i niebezpieczne. W niektórych też sytuacjach, przykładowo w logice formalnej, intuicja nie wystarcza – potrzebne są dowody.

Ostatecznie ocena przydatności intuicji zależy od kontekstu. Czy ważne decyzje opierać na intuicji czy sumiennej analizie? Czy moralność jest intuicyjna, skoro różne kultury różnie definiują „dobre" działanie? W końcu, czy wystarczy, że eksperyment myślowy będzie przekonujący dla nas, czy istnieją też bardziej intersubiektywne kryteria jego oceny?

Skoro eksperymenty myślowe z powodzeniem stosuje się w nauce, można bezpiecznie założyć, że przynajmniej niektóre są też wiarygodne. Powiemy więc, że eksperyment myślowy jest wiarygodny, o ile odwołuje się do wiarygodnych intuicji[51]. Celowo posługuję się tu terminem *wiarygodność*, a nie *prawdziwość* czy *poprawność*. Każda nasza intuicja jest poprawna w tym sensie, że jest zgodna z naszą nieświadomą oceną. Każda jest też prawdziwa w tym sensie, iż faktycznie odczuwamy, że ją posiadamy. Intuicje są zaś wiarygodne, o ile są uzasadnione. Rekonstruując wnioskowanie prowadzone na naszych intuicjach, znajdziemy więcej argumentów za konkluzją niż przeciw niej. Jednakże przebieg eksperymentu myślowego zależy zawsze od osoby go przeprowadzającej: skoro myślenie intuicyjne jest nieświadomym procesem jedynie dostarczającym świadomych wniosków, to nie możemy być pewni, czy to, co nazwaliśmy wcześniej pozaformalnym kontekstem myślenia warunkowego, nie szkodzi naszej ocenie konkluzji eksperymentu. W przypadku eksperymentów myślowych istnieje ryzyko pomylenia wiarygodności z jedynie przekonującym charakterem intuicji.

[51] Prezentuję tu bardzo pragmatyczne rozumienie wiarygodności. O wiarygodności intuicji w eksperymentach myślowych z czysto epistemologicznej perspektywy pisał Manhal Hamdo (2023).

Opisując mechanizm poznawczy stojący za eksperymentowaniem w myślach, wskazaliśmy na trudność w przekształcaniu ich do układu przesłanek i wniosków. Scharakteryzowaliśmy też pozaformalne czynniki stanowiące o tym, na ile przekonująca będzie dla nas konkluzja eksperymentu. Już przekonujący eksperyment myślowy jest wartościowy, dostarcza bowiem wstępnego rozwiązania problemu przedstawionego w narracji. To już może nam wystarczyć, szczególnie jeśli traktujemy eksperyment jako narzędzie do dalszego myślenia. Jednak, aby ocenić wiarygodność eksperymentowania w myślach, musimy przeanalizować zarówno strukturę jego wnioskowania, jak i wiarygodność intuicji, do których się odwołuje[52].

Tak naprawdę nie mamy alternatywy dla intuicji:

> Jeśli bowiem będziemy analizowali myślenie nieintuicyjne, cofając się do poprzednich stadiów opartych na świadomym wnioskowaniu, wcześniej czy później dojdziemy do takiego etapu, który nie będzie już rezultatem świadomego wnioskowania, czyli będzie myśleniem intuicyjnym (Williamson, 2019, s. 77).

Biorąc więc pod uwagę, że intuicja stanowi podwalinę dla myślenia nieintuicyjnego, pytanie „Czy intuicyjność eksperymentu myślowego może być wiarygodna?" jest źle postawione. Powinniśmy raczej pytać: „Dla jakich eksperymentów myślowych możemy ocenić wiarygodność intuicji?". Kluczowe jest, jak sądzę, spełnienie trzech kryteriów: (1) narracja musi przedstawiać możliwą sytuację, (2) konkluzja eksperymentu musi być treściowo związana z sytuacją przedstawioną w narracji i (3) uznana rekonstrukcja wnioskowania zawartego w eksperymencie myślowym wskazuje na jego dedukcyjny lub uprawdopodabniający charakter. Dopiero po upewnieniu się co do spełnienia tych trzech elementów ma sens prowadzenie dalszej analizy eksperymentu, poszczególnych jego przesłanek, wpływu czynników pozaformalnych na jego odbiór i ostatecznie dokonanie oceny jego wiarygodności. Powyższe punkty omawiam poniżej.

---

[52] We współczesnej metafilozofii zagadnieniem wzbudzającym wiele kontrowersji jest kwestia tego, na ile myślenie filozoficzne opiera się na czymś innym niż intuicjach. Nawet jeśli uznamy traktowanie filozofowania jako ufundowanego na intuicjach za kontrowersyjne stanowisko, to dziś niewątpliwie żywo debatuje się nad statusem wiarygodności osądów filozoficznych i rolą intuicji (zob. Cappelen, 2012; Climenhaga, 2018; Colaço i Machery, 2017; Deutsch, 2015). Wyniki eksperymentalne nie wskazują ponadto, żeby sama refleksja albo trening filozoficzny zmieniały w znaczącym stopniu intuicyjnie wydawane sądy filozoficzne (Kneer i in., 2021; Maćkiewicz i in., 2023).

5.1 KREŚLENIE MOŻLIWEJ SYTUACJI

Wiarygodność eksperymentu wymaga jasnej narracji. Musi być zrozumiałe, *co* mamy sobie wyobrazić, *jakie cechy* posiada opisywana sytuacja, i w ogóle ma ona *dać się* wyobrazić. W przeciwnym wypadku istnieje ryzyko stworzenia nieadekwatnego modelu mentalnego, nieoddającego intencji autora eksperymentu.

W rozdziale o granicach wyobraźni wskazałem, że niewyobrażalność oznacza niepojmowalność, czyli sprzeczność z prawami logiki, posiadanymi pojęciami i innymi zasobami epistemicznymi. Jednak nie każda pojmowalna treść gwarantuje wiarygodność eksperymentu. By eksperymenty myślowe zbliżały nas do prawdy lub przynajmniej prawdopodobnych ustaleń, ich scenariusze muszą być nie tylko pojmowalne, ale i możliwe w określonym sensie.

W literaturze dotyczącej modalności wyróżnia się zwykle co najmniej dwa rodzaje możliwości[53]: logiczną, czyli zgodność z prawami logiki, oraz fizyczną (lub nomologiczną), czyli zgodność z prawami fizyki i innych nauk przyrodniczych. Analogicznie, konieczność oznacza albo wynikanie z praw logiki (konieczność logiczna), albo wynikanie z praw fizyki i innych nauk przyrodniczych (konieczność fizyczna). I tak, upakowanie masy Słońca w kuli o średnicy dwóch kilometrów jest możliwe logicznie, ale też niemożliwe fizycznie – taki obiekt zapadłby się w czarną dziurę. Podobnie, machanie rękami, aby wznieść się na 12. piętro, jest logicznie możliwe, lecz fizycznie niemożliwe – nasz układ mięśniowo-kostny na wystarczająco szybkie ruchy nie pozwala. Naturalnie, jako że nie posiadamy ostatecznej teorii fizyki, nie możemy być stuprocentowo pewni tego rodzaju możliwości – nie sposób wykluczyć, że mylimy się na temat przynajmniej niektórych praw nauki. Możliwość logiczna jest najszerszym rozumieniem możliwości i odnosi się do spójności – coś, co jest niemożliwe logicznie, jest też niepojmowalne, jak przykładowo jednoczesne urodzenie się i nieurodzenie w Kioto[54].

Rzeczy fizycznie możliwe są zawsze logicznie możliwe, ale pośród rzeczy możliwych logicznie jest nieskończenie wiele rzeczy niemożliwych fizycznie[55].

---

[53] Pojęcia modalne takie jak pojmowalność czy możliwość są współcześnie intensywnie badane i traktuje o nich dziś cały dział logik i epistemologii modalnych (zob. Mallozzi i in., 2023).

[54] Kolejnym krokiem byłoby naturalnie dookreślenie, o prawa jakiego rodzaju logiki tutaj chodzi.

[55] Jest to tzw. zagnieżdżające ujęcie relacji między możliwościami (zob. Mallozzi i in., 2023).

W przypadku eksperymentów myślowych dotyczących sytuacji fizycznych jesteśmy w stanie ocenić możliwość fizyczną po uchyleniu zastosowanych abstrakcji i idealizacji, lecz inne eksperymenty dotyczą kwestii abstrakcyjnych bądź pojęciowych. W takich sytuacjach pozostaje nam ocena możliwości logicznej narracji. Od nas zależy jednak, czy uznamy przydatność rozważania, przykładowo, sprawiedliwości w oderwaniu od fizycznej możliwości przebiegu sytuacji z eksperymentu z pierścieniem Gygesa. W końcu tylko sama kwestia sprawiedliwości jest pojęciowa, reszta scenariusza podlega ocenie możliwości fizycznej. Nie ma powodu, abyśmy od eksperymentów myślowych dotyczących nawet spraw abstrakcyjnych nie mogli oczekiwać osadzenia w okolicznościach możliwych fizycznie[56].

Co jednak z pojmowalnymi sytuacjami możliwymi logicznie, ale nie fizycznie? Kluczowe jest tu pojęcie możliwości metafizycznej, która wykracza poza fizyczną, lecz nie przekracza logicznej. Przykładem może być istnienie pierwiastka duchowego lub wartości stałych fizycznych innych niż w naszym wszechświecie. Możliwość metafizyczna bywa rozumiana jako zgodność z istotą rzeczy bądź prawami metafizycznymi[57]. Dla ilustracji, zdanie „Sokrates był mężczyzną" jest konieczne metafizycznie, bo historyczny Sokrates był w istocie mężczyzną, ale logicznie to samo zdanie konieczne nie jest – imię „Sokrates" może odnosić się do dowolnego bytu.

W analizach możliwości metafizycznej często korzysta się z idei światów możliwych, sięgającej co najmniej Leibniza, która zyskała popularność w latach 60. XX wieku. Formalnie światy możliwe są narzędziem do wyrażania i oceniania poprawności rozumowań opartych na pojęciach modalnych (Menzel, 2023). Nasz świat („świat aktualny") jest zbiorem prawdziwych stwierdzeń o rzeczywistości, podczas gdy światy możliwe względem naszego i z niego dostępne to zbiory stwierdzeń opisujące alternatywne wizje rzeczywistości[58]. Możliwość i konieczność metafizyczne można w tym rozumieniu wyrazić następująco:

[56] Z dzisiejszymi technologiami inteligentnego maskowania wydaje się całkiem możliwe fizycznie korzystanie z przedmiotów gwarantujących wystarczający dla Gygesa poziom niewykrywalności.

[57] Te wyrażać mają zależności ontologiczne, takie jak chociażby relacje fundujące (*grounding*). O tych dwóch sensach modalności i metafizyki pisali Justin Clarke-Doane (2021) oraz Jonathan Schaffer (2009).

[58] Niektórzy myślą też o światach możliwych realistycznie. Takim znanym realistą modalnym był David Lewis (1986).

- X jest możliwe metafizycznie wtedy i tylko wtedy, gdy X jest prawdziwe w przynajmniej jednym świecie możliwym.
- X jest konieczne metafizycznie wtedy i tylko wtedy, gdy X jest prawdziwe we wszystkich światach możliwych.

Różne systemy logiki modalnej definiują, które światy są dostępne ze świata aktualnego. Na przykład w systemie T, którego charakterystycznym aksjomatem jest stwierdzenie, że to, co istnieje, istnieje koniecznie, takich światów jest niewiele, podczas gdy w systemie S5 wszystkie światy są dostępne ze świata aktualnego. Charakterystycznym aksjomatem tego ostatniego jest stwierdzenie, że to, co możliwe, jest koniecznie możliwe[59].

Możliwość i konieczność metafizyczna są intensywnie dyskutowanymi zagadnieniami modalnymi, które w kontekście eksperymentów myślowych warto analizować w relacji do pojmowalności. Niektórzy twierdzą, że pojmowalność nie gwarantuje nam takiej możliwości (Hill i Mclaughlin, 1999), ale najbardziej wpływowe w debacie są głosy Stephena Yablo (1993) i Davida Chalmersa (2002). Yablo proponuje podejście epistemiczne, uzależniając pojmowalność od podmiotu pojmującego, natomiast Chalmers nieepistemiczne, postulując przejście od pojmowalności do możliwości metafizycznej dla idealnego pojmującego (Roca-Royes, 2011).

Yablo traktuje pojmowalność jako zdolność wyobrażenia sobie świata możliwego, w którym dany stan rzeczy zachodzi. Taka pojmowalność wywołuje wrażenie możliwości metafizycznej zajścia tego stanu rzeczy, które samo jest już *prima facie* uzasadnieniem tej możliwości. Jednak to wrażenie może być błędne. Osoba nieświadoma, że Gwiazda Wieczorna i Gwiazda Poranna to Wenus, może uznać zdanie „Gwiazda Wieczorna nie jest Gwiazdą Poranną" za pojmowalne i wstępnie stwierdzić, że wyraża pewną możliwość.

Można wątpić, czy takie wstępne uzasadnianie możliwości przez wrażenie jest wystarczające w kontekście wiarygodności rozumowania modalnego. Chalmers twierdzi, że jeśli myślimy o stanie rzeczy pojmowalnym idealnie (przez podmiot idealnie poznający), pozytywnie (wykształca się spójny obraz tego stanu rzeczy) i pierwotnie, to jest on także możliwy metafizycznie i to tylko nie *prima facie* (zob. Chalmers, 2002, 2008, 2010b). Co oznacza termin *pierwotna pojmowalność*? Unikając formalizmów, oznacza on, że stan rzeczy jest dla nas pojmowalny apriorycznie, bez dodatkowej wiedzy

[59] Bardziej systematyczne wprowadzenie w logiki modalne można znaleźć w pracy Kazimierza Świrydowicza (2004).

empirycznej. Mówiąc prościej, taki stan rzeczy nie jest sprzeczny z tym, co już wiemy o naszym świecie. Na przykład zdanie „woda nie jest H_2O" jest pierwotnie pojmowalne, bo bez dodatkowej wiedzy na temat chemii wiedzielibyśmy tylko, że „woda" to mokra substancja w oceanach, morzach, rzekach czy kałużach.

Ten rodzaj opisu, wychwytywania elementów świata, do których odnosi się nazwa „woda", nazywamy intensją pierwotną (stąd: pierwotna pojmowalność). Pierwotna pojmowalność odzwierciedla intuicję, że nasz świat może być inny, niż nam się wydaje na podstawie naszej potocznej, niespecjalistycznej wiedzy uformowanej codziennymi interakcjami ze światem (tego, że używamy wody/tej mokrej substancji do mycia, picia itd.). Wtórna pojmowalność natomiast odnosi się do intuicji, że świat jest dokładnie taki, jaki jest: woda jest w nim H_2O niezależnie od stanu naszej wiedzy. Wtórnie niepojmowalne jest zatem zdanie „Woda nie jest H_2O". Określenie „wtórna" bierze się od pojęcia intensji wtórnej – takiego sposobu opisu, który wskazuje daną nazwą te rzeczy, które faktycznie są tą rzeczą[60].

Niektórzy, jak Chalmers (2002), argumentują, że metafizyczna możliwość to najszersza, a nawet jedyna kategoria możliwości[61]. Kwestią sporną pozostaje choćby to, na ile, szacując idealną pojmowalność, jesteśmy w stanie wziąć w nawias tendencyjność i inne niedoskonałości naszego myślenia. Istnieją też alternatywne ujęcia modalności, niewychodzące od kategorii pojmowalności (np. Williamson, 2007). Nie będziemy tutaj rozstrzygać tej debaty, lecz będziemy traktować powyższe modalności jako narzędzia do systematyczniejszej oceny możliwości sytuacji w narracjach eksperymentów myślowych. Taki jest zresztą uzus filozoficznych analiz eksperymentów myślowych w filozofii umysłu (zob. Chalmers, 2003a).

Jeśli tylko nie dostrzegamy w narracji wewnętrznej sprzeczności, możemy mówić o możliwości logicznej. Taka możliwość pozwala zbadać zależności między pojęciami, choć nie mówi nic bezpośrednio o świecie. Jeśli możliwość oceniamy jako metafizyczną, to racje pozytywnie opisujące zjawiska

[60] Ujmując rzecz bardziej technicznie, możemy powiedzieć, że z wtórną intensją i pojmowalnością mamy do czynienia, kiedy ustalimy – najczęściej drogą doświadczenia – czym faktycznie są obiekty, do których się dana nazwa odnosi. W duchu Saula Kripkego powiedzielibyśmy też, że wtórna intensja związana jest z koniecznościami *a posteriori*. Bardziej generalny opis semantyk dwuwymiarowych można znaleźć w pracy Laury Schroeter (2021).

[61] Do tematyki relacji między możliwością a pojmowalnością powrócimy w rozdziale poświęconym eksperymentowi myślowemu z zombie.

są atrakcyjniejsze niż te, które jedynie odmawiają zjawiskom danych cech. Dla oceny takiej możliwości ważne jest też, aby racje w eksperymentach myślowych opisane były zgodnie z co najmniej pierwotnymi intensjami użytych terminów, dla których jesteśmy w stanie ocenić ryzyko błędów poznawczych i innych tendencji w myśleniu. Tym samym bowiem będziemy mogli przybliżyć się możliwie blisko idealnej pojmowalności. Z kolei w odniesieniu do sytuacji możliwych fizycznie będą nas interesować racje sprzęgnięte z przyrodoznawczym obrazem rzeczywistości. Szczególnie w tym ostatnim przypadku będą więc nas zajmować racje prawdziwe w odczytaniu zgodnym z wtórnymi intensjami użytych terminów.

5.2 TREŚCIOWY ZWIĄZEK

Zgodnie z kryterium (2) wniosek eksperymentu myślowego musi dotyczyć rozważanej sytuacji. Gdyby w eksperymencie Galileusza o spadku swobodnym skupić się na kolorze kamieni zamiast na ich złączeniu, nie udałoby się obalić tezy o szybszym spadku cięższych ciał. Niespełnione byłoby też jedno z kryteriów Beckera przywołanych w rozdziale 1 – przedstawiona fikcja eksperymentu nie byłaby argumentacyjnie niezbędna.

Aby zrozumieć, jak sposób przedstawienia sytuacji przekłada się na skłonność do akceptacji wniosku, spójrzmy na eksperyment znany jako młyn Leibniza:

> […] postrzeżenie i to, co od niego zależy, nie da się wytłumaczyć racjami mechanicznymi, tzn. przez kształty i ruchy. Przypuściwszy zaś, że istnieje maszyna, której budowa pozwala, aby myślała, czuła, miewała postrzeżenia, będzie można pomyśleć ją, z zachowaniem tych samych proporcji, tak powiększoną, by można do niej wejść jak do młyna. Założywszy to, odnaleźlibyśmy wewnątrz przy zwiedzaniu jej tylko części, które popychają siebie wzajemnie, nigdy jednak nic, co tłumaczyłoby postrzeżenie. Toteż trzeba szukać tego właśnie w substancji prostej, a nie w rzeczy złożonej, czy też w machinie (Leibniz, 1969, § 17).

Leibniz przekonuje, że zrozumienie mechanizmu działania mózgu nie wyjaśnia aktywności umysłowych. Spacerując po mózgu powiększonym do rozmiarów młyna, nie zauważymy przecież nigdzie myśli ani odczuć – są one jego zdaniem własnościami substancji niepodzielnej, prostej. Czy wniosek jest faktycznie powiązany z nakreśloną sytuacją? Można mieć wątpliwości: równie dobrze możemy wyobrazić sobie spacer po mechanizmach tłumaczących inne zjawiska, także te nieuznawane za tak problematyczne jak aktywność

umysłowa. Zmodyfikujmy nieco przebieg eksperymentu, nie naruszając przy tym istoty opisywanej sytuacji (nowy fragment pisany kursywą):

> Przypuściwszy zaś, że istnieje maszyna, której budowa pozwala, aby myślała, czuła, miewała postrzeżenia, będzie można pomyśleć ją, z zachowaniem tych samych proporcji, tak powiększoną, by można do niej wejść jak do młyna. *Tak jak za powstanie mąki odpowiada proces mielenia ziaren mlewnikiem poruszanym siłą spadającej wody lub wiatru, tak samo za postrzeganie kota odpowiada w maszynie mechanizm odwzorowujący wygląd kota poprzez skomplikowane układanie klocków, wywołujące wtórnie ucisk na elementy odpowiadające za generowanie przez maszynę dźwięku brzmiącego jak ludzki głos mówiący „Widzę kota!"*. Założywszy to, odnaleźlibyśmy wewnątrz przy zwiedzaniu maszyny jej tylko części, które popychają siebie wzajemnie, nigdy jednak nic, co tłumaczyłoby postrzeżenie bądź powstanie mąki.

Jeśli Leibniz ma rację, że postrzeżenia (kota, dla ilustracji) nie sposób zobaczyć w maszynie, to analogicznie nie można zobaczyć w młynie produkcji mąki. Sęk w tym, że da się ją zobaczyć, z tym że nie jest ona elementem młyna, tylko realizowanym w nim procesem. Podobnie spostrzeganie kota jest procesem zachodzącym w maszynie, którego wynikiem jest stan, w którym jest ona „przekonana", że widzi kota. Każdy krok obu procesów można ujrzeć, przebywając w maszynie[62]. W zmodyfikowanej wersji narracji nie zmieniliśmy przebiegu sytuacji (spacer po młynie/maszynie). Nie dodaliśmy żadnych informacji, które zniekształciłyby analogię młyna i maszyny, dodaliśmy jedynie przykład zgodny z oryginalną narracją.

Można zarzucić, że mimo wszystko takie obudowanie większą ilością informacji zmienia wydźwięk eksperymentu w nieżyczliwy dla Leibniza sposób. Podejdźmy więc do sprawy nieco inaczej i zamiast doprecyzowania młyna rozważmy kroplę wody:

> Wyobraź sobie kroplę wody powiększoną do rozmiaru młyna. Przechadzając się po tej ogromnej cząsteczce możesz ujrzeć wiele niesamowitych rzeczy, ale niczego mokrego. Ale to wcale nie pokazuje, że woda nie składa się jedynie z H_2O (Cole, 1984, s. 432).

Z pewnością, spacerując po wodzie w skali atomowej, nie dostrzeżemy niczego mokrego. Jednak w skali makro woda jest niewątpliwie mokra. Czy wyobrażenie sobie takiego „suchego" spaceru skłoni nas do myślenia, że nie

[62] Opis, jak maszyna „spostrzega" kota, nie jest oczywiście z dzisiejszej perspektywy adekwatnym obrazem tego, jak nasz mózg identyfikuje obiekty w otoczeniu z informacji wzrokowej (Masland, 2020), natomiast jest odbiciem lustrzanym mechanizmu młyna.

ma nic w strukturze cząsteczki wody, co tłumaczyłoby mokrość? Zapewne nie, jesteśmy mocno przyzwyczajeni do zwyczajności tej własności[63].

Leibniz chciał pokazać, dlaczego mechanicyzm nie wyjaśnia natury umysłu. Aby uznać ten wniosek, musielibyśmy jednak stwierdzić, że podobnie tajemnicza jest natura produkcji mąki i mokrości wody. Tego raczej nie uznamy, natomiast eksperyment Leibniza pozostaje wartościowy dydaktycznie, ilustruje bowiem, na czym polegają własności emergentne, czyli takie, które ma pewna całość, ale nie mają ich jej części: młyn produkuje mąkę, samochód transportuje, a mózg generuje postrzeżenia, choć żaden z elementów składowych tych systemów nie ma tych cech. Czy możemy więc potraktować młyn Leibniza jako wiarygodny eksperyment myślowy w zakresie emergencji, a nie argument antymechanicystyczny? Nie do końca, bo zmieniamy przecież oryginalną konkluzję eksperymentu, a to już jest zmiana zasadnicza. W konsekwencji mamy do czynienia z innym eksperymentem myślowym. Jego wartość polega na ukazaniu istnienia własności emergentnych, a nie odrzuceniu mechanicyzmu[64].

5.3 STRUKTURA WNIOSKOWANIA W EKSPERYMENCIE

Kryterium (3) dotyczy struktury wnioskowania w eksperymencie myślowym. Jeśli przejście do konkluzji ma charakter dedukcyjny, możemy ją przyjąć za prawdziwą, o ile narracja wydaje się przedstawiać prawdopodobny rozwój sytuacji wyjątkowej. W przypadku wnioskowań zawodnych możemy jedynie mniej lub silniej przypuszczać, że konkluzja jest prawdziwa, a sam eksperyment wiarygodny. Jak powiedzieliśmy w rozdziale o modelach mentalnych, rekonstrukcja struktury wnioskowania eksperymentu bywa trudna – opisy eksperymentów często są rozproszone w tekście, bez wyraźnie zarysowanych granic. Ponadto narracja przedstawia zwykle pewien porządek czasowy i ciąg przyczynowo-skutkowy, co wspiera rozwój stosownego modelu mentalnego, ale może też być on zgodny z wieloma wnioskowaniami.

[63] Ponadto do zaobserwowania mokrości w skali makro potrzeba czegoś więcej niż jedna cząstka wody, potrzeba jeszcze interakcji tych cząsteczek.

[64] Rzadko kiedy w filozofii udaje się tak prosto przekonać innych, ominęliśmy zresztą sporo niuansów pojęciowych. Przykładowo, niektóre rodzaje emergencji są kłopotliwe dla mechanicyzmu, inne zaś nie. Zainteresowanych odsyłam do bardziej systematycznych analiz eksperymentu Leibniza (np. Duncan, 2012), jak i dyskusji nad emergencją (Poczobut, 2009).

Posiadanie wielu rekonstrukcji jest trudnością, lecz nie nieprzezwyciężalną. Kluczowe jest ustalenie kontekstu i pytania, na które odpowiada konkluzja eksperymentu. Podobnie jak badania empiryczne podlegają ocenie społeczności naukowej, tak wiarygodność eksperymentu myślowego podlega ocenie społeczności osób je analizujących. Większe zainteresowanie eksperymentem sprzyja lepszym rekonstrukcjom, a i niekiedy sam autor eksperymentu myślowego sygnalizuje wprost strukturę wnioskowania[65].

Aby zilustrować różnicę między wnioskowaniem a narracją, przyjrzyjmy się eksperymentowi Johna Searle'a (1995) znanemu jako chiński pokój. Jest on jednym z najważniejszych argumentów wymierzonym przeciw mocnemu programowi sztucznej inteligencji (SI), głoszącemu, że programy wykonujące formalne przekształcenia znaków faktycznie rozumieją język i mają inne stany poznawcze[66]. Narracja eksperymentu przedstawia sytuację, w której Searle, nieznający języka chińskiego, znajduje się w zamkniętym pokoju z trzema plikami chińskich znaków i instrukcjami w języku angielskim. Instrukcje pozwalają mu korelować po kształtach znaki z pierwszego pliku z drugim, a następnie odpowiadać na pytania z trzeciego, tworząc komunikaty nieodróżnialne od odpowiedzi natywnego użytkownika języka chińskiego:

> […] po pewnym czasie osiągam taką sprawność w stosowaniu instrukcji posługiwania się chińskimi symbolami […], że z zewnętrznego punktu widzenia – to znaczy z punktu widzenia kogoś przebywającego poza pokojem, w którym jestem zamknięty – moje odpowiedzi na owe pytania są absolutnie nieodróżnialne od odpowiedzi człowieka mówiącego po chińsku od urodzenia. […] Jeśli chodzi o język chiński, to zachowuję się po prostu jak komputer; wykonuję obliczenia na elementach określonych w sposób formalny. Na użytek języka chińskiego jestem po prostu przykładem programu komputerowego (Searle, 1995, s. 301–304).

Narracja jest zdecydowanie wyobrażalna[67], sytuacją wyjątkową jest w niej zaś rozpoznanie jako biegłego w języku chińskim umieszczonego w pokoju

[65] Jak zobaczymy w kolejnym rozdziale, tak robi chociażby Chalmers (2003a), rekonstruując wnioskowanie w eksperymencie myślowym z zombie. Ta rekonstrukcja jest powszechnie akceptowana, nawet przez krytyków (np. Stoljar, 2007).

[66] „Ale czy coś mogłoby myśleć, rozumieć […] wyłącznie za sprawą tego, że jest komputerem o właściwym programie? Czy to, że coś ucieleśnia jakiś program – rzecz jasna, właściwy program – może być samo przez się wystarczającym warunkiem rozumienia?" (Searle, 1995, s. 319). Zainteresowanych dyskusją wokół chińskiego pokoju odsyłam do pracy Davida Cole'a (2023).

[67] Ograniczeniem może być tempo przetwarzania znaków wprowadzanych przez osobę znajdującą się wewnątrz pokoju. Pamiętajmy jednak, że opisana narracja jest jedynie

Searle'a przez osoby poza pomieszczeniem. Konsekwentny zwolennik myślenia o mocnej SI powinien na pierwszy rzut oka powiedzieć, że osobnik w pomieszczeniu faktycznie rozumie język chiński. Niemniej trudno się zgodzić z tym stanowiskiem w sytuacji, kiedy znamy powody, dla których Searle potrafi formułować poprawne komunikaty w obcym dla siebie języku, i wiemy, że nie zna on języka chińskiego.

Eksperyment chińskiego pokoju, choć dość wiekowy, nadal rezonuje w kontekście generatywnych SI, takich jak ChatGPT, opartych na dużych modelach językowych. Samo wnioskowanie ma całkiem prostą postać (za: Cole, 2023):

1. Jeśli mocny program SI jest prawdziwy, to istnieje program rozumiejący język chiński poprzez przekształcanie znaków w sposób poprawny z perspektywy natywnego użytkownika.
2. Osoba może wykonywać kroki tego programu, nie rozumiejąc języka chińskiego.

Konkluzja: [Zgodnie ze stwierdzeniem 2.] Mocny program SI jest fałszywy.

Przesłanka 1. wynika z definicji mocnego programu SI, utożsamiającego umysł ze zbiorem programów obliczeniowych, 2. zaś jest niepodważalna w świetle narracji eksperymentu. Nawet najsprawniejsze dziś SI działają w oparciu o czysto formalne zależności[68].

Dodajmy, że zdaniem Searle'a, aby móc autentycznie *rozumieć*, potrzeba intencjonalności. Stany umysłowe człowieka są intencjonalne, bo odnoszą się ku konkretnym stanom rzeczy[69]. Kiedy pomyślimy o leżącym na blacie kuchennym jabłku, wyobrażając sobie jego kształt, kolor, twardość i soczystość, nie mamy żadnych wątpliwości, do czego ta treść umysłowa się odnosi. Analogicznie dzieje się, kiedy widzimy nazwę „jabłko". Gdy jednak Searle w chińskim pokoju zgodnie z regułami maluje znaki języka chińskiego, jego stany umysłowe nie odnoszą się do stanów rzeczy faktycznie oznaczanych przez te znaki w tym języku. Program komputerowy przekształcający symbole

uproszczoną reprezentacją algorytmu działającego na składni językowej. Możemy też wyobrazić sobie wdrożenie takiego algorytmu w konkretnym podmiocie, np. w Searle'u.

[68] A dokładniej, częstość występowania wyrazów w swoim pobliżu (zob. OpenAI, 2023).

[69] Łatwo tu pomieszać dwa porządki: językowy i umysłowy. Sposób wskazywania odniesień przedmiotowych wyrażeń językowych (intensja) i bycie nakierowanym na przedmiot myśli (intencjonalność) to dwa różne zagadnienia (zob. Searle, 2010, s. 176–179, 250).

analogicznie do Searle'a w pokoju również stanów intencjonalnych nie posiada, a więc nie może rozumieć języka chińskiego[70].

Wiele słynnych eksperymentów myślowych skrywa podobnie proste wnioskowania, choćby młyn Leibniza z poprzedniego rozdziału:

1. Jeśli postrzeżenia są złożonej natury, to są mechaniczne.
2. To, co mechaniczne, można wytłumaczyć pracą jego części.
3. Obserwując pracę części maszyny mającej postrzeżenia, nie dostrzegamy postrzeżeń.

Konkluzja: Postrzeżenia nie są złożonej natury.

Wskazując niewystarczający związek treściowy pomiędzy narracją a konkluzją eksperymentu, dodaliśmy więcej detali (produkcję mąki, spostrzeganie kota) do wyobrażanej sytuacji, nie negując tego, co już było w pierwotnej narracji. Wnioskowanie jednak zmieniło swoją postać. Rozważmy dwie jego możliwe rekonstrukcje:

1. Jeśli postrzeżenia są złożonej natury, to są mechaniczne.
2. Jeśli powstawanie mąki jest złożonej natury, to jest mechaniczne.
3. To, co mechaniczne, można wytłumaczyć pracą jego części.
4. Obserwując pracę części maszyny mającej postrzeżenia, nie dostrzegamy postrzeżeń.
5. Obserwując pracę części produkującej mąkę maszyny, nie dostrzegamy powstania mąki.

Konkluzja: Postrzeżenia i powstawanie mąki nie są złożonej natury.

Strukturalnie rekonstrukcja może się wydać bliska naszemu (jedynie) wzbogaceniu narracji, niemniej jest niepoprawna. Zakwestionować możemy nie tylko konkluzję (analogiczną do oryginalnej), ale i przesłankę 5. – na końcu procesu powstawania mąki dostrzeżemy w istocie mąkę. Osoba spacerująca tymczasem po postrzegającej maszynie nie dostrzeże samego postrzeżenia kota.

[70] Zdaniem Searle'a (2010, rozdz. 4) intencjonalność może towarzyszyć tylko maszynom biologicznym skonstruowanym analogicznie do człowieka. Przy okazji warto zauważyć, że również w eksperymentach myślowych rozumienie odgrywa istotną rolę. Swobodniejsza narracja eksperymentu czy brak jednoznaczności terminologicznej mogą prowadzić do jego błędnej interpretacji. Powinniśmy także brać pod uwagę możliwość, że współczesne generatywne SI mogą symulować poprawne przeprowadzenie eksperymentu myślowego – tj. trafnie uznać wniosek i jego uzasadnienie – nawet jeśli wciąż jedynie symulują rozumienie języka.

Istota naszego rozbudowania narracji była inna. Chcieliśmy zwrócić uwagę na to, że układy złożone (w tym i mechanizmy) miewają własności emergentne, nieposiadane przez ich części. Produkcja mąki i postrzeganie są takimi własnościami. Rekonstrukcja lepiej oddająca nasz zamysł będzie więc następująca:

1. Jeśli postrzeżenia są złożonej natury, to są mechaniczne.
2. To, co mechaniczne, można wytłumaczyć pracą jego części.
3. Mechanizm może posiadać własności, których nie posiadają jego części.
4. Obserwując pracę części maszyny mającej postrzeżenia, nie dostrzegamy postrzeżeń.

Konkluzja: Postrzeżenia są własnościami emergentnymi mechanizmu.

Przy tak zmodyfikowanym wnioskowaniu, nawet jeśli zaakceptujemy wszystkie przesłanki, prawdziwość stwierdzenia „postrzeżenia nie są złożonej natury" pozostaje kwestią otwartą. Być może własności emergentne są prostej natury. Wcale więc nie musimy się zgadzać z konkluzją oryginalnego sformułowania Leibniza.

5.4 WIARYGODNOŚĆ INTUICJI

Jeśli narracja eksperymentu myślowego przedstawia możliwą sytuację, konkluzja jest treściowo z nią związana, a wnioskowanie ma dedukcyjny lub uprawdopodabniający charakter, to możemy przejść do oceny wiarygodności intuicji, na których się opiera. Tu pojawia się kluczowe pytanie: na ile intuicje te wynikają z wiedzy, a na ile z przekonań?

Warto już na początku tych rozważań pozbyć się wrażenia, że wiedzę można łatwo odróżnić od przekonań. Klasyczne ujęcie wiedzy jako prawdziwego i uzasadnionego przekonania jest problematycznie niewystarczające, co dobrze ilustruje słynny eksperyment myślowy Gettiera (1963, s. 122). Wyobraźmy sobie, że Smith i Jones starają się o tę samą posadę. Smith uważa, że prawdziwe jest następujące zdanie: „Jones dostanie posadę i Jones ma dziesięć monet w kieszeni". Jest to dla niego pewne, choćby dlatego, że prezes firmy go zapewnił, że to Jones dostanie pracę, i policzył też monety w kieszeni Jonesa. Smith konkluduje, że posadę zatem otrzyma osoba, która ma dziesięć monet w kieszeni. Niewątpliwie to jego przekonanie jest uzasadnione. Pod koniec rekrutacji Smith dowiaduje się, że to on dostał pracę, a nie Jones. Jego przekonanie jest jednak prawdziwe, gdyż on również ma dziesięć monet w kieszeni, choć akurat o tym nie wie. Czy zatem Smith *wie*, że posadę

otrzyma osoba z dziecięcioma monetami w kieszeni? Zgodnie z klasyczną koncepcją wiedzy powinniśmy odpowiedzieć twierdząco. Jednocześnie jednak czujemy, że sensownie jest powiedzieć, że Smith *nie wie*, ponieważ jego uzasadnienie odnosi się do monet w kieszeni Jonesa, a nie jego własnych.

Paradoks w intuicyjnych ocenach tego typu scenariuszy (nazywanych od nazwiska filozofa scenariuszami gettierowskimi) doprowadził do zintensyfikowania prac w obrębie epistemologii (Ichikawa i Steup, 2017). Lecz choć proponuje się inne sposoby odróżniania wiedzy od przekonań (np. Phillips i in., 2021), to do dziś nie ma zgody co do kryteriów wiedzy. Zgodnie z ostatnią ankietą PhilSurvey z 2020 roku 23% filozofów akademickich skłania się ku klasycznej koncepcji, 32% ku odmiennym analizom pojęcia wiedzy, a 31% ku jego nieanalizowalności (Bourget i Chalmers, 2023).

Często wzorcowej pewności informacji szukamy w naukach formalnych, takich jak matematyka czy logika, gdzie procedura dowodowa gwarantuje prawdziwość twierdzeń (zob. Antonutti Marfori, 2010). Jednak i w tych naukach są przyjmowane bez dowodu założenia zwane aksjomatami. Co więcej, nawet w nieaksjomatycznych i innych nietradycyjnych matematycznych i logicznych systemach musimy założyć definicje elementów i dopuszczalnych operacji (Wang, 2013). Interpretacje rzeczywistości nauk przyrodniczych oparte są zaś na analizach danych wskazujących różny stopień prawdopodobieństwa nieprzypadkowości zaobserwowanych zależności[71]. Uznajemy te nauki za wiedzotwórcze, ponieważ stosują powtarzalne, intersubiektywne i wystandaryzowane praktyki potwierdzania i odrzucania hipotez. Stopień ufności w ustalenia naukowe jest tym większy, im więcej potwierdzeń eksperymentalnych za nimi stoi, przy jednoczesnym braku przekonującej większość społeczności naukowej ich falsyfikacji[72].

[71] To, jaki stopień prawdopodobieństwa jest akceptowalny, zależy od danej nauki empirycznej. Dodać trzeba, że oszacowanie przypadkowości zaobserwowanych zależności nie kończy analizy. Powszechna postawa badaczy sprowadzająca analizę danych do poszukiwań istotnych statystycznie (nieprzypadkowych) zależności stanowi jedną z przyczyn kryzysu replikacyjnego dzisiejszej nauki i jest potępiana przez naukowe towarzystwa statystyczne (zob. Camerer i in., 2018; Matthews i in., 2017).

[72] Jeśli hipoteza jest całkowicie nietestowalna eksperymentalnie, to trudno uznać ją za naukową. W praktyce naukowej często jednak formułuje się hipotezy dotyczące zależności testowalnych jedynie w określonych warunkach – np. w przewidywalnych, lecz jeszcze nieosiągalnych technologicznie eksperymentach bądź w specyficznych kontekstach sytuacyjnych. Humanistykę można natomiast opisać jako niekończący się cykl (re)interpretacji i rozumienia kulturowego. To dalece uproszczony i normatywny obraz szeroko

Każda informacja, którą uznajemy za wiedzę, opiera się ostatecznie na przyjęciu określonych założeń. Niemniej celem nie jest tu wcale szamotanie się ze skrajnym sceptycyzmem poznawczym, a zwrócenie uwagi na społeczny charakter praktyk wiedzotwórczych. Podstawą tych praktyk jest rozsądek, struktura stanów epistemicznych kształtowana przez filogenezę (rozwój gatunkowy), częściowo wspólną dla większości z nas ontogenezę (rozwój osobniczy) oraz kulturę. Za Timothym Williamsonem (2019, s. 21) możemy powiedzieć, że wiedzą zdroworozsądkową jest to, co wie większość członków danej społeczności, przekonaniami zdroworozsądkowymi zaś to, w co jedynie wierzą.

Przynależność informacji do korpusu wiedzy zdroworozsądkowej jest częściowo zależna kulturowo – choćby znaczenia słów danego języka są niedostępne osobom go nieznającym – ale należą też do niego prawdy dość uniwersalne: wszyscy ludzie mają głowy, a słońce jest widoczne za dnia. Naturalnie, wprowadzenie perspektywy zdroworozsądkowej nie eliminuje trudności z odróżnieniem wiedzy od przekonań, przenosi ją jedynie z poziomu indywidualnego na grupowy.

Bo chociaż rozsądek nie jest niezawodny, to jego wprowadzenie umożliwia odróżnienie jedynie indywidualnych przekonań od tych podzielanych przez społeczność. Najlepszym do tego narzędziem jest szczegółowa analiza rozumowania, skupiona na wykryciu niedopowiedzeń, różnic znaczeniowych oraz innych potencjalnych czynników zakłócających ocenę wiarygodności eksperymentu myślowego. Wiedzą jest więc takie przekonanie zdroworozsądkowe lub jego konsekwencja, dla których jako społeczność znajdujemy wyjątkowo dużo racji, a jednocześnie nie jesteśmy w stanie obronić żadnej racji zaprzeczającej.

Oznacza to, że nasze uznanie konkluzji eksperymentów myślowych opartych na rozsądku może ulegać zmianie. Analogiczne zjawisko zachodzi w naukach empirycznych; przyszłe badania czy teorie mogą podważyć wcześniejszą metodologię lub interpretację danych. Lobotomia, przecięcie większości połączeń między przedczołowymi obszarami kory a pozostałymi częściami mózgowia, była począwszy od połowy lat 30. do końca lat 40. XX wieku uznawana przez środowiska naukowe za procedurę medyczną leczącą zaburzenia

rozumianej naukowości. Nauka jest również działalnością międzyludzką i na faktyczną ocenę naukowości mają także wpływ czynniki społeczne, kulturowe i technologiczne (zob. Pigliucci i Boudry, 2013).

psychiczne. Z czasem została zdelegalizowana ze względu na swoje destrukcyjne skutki. Tak samo jak w medycynie pewne procedury uznawane za skuteczne mogą zostać odrzucone z uwagi na późniejsze dane, tak i z postępem naukowym ocena metod pomiaru, teorii czy pojedynczych konstruktów teoretycznych może ulec zmianie. Przykładów można by podać więcej: kiedyś myślano, że kształt czaszki odzwierciedla cechy psychologiczne człowieka (frenologia), światło rozchodzi się w hipotetycznym eterze, a spalanie umożliwia substancja zwana flogistonem.

Zwróćmy uwagę, że dynamika zmian w korpusie wiedzy charakteryzuje się metodą niewielkich kroków. Hipotezy badawcze w nauce i sytuacje wyjątkowe w eksperymentach myślowych powstają z potrzeby wypełnienia luk w naszej wiedzy. Śmiałość w wysuwaniu hipotez i kreśleniu sytuacji wyjątkowych polega na przekraczaniu naszego obecnego rozumienia, ale tylko o jeden krok. Dzięki temu testowanie hipotez oraz rozważanie rozwoju sytuacji wyjątkowych pozostaje możliwe w ramach uznanej metodologii i dostępnej aparatury pomiarowej (dla hipotez eksperymentalnych), a także prowadzi do wyjaśnienia i wzrostu rozumienia w sposób zgodny z naszym zasobem epistemicznym (dla eksperymentów myślowych). Jeśli zrobimy za duży krok naprzód, możemy stracić kryteria oceny przedsięwzięcia badawczego i intelektualnego.

O ile eksperyment myślowy opiera się na istniejących teoriach naukowych i poczynionych już problematycznych obserwacjach, stosunkowo łatwo zachować metodę niewielkich kroków. W przypadku pozostałych eksperymentów myślowych, bazujących głównie na intuicjach pojęciowych, jest to dużo trudniejsze. Ocena ich trafności wymaga szczególnej uwagi. Choć w obrębie jednej kultury posługujemy się podobnymi pojęciami – co gwarantują procesy akwizycji języka, enkulturacji i socjalizacji – to okazuje się, że w celu poprawnej rekonstrukcji pojęć potocznych nie zawsze wystarczy sięgnąć do własnej głowy. Nawet filozofowie, uznający swoje myślenie za bardziej systematyczne i wiarygodne niż potoczne (Egler i Ross, 2020; Kauppinen, 2007; Ludwig, 2007), są podatni na błędy poznawcze.

Badania wskazują, że sądy filozofów różnicuje płeć (Buckwalter i Stich, 2014), czynniki kulturowe (Byrd, 2023; Kerr, 2019; Kim i Yuan, 2015; Weinberg i in., 2001; Yuan i Kim, 2021), warunki socjoekonomiczne (Nichols i in., 2003) czy poziom znajomości poszczególnych obszarów językoznawstwa (Machery, 2012). Filozofowie są też podatni na inne błędy poznawcze, jak chociażby wpływ kolejności prezentowania bodźców w badaniach czy stronniczości opisu bodźców (Schwitzgebel i Cushman, 2012, 2015; Tobia i in., 2013).

Wskazywanie na tego rodzaju wyniki prowadzi nie do relatywizmu, lecz do uznania potrzeby współpracy możliwie różnorodnych grup filozofów zajmujących się daną materią. To praca w kierunku wypracowywania konsensusu filozoficznego umożliwia odróżnienie przekonań (tylko) indywidualnych od takich, które mają szerszą, intersubiektywną podstawę.

Może jednak nie ma znaczenia, kto spośród członków wspólnoty pojęciowej odwołuje się do intuicji pojęciowych? Joshua Knobe (2019), pionier filozofii eksperymentalnej, przywołuje 30 badań obejmujących prawie 13 tysięcy osób i wskazujących na wysoką odporność intuicji pojęciowych dotyczących kwestii filozoficznych na czynniki demograficzne. Jednak Stephen Stich i Edouard Machery (2023), zarzucając Knobe'owi wybiórczość, przedstawili listę 100 badań z udziałem ponad 40 milionów osób, które wskazują na demograficzną zmienność takowych intuicji[73]. To potwierdza znaczenie pracy nad konsensusami filozoficznymi z wykorzystaniem różnorodnych metod, zarówno teoretycznych, jak i empirycznych, jako narzędzi minimalizujących wpływ błędów poznawczych. Liczbowo hipoteza o zmienności intuicji filozoficznych przeważa. Niemniej, niezależnie od przyszłych ustaleń, ferwor badań metafilozoficznych[74] już przyczynił się do bardziej świadomego stosowania przesłanek probabilistycznych przez filozofów, w tym statystyki i wyników eksperymentalnych (Fletcher i in., 2021), nowych podejść do filozoficznej metodologii (np. Machery, 2017), jak i bardziej zniuansowanego rozumienia eksperckości filozofów (Egler i Ross, 2020).

Kolejnym zagrożeniem dla poprawnej interpretacji intuicji pojęciowych jest błąd językowej istotności (*linguistic salience bias*). Wynika on z faktu, że słowa często mają stereotypowe znaczenia, które pierwsze przychodzą nam na myśl (Giora, 2003). Niektóre są zupełnie neutralne. Słowo „laptop" może automatycznie przywołać obraz komputera z systemem Windows, a nie macOS czy Linux. Jednak inne stereotypy mogą prowadzić

[73] Autorzy nadmieniają, że jest to dramatycznie wysoka liczba, biorąca się przede wszystkim z trzech obszernych badań, ale nawet jeśli je pominiemy, to pozostałe eksperymenty objęły ponad 93 tysiące osób (Stich i Machery, 2023, s. 450).

[74] Nie wszyscy tak optymistycznie podchodzą do tej dyskusji. Weasley Buckwalter (2022) rozważa np., na ile możemy mówić o replikacyjnym kryzysie w filozofii, podobnym do zaobserwowanych w ostatnich latach problemów z rzetelnością, powtarzalnością wyników badań psychologicznych (Open Science Collaboration, 2015) i nauk społecznych (Camerer i in., 2018) czy biologii nowotworowej (Errington i in., 2021; zob. też Cova i in., 2021).

do istotnych zakłóceń, szczególnie w kontekście społecznych i filozoficznych debat[75].

Stereotypy językowe odgrywają niebagatelną rolę w spontanicznych rozumowaniach. Interpretujemy wypowiedzi innych zgodnie ze stereotypowymi cechami obiektów i zjawisk, o których traktują (Levinson, 2000). Im bardziej dominujący (*salient*, *dominant*) jest sens przypisywany danemu słowu[76], tym szybciej i silniej jest on przywoływany (Fein i in., 2015). Separowanie takich najsilniejszych znaczeń od innych, preferowanych przez osobę używającą danego słowa w określonym kontekście, wymaga od obu stron komunikacji celowych starań. Należy uważać na kontekst, hamować niepasujące skojarzenia czy chociażby bezpośrednio wskazywać specjalne użycia danego słowa (E. Fischer i Engelhardt, 2017, s. 414; Givoni i in., 2013).

W serii artykułów Eugen Fischer i Paul Engelhardt (2016, 2017, 2019a, 2019b) wskazali, jak aktywacja stereotypowych znaczeń negatywnie wpływa na ocenę wiarygodności eksperymentów myślowych i argumentacji filozoficznej. Twierdzą oni, że filozoficzne argumenty, w tym eksperymenty myślowe, często są obarczone kontekstowo nieprawidłowym wnioskowaniem stereotypicznym (*contextually inappropriate stereotypical inferences*). Z uwagi na opisywanie sytuacji wyjątkowych często stosowane są w narracji eksperymentów sensy niedominujące. Dominujące znaczenia są jednak także aktywowane, zgodnie ze swoją stereotypową naturą, i mogą zakłócać ocenę eksperymentu myślowego bądź innej argumentacji (E. Fischer i Engelhardt, 2019b, s. 420–421). Przykładowo, czasownik „widzieć" (*to see*) w wypowiedziach typu „S widzi X" ma dominujące znaczenie dotyczące relacji przestrzennej. To znaczenie jest automatycznie aktywowane, nawet kiedy kontekst użycia tego słowa jest wyraźnie pozbawiony traktowania o przestrzenności, tak jak w filozofii, gdzie „widzieć" to postrzegać, rozumieć, ale też uświadamiać sobie (E. Fischer i Engelhardt, 2017, s. 422).

[75] Na przykład wiele nazw zawodów domyślnie sugeruje płeć osób je wykonujących (m.in. „adwokat", „sprzątaczka"), a w niektórych przypadkach, jak „sekretarz" i „sekretarka", mogą one odnosić się do zupełnie różnych stanowisk.

[76] Używam określeń „sens" i „znaczenie" z grubsza zamiennie, na oznaczenie treści pamięciowej aktywowanej przez napotkanie danego słowa. W literaturze dotyczącej *salience bias* autorzy również posługują się anglojęzycznymi odpowiednikami, *sense* i *meaning*, dość zamiennie. Istotność wyznaczana jest przez częstość stykania się z określonym kontekstem znaczeniowym słowa i prototypowości przynależenia nazywanych tym słowem obiektów/zjawisk do kategorii pojęciowej (Giora, 2003).

Co wynika z tego dla oceny wiarygodności eksperymentów myślowych? Aby móc uznać potoczne intuicje pojęciowe zawarte w narracji eksperymentu za wiarygodne, nie wystarczy opinia wąskiej grupy ekspertów. Konieczne jest dodatkowe pokazanie, na ile narracja trafnie oddaje powszechne, zdroworozsądkowe rozumienie rzeczywistości i w jakim stopniu uniknęła zniekształceń wynikających z błędów poznawczych autora eksperymentu.

Przykładowo, spójrzmy na narrację eksperymentu myślowego Dereka Parfita z teleporterem (pełna wersja znajduje się w: Parfit, 1984, rozdz. 10)[77]:

> Wchodzę do maszyny, która po naciśnięciu guzika pozbawi mnie świadomości, zeskanuje wszystkie komórki w moim ciele i po drodze je zniszczy. Pełna informacja na temat moich komórek będzie przekazana z prędkością światła na Marsa, gdzie druga maszyna wyprodukuje idealną, organiczną replikę mnie. Jestem pełen nerwów, ale moja żona, która wielokrotnie już się teleportowała, powiedziała mi, że nic złego jej się nie stało. Mózg repliki będzie zresztą dokładnie taki sam jak mój w momencie naciśnięcia guzika, replika będzie więc psychologicznie identyczna. Naciskam guzik. Po kilku latach maszyna jest udoskonalona. Będąc z powrotem na Ziemi, wchodzę do tej nowszej maszyny. Nie tracę świadomości, więc pytam technika, czy maszyna się nie zepsuła. Ten odpowiada, że nie i że jeśli poczekam chwilę, to uruchomi połączenie wideo z Marsem i będę mógł ze sobą porozmawiać. Ale chwila, jeśli jestem tutaj, to nie mogę być też na Marsie! Technik uruchamia połączenie z Marsem. Czy replika jest mną?

Łatwiej (ale niekoniecznie łatwo!) zaakceptować pierwszą teleportację, niszczącą nasze oryginalne ciało. Druga stawia trudniejsze pytanie o to, czy „ja" to unikalny podmiot. Jesteśmy tylko ciałami z mózgami, które myślą i mają określoną psychikę. Parfit wykorzystuje przedstawioną narrację, by argumentować, że pytanie o tożsamość osobową – pozostawanie tą samą osobą – jest źle postawione, bo nie ma jednej poprawnej odpowiedzi na pytanie o replikę. Zamiast tego proponuje skupienie się na relacji psychologicznej ciągłości, pamiętaniu wcześniejszych przeżyć[78].

---

[77] Podobne pomysły pojawiały się też wcześniej, chociażby u Stanisława Lema w 1956 roku, w opowiadaniu *Z dziennika gwiezdnego Ijona Tichego*, które dziś odnajdziemy w zbiorze *Dzienniki gwiazdowe*.

[78] Ujmując to precyzyjniej: Parfit mówi o psychicznym powiązaniu i ciągłości. Jestem psychicznie powiązany ze sobą w określonej chwili z przeszłości, bo mam wspomnienia przeżyć w tej przeszłej chwili. Mam psychiczną ciągłość ze sobą w określonej chwili z przeszłości, kiedy nie mam bezpośrednich wspomnień przeżyć z tej chwili, ale mam wspomnienia przeżyć z nieco bliższej w czasie chwili, kiedy to pamiętałem przeżycia jeszcze wcześniejsze.

Czy w eksperymencie Parfita dokonujemy małego kroku, trafnie rekonstruując potoczny sens „pozostawania tą samą osobą"?[79] Nad potocznymi intuicjami pozostawania tą samą osobą przeprowadzono szereg badań. Wczesne wyniki potwierdzały znaczenie psychologicznej ciągłości (np. Blok i in., 2005; Nichols i Bruno, 2010), późniejsze wykazały jednak istotne niuanse. Przykładowo, badani chętniej uznają ciągłość osoby, gdy zmiany psychologiczne są pozytywne, a nie negatywne (Tobia, 2015). Rezultaty Renatasa Berniūnasa i Viliusa Dranseiki (2016) sugerują z kolei, że w scenariuszach podobnych do eksperymentu Parfita laicy rozróżniają „osobę" od „indywiduum". To drugie jest postrzegane w kontekście czasowym, co sprawia, że zrozumiałe dla laików jest pozostawanie tym samym indywiduum, przy stawaniu się inną osobą. Potocznie rozróżnialne okazują się także przypadki tożsamości numerycznej i jakościowej – czyli bycia dokładnie tym jednym jedynym obiektem (tożsamość numeryczna) oraz posiadania dokładnie tych samych cech (tożsamość jakościowa).

Bycie osobą obejmuje szereg rozmytych znaczeń, związanych zarówno z psychologią, cielesnością, moralnością, świadomością, relacjami społecznymi, jak i bardziej duchowo rozumianą Jaźnią. Nawet jeśli laicy nie traktują podmiotów w eksperymentach myślowych jako tę samą osobę, często dalej uznają ich tożsamość indywidualną, co pozwala na kontynuowanie używania ich pierwotnych nazw własnych (zob. Dranseika i in., 2023; Schwenkler i in., 2022)[80]. Trudno więc uznać eksperyment Parfita za wiarygodny argument przemawiający za wyższością psychologicznej ciągłości nad innymi kryteriami.

Ocena wiarygodności tego eksperymentu opiera się na wspólnocie pojęciowej, ponieważ nie istnieje ekspercki standard myślenia o osobach. Pojęcie osoby jest abstrakcyjne i prawdopodobnie nie opisuje rodzaju naturalnego – rzeczy istniejących niezależnie od podmiotu poznającego, zwykle wyróżnianych przez nauki[81]. Planety, drzewa i organizmy są ewidentnie rodzaju naturalnego. O niektórych „rzeczach" trudno orzec, czy są rodzaju naturalnego – np. czy emocje takie jak żal czy wstyd lub taksonomiczne kategorie

[79] To, że o takowe mu właśnie chodzi, jest zupełnie jasne dla czytelnika jego książki, zresztą eksperyment myślowy z teleporterem rozpoczyna rozdział zatytułowany *What we believe ourselves to be*, w luźnym tłumaczeniu *Czym wierzymy, że jesteśmy?*.

[80] Zob. tradycyjne filozoficzne ujęcie Petera Strawsona w jego pracy *Indywidua* (1980).

[81] To metafizycznie dość ostrożna charakterystyka rodzajów naturalnych. Dla pogłębienia tematu warto zajrzeć do pracy Alexandra Birda i Emmy Tobin (2023).

gromad, rzędów i rodzin nie są jedynie konwencjami klasyfikacyjnymi? Istnienie wielu rzeczy jest też całkowicie zależne od podmiotu poznającego – ulubiony smak pączka, sentymentalne piosenki, ale i odwaga, sprawiedliwość czy przekonania religijne[82]. Procesy przyrody tworzą organizmy, osoby należą zaś do porządku bardziej kulturowego lub metafizycznego aniżeli naturalnego[83].

W tym też tkwi pewne ograniczenie możliwości oceny wiarygodności intuicji w eksperymentach myślowych. Możliwe, że na niektóre tematy *nie istnieje* żadna wiedza – ani ekspercka, ani zdroworozsądkowa. Przekonania moralne stanowią dobry przykład. To, iż mamy konkretne poglądy moralne, samo jeszcze nie oznacza, że *powinniśmy* je mieć ani że stwierdzają coś obiektywnie prawdziwego o świecie. Wielu filozofów zgadza się z gilotyną Hume'a: z tego, jak jest, nie wynika, jak być powinno. W takich przypadkach ocena intuicji może pozostać nieuchwytna, szczególnie jeśli jesteśmy zainteresowani nie samymi relacjami pomiędzy pojęciami, a stwierdzeniami na temat świata.

5.5 PODSUMOWANIE

- Intuicje – nieświadome rozumowania prowadzące do świadomych wniosków – są nieodłącznym elementem naszego myślenia. Stanowią fundament zarówno codziennych decyzji, jak i najbardziej zaawansowanych ustaleń naukowych. Eksperymenty myślowe mogą dostarczać wiarygodnych wniosków, o ile są oparte na wiarygodnych intuicjach. Ich ocena wymaga jednak systematycznej analizy.
- Aby móc ocenić wiarygodność intuicji w eksperymencie myślowym, ten musi spełnić trzy warunki. Po pierwsze, narracja musi przedstawiać możliwą sytuację, zgodną z wybraną kategorią możliwości. Po drugie, konkluzja powinna być treściowo związana z zaprezentowaną w narracji sytuacją. Po trzecie, uznana rekonstrukcja zawartego w nim wnioskowania ma dedukcyjny lub uprawdopodabniający charakter.
- Podczas dokonywania oceny wiarygodności eksperymentu myślowego niezbędna jest decyzja co do oczekiwanego sensu możliwości sytuacji prezentowanej w narracji. Możliwość logiczna oznacza spójność z naszymi

[82] Osoby platońsko zorientowane, dla których wartości i wszelkie rzeczy abstrakcyjne istnieją obiektywnie, mogą być w tym miejscu nieusatysfakcjonowane, ale przyjmuję tu perspektywę naturalistyczną.

[83] Dostrzegalne jest to także w debatach społecznych na temat prawnego definiowania osoby ludzkiej, chociażby w kontekście legalizacji aborcji.

pojęciami, jej spełnienie więc jest warunkiem koniecznym dla każdego innego rozumienia możliwości. Możliwość fizyczna wymaga zgodności z wiedzą naukową, natomiast możliwość metafizyczna – trudniejsza do oszacowania – pozwala precyzyjnie związać rodzaj pojmowalności narracji ze stopniem ufności we wniosek eksperymentu.

- Rola eksperckich i zdroworozsądkowych zasobów epistemicznych w ocenie wiarygodności eksperymentów myślowych zależy od ich charakteru. Wiedza ekspercka jest kluczowa w przypadku naukowych eksperymentów myślowych, na myśleniu zdroworozsądkowym możemy zaś polegać w eksperymentach dotyczących kwestii abstrakcyjnych. Oba sposoby myślenia wymagają jednak ostrożności, aby uniknąć błędów poznawczych i pochopnych rekonstrukcji pojęciowych.
- Ocena wiarygodności intuicji staje się dodatkowo trudniejsza ze względu na ich zmienność kulturową, demograficzną czy językową. Istnieją również tematy – takie jak przekonania moralne – których być może nie sposób jednoznacznie rozstrzygnąć ani na podstawie wiedzy eksperckiej, ani zdroworozsądkowej. Wiarygodność eksperymentów myślowych opiera się więc nie tylko na jakości argumentacji, ale także na możliwości osiągnięcia szerokiego konsensusu pojęciowego.

6 STUDIUM FILOZOFICZNYCH ZOMBIE

W ramach ostatniego rozdziału przyjrzyjmy się filozoficznym zombie. Stanowią one jeden z najbardziej znanych i kontrowersyjnych eksperymentów myślowych w filozofii umysłu. Rozważenie konsekwencji wyobrażenia sobie zombie skłania nas bowiem do refleksji zarówno nad intuicjami na temat świadomości, jak i nad granicami wyobraźni. Jego analiza więc nie tylko pozwala docenić złożoność problemu świadomości, ale też stanowi doskonałą okazję do wykorzystania omówionych w poprzednich rozdziałach uwag na temat metody eksperymentów myślowych i innych aspektów tego działania poznawczego.

Filozoficzna idea istot do nas podobnych, lecz pozbawionych świadomości, sięga co najmniej czasów Kartezjusza[84]. W *Rozprawie o metodzie* rozważał on różnicę między człowiekiem a zwierzętami, pisząc, że zwierzę jest niczym maszyna, ale maszyna nie jest jak człowiek, bo nawet jeśliby go przypominała w budowie i działaniach, to bylibyśmy w stanie ją odróżnić. Nie byłaby zdolna posługiwać się tak swobodnie językiem jak my i przede wszystkim nie miałaby świadomości[85]. Dzisiejsze zombie rozwijają intuicję Kartezjusza, jednak w oderwaniu od dualizmu – są narzędziem krytyki fizykalizmu[86]. Aby docenić wagę wyzwania, jakie stanowi argument z zombie,

---

[84] Skupiam się w tym miejscu na tropie filozoficznym. Kulturowy trop zombie jako „żywych trupów" ma znacznie bogatszą historię, najbardziej bezpośrednio sięgającą do wierzeń afrykańskich i niewolników w Nowym Świecie (zob. Thomas, 2010).

[85] Zaznaczyć trzeba, że Tadeusz Boy-Żeleński przełożył na *świadomość* starofrancuskie *connoissance* (dzisiejsze *connaissance*), którego bardziej literalne tłumaczenie to *wiedza*. Niemniej w filozofii Kartezjusza treść umysłu czy ducha (*esprit*) jest zawsze świadoma: „Przez nazwę «myślenie» rozumiem to wszystko, co w nas zachodzi, gdy jesteśmy świadomi, o ile jest w nas właśnie świadomość tych rzeczy. Tak więc nie tylko rozumienie, chcenie, wyobrażanie sobie, ale także czucie jest tutaj tym samym, co myślenie" (Descartes, 2001, s. 31).

[86] Dodajmy, że jeszcze za życia Kartezjusza dostrzeżona była trudność dualizmu, tzw. problem interakcji: skoro mamy do czynienia z dwoma niesprowadzalnymi do siebie substancjami, to w jaki sposób umysł czy dusza może wpływać na ciało (i odwrotnie)?

warto wpierw zrozumieć, na czym polega współczesny fizykalizm, będący dominującym dziś rozumieniem relacji umysł – ciało.

6.1 FIZYKALIZM A TRUDNY PROBLEM ŚWIADOMOŚCI

Terminy *fizykalizm* i *materializm* będziemy rozumieć zamiennie (choć zob. Stoljar, 2023), jako stanowisko, wedle którego wszystko, co istnieje, ma charakter fizyczny czy też materialny. Popularne dzięki sukcesom nauk przyrodniczych opiera się na dwóch tezach: ontologicznej (wszystko, co istnieje, jest fizyczne) i epistemologicznej (zjawiska wyższego poziomu można w pełni wyjaśnić w kategoriach fizycznych). Pierwsza głosi więc redukcję zjawisk umysłowych do fizycznych i to pomimo że aktualnie nie wiemy, czym dokładnie jest to, co fizyczne, bo nie posiadamy ostatecznej, fizycznej teorii fundamentalnej. W drugiej natomiast redukcja ma charakter opisowy (zob. też Carnap, 2000).

Powstało wiele wariantów fizykalizmu. Jednymi z nich były teorie identyczności typów i egzemplarzy, rozwijane od lat 50. XX wieku. Te pierwsze zakładały, że określone rodzaje stanów umysłowych są ściśle identyczne z określonymi rodzajami stanów mózgowych, podobnie jak błyskawica jest ruchem ładunków elektrycznych (Place, 1956; Smart, 1959, 1995). Zgodnie z teoriami identyczności egzemplarzy natomiast każde indywidualne zdarzenie umysłowe ma swoją realizację fizyczną, ale niekoniecznie według ścisłej reguły (Davidson, 1992). Oba podejścia miały problem z tezą epistemologiczną – teorie identyczności typów nie wyjaśniały reguł decydujących, czy i dlaczego te same stany umysłowe mogą być realizowane w tych samych układach neuronalnych, identyczności egzemplarzy nie dopuszczały zaś uniwersalnych prawidłowości w tym zakresie, a więc i możliwości pełniejszego zrozumienia identyczności tego, co fizyczne, z tym, co umysłowe. Materializm eliminatywistyczny z kolei odrzucał język psychologii potocznej, z terminami takimi jak *umysł* czy *świadomość*, jako wadliwy i nieprzekładalny na naukowy (Churchland, 1981).

Ogólnie rzecz ujmując, takie fizykalizmy tłumaczące stany umysłu poprzez ich identyczność ze stanami mózgowymi lub poprzez eliminację języka psychologii potocznej przyjęło nazywać się redukcyjnymi. Jednak wielu

---

Zob. korespondencję Elżbiety Czeskiej z Kartezjuszem (Princess Elisabeth of Bohemia, Descartes, 1645/2007).

dzisiejszych filozofów twierdzi, że stany umysłowe mają inne jakości niż stany fizyczne. Taka emergencja bywa rozmaicie dookreślana. W słabszych wariantach własności wyższego rzędu (takie jak umysłowe), choć mogą być zaskakujące, to wynikają z organizacji systemu na niższym poziomie (takim jak mózgowy), w silniejszych podkreśla się zaś, że nie istnieją prawa psychofizyczne pozwalające na redukcję stanów umysłowych do fizycznych. O ile jednak emergencję funduje się na obiektach fizykalnych, to pozostaje fizykalizmem (Poczobut, 2009; van Gulick, 2001). Fizykalistycznie rozumie się dziś też zwykle funkcjonalizm, dla którego stany umysłowe definiuje ich rola przyczynowa w systemach poznawczych. Często kojarzy się funkcjonalizm ze sloganem: stany umysłowe to *software*, mózg zaś to *hardware*. Kluczową cechą funkcjonalizmu jest wieloraka realizowalność – te same stany umysłowe mogą być realizowane na różnych podłożach, nie tylko mózgowym, ale także np. krzemowym (Putnam, 1960).

Zamiast emergencji i redukcji niektórzy mówią też o superweniencji, czyli o tym, że własności umysłowe zależą od fizycznych i nie mogą się zmienić bez jednoczesnych zmian własności fizycznych (tzw. zmiana zależna) (Poczobut, 2012, s. 89). Własności fizyczne są rozumiane jako bazowe, determinujące wystąpienie własności wyższego rzędu, tj. umysłowych[87]. Relacja redukcji psychofizycznej jest więc jednocześnie superwenientną, ale w drugą stronę niekoniecznie. Możemy mieć bowiem do czynienia z niewyjaśnialnym naukowo superweniowaniem (zob. J. Kim, 1990; Wilson, 2005).

Wraz z rozwojem fizykalizmu powróciła idea nieświadomych, choć łudząco podobnych do nas istot. W latach 30. XX wieku George F. Stout pisał o zgodnej z materializmem, ale nie ze zdrowym rozsądkiem wizji nieistnienia „doznających jednostek" ludzkich (1931, s. 138–139). Keith Campbell w latach 70. ukuł na takie istoty określenie „człowiek imitacja" (1970, s. 100–101). Najważniejszymi zaś dla filozoficznych zombie były prace Roberta Kirka oraz Saula A. Kripkego. Ten pierwszy wprowadził zresztą nazwę „zombie" do dyskursu filozoficznego (Kirk, 1974; Kirk i Squires, 1974), atakując behawioryzm analityczny Gilberta Ryle'a (1970), zgodnie z którym wyrażenia z terminami umysłowymi można przełożyć na opisy behawioralne. Filozoficzne zombie Kirka miały obalać każde stanowisko, w którym „zdania na temat

[87] „Powiedzieć, że A-własności lub fakty superweniują na B-własnościach lub faktach, to powiedzieć tyle, że A-fakty są w pewnym sensie redundantne, ponieważ są one całkowicie wyznaczone już wtedy, gdy wyznaczone zostały wszystkie B-fakty" (Stalnaker, 2008, s. 99).

doznań zmysłowych są analizowalne lub przekładalne na wyrażenia tematycznie neutralne [a zatem wyrażenia nauk przyrodniczych]" (Kirk, 1974, s. 59). Kripke (1972, 2001, wykład III) natomiast argumentował na rzecz pojmowalności tego, że nasze doznania bólowe nie są identyczne z pobudzeniem neuronów. Jego argumentacja nie odnosiła się więc do całych istot pozbawionych świadomości, ale strukturalnie należy do tej samej klasy rozumowań, co argument z zombie, czyli tzw. argumentów modalnych czy też argumentów z pojmowalności.

W nauce o świadomości zombie są obecne od samego początku. Mimo rozwoju kognitywistyki i interesowania się nią na początku historii psychologii badania świadomości zaczęto dopuszczać dopiero w latach 90. XX wieku (Dehaene, 2023, s. 16–18; Hoel, 2023, rozdz. 5), za sprawą noblistów Geralda Edelmana (odkrywca struktury przeciwciał) i Francisa Cricka (współodkrywca struktury DNA) (Crick i Koch, 1990; Edelman, 1989b). Choć głosili oni odmienne teorie dotyczące tego, jakie mechanizmy neuronalne tłumaczą świadomość, obaj zgadzali się, że jest ona zjawiskiem naturalnym, wynikającym wprost z naszej biologii (zob. Crick, 1994, polska wersja 1997; Edelman, 1989a).

Nauka o świadomości zmaga się jednak z tzw. trudnym problemem, jak to określił Chalmers (1995, s. 201). Standardowe metody naukowe wyjaśniają procesy uwagowe i integrację informacji celem kontroli zachowania, co nazywamy funkcjonalnym aspektem świadomości (świadomością dostępu, przytomnością, zob. Block, 1995; Chalmers, 2010a, s. 375–381). Jednak nie dotykają one istoty zjawiska, czyli tego, że subiektywnie doświadczamy (*experience*). Każdy z nas *jakoś* doświadcza procesu myślowego i percepcyjnego. Doświadczenia pełne są unikalnych jakości dostępnych tylko osobie je przeżywającej – zwanych qualiami lub własnościami fenomenalnymi (zob. Crane, 2000; C.I. Lewis, 1929; Peirce, 1982).

Qualia, choć osobliwe, są nam najbliższe w indywidualnym doświadczeniu. Możemy znać wiele faktów o nietoperzach, ale żadna ilość danych obserwacyjnych nie powie nam, jak to jest być nietoperzem (Nagel, 1974). Podobnie, nie możemy zobaczyć qualiów drugiej osoby – zakładamy ich istnienie na podstawie podobieństwa, a nie neuronaukowych danych obserwacyjnych.

Funkcjonalnie jesteśmy świadomi, kiedy sięgamy po jabłko lub rozmawiamy o jego barwie. Gdy mówimy „Podaj mi czerwone jabłko", odnosimy się do barwy jako cechy obiektu, a nie naszego stanu umysłowego. Jednak

kiedy spostrzegamy czerwone jabłko, jest nam też *jakoś*[88], doznajemy fenomenalnej czerwieni, jakościowego charakteru świadomego spostrzegania czerwonego jabłka.

Qualia są niewidoczne w zachowaniu i mechanizmach neuronalnych, wymykają się więc naukom przyrodniczym, badającym przecież zjawiska trzecio-, a nie pierwszoosobowe (Goff, 2022). Trudny problem świadomości polega zatem na zapełnieniu luki eksplanacyjnej[89], na zbudowaniu wyjaśnienia pomiędzy tym, co fizyczne, a tym, co fenomenalne. Jest więc wyzwaniem dla fizykalizmu, bo nie wiadomo, jak świadomość mogłaby być fizyczna, skoro ma aspekt fenomenalny. Eksperyment myślowy z zombie jest zaś jednym z kluczowych argumentów tę słabość fizykalizmu eksponujących.

6.2 EKSPERYMENT MYŚLOWY Z ZOMBIE

Naświetliwszy tło filozoficzne popularności, którą zyskał eksperyment myślowy z zombie, możemy przejść do jego szczegółowej analizy. Najbardziej wpływowe sformułowanie tego eksperymentu pochodzi z książki Chalmersa z 1996 roku (w Polsce wydanej w 2010 roku):

> Zastanówmy się więc nad moim bliźniakiem-zombim. To stworzenie jest identyczne ze mną co do cząsteczki, a także identyczne pod względem wszystkich własności niskiego poziomu postulowanych przez kompletną fizykę, ale zupełnie brak mu przeżyć świadomych. […] możemy sobie wyobrazić, że teraz wyglądam przez okno, przeżywając przyjemne doznania zieleni płynące z zielonych drzew na zewnątrz, mając przyjemne doznania smakowe dzięki kawałkowi czekolady […]. Co się dzieje u mojego bliźniaka-zombiego? Jest ze mną fizycznie identyczny i możemy również przypuszczać, że znajduje się w identycznym środowisku. Z pewnością będzie identyczny ze mną funkcjonalnie: będzie przetwarzać tego samego rodzaju informacje, reagować podobnie na dane wejściowe, a jego wewnętrzne konfiguracje będą odpowiednio modyfikowane, co doprowadzi do nieodróżnialnych zachowań. Będzie ze mną identyczny psychologicznie […]. Będzie postrzegał drzewa na zewnątrz w sensie funkcjonalnym i smakował czekoladę w sensie psychologicznym. […] będzie na jawie, będzie w stanie zdawać sprawę z treści swoich stanów wewnętrznych, będzie w stanie skupić uwagę na różnych sprawach i tak dalej. Tyle że temu

[88] Idąc za Andrzejem Klawiterem (2012), moglibyśmy też powiedzieć „Jest mi czerwono", kiedy widzimy czerwone jabłko, analogicznie do tego, gdy mówimy „Jest mi gorąco", kiedy odczuwamy temperaturę swojego ciała.

[89] „[…] nie posiadamy żadnego sposobu wyznaczenia, które dokładnie zdania na temat identyczności psychofizycznej są prawdziwe" (Levine, 1983, s. 334).

> funkcjonowaniu nie będą towarzyszyć w ogóle rzeczywiste przeżycia świadome. Nie będzie poczucia fenomenalnego. Nie istnieje coś takiego, jak to jest być zombim (Chalmers, 2010a, s. 173, 175).

Sytuacja wyjątkowa polega więc na wyobrażeniu sobie świata, w którym istnieją nasze duplikaty doskonale do nas podobne pod względem fizycznym (a więc i biologicznym), psychologicznym i behawioralnym (w sensie wcześniej przywoływanej świadomości dostępu), ale które nie doznają swojego istnienia tak jak my. Jeśli taki świat jest możliwy logicznie, to – twierdzi Chalmers – materializm nie może być prawdziwy, a zatem istnieje luka eksplanacyjna dotycząca świadomości, trudny problem świadomości[90]. Widać to wyraźnie w kontekście, w którym eksperyment ten pojawia się u Chalmersa i innych autorów (zob. Balog, 1999; Stoljar, 2007):

> Aby argumentować przeciwko wyjaśnianiu redukcyjnemu (materialistycznemu), musimy pokazać, że świadomość nie jest superwenientna logicznie na tym, co fizyczne. […] Jak możemy [to] pokazać […]? Najbardziej oczywistą metodą […] jest rozważenie możliwości logicznej zombiego: kogoś lub czegoś fizycznie identycznego ze mną (lub z inną istotą świadomą), ale zupełnie pozbawionego przeżyć świadomych. Na poziomie globalnym możemy rozważyć możliwość logiczną świata zombich: świata fizycznie identycznego z naszym, w którym jednak nie ma w ogóle przeżyć świadomych. W takim świecie każdy jest zombim (Chalmers, 2010a, s. 171–173).

Chalmers rozumie fizykalizm przez pryzmat superweniencji logicznej, czyli doprecyzowując nasze wcześniejsze rozumienie tej relacji: bez zmian fizycznych niemożliwe logicznie są zmiany umysłowe (Chalmers, 2010a, s. 79). Zgodnie z tym rozumieniem o ile tylko dany zbiór wyrażeń jest spójny, to opisuje pewien świat możliwy, a zatem wyraża możliwość logiczną. Jeśli świat zombich jest pojęciowo spójny, to będzie więc nie tylko pojmowalny, ale i możliwy logicznie.

Pojęcie superweniencji logicznej jest silniejsze od przyrodniczej, która mówi jedynie o zależnościach zachodzących w naszym świecie przyrody, bez zmiany praw fizycznych (Chalmers, 2010a, s. 81). Dla Chalmersa superweniencja logiczna najlepiej oddaje ideę redukcyjnego fizykalizmu, dla którego „fakty fizyczne w pewnym sensie wyczerpują cały zakres faktów o świecie" (Chalmers, 2010a, s. 93). Jeśli fizykalizm jest prawdziwy, to nie może być

[90] Należy pamiętać, że według Chalmersa możliwość metafizyczna pokrywa się z logiczną.

żadnej wątpliwości, dlaczego własności wyższego poziomu (fenomenalne) towarzyszą własnościom niższego poziomu (fizycznym). Nie będzie takich wątpliwości, jeżeli zachodzi między nimi superweniencja logiczna, byłoby to bowiem sprzeczne logicznie, a nie tylko niemożliwe w świetle praw fizycznych naszego świata.

W kolejnych podrozdziałach rozpatrzymy m.in. wartość potrzeby tak silnego rozumienia fizykalizmu, w tym momencie natomiast rozważmy, jak tak sformułowany eksperyment z zombie spełnia kryteria poprawności eksperymentów myślowych Beckera.

Funkcja epistemiczna: Eksperyment z zombie ma nas przekonać, że fizykalizm jest fałszywy, ponieważ nie wyjaśnia świadomości. Celem jest zatem stwierdzenie czegoś prawdziwego na temat świata.

Argumentacyjna niezbędność narracji: Narracja o zombie jest kluczowa dla argumentu – tylko wyobrażalne zombie prowadzić mają do problemu dla fizykalizmu. Narracja nie jest więc ilustracją, lecz kluczowym elementem argumentu.

Rozstrzygalność: Aby ocenić, czy zgadzamy się z konkluzją argumentu, musimy być w stanie wyobrazić sobie zombie. Ponadto narracja eksperymentu wprost odwołuje się do pojmowalności i możliwości zombie. W świetle rozdziału 5 widzimy, że intuicje wywoływane przez narracje są podatne na błędy poznawcze, takie jak błąd językowej istotności, i uproszczenia w konstrukcji modeli mentalnych. Potencjalny wpływ tych czynników jest kluczowy w ocenie pojmowalności filozoficznych zombie, szczególnie (jak zobaczymy) w kontekście tego, na ile pojęciowo, a na ile percepcyjnie jesteśmy w stanie wyobrazić sobie zombie. Jeśli nie możemy stworzyć spójnego modelu mentalnego świata zombie, możemy uznać, że eksperyment z zombie jest niepoprawnie skonstruowany. Tylko przystając na to, że zombie są pojmowalne, możemy rozważać argumentację Chalmersa do jej końca.

O rozumowaniu na temat zombie można myśleć na dwa sposoby: z narracją lub bez niej. W tym drugim przypadku mówi się zwykle o argumencie modalnym bądź argumencie z pojmowalności. Różnicę widać w przyjętych w literaturze rekonstrukcjach wnioskowania („P" to suma wszystkich prawd na temat fizyki, „Q" jest dowolną prawdą fenomenalną, „∧" reprezentuje „i", a „¬" to symbol oznaczający „nie", Chalmers, 2008, s. 448)[91]:

[91] Oba sformułowania są do znalezienia także w pracy Chalmersa (2003a).

(A) Wersja zachowująca narrację:

1. Istnienie zombie jest pojmowalne.
2. Jeżeli istnienie zombie jest pojmowalne, to istnienie zombie jest metafizycznie możliwe.
3. Jeśli istnienie zombie jest metafizycznie możliwe, to świadomość jest niefizyczna.

Konkluzja: Świadomość jest niefizyczna.

(B) Wersja pozbawiona elementów narracji:

4. To, że $P \wedge \neg Q$ jest pojmowalne.
5. Jeżeli $P \wedge \neg Q$ jest pojmowalne, to jest metafizycznie możliwe.
6. Jeśli jest metafizycznie możliwe, że $P \wedge \neg Q$, to materializm jest fałszywy.

Konkluzja: Materializm jest fałszywy.

Korzystając z nomenklatur z rozdziału 2, możemy stwierdzić, że obie rekonstrukcje mają strukturę dedukcyjną, a więc są „Typu 1" (Norton). Ponadto są one destruktywne (Brown), bo prowadzą do odrzucenia fizykalizmu. W przypadku podziału Sorensena argument z zombie to zaś przede wszystkim podważacz konieczności: podważa, że fizykalizm koniecznie tłumaczy świadomość.

Na pierwszy rzut oka może się wydawać, że (A) i (B) są równoznaczne, bo zombie są realizacją przecież $P \wedge \neg Q$: są fizyczne, ale nie zawierają treści fenomenalnych. Sformułowanie (B) jest jednak bardziej abstrakcyjne od (A), bo nie precyzuje, co ma realizować warunek $P \wedge \neg Q$: zombie, człowiek-imitacja, duplikat, klon, a może jakiś niecodzienny gatunek jeża. Do (B) można więc przyporządkować starsze niż zombie przykłady argumentu z pojmowalności. Jak wskazaliśmy w rozdziale poświęconym modelom mentalnym, sposób opisu sytuacji wyjątkowej ma bezpośrednie przełożenie na szczegółowość konstruowanego modelu, natomiast jak stwierdziliśmy, omawiając kwestię wiarygodności intuicji, obierany język potrafi mimowolnie wpływać na nasze intuicyjne oceny. Dlatego, analizując eksperyment myślowy, nie będziemy pomijać elementów narracyjnych i skorzystamy z pierwszej rekonstrukcji, pamiętając, że (A) i (B) mogą dla tych samych osób być przekonujące w różnym stopniu.

Czy jesteśmy w stanie stworzyć model poznawczy świata zombie i, jeśli tak, to czy faktycznie podważa on zdolność współczesnej nauki do wyjaśnienia świadomości? Prześledźmy wątpliwości, które wyrosły na kanwie tych dwóch pytań.

6.3 ZASTRZEŻENIA CO DO POJMOWALNOŚCI ZOMBIE

Jak omówiłem w rozdziale 4, konstrukcja modeli mentalnych opiera się na integracji informacji percepcyjnych i pojęciowych. W przypadku wyobrażania sobie zombie model mentalny wymaga integracji wyobrażenia identyczności fizycznej i psychologicznej z jednoczesnym brakiem fenomenalnego doświadczania. Utworzenie takiego modelu jest szczególnie trudne, ponieważ nie możemy (quasi)percepcyjnie wyobrazić sobie, *zobaczyć* czy *wczuć się* w taką istotę, skoro ma jej brakować doświadczeniowego komponentu jej stanów umysłowych. Jak zobaczyliśmy wcześniej, granice wyobraźni często wynikają z niedostatecznej precyzji pojęciowej i nadmiernego polegania na percepcyjnych aspektach modeli. Nasza początkowa intuicja na temat pojmowalności zombie może być więc tylko pozorna, stąd warto ostrożnie rozważyć wątpliwości dotyczące tej pierwszej przesłanki eksperymentu.

6.3.1 Słownik i przekonania a częściowy zombie

W eksperymencie z zombie pojmowalność jest zawarta w samym wnioskowaniu. Jeśli idea zombie jest niepojmowalna, to pierwsza przesłanka jest fałszywa, a sama narracja niepojmowalna vel niewyobrażalna. Zatem, negując pierwszą przesłankę, negujemy *de facto* nie jedynie przesłankę, ale i poprawność konstrukcji samego eksperymentu. Czy zatem idea zombie jest pojmowalna? Zdaniem niektórych filozofów nie (Braddon-Mitchell, 2003; Carruth, 2016; Dennett, 1995; Márton, 2019; Moody, 1994; Nigel, 1998)[92].

Pierwsza wątpliwość dotyczy języka w świecie zombie (Moody, 1994), a dokładniej asymetrii pomiędzy treścią naszego słownika a treścią słownika zombie. Choć praktyka językowa zombie pokrywałaby się z naszą, to terminy opisujące świadome przeżycia miałyby inne odniesienia. Gdy opowiadamy o intensywnym śnie, odnosimy się opisem zarówno do wspomnień po przebudzeniu, jak i samego procesu przeżywania. Zombie może mówić tylko o wspomnieniach. Podobnie byłoby z każdym terminem opisującym stan umysłowy, który my przeżywamy świadomie fenomenalnie. Jako że zombie nie doświadczają swoich stanów umysłowych w żaden fenomenalny sposób – nie jest im *jakoś*, kiedy je mają – język musi się różnić od naszego.

[92] Chalmers (2003a, s. 108–112, 117) nazywa krytyków pierwszej przesłanki eksperymentu materialistami typu A.

Brak symetrii może wydawać się mało istotny dla wyobrażalności zombie, bo efektywna różnica w znaczeniach terminów opisujących stany umysłowe w obu światach byłaby niewykrywalna (Flanagan i Polger, 1995). Nie jest to jednak specjalnie obiecujące odrzucenie kontrargumentu. Przypomnijmy, że nasza własna siatka pojęciowa, podobnie jak wiedza i przekonania, wpływa na treść tworzonych przez nas modeli mentalnych. To może utrudnić dostrzeżenie, że wyobrażenie sobie zombie oznacza również konieczność wyobrażenia sobie innych pojęć niż nasze. Nasze rozumienie „widzenia" czy „czucia" zawiera odniesienia do przeżyć świadomych. Skoro te przeżycia nie występują u zombie, to ich pojęcia są inne niż nasze. Jest to problem, bo nie bardzo wiadomo, jak zatem mogą być identyczne z nami pod względem psychologicznym, czyli być przekonane, że jest im *jakoś* być w świadomych stanach umysłowych, mimo że tak naprawdę im nie jest (Nigel, 1998). Skoro mają być identyczne psychologicznie, to muszą mieć więc w swoim słowniku pojęcie świadomości fenomenalnej. Ale w świecie zombie nie ma niczego, do czego mogłoby się takie pojęcie odnosić. To nie to samo, co mówienie o ulubionych smakach, które można przełożyć na kategorie funkcjonalne.

Analogiczna do słownikowej jest wątpliwość dotycząca przekonań fenomenalnych. Mając świadome przeżycia, nabywamy też na ich temat przekonań. Gdy wchodząc do pomieszczenia, zauważamy krzesło, nabywamy nie tylko przekonania „Tutaj znajduje się krzesło", ale także przekonania fenomenalnego „Widzę krzesło". Z oczywistych względów przekonań drugiego rodzaju nie będzie miał zombie wchodzący do pokoju. A skoro tak, nie sposób jednocześnie nazywać go psychologicznie identycznym (Shoemaker, 1975, 1999).

Zwolennik eksperymentu myślowego z zombie może, jak Chalmers (2003b), stwierdzić, że w świecie zombie nazwy dla fenomenalnych treści są po prostu puste, a przekonania jego mieszkańców na kwestie fenomenalne błędne. Nie wyjaśni w ten sposób, dlaczego w zombie miało w ogóle powstać przekonanie o nim samym, że ma świadomość fenomenalną, skoro tak naprawdę nigdy nie było mu *jakoś*, ale też nie musi tego wyjaśniać. W idei świata zombie nie chodzi przecież o bogatą historię powstania zombie, tylko o pojęciową spójność. Trudnością pozostaje jednak to, że jeśli zombie mogą mieć fałszywe przekonania na temat własnej świadomości, to być może my również jesteśmy zombie (Hoel, 2023, rozdz. 7; Rose, 2013). Jesteśmy całkowicie pewni swojej świadomości i zakładamy, iż ta pewność ma pokrycie w stanie świata. Zombie też jest przekonany, tylko że błędnie. W myśl założenia o identycznościach fizycznych, biologicznych itd. nasze naukowe dane

obserwacyjne nie różnią się nijak od tych w świecie zombie. Pozostaje tylko nasza pewność. Pytaniem otwartym jest to, czy nasza pewność jest warta więcej niż pewność zombie. Powrócimy do tej kwestii w dalszej części tekstu.

Inny argument przeciw pierwszej przesłance dotyczy częściowych zombie, naszych kopii pozbawionych tylko części świadomości fenomenalnej (Tye, 2006). Jeśli całkowite zombie są pojmowalne, to świadomość fenomenalna nie jest istotna dla działania świata, bo jej brak niczego nie zmienia, jest epifenomenalna (zob. Chalmers, 2003a, s. 105, 2010a, s. 269–270; Jackson, 1982, s. 133). Zatem to, czy świat zombie pozbawiliśmy jedynie części świadomości fenomenalnej czy całości, nie powinno mieć żadnego znaczenia.

Wyobraźmy sobie świat częściowych zombie, które nie mają świadomych przeżyć w czasie rzeczywistym, ale zachowują wspomnienia tych przeżyć. Takim zombie „nie jest czerwono", gdy patrzą na czerwone jabłko, lecz „jest im czerwono", gdy przypominają sobie czerwone przedmioty. Kiedy tylko jednak zaczną wspominać takie dawne sytuacje, od razu dostrzegą przecież różnicę między ich umysłowym stanem wspominania (z qualiami) a aktualnym stanem widzenia jabłka (bez qualiów). Częściowe zombie nie może więc być funkcjonalnie z nami identyczne, bo zareaguje inaczej na te dwa stany. Zauważy różnicę, której nie dostrzegają ani pełne zombie, ani my. Zwolennicy pojmowalności zombie nie powinni poprzestać na pojmowalności całkowitych zombie. Jeśli bowiem usunięcie części świadomości fenomenalnej jest niemożliwe, to istnieją prawidłowości rządzące ich posiadaniem – a przecież zwolennik pojmowalności zombie twierdzi, że żadnych takich prawidłowości nie ma.

Zarzut wobec częściowych zombie wydaje się jednak wynikać z nieporozumienia. Po pierwsze, nie jest jasne, dlaczego zwolennik idei zombie miałby traktować świadomość epifenomenalnie. Świadomość fenomenalna nie ma dla niego znaczenia jedynie dla interakcji czysto fizycznych. Chalmers (1996, rozdz. 6, 7 i 8) sam ostrożnie argumentuje, że świadomość jest fundamentalną cechą świata, podobnie jak grawitacja czy oddziaływania słabe, aczkolwiek nie obejmujemy świadomości, opisując świat fizykalnie. Po drugie, usunięcie niektórych doznań nie oznacza usunięcia świadomości. Częściowym zombie brakowałoby pewnych treści, ale co do zasady będą one zdolne do przeżyć fenomenalnych. Nie są więc zombie *per se*, bo posiadają świadomość fenomenalną, po prostu nie mają niektórych jej stanów. Nie mieć zaś dzwonka w rowerze to coś zupełnie innego niż nie mieć roweru. W końcu, jak już wspomnieliśmy, zwolennik eksperymentu może też odpowiedzieć,

że częściowy zombie po prostu będzie się mylić co do rzekomej fenomenalności niektórych swoich stanów.

6.3.2 Zombie niczego nie brakuje

Wróćmy do kwestii pewności, że nie jesteśmy zombie. Zdaniem niektórych poważne potraktowanie psychologiczno-behawioralnej identyczności zombie wymaga wyobrażenia sobie nas samych. Zombie mają naprawdę być w stanie dać się pochłonąć fantazjom seksualnym i unikać zachowań przynoszących ból, jeśli zaś „wyobrażasz je sobie inaczej, nie wyobrażasz ich sobie poprawnie" (Dennett, 1995, s. 324). Dla cytowanego tu Dennetta zombie musi mieć wszystkie nasze dyspozycje, więc różnica między nami a nim będzie wyłącznie pozorna. Analogicznie jedynie pozornie (*prima facie*) możemy wyobrazić sobie fizycznie, biologicznie, psychologicznie i behawioralnie identyczny z nami duplikat, który jednak pozbawiony jest zdrowia. Jeśli uznamy, że taki duplikat jest pojmowalny, to błędnie rozumiemy zdrowie nie jako poprawne funkcjonowanie ciała, lecz jako jakiś dodatkowy moduł, który może być obecny w ciele albo nie (Dennett, 1995, s. 325).

Dennett podchodzi do kwestii świadomości eliminatywistycznie, twierdząc, że świadomość fenomenalna nie istnieje. Nie jest dla niego problemem żyć w świecie zombie, bo jego zdaniem nie ma innego. Dennett zasłynął swoją krytyką qualiów jako nieopisywalnych, wewnętrznych, prywatnych i bezpośrednio uchwytnych jedynie w trakcie doświadczenia własności (Dennett, 1991, 2016). Zamiast bezpośredniego wskazywania na te cechy dziś częściej podkreśla się, że jest nam *jakoś* być w świadomych stanach umysłowych: jest nam *jakoś* widzieć czerwony (jak? czerwono), czuć radość (jak? radośnie), ból (jak? boleśnie) itd. Odwołanie do tego zwrotu jest dzisiaj wręcz standardem definiowania świadomości fenomenalnej[93]. Takie podejście ma rozwiązywać trudność podania bardziej szczegółowej charakterystyki, ale nie brakuje kwestionujących je filozofów, dziś nazywanych zwykle iluzjonistami (zob. Frankish, 2012, 2016; Irvine i Sprevak, 2020). Jeśli własności fenomenalne sprowadzimy wyłącznie do tego, że czynią różnicę w doświadczaniu stanów umysłowych – że jest nam *jakoś* je mieć – to wydaje się, że jedynym uzasadnieniem ich faktycznego istnienia pozostaje intuicja. Stąd wątpliwość

[93] „[…] jest nią [świadomością fenomenalną] własność posiadana przez stany, zdarzenia i procesy umysłowe wtedy i tylko wtedy, gdy jest jakoś ich podmiotom je mieć lub być w nich" (Kriegel, 2006, s. 58).

iluzjonistów jest uzasadniona: czy faktycznie sam tego rodzaju zwrot i zwracanie uwagi na własności fenomenalne nie są trikiem, w który wpędzili się sami filozofowie (Ozdemir, 2022; zob. Farrell, 2016; Liu, 2022; Owesen, 2023).

Zakładając, iż nie jesteśmy w świecie zombie, dlaczego ma nam się jedynie wydawać, że wyobrażamy sobie identyczne z nami istoty pozbawione świadomości fenomenalnej? Mylić mamy to wyobrażanie z innym, z wyobrażaniem sobie istot bez uwzględniania ich świadomości (Marcus, 2004). Możemy wyobrazić sobie świat bez pewnych elementów, jak pokój bez słonia, ale nie możemy wyobrazić sobie braku, którego nie sposób zaobserwować. Zamiast słonia możemy wyobrazić sobie podłogę ograniczoną z każdej ze stron ścianami – brak słonia oznacza obecność niezasłoniętej podłogi. Nie możemy jednak wyobrazić sobie naszego zombie bez świadomości fenomenalnej. Przecież z zewnątrz nic w świecie zombie się nie zmienia. Ciała naszych zombie są takie same jak nasze. Nie ma żadnej treści, którą moglibyśmy wyobrazić sobie w miejsce świadomości fenomenalnej. Wyobrażając sobie zombie, zatrzymujemy się więc na etapie wyobrażania sobie ich trzecioosobowych aspektów, ich ciał i zachowań – wyobrażamy sobie świat identyczny fizycznie z naszym bez wyobrażania sobie świadomości (Marcus, 2004, s. 487). Jeden mały krok wystarczy jednak, żeby pokazać sobie, że ten wyobrażony świat nie jest światem zombie. Wystarczy dodać wyobrażenie, jak to jest w tym identycznym fizycznie świecie naszym duplikatom smakować jabłko – z pewnością soczyście[94].

Powyższa argumentacja sugeruje percepcyjne rozumienie wyobraźni, które – jak wskazałem w rozdziale 3 – jest bardzo wąskie. Ignoruje podobieństwa funkcjonalne, podobieństwa neuronalnych mechanizmów różnych form myślenia na temat rzeczy nieobecnych w aktualnym środowisku oraz zależności odkryte u osób z zaburzeniami pamięci, prospekcji i z afantazją. Nawet jednak posługując się tym bogatszym rozumieniem wyobraźni, powyższa argumentacja jest wartościowa. Zdajemy sobie teraz sprawę, że narracja eksperymentu musi być wyobrażalna/pojmowalna po części percepcyjnie i pojęciowo. Ta część zombie, która odróżnia go od nas, podlega wyłącznie wyobrażeniu pojęciowemu. Prowadzi to do pytania – czy taki eksperyment

[94] Nieco podobnie w przypadku rozróżnienia między aktem wyobrażania sobie identycznego fizycznie świata a niewyobrażaniem sobie świadomości postępuje David Braddon-Mitchell (2003). Jego propozycja głosi, że zwolennicy eksperymentu z zombie mylą pojmowalność idei zombie z możliwością jej pojmowalności.

myślowy pomaga nam zrozumieć fenomenalne aspekty świadomości, czy też jedynie ilustruje granice naszej pojęciowej wyobraźni?

Wyobrażalność zombie jest więc zależna od pojęć posiadanych przez osobę przeprowadzającą eksperyment, a nie implicytnego rozumienia czasowo-przestrzennych prawidłowości w rozwoju sytuacji w środowisku. Wyobrażając sobie ciężką kulę spychaną ze szczytu wzgórza, wiem, że konsekwencją będzie znalezienie się tej kuli u podstawy wzgórza, nawet jeśli niewiele pamiętam z lekcji fizyki. Natomiast skutek braku świadomości fenomenalnej zależy od tego, jakie dokładnie mam pojęcie świadomości. Dualista co do relacji psychofizycznej, prawdopodobnie, mimowolnie rozumie inaczej świadomość niż konsekwentny materialista. Co zaś, jeśli w ogóle nie mamy pojęcia świadomości fenomenalnej? Czy udane przeprowadzenie eksperymentu myślowego z zombie wymaga wiedzy, którą można nazwać ekspercką?

Jak wskazałem w rozdziale 5, intuicje są szczególnie podatne na wpływ błędów poznawczych i efektów językowych, takich jak błąd językowej istotności. Dotyczy on mimowolnej aktywacji stereotypowych znaczeń słów w nieadekwatnych kontekstach i przekładania się tej aktywacji na osądy o intuicyjności twierdzeń. W ostatnich latach Eugen Fischer i Justin Sytsma (2021) przeprowadzili badanie pokazujące, jak skojarzenia wpływają na rozumienie eksperymentu z zombie. Rezultaty sugerują, że uznanie zombie za przynajmniej częściowo pojmowalne spowodowane jest właśnie błędem językowej istotności, wywoływanym dla słowa „zombie".

Dominujące rozumienie zombie zgodne jest – co mało zaskakujące – z filmowym obrazem ożywionych ciał. W największym korpusie współczesnego języka angielskiego, COCA (*Corpus of Contemporary American English*, M. Davies, 2008), zombie występuje przede wszystkim w kontekstach tyczących się nieumarłych, wampirów i innych postaci oraz motywów rodem z horrorów. Rozumienie zombie jako gnijących, ale nieumarłych, ożywionych ciał żywiących się ludzkim mięsem nie tylko dominuje w danych korpusowych, lecz także jest powszechne w swobodnych wypowiedziach badanych na temat zombie. W obu przypadkach filmowe znaczenie obecne jest w około 80% zdań[95].

Dlaczego takie rozumienie miałoby wpływać na ocenę pojmowalności filozoficznego zombie? Te ostatnie wyglądają inaczej niż filmowe – ich ciała są identyczne z naszymi, ciała filmowych zombie zaś ulegają rozkładowi.

[95] W przypadku korpusu wygenerowanych zostało 500 losowych zdań zawierających słowo „zombie".

Zachowują się także inaczej, filmowe atakują ludzi, filozoficzne natomiast robią dokładnie to samo co my. Oba rodzaje istot mają wspólną cechę: nie posiadają świadomości. Usłyszenie w kontekście filozoficznych zombie, że mają identyczne z naszymi ciała i wzorce zachowań, mimowolnie aktywuje w nas przypisanie świadomości, ale jednocześnie podawana jest informacja, że tej świadomości one nie posiadają. Sprzeczność między informacją o braku świadomości u filozoficznych zombie a naszym automatycznym przypisaniem im jej ze względu na ich podobieństwo do ludzi jest maskowana przez dominujące, filmowe rozumienie zombie, które są świadomości pozbawione. To sprawia, że badani łatwiej akceptują ideę zombie jako pojmowalną.

Kiedy w opisie (podobnym do narracji omawianego eksperymentu myślowego) użyto słowa „zombie", badani częściej przypisywali istotom cechy charakterystyczne dla filmowych zombie – w tym brak świadomości – niż gdy użyto neutralnego określenia „duplikat". Wyniki Fischera i Sytsmy (2021) sugerują, że błąd językowej istotności wyjaśnia około połowy przypadków, w których badani uznawali zombie za nieświadome. Osób uznających „duplikaty" za nieświadome było dwa razy mniej. Co więcej, w przypadku obu wariantów większość badanych uznawała, że istoty są świadome mimo narracji wprost stwierdzającej brak życia wewnętrznego.

Zwolennicy eksperymentu często zakładają, że intuicje na temat trudnego problemu świadomości są powszechne, a zatem, iż większość ludzi bez trudu uchwyci ideę zombie (Chalmers, 2018, 2020). Badania empiryczne o intuicjach zgodnych z trudnym problemem są dziś nieliczne i niejednoznaczne (Díaz, 2021; zob. Gregory i in., 2022). Jednakże badań dotyczących potocznych i ogólniejszych intuicji o świadomości jest coraz więcej i sugerują one raczej, że nie-filozofowie mogą nie dostrzegać fenomenalności świadomości jako takiej. Osoby takie nie znają oczywiście terminu *świadomość fenomenalna*, ale co ważniejsze, nie dokonują rozróżnień między stanami mentalnymi w sposób sugerujący, że posiadają intuicyjne pojęcie fenomenalności (Sytsma, 2012; Sytsma i Machery, 2010; Sytsma i Ozdemir, 2019). Prawda na temat potocznej psychologii świadomości jest z pewnością zniuansowana (Reuter, 2024), lecz również dane językowe sugerują ostrożność. Zaobserwowano chociażby, że w potocznym użyciu słowa „doświadczanie" (*experience*) brakuje takiego sensu, który mógłby objąć świadomość fenomenalną (Sytsma i Fischer, 2023).

Dlaczego dane na temat skojarzeń osób badanych powinniśmy traktować poważnie przy ocenie pierwszej przesłanki eksperymentu? Ukazują one

to, co wskazałem w rozdziale na temat modeli mentalnych: nie rozumujemy w izolacji. W konstrukcję i przekształcanie modeli mentalnych angażowane są nie tylko informacje wprost podane w narracji eksperymentu, ale także nasze przekonania, wiedza i posiadana siatka pojęciowa. O ile elementy te są nam świadomie dostępne, możemy je poddać refleksji. Tak też przebiegała dotychczasowa dyskusja na temat zombie – poszczególni filozofowie dostrzegali trudności w spójnym wyobrażeniu sobie sytuacji bądź jej konsekwencji dla materializmu. Pomijano jednak czynniki nieuświadamiane, a tym samym niedostępne świadomej kontroli. Zaobserwowanie błędu językowej istotności ujawnia, że także takie czynniki odgrywają rolę przy ocenie eksperymentu myślowego[96].

Uzależnienie skuteczności argumentu z zombie od wiedzy i pojęć osoby zapoznającej się z nim nie oznacza od razu, że jest on niewiarygodny. Jak zaznaczyłem w rozdziale 5, intuicje są podatne na uproszczenia i błędy poznawcze, szczególnie kiedy dotyczą kwestii zdroworozsądkowych i pojęciowych. Dla przeciwników eksperymentu myślowego wygodniejsze może być skupienie się na potocznej nieintuicyjności idei zombie, jednak ani narracja, ani struktura argumentu z pojmowalności nie zobowiązuje nas do polegania na myśleniu zdroworozsądkowym. Konkluzja eksperymentu dotyczy nie samych zombie, tylko fałszywości fizykalizmu. Wykorzystanie przekonań czy wiedzy zdroworozsądkowej może prowadzić do fałszywych wniosków, szczególnie w przypadku osób niemających pojęcia świadomości fenomenalnej. Być może dostrzeżenie fenomenalności naszego przeżywania wymaga takiej uważności skierowanej do wewnątrz naszego życia umysłowego, którą spontanicznie przejawia stosunkowo niewielu, a która jest wspierana przez trening z zakresu szeroko pojętej filozofii świadomości.

Ocena eksperymentu z zombie wymaga więc łączenia obu typów zasobów epistemicznych: filozoficznych i zdroworozsądkowych. W tym sensie argument z zombie może służyć za tester posiadanych intuicji zdroworozsądkowych, stopnia ich zgodności ze stanowiskami filozoficznymi. Jednak to, czy uznamy pojęcie świadomości fenomenalnej funkcjonujące w debacie filozoficznej za poprawnie opisujące to, jak treści świadomości się nam jawią – jako

[96] Zob. polemiki Davida Chalmersa, Keitha Frankisha, Michelle Liu i Edouarda Machery'ego wokół badania Fischera i Sytsmy, jak również odpowiedź autorów zamieszczoną na stronie *The Brains Blog* (https://philosophyofbrains.com/2021/09/21/symposium-on-fischer-sytsmas-zombie-intuitions.aspx, dostęp: 6.12.2024).

jakości dostępne jedynie w przeżyciu – wydaje się zależeć bardziej od myślenia eksperckiego. Dla pozytywnej oceny tego jawienia się nie potrzebujemy nieomylnej introspekcji, jeśli tę ostatnią rozumielibyśmy jako pewność co do przeżywanych treści. Wręcz przeciwnie, taka introspekcja często nas zawodzi (Schwitzgebel, 2008). Badacze przykładowo debatują, czy abstrakcyjnym myślom towarzyszą specyficzne własności fenomenalne, a wiele osób tak z akademii, jak i spoza niej myli się co do dokładnej treści stanów emocjonalnych, ze szczerym oburzeniem krzycząc: „Wcale nie jestem zdenerwowany!". Potrzebujemy natomiast być realistami fenomenalnymi, uznać, że świadomość jest dokładnie taka, jak nam się prezentuje: fenomenalna (G. Strawson, 2019). Realista fenomenalny zgodzi się, że możemy mylić się odnośnie do trafności treści świadomości, ale powie też, że wszystko, co odczuwamy jako świadome, jest istotnie świadome (Chalmers, 2010b, s. 149–150), nawet jeśli niepoprawnie nazywamy przedmiot, którego jesteśmy w danym momencie świadomi.

6.4 JEŚLI NAWET POJMOWALNE, TO I TAK NIEMOŻLIWE

Załóżmy, że zombie są pojmowalne. Czy oznacza to automatycznie, że są możliwe? Niektórzy argumentują, że podczas gdy pojmowalność zależy od naszego procesu poznawczego, to ocena możliwości co do relacji między świadomością a stanami fizycznymi wykracza poza nasze możliwości apriorycznego myślenia[97]. Tak jak wykracza poza aprioryczne ustalenie to, czy woda równa się H_2O. Potrzeba nam faktów na temat świata. A skoro zgodnie z nauką za stany świadomości muszą odpowiadać fizyczne stany mózgu, to istnienie zombie jest niemożliwe, podobnie jak niemożliwe jest to, aby woda nie była H_2O (Block, 2007; Block i Stalnaker, 1999; Hill, 1997; Hill i Mclaughlin, 1999). Pytanie jednak, czy nauka faktycznie to sugeruje? W 2023 roku 25-letni zakład – o skrzynkę wina – o to, czy nauka znajdzie biologiczne wyjaśnienie mechanizmu świadomości, wygrał sceptyczny Chalmers, a nie optymistyczny Christof Koch, jeden z pionierów i najważniejszych neurobadaczy świadomości (Lenharo, 2023). Istotnie, przy braku jednoznacznych kryteriów świadomości i zrozumienia, o jakiego poziomu zjawisku w zasadzie mówimy – biologicznym, fizycznym, psychologicznym – oraz przy niewielkim pojęciowym konsensusie pomiędzy najpoważniejszymi dziś empirycznymi

[97] Zwolenników odrzucenia drugiej przesłanki eksperymentu Chalmers nazywa materialistami typu B (2003a, s. 112–115).

teoriami świadomości (Francken i in., 2022), jest całkiem rozsądne być bardziej sceptycznym na temat świadomości niż na temat wody.

Popularną reakcją na drugą przesłankę jest tzw. strategia z pojęć fenomenalnych (Balog, 1999, 2012, 2019; Loar, 1997; Papineau, 2007; Tye, 2003). Nabywamy pojęć fenomenalnych w momencie przeżywania danego rodzaju treści i wykorzystujemy je do rozumienia stanów świadomych, na które introspekcyjnie zwracamy uwagę. Przykładowo, pojęcie fenomenalne bólu pozwala mi rozpoznać w moim przeżyciu, na czym polega owe wyjątkowe „jest mi *jakoś*", kiedy mnie boli. Pojęcia fenomenalne różnią się od typowych teoretycznych wykorzystywanych w naukach przyrodniczych – wprost wyrażają to, czego doznajemy. Można powiedzieć, że są one raczej jak znajomość samego przeżycia świadomego, niezależnie od zdolności do jego opisania. Ale choć pojęcia fenomenalne różnią się od teoretycznych, nie wynika z tego, że odnoszą się do czegoś niefizycznego. Własności stanów świadomych generujące w nas pojęcia fenomenalne mogą przecież być stanami mózgowymi, fizycznymi. Stąd, nawet jeśli idea zombie jest pojmowalna, nie oznacza to, że możliwe są niefizyczne własności. Luka eksplanacyjna, tym samym, ma charakter wyłącznie pojęciowy, a nie ontologiczny.

Zwróćmy jednak uwagę, że stwierdzenie, iż pojęcia fenomenalne są w jakimś silnym związku z pojęciami odnoszącymi się do tego, co fizyczne, jest najzwyklejszym założeniem. W przeżyciu stanów świadomych – nabywaniu pojęć fenomenalnych – takiego związku nie doświadczamy. Nie jest elementem przeżycia, jakobyśmy nie byli w stanie dostrzec jakiejś ukrytej, fizycznej natury stanów fenomenalnych. Wręcz przeciwnie, wydają się one kierować wprost na swoją jakościową, doświadczeniową naturę. Zatem, myśląc o pojęciach fenomenalnych, tym bardziej wskazujemy na nietypowość świadomości fenomenalnej względem tego, co fizyczne (Goff, 2011, 2017, w szczególności rozdz. 4 i 5)[98].

Kolejna wątpliwość głosi, że równie pojmowalne co zombie są antyzombie. Są to nasze fizyczne duplikaty, które są świadome fenomenalnie jedynie dzięki własnościom identycznym z naszymi fizycznymi własnościami (Frankish, 2007). Tak jak w przypadku zwykłych zombie odbieramy im

[98] Odwołałem się tutaj bardziej do toku myślenia Philipa Goffa niż samego Chalmersa, z uwagi na to, że odpowiedź tego drugiego na strategię z pojęć fenomenalnych i – ogólniej – akceptację pierwszej przesłanki eksperymentu przy jednoczesnym odrzuceniu drugiej jest w większym stopniu techniczna (zob. Chalmers, 2003a, s. 115–119, 2003b, 2010b; Polcyn, 2022).

świadomość fenomenalną, tak antyzombie pozbawiamy wszelkich domniemanych własności niefizycznych. Można zaryzykować nawet stwierdzenie, że antyzombie są silniej pojmowalne niż zombie, bo skoro nauka nie ma pewności, czy mieliśmy kiedykolwiek kontakt z własnościami niefizycznymi, to zasadniczo możemy sobie wyobrazić jako antyzombie zwykłego biologicznego człowieka. Czy jest on pozytywnie pojmowalny? Zdaniem orędowników antyzombie jest pojmowalny tak samo jak biologiczny i świadomy człowiek. Czy jest idealnie pojmowalny? Możemy być tego pewni co najwyżej w takim samym stopniu jak tego, że zwykły zombie jest idealnie pojmowalny. Antyzombie jest problemem dla zwolennika drugiej przesłanki eksperymentu z zombie, bo tylko jedno z dwojga może być na ten sam sposób możliwe: (1) wszystkie fakty fizyczne na temat świata sprawiają, że antyzombie są świadome, albo (2) wszystkie i te same fakty nie wystarczają, aby zombie były świadome. Ta na pierwszy rzut oka porównywalna pojmowalność zombie i antyzombie – zdaniem zwolenników tych drugich – sugeruje, że zwolennicy zombie już na starcie zakładają naturę własności fenomenalnych – jej niefizyczność – trywializując tym samym całe rozumowanie.

Nieprzypadkowo odnieśliśmy się w tym miejscu do idealnej i pozytywnej pojmowalności. Jak wskazano w rozdziale 5, według Chalmersa jeden typ pojmowalności daje nam najpewniejsze przejście do orzekania o możliwości idealnej pierwotnej pozytywnej pojmowalności. I choć nie jesteśmy idealnymi rozumującymi, to przecież niekiedy rozumujemy równie dobrze jak idealne istoty – nawet taka istota musiałaby stwierdzić, że $2 + 2 = 4$. W przeciwieństwie jednak, zdaniem Chalmersa, do zombie argumenty takie jak antyzombie (i inne podobne, które można zbiorczo określić argumentami z pojmowalności materializmu) są co najwyżej *prima facie* negatywnie pojmowalne (Chalmers, 2010b, s. 180). Między zombie a antyzombie jest mianowicie swoista asymetria (Prelević, 2015). Świat zombie jest scenariuszem otwartym, dopuszczającym niefizyczne wyjaśnienie świadomości, choćby poprzez istnienie duchów albo magii, świat antyzombie jest zaś zamknięty, w którym wszystko jest z założenia fizyczne, w tym świadomość. Wyobrażenie otwartego scenariusza jest w pewnym sensie łatwiejsze niż zamkniętego, bo podczas gdy pierwszy wymaga jedynie założenia braku jednego zjawiska w świecie, ten drugi wymaga poczynienia założenia na temat całej rzeczywistości. Pierwszy wychodzi od braku związku pomiędzy pojęciem świadomości fenomenalnej a pojęciami opisującymi świat fizyczny, ten drugi natomiast od głębokiego w filozoficznych konsekwencjach stanowiska metafizycznego.

Innymi słowy, antyzombie mają być co najwyżej *prima facie* negatywnie pojmowalne, gdyż bazują na mocnej deklaracji metafizycznej oraz założeniu lepszego zrozumienia związku między fizyką a świadomością, niż faktycznie posiadamy.

6.5 PRZYDATNOŚĆ EKSPERYMENTU MYŚLOWEGO Z ZOMBIE

Mimo licznych zastrzeżeń eksperyment myślowy z zombie odegrał i nadal odgrywa istotną rolę w filozofii umysłu i nauce o świadomości. Choć nie rozstrzyga jednoznacznie debaty nad naturą świadomości i jej relacji do świata fizycznego, wciąż znajduje zwolenników – przyjmuje go dziś mniej więcej co czwarty filozof akademicki i co piąty filozof umysłu (Bourget i Chalmers, 2023) – a jednocześnie stanowi doskonały tester zdroworozsądkowych intuicji. Trudno dziś wyobrazić sobie wprowadzenie do tematyki świadomości bez odwołania się do tego eksperymentu.

Idea zombie pomogła skupić uwagę badaczy na świadomości fenomenalnej i wyrazić związaną z nią trudność dla trzecioosobowego, fizykalnego opisu świata. Pomogła również zniuansować epistemologię modalną (granice pojmowalności i możliwości), jak i samą filozofię świadomości (dzisiejsza filozofia świadomości to w znacznej części filozofia świadomości fenomenalnej). Doprowadziła także do ożywienia monistycznych alternatyw dla fizykalizmu, takich jak panpsychizm i monizm neutralny[99], traktujących świadomość jako fundamentalny element rzeczywistości (Brüntrup, 2017; Goff, 2022; Seager, 2020). Czas dopiero oczywiście pokaże, czy jest to chwilowe zauroczenie alternatywami, czy też przejaw odczarowywania światopoglądu materialistycznego.

Sukces nauk przyrodniczych nie sprawił, że eksperymenty myślowe stały się mniej istotne, i zombie są tego świetnym przykładem. Choć Kuhn powiedziałby zapewne, że popularność zombie świadczy o kryzysie naukowego rozumienia świadomości, to mamy do czynienia ze stosunkowo młodą dziedziną, być może jeszcze w stanie przedparadygmatycznym (Hoel, 2023). Co najmniej od Kartezjusza filozofowie miewali przeczucie, że świadomość nie pasuje do trzecioosobowego, fizykalnego opisu świata. Jednak dopiero

[99] Dla tych pierwszych elementarna forma doświadczania, świadomości fenomenalnej, może być jedynym budulcem rzeczywistości – albo przynajmniej cechą tego budulca równie fundamentalną jak właściwości fizyczne. Dla tych drugich natomiast natura rzeczywistości nie jest ani fizykalna, ani doświadczeniowa, jedynie jawi się nam na te dwa sposoby.

z powstaniem nauki o świadomości pod koniec XX wieku idea zombie nabrała skrzydeł zarówno w popularności, jak i oddźwięku w środowiskach pozafilozoficznych. Obecnie jest ona jedną z bezpośrednich inspiracji dla wpływowych teorii naukowych (Koch, 2019; Tononi, 2012). Od tego dopiero też momentu, wyjścia poza wąskie ramy filozofii akademickiej, możemy powiedzieć, że świadomość fenomenalna jest problemem iście kognitywistycznym – angażującym badaczy z różnych dyscyplin poznawczych i wymagającym pomysłowości nie tylko eksperymentalnej, lecz również teoretycznej w obliczu nieustannie pojawiających się kolejnych problemów badawczych (Klawiter, 2004).

Eksperyment myślowy z zombie doskonale ilustruje zarówno metodologiczne wyzwania omawiane w poprzednich rozdziałach, jak i fundamentalne pytania dotyczące istoty naszej świadomości. Jego popularność wynika nie tylko z argumentacyjnej siły – jak pokazaliśmy, łatwo można zbłądzić w myśleniu na temat zombie – ale także tego, że prowokuje refleksję nad ograniczeniami naukowej metodologii w konfrontacji z problemami o rodowodzie filozoficznym. W tym właśnie zakresie argument z zombie odniósł ogromny sukces, podobnie jak inne eksperymenty myślowe z filozofii umysłu, jak chociażby argument z wiedzy (Gregory i in., 2022; Jackson, 1982). Sukces ten jest mierzalny – dziś ponad 60% filozofów akademickich jest przekonanych, że trudny problem świadomości istnieje, podobny odsetek odnotowano też wśród badaczy empirycznych (Bourget i Chalmers, 2023; Francken i in., 2022).

Osoba szczególnie mocno zauroczona sukcesami przyrodoznawstwa mogłaby zgryźliwie podsumować, że eksperymenty myślowe pomogły zachować tajemniczość świadomości mimo rodzącej się o niej nauki. Być może jednak potrzeba kolejnych *Gedankenexperimente*, aby tę tajemniczość rozwiać albo przynajmniej lepiej zrozumieć jej konsekwencje.

ZAKOŃCZENIE

Omawiając świat eksperymentów myślowych, staraliśmy się zrozumieć ich naturę, mechanizm działania oraz znaczenie dla ludzkiego myślenia. Okazały się narzędziem o uniwersalnym charakterze, o dużym potencjale zarówno poznawczym, jak i argumentacyjnym. Stąd nie dziwi, że eksperymenty myślowe towarzyszą ludziom od starożytności.

Zgodnie z proponowanym tu kognitywistycznym ich ujęciem są nie tylko układem przesłanek i wniosków, ale przede wszystkim procesem aktywnego konstruowania i przekształcania modeli mentalnych w umyśle. Te reprezentacje wyjątkowych sytuacji warunkowych pozwalają nam wyciągać wnioski i odkrywać nowe, nieoczywiste konsekwencje rozważanych scenariuszy.

Eksperymenty myślowe to nie tylko abstrakcyjne, intelektualne ćwiczenia – odgrywały i nadal odgrywają kluczową rolę w naszym zrozumieniu rzeczywistości. W nauce są używane do testowania teorii, tworzenia nowych hipotez, a także przewidywania wyników konkretnych eksperymentów empirycznych. W filozofii służą do analizy pojęć, rozwiązywania pozornych paradoksów i pogłębiania odpowiedzi na fundamentalne pytania dotyczące natury świata i umysłu. W literaturze pomagają lepiej zrozumieć motywacje, emocje i wybory bohaterów, jak również odkrywać nowe perspektywy interpretacyjne. Mogą być też stosowane wszędzie tam, gdzie pragniemy kogoś przekonać do naszych racji.

Jak każde narzędzie, eksperymenty myślowe mają swoje ograniczenia. Ich wiarygodność jest zależna od struktury wnioskowania i możliwości stworzenia modelu mentalnego, jego zgodności z naszą bądź ekspercką wiedzą na temat rzeczywistości, a także od posiadanych intuicji i siatki pojęciowej. Należy więc podchodzić do nich z krytycznym namysłem, analizując założenia, konsekwencje i potencjalne pozamerytoryczne czynniki wpływające na naszą ich ocenę. Dlatego pochylaliśmy się nad różnymi sposobami oceny wiarygodności eksperymentów myślowych. Ocena ta nie zawsze jest jednoznaczna i może zależeć od kontekstu, dziedziny wiedzy czy nawet indywidualnych

przekonań. Niemniej jednak analiza struktury eksperymentu myślowego, jego założeń i konsekwencji może zwiększyć świadomość mocnych oraz słabych stron naszego myślenia o danej, konkretnej sytuacji wyjątkowej. Tym samym możemy zrozumieć, czy narracja charakteryzuje faktycznie nietypową w świetle dostępnych nam informacji sytuację. Tę potrzebę wieloaspektowej analizy, pomocnej dla przemyślanej oceny wiarygodności eksperymentu myślowego, stosunkowo szeroko zilustrowaliśmy na przykładzie studium idei filozoficznych zombie. Eksperyment ten szczególnie wyraźnie pokazuje, jaką wagę może mieć myśl filozoficzna. Jeśli filozoficzny zombie jest empirycznie nieodróżnialny od świadomej istoty ludzkiej, to nie istnieje – i nie może istnieć – żaden obiektywny test pozwalający odróżnić świadomość od jej braku. Świadomość możemy rozpoznać jedynie subiektywnie – każdy z nas doświadcza siebie jako siebie. Nie jest to fakt empiryczny, ale trudno nam zakwestionować rzeczywistość naszej subiektywności. Jeśli uznamy jej realność, musimy odrzucić fizykalizm.

Eksperymenty myślowe to nie tylko narzędzie do rozwiązywania problemów, lecz także źródło inspiracji i kreatywności. Pozwalają nam na swobodne eksplorowanie przestrzeni możliwości, wykraczanie poza utarte schematy myślowe oraz odkrywanie nowych, nieoczywistych rozwiązań. Dlatego nie bójmy się samodzielnie eksperymentować w myślach, rozważać warunkowe scenariusze „Co jeśli…?" i poszukiwać na nie odpowiedzi. Pamiętajmy przy tym, że wartość poznawcza tego narzędzia zależy także od tego, czy nasz problemowy scenariusz faktycznie jest z perspektywy dostępnej nam wiedzy wyjątkowy i nietypowy, czy tylko niejasny i fikcyjny.

BIBLIOGRAFIA

Abraham, A. (2016). The Imaginative Mind: The Imaginative Mind. *Human Brain Mapping*, *37*(11), 4197–4211. https://doi.org/10.1002/hbm.23300

Abraham, A. (red.). (2020). *The Cambridge Handbook of the Imagination*. Cambridge University Press. https://doi.org/10.1017/9781108580298

Agostini, G. (2019). Projecting Spaces of Thought: The Geometrical Figure in the Works of Samuel Beckett and Julio Cortázar. W: F. Bornmüller, J. Franzen, M. Lessau (red.), *Literature as Thought Experiment? Perspectives from Philosophy and Literary Studies* (s. 139–154). Brill | Fink.

Ajdukiewicz, K. (1985). Klasyfikacja rozumowań – referat wygłoszony na I Konferencji Logików odbytej w Warszawie w dniach 13–15 XII 1952. W: K. Ajdukiewicz, *Język i poznanie. Tom II* (s. 206–225). PWN.

Alquist, J.L., Ainsworth, S.E., Baumeister, R.F., Daly, M., Stillman, T.F. (2015). The Making of Might-Have-Beens: Effects of Free Will Belief on Counterfactual Thinking. *Personality and Social Psychology Bulletin*, *41*(2), 268–283. https://doi.org/10.1177/0146167214563673

Andersen, R. (2012). *Has Physics Made Philosophy and Religion Obsolete?* The Atlantic. https://www.theatlantic.com/technology/archive/2012/04/has-physics-made-philosophy-and-religion-obsolete/256203/

Andrews-Hanna, J.R., Grilli, M.D. (2021). Mapping the Imaginative Mind: Charting New Paths Forward. *Current Directions in Psychological Science*, *30*(1), 82–89. https://doi.org/10.1177/0963721420980753

Andrews-Hanna, J.R., Reidler, J.S., Sepulcre, J., Poulin, R., Buckner, R.L. (2010). Functional-Anatomic Fractionation of the Brain's Default Network. *Neuron*, *65*(4), 550–562. https://doi.org/10.1016/j.neuron.2010.02.005

Andrews-Hanna, J.R., Smallwood, J., Spreng, R.N. (2014). The Default Network and Self-Generated Thought: Component Processes, Dynamic Control, and Clinical Relevance: The Brain's Default Network. *Annals of the New York Academy of Sciences*, *1316*(1), 29–52. https://doi.org/10.1111/nyas.12360

Antonutti Marfori, M. (2010). Informal Proofs and Mathematical Rigour. *Studia Logica*, *96*(2), 261–272. https://doi.org/10.1007/s11225-010-9280-4

Arana, A. (2016). Imagination in Mathematics. W: A. Kind (red.), *The Routledge Handbook of Philosophy of Imagination* (s. 463–477). Routledge.

Arcangeli, M. (2020). The Two Faces of Mental Imagery. *Philosophy and Phenomenological Research*, *101*(2), 304–322. https://doi.org/10.1111/phpr.12589

Arthur, R.T.W. (2018). Thought Experiments in Newton and Leibniz. W: M.T. Stuart, Y. Fehige, J.R. Brown (red.), *The Routledge Companion to Thought Experiments* (s. 111–127). Routledge.

Arystoteles. (1980). *O Niebie*, tłum. P. Siwek. Państwowe Wydawnictwo Naukowe.

Aspect, A., Grangier, P., Roger, G. (1982). Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities. *Physical Review Letters*, *49*(2), 91–94.

Bacon, A.M., Walsh, C.R., Martin, L. (2013). Fantasy Proneness and Counterfactual Thinking. *Personality and Individual Differences*, *54*(4), 469–473. https://doi.org/10.1016/j.paid.2012.09.023

Baddeley, A. (2003). Working Memory: Looking Back and Looking Forward. *Nature Reviews Neuroscience*, *4*(10), 829–839. https://doi.org/10.1038/nrn1201

Balcerak Jackson, M. (2016). On the Epistemic Value of Imagining, Supposing, and Conceiving. W: A. Kind, P. Kung (red.), *Knowledge Through Imagination* (s. 41–60). Oxford University Press.

Balog, K. (1999). Conceivability, Possibility, and the Mind-Body Problem. *The Philosophical Review*, *108*(4), 497–528. https://doi.org/10.2307/2998286

Balog, K. (2012). Acquaintance and the Mind–Body Problem. W: S. Gozzano, C.S. Hill (red.), *New Perspectives on Type Identity* (s. 16–42). Cambridge University Press.

Balog, K. (2019). Hard, Harder, Hardest. W: A. Sullivan (red.), *Sensations, Thoughts, Language. Essays in Honour of Brian Loar* (s. 265–288). Taylor and Francis.

Beck, S.R., Weisberg, D.P., Burns, P., Riggs, K.J. (2014). Conditional Reasoning and Emotional Experience: A Review of the Development of Counterfactual Thinking. *Studia Logica*, *102*(4), 673–689. https://doi.org/10.1007/s11225-013-9508-1

Becker, A. (2018). Thought Experiments in Plato. W: M.T. Stuart, Y. Fehige, J.R. Brown (red.), *The Routledge Companion to Thought Experiments* (s. 44–56). Routledge.

Bermúdez, J.L. (2020). *Cognitive Science: An Introduction to the Science of the Mind* (wyd. 3). Cambridge University Press.

Berniūnas, R., Dranseika, V. (2016). Folk Concepts of Person and Identity: A Response to Nichols and Bruno. *Philosophical Psychology*, *29*(1), 96–122. https://doi.org/10.1080/09515089.2014.986325

Bird, A., Tobin, E. (2023). Natural Kinds. W: E.N. Zalta, U. Nodelman (red.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Birdsell, D.S., Groarke, L. (2007). Outlines of a Theory of Visual Argument. *Argumentation and Advocacy*, *43*(3–4), 103–113.

Blanchette, I., Caparos, S. (2013). When Emotions Improve Reasoning: The Possible Roles of Relevance and Utility. *Thinking & Reasoning*, *19*(3–4), 399–413. https://doi.org/10.1080/13546783.2013.791642

Blanchette, I., Richards, A. (2004). Reasoning About Emotional and Neutral Materials: Is Logic Affected by Emotion? *Psychological Science*, *15*(11), 745–752. https://doi.org/10.1111/j.0956-7976.2004.00751.x

Block, N. (1995). On a Confusion about a Function of Consciousness. *Behavioral and Brain Sciences*, *18*(2), 227–247. https://doi.org/10.1017/S0140525X00038188

Block, N. (2007). Max Black's Objection to Mind-Body Identity. W: T. Alter, S. Walter (red.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (s. 249–306). Oxford University Press.

Block, N., Stalnaker, R. (1999). Conceptual Analysis, Dualism, and the Explanatory Gap. *The Philosophical Review*, *108*(1), 1–46. https://doi.org/10.2307/2998259

Blok, S., Newman, G., Rips, L.J. (2005). Individuals and Their Concepts. W: W. Ahn, R.L. Goldstone, B.C. Love, A.B. Markman, P. Wolff (red.), *Categorization Inside and Outside the Laboratory: Essays in Honor of Douglas L. Medin* (s. 127–149). American Psychological Association.

Bornmüller, F., Lessau, M. (2019). Are Thought-Experiments Always Arguments and Does Literature Literally Re-present? W: F. Bornmüller, J. Franzen, M. Lessau (red.), *Literature as Thought Experiment? Perspectives from Philosophy and Literary Studies* (s. 43–56). Brill | Fink.

Boudry, M., Pigliucci, M. (red.). (2018). *Science Unlimited?: The Challenges of Scientism*. University of Chicago Press.

Bourget, D., Chalmers, D. (2023). Philosophers on Philosophy: The 2020 PhilPapers Survey. *Philosophers' Imprint*, *23*(11). https://doi.org/10.3998/phimp.2109

Braddon-Mitchell, D. (2003). Qualia and Analytical Conditionals. *The Journal of Philosophy*, *100*(3), 111–135. http://www.jstor.org/stable/3655591

Brecevic, C. (2021). The Role of Imagination in Ernst Mach's Philosophy of Science: A Biologico-economical View. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, *11*(1), 241–261. https://doi.org/10.1086/712974

Bremer, J. (2013). Pojęcie „bólu" zdobyłeś wraz z językiem. *Przegląd Filozoficzny – Nowa Seria*, *2*(86), 401–416.

Broomhall, A.G., Phillips, W.J., Hine, D.W., Loi, N.M. (2017). Upward Counterfactual Thinking and Depression: A Meta-Analysis. *Clinical Psychology Review*, *55*, 56–73. https://doi.org/10.1016/j.cpr.2017.04.010

Brown, J.R. (2004). Why Thought Experiments Transcend Empiricism. W: C. Hitchcock (red.), *Contemporary Debates in Philosophy of Science* (s. 23–43). Blackwell Publishing Ltd.

Brown, J.R. (2011). *The Laboratory of the Mind. Thought Experiments in the Natural Sciences* (wyd. 2). Routledge.

Brown, J.R., Fehige, Y. (2022). Thought Experiments. W: E.N. Zalta (red.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Brüntrup, G. (red.). (2017). *Panpsychism: Contemporary Perspectives*. Oxford University Press.

Buckwalter, W. (2022). The Replication Crisis and Philosophy. *Philosophy and the Mind Sciences*, *3*. https://doi.org/10.33735/phimisci.2022.9193

Buckwalter, W., Stich, S. (2014). Gender and Philosophical Intuition. W: J. Knobe, S. Nichols (red.), *Experimental Philosophy: Volume 2* (s. 307–346). Oxford University Press.

Butler, T. (2003). Empeiria in Aristotle. *The Southern Journal of Philosophy*, *41*(3), 329–350. https://doi.org/10.1111/j.2041-6962.2003.tb00956.x

Butterfield, H. (1958). *The Origins of Modern Science 1300–1800*. G. Bell and Sons.

Buzzoni, M. (2018a). Pierre Duhem and Ernst Mach on Thought Experiments. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, *8*(1), 1–27. https://doi.org/10.1086/695720

Buzzoni, M. (2018b). Kantian Accounts of Thought Experiments. W: M.T. Stuart, Y. Fehige, J.B. Brown (red.), *The Routledge Companion to Thought Experiments* (s. 327–341). Routledge.

Byrd, N. (2023). Great Minds Do Not Think Alike: Philosophers' Views Predicted by Reflection, Education, Personality, and Other Demographic Differences. *Review of Philosophy and Psychology*, *14*(2), 647–684. https://doi.org/10.1007/s13164-022-00628-y

Byrne, R.M.J. (2005). *The Rational Imagination*. The MIT Press.

Byrne, R.M.J. (2018). Counterfactual Reasoning and Imagination. W: L.J. Ball, V.A. Thompson (red.), *The Routledge International Handbook of Thinking and Reasoning* (s. 71–87). Routledge.

Byrne, R.M.J., Johnson-Laird, P.N. (1989). Spatial Reasoning. *Journal of Memory and Language*, *28*(5), 564–575. https://doi.org/10.1016/0749-596X(89)90013-2

Camerer, C.F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B.A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., …, Wu, H. (2018). Evaluating the Replicability of Social Science Experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.-R., Sirigu, A. (2004). The Involvement of the Orbitofrontal Cortex in the Experience of Regret. *Science*, *304*(5674), 1167–1170. https://doi.org/10.1126/science.1094550

Camilleri, K. (2014). Toward a Constructivist Epistemology of Thought Experiments in Science. *Synthese*, *191*(8), 1697–1716. https://doi.org/10.1007/s11229-013-0358-1

Campbell, K. (1970). *Body and Mind*. The Macmillan Press Ltd.

Cappelen, H. (2012). *Philosophy without Intuitions*. Oxford University Press.

Cappelen, H., Gendler, T.S., Hawthorne, J. (red.). (2016). *The Oxford Handbook of Philosophical Methodology*. Oxford University Press.

Carnap, R. (2000). Język fizykalny jako uniwersalny język nauki. W: A. Koterski (red.), *Spór o zdania protokolarne. Erkenntnis i Analysis 1932–1940* (s. 7–38). Fundacja Aletheia.

Carroll, N. (2002). The Wheel of Virtue: Art, Literature, and Moral Knowledge. *Journal of Aesthetics and Art Criticism*, *60*(1), 3–26. https://doi.org/10.1111/1540-6245.00048

Carruth, A. (2016). Powerful Qualities, Zombies and Inconceivability. *The Philosophical Quarterly*, *66*(262), 25–46. https://doi.org/10.1093/pq/pqv055

Chalmers, D.J. (1995). Facing up to the Problem of Consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.

Chalmers, D.J. (1996). *The Conscious Mind. In Search of a Fundamental Theory*. Oxford University Press.

Chalmers, D.J. (2002). Does Conceivability Entail Possibility? W: T.S. Gendler, J. Hawthorne (red.), *Conceivability and Possibility* (s. 145–200). Oxford University Press. https://consc.net/papers/conceivability.html

Chalmers, D.J. (2003a). Consciousness and Its Place in Nature. W: S.P. Stich, T.A. Warfield (red.), *The Blackwell Guide to Philosophy of Mind* (s. 102–142). Blackwell Publishing.

Chalmers, D.J. (2003b). The Content and Epistemology of Phenomenal Belief. W: Q. Smith, A. Jokic (red.), *Consciousness: New Philosophical Perspectives* (s. 220–272). Oxford University Press.

Chalmers, D.J. (2008). Świadomość i jej miejsce w naturze. W: M. Miłkowski, R. Poczobut (red.), *Analityczna metafizyka umysłu. Najnowsze kontrowersje* (s. 442–494). Wydawnictwo IFiS PAN.

Chalmers, D.J. (2010a). *Świadomy umysł. W poszukiwaniu teorii fundamentalnej*. Wydawnictwo Naukowe PWN.

Chalmers, D.J. (2010b). The Two-Dimensional Argument Against Materialism. W: D.J. Chalmers, *The Character of Consciousness* (s. 141–205). Oxford University Press.

Chalmers, D.J. (2018). The Meta-Problem of Consciousness. *Journal of Consciousness Studies*, *25*(9–10), 6–61.

Chalmers, D.J. (2020). Is the Hard Problem of Consciousness Universal? *Journal of Consciousness Studies*, *27*(5–6), 227–257.

Chapman, S. (2017). The Experimental and the Empirical: Arne Naess' Statistical Approach to Philosophy. *British Journal for the History of Philosophy*, *26*(5), 961–981. https://doi.org/10.1080/09608788.2017.1336075

Cheke, L.G., Clayton, N.S. (2010). Mental Time Travel in Animals. *WIREs Cognitive Science*, *1*(6), 915–930. https://doi.org/10.1002/wcs.59

Churchland, P.M. (1981). Eliminative Materialism and Propositional Attitudes. *The Journal of Philosophy*, *78*(2), 67–90.

Clarke-Doane, J. (2021). Metaphysical and Absolute Possibility. *Synthese*, *198*(8), 1861–1872. https://doi.org/10.1007/s11229-019-02093-0

Cleve, J. van (1983). Conceivability and the Cartesian Argument for Dualism. *Pacific Philosophical Quarterly*, *64*(1), 35–45. https://doi.org/10.1111/j.1468-0114.1983.tb00182.x

Climenhaga, N. (2018). Intuitions are Used as Evidence in Philosophy. *Mind*, *127*(505), 69–104. https://doi.org/10.1093/mind/fzw032

Cocquyt, C.M., Palombo, D.J. (2023). Emotion in the Mind's Eye: Imagination for Adaptive Cognition. *Annals of the New York Academy of Sciences*, *1526*(1), 59–72. https://doi.org/10.1111/nyas.15011

Colaço, D., Machery, E. (2017). The Intuitive Is a Red Herring. *Inquiry. An Interdisciplinary Journal of Philosophy*, *60*(4), 403–419. https://doi.org/10.1080/0020174X.2016.1220638

Cole, D. (1984). Thought and Thought Experiments. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *45*(3), 431–444. http://www.jstor.org/stable/4319683

Cole, D. (2023). The Chinese Room Argument. W: E.N. Zalta, U. Nodelman (red.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Corcilius, K. (2018). Aristotle and Thought Experiments. W: M.T. Stuart, Y. Fehige, J.R. Brown (red.), *The Routledge Companion to Thought Experiments* (s. 57–76). Routledge. https://doi.org/10.4324/9781315175027-4

Cova, F., Strickland, B., Abatista, A., Allard, A., Andow, J., Attie, M., Beebe, J., Berniūnas, R., Boudesseul, J., Colombo, M., Cushman, F., Diaz, R., Dongen, N.N.N. van, Dranseika, V., Earp, B., Torres, A.G., Hannikainen, I., Hernández-Conde, J.V., Hu, W., …, Zhou, X. (2021). Estimating the Reproducibility of Experimental Philosophy. *Review of Philosophy and Psychology*, *12*, 9–44. https://doi.org/10.1007/s13164-018-0400-9

Craik, K. (1943). *The Nature of Explanation*. Cambridge University Press.

Crane, T. (2000). The Origins of Qualia. W: T. Crane, S. Patterson (red.), *History of the Mind-Body Problem* (s. 177–202). Routledge.

Crease, R.P. (2003). *The Prism and the Pendulum: The Ten Most Beautiful Experiments in Science.* Random House.

Crespi, B., Leach, E., Dinsdale, N., Mokkonen, M., Hurd, P. (2016). Imagination in Human Social Cognition, Autism, and Psychotic-Affective Conditions. *Cognition*, *150*, 181–199. https://doi.org/10.1016/j.cognition.2016.02.001

Crick, F. (1994). *Astonishing Hypothesis: The Scientific Search for the Soul*. Macmillan Publishing Company.

Crick, F. (1997). *Zdumiewająca hipoteza, czyli nauka w poszukiwaniu duszy*. Prószyński i S-ka.

Crick, F., Koch, C. (1990). Towards a Neurobiological Theory of Consciousness. *Seminars in the Neurosciences*, *2*, 263–275.

Currie, G., Ravenscroft, I. (2002). *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford University Press.

Dance, C.J., Ipser, A., Simner, J. (2022). The Prevalence of Aphantasia (Imagery Weakness) in the General Population. *Consciousness and Cognition*, *97*, 103243. https://doi.org/10.1016/j.concog.2021.103243

Davey, C.G., Pujol, J., Harrison, B.J. (2016). Mapping the Self in the Brain's Default Mode Network. *NeuroImage*, *132*, 390–397. https://doi.org/10.1016/j.neuroimage.2016.02.022

Davidson, D. (1992). Zdarzenia mentalne. W: D. Davidson, *Eseje o prawdzie, języku i umyśle* (s. 163–193). Wydawnictwo Naukowe PWN.

Davies, D. (2018). Art and Thought Experiments. W: M.T. Stuart, Y. Fehige, J.R. Brown (red.), *The Routledge Companion to Thought Experiments* (s. 512–525). Routledge.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 560 Million Words, 1990-Present*. https://corpus.byu.edu/coca/

Dawes, A.J., Keogh, R., Robuck, S., Pearson, J. (2022). Memories with a Blind Mind: Remembering the Past and Imagining the Future with Aphantasia. *Cognition*, *227*, 105192. https://doi.org/10.1016/j.cognition.2022.105192

De Neys, W. (red.). (2018). *Dual Process Theory 2.0*. Routledge.

De Neys, W. (2023). Advancing Theorizing about Fast-and-Slow Thinking. *Behavioral and Brain Sciences*, *46*, e111. https://doi.org/10.1017/S0140525X2200142X
De Neys, W., Pennycook, G. (2019). Logic, Fast and Slow: Advances in Dual-Process Theorizing. *Current Directions in Psychological Science*, *28*(5), 503–509. https://doi.org/10.1177/0963721419855658
Dehaene, S. (2023). *Świadomość i mózg. Odczytywanie kodu naszych myśli*. Copernicus Center Press.
Dennett, D.C. (1991). *Consciousness Explained*. Little, Brown and Company.
Dennett, D.C. (1995). The Unimagined Preposterousness of Zombies: Commentary on Moody, Flanagan, and Polger. *Journal of Consciousness Studies*, *2*(4), 322–325.
Dennett, D.C. (1997). *Natura umysłów*. CiS.
Dennett, D.C. (2015). *Dźwignie wyobraźni i inne narzędzia do myślenia*. Copernicus Center Press.
Dennett, D.C. (2016). *Świadomość*. Copernicus Center Press.
Dere, E., Dere, D., Souza Silva, M.A. de, Huston, J.P., Zlomuzica, A. (2018). Fellow Travellers: Working Memory and Mental Time Travel in Rodents. *Behavioural Brain Research*, *352*, 2–7. https://doi.org/10.1016/j.bbr.2017.03.026
Descartes, R. (1948). *Medytacje o pierwszej filozofii*, tłum. M. Ajdukiewicz. Polska Akademia Umiejętności.
Descartes, R. (2001). *Zasady filozofii*, tłum. I. Dąbska. Wydawnictwo Marek Derewiecki.
Descartes, R. (2005). *Medytacje o filozofii pierwszej*, tłum. J. Hartman. Zielona Sowa.
Descartes, R. (2012). *Rozprawa o metodzie dobrego powodowania swoim rozumem i szukania prawdy w naukach*, tłum. T. Boy-Żeleński. Fundacja Nowoczesna Polska. https://wolnelektury.pl/media/book/pdf/rozprawa-o-metodzie.pdf
Deutsch, M. (2015). *The Myth of the Intuitive. Experimental Philosophy and Philosophical Method*. The MIT Press.
Díaz, R. (2021). Do People Think Consciousness Poses a Hard Problem? Empirical Evidence on the Meta-Problem of Consciousness. *Journal of Consciousness Studies*, *28*(2–4), 55–75.
Dolan, R.J. (2002). Emotion, Cognition, and Behavior. *Science*, *298*(5596), 1191–1194. https://doi.org/10.1126/science.1076358
D'Olimpio, L. (2019). *Ethics Explainer: Plato's Cave*. The Ethics Centre. https://ethics.org.au/ethics-explainer-platos-cave/
Drake, S. (2003). *Galileo at Work: His Scientific Biography*. Dover Publishing.
Dranseika, V., Nichols, S., Strohminger, N. (2023). Which Kind of Sameness? Disambiguating Two Senses of Identity with a Novel Linguistic Task. *Cognition*, *238*, 105545. https://doi.org/10.1016/j.cognition.2023.105545
Duncan, S. (2012). Leibniz's Mill Argument Against Materialism. *Philosophical Quarterly*, *62*(247), 250–272. https://doi.org/10.1111/j.1467-9213.2011.00017.x
Eckardt, B. von (1995). *What Is Cognitive Science?* The MIT Press.
Edelman, G.M. (1989a). *Neural Darwinism: The Theory of Neuronal Group Selection*. Oxford University Press.

Edelman, G.M. (1989b). *The Remembered Present: A Biological Theory of Consciousness*. Basic Books.

Egan, D. (2016). Literature and Thought Experiments. *The Journal of Aesthetics and Art Criticism*, *74*(2), 139–150. https://doi.org/10.1111/jaac.12270

Egler, M., Ross, L.D. (2020). Philosophical Expertise under the Microscope. *Synthese*, *197*(3), 1077–1098. https://doi.org/10.1007/s11229-018-1757-0

Einstein, A. (1951). Autobiographical Notes. W: P.A. Schilpp (red.), *Albert Einstein-Philosopher Scientist* (wyd. 2, s. 2–95). Tudor Publishing.

Einstein, A., Infeld, L. (1967). *Evolution of Physics*. Touchstone.

Einstein, A., Podolsky, B., Rosen, N. (1935). Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Physical Review*, *47*, 777–780.

Elgin, C.Z. (1993). Understanding: Art and Science. *Synthese*, *95*, 13–28.

Elgin, C.Z. (2007). The Laboratory of the Mind. W: W. Huemer, J. Gibson, L. Pocci (red.), *A Sense of the World: Essays on Fiction, Narrative and Knowledge* (s. 43–54). Routledge.

Elgin, C.Z. (2014). Fiction as Thought Experiment. *Perspectives on Science*, *22*(2), 221–241. https://doi.org/10.1162/POSC_a_00128

Emery, N.J., Clayton, N.S. (2004). The Mentality of Crows: Convergent Evolution of Intelligence in Corvids and Apes. *Science*, *306*(5703), 1903–1907. https://doi.org/10.1126/science.1098410

Errington, T.M., Denis, A., Perfito, N., Iorns, E., Nosek, B.A. (2021). Reproducibility in Cancer Biology: Challenges for Assessing Replicability in Preclinical Cancer Biology. *eLife*, *10*, e67995. https://doi.org/10.7554/eLife.67995

Espino, O., Santamaría, C., Byrne, R.M.J. (2009). People Think about What Is True for Conditionals, Not What Is False: Only True Possibilities Prime the Comprehension of „If". *Quarterly Journal of Experimental Psychology*, *62*(6), 1072–1078. https://doi.org/10.1080/17470210802602375

Evans, J.S.B.T. (2017). *Thinking and Reasoning: A Very Short Introduction*. Oxford University Press.

Farrell, J. (2016). „What It Is Like" Talk Is Not Technical Talk. *Journal of Consciousness Studies*, *23*(9–10), 50–65.

Fehige, Y. (2023). *Thought Experiments, Science, and Theology*. Brill.

Fein, O., Yeari, M., Giora, R. (2015). On the Priority of Salience-Based Interpretations: The Case of Sarcastic Irony. *Intercultural Pragmatics*, *12*(1), 1–32. https://doi.org/10.1515/ip-2015-0001

Ferrante, D., Girotto, V., Stragà, M., Walsh, C. (2013). Improving the Past and the Future: A Temporal Asymmetry in Hypothetical Thinking. *Journal of Experimental Psychology: General*, *142*(1), 23–27. https://doi.org/10.1037/a0027947

Fischer, A. (2019). Guinea Pigs in a Terrarium? Albert Camus' The Plague as a Thought Experiment. W: F. Bornmüller, J. Franzen, M. Lessau (red.), *Literature as Thought Experiment? Perspectives from Philosophy and Literary Studies* (s. 155–169). Brill | Fink.

Fischer, E., Engelhardt, P.E. (2016). Intuitions' Linguistic Sources: Stereotypes, Intuitions and Illusions. *Mind & Language*, *31*(1), 67–103. https://doi.org/10.1111/mila.12095

Fischer, E., Engelhardt, P.E. (2017). Stereotypical Inferences: Philosophical Relevance and Psycholinguistic Toolkit. *Ratio*, *30*(4), 411–442. https://doi.org/10.1111/rati.12174

Fischer, E., Engelhardt, P.E. (2019a). Eyes as Windows to Minds: Psycholinguistics for Experimental Philosophy. W: E. Fischer, M. Curtis (red.), *Methodological Advances in Experimental Philosophy* (s. 43–100). Bloomsbury Academic.

Fischer, E., Engelhardt, P.E. (2019b). Lingering Stereotypes: Salience Bias in Philosophical Argument. *Mind and Language*, *35*(4), 415–439. https://doi.org/10.1111/mila.12249

Fischer, E., Sytsma, J. (2021). Zombie Intuitions. *Cognition*, *215*, 104807. https://doi.org/10.1016/j.cognition.2021.104807

Flanagan, O.J., Polger, T. (1995). Zombies and the Function of Consciousness. *Journal of Consciousness Studies*, *2*(4), 313–321.

Fletcher, S.C., Knobe, J., Wheeler, G., Woodcock, B.A. (2021). Changing Use of Formal Methods in Philosophy: Late 2000s vs. Late 2010s. *Synthese*, *199*(5), 14555–14576. https://doi.org/10.1007/s11229-021-03433-9

Foot, P. (1967). The Problem of Abortion and the Doctrine of the Double Effect. *Oxford Review*, *5*, 5–15.

Francken, J., Beerendonk, L., Molenaar, D., Fahrenfort, J.J., Kiverstein, J., Seth, A., Gaal, S. van (2022). An Academic Survey on Theoretical Foundations, Common Assumptions and the Current State of the Field of Consciousness Science. *Neuroscience of Consciousness*, *2022*(1), niac011. https://doi.org/10.1093/nc/niac011

Frankish, K. (2007). The Anti-Zombie Argument. *The Philosophical Quarterly*, *57*(229), 650–666. https://doi.org/10.1111/j.1467-9213.2007.510.x

Frankish, K. (2012). Quining Diet Qualia. *Consciousness and Cognition*, *21*(2), 667–676. https://doi.org/10.1016/j.concog.2011.04.001

Frankish, K. (2016). Illusionism as a Theory of Consciousness. *Journal of Consciousness Studies*, *23*(11–12), 11–39.

Fuchs, T. (2016). The Embodied Development of Language. W: G. Etzelmüller, C. Tewes (red.), *Embodiment in Evolution and Culture* (s. 107–128). Mohr Siebeck.

Fuentes, A. (2018). How Humans and Apes Are Different, and Why It Matters. *Journal of Anthropological Research*, *74*(2), 151–167. https://doi.org/10.1086/697150

Galilei, G. (1930). *Rozmowy i dowodzenia matematyczne w zakresie dwóch nowych umiejętności dotyczących mechaniki i ruchów miejscowych*. Wydawnictwo Kasy im. Mianowskiego, Instytutu Popierania Nauki.

Gallagher, S., Zahavi, D. (2015). *Fenomenologiczny umysł*, tłum. M. Pokropski. Wydawnictwo Naukowe PWN.

Ganis, G., Thompson, W.L., Kosslyn, S.M. (2004). Brain Areas Underlying Visual Mental Imagery and Visual Perception: An fMRI Study. *Cognitive Brain Research*, *20*(2), 226–241. https://doi.org/10.1016/j.cogbrainres.2004.02.012

Gendler, T.S. (1998). Galileo and the Indispensability of Scientific Thought Experiment. *The British Journal for the Philosophy of Science*, *49*(3), 397–424.

Gendler, T.S. (2000). *Thought Experiment. On the Powers and Limits of Imaginary Cases*. Garland Publishing Inc.

Gendler, T.S., Liao, S. (2016). The Problem of Imaginative Resistance. W: N. Carroll, J. Gibson (red.), *The Routledge Companion to Philosophy of Literature* (s. 405–418). Routledge.

Gettier, E.L. (1963). Is Justified True Belief Knowledge? *Analysis*, *23*(6), 121–123.

Gigerenzer, G. (2008). Why Heuristics Work. *Perspectives on Psychological Science*, *3*(1), 20–29. https://doi.org/10.1111/j.1745-6916.2008.00058.x

Giora, R. (2003). *On Our Mind: Salience, Context, and Figurative Language*. Oxford University Press.

Girotto, V., Ferrante, D., Pighin, S., Gonzalez, M. (2007). Postdecisional Counterfactual Thinking by Actors and Readers. *Psychological Science*, *18*(6), 510–515. https://doi.org/10.1111/j.1467-9280.2007.01931.x

Girotto, V., Legrenzi, P., Rizzo, A. (1991). Event Controllability in Counterfactual Thinking. *Acta Psychologica*, *78*(1–3), 111–133. https://doi.org/10.1016/0001-6918(91)90007-M

Givoni, S., Giora, R., Bergerbest, D. (2013). How Speakers Alert Addressees to Multiple Meanings. *Journal of Pragmatics*, *48*(1), 29–40. https://doi.org/10.1016/j.pragma.2012.11.011

Goff, P. (2011). A Posteriori Physicalists Get Our Phenomenal Concepts Wrong. *Australasian Journal of Philosophy*, *89*(2), 191–209. https://doi.org/10.1080/00048401003649617

Goff, P. (2017). *Consciousness and Fundamental Reality*. Oxford University Press.

Goff, P. (2022). *Błąd Galileusza. Fundamenty nowej nauki o świadomości*, tłum. J. Jarocki. Wydawnictwo Naukowe PWN.

Goodman, N. (1965). The New Riddle of Induction. W: N. Goodman, *Fact, Fiction, and Forecast*. Oxford University Press.

Grant, E. (2001). *God & Reason in the Middle Ages*. Cambridge University Press.

Grant, E. (2004). Scientific Imagination in the Middle Ages. *Perspectives on Science*, *12*(4), 394–423. https://doi.org/10.1162/1063614042776021

Greene, J., Cohen, J. (2004). For the Law, Neuroscience Changes Nothing and Everything. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1451), 1775–1785. https://doi.org/10.1098/rstb.2004.1546

Gregory, D., Hendrickx, M., Turner, C. (2022). Who Knows What Mary Knew? An Experimental Study. *Philosophical Psychology*, *35*(4), 522–545. https://doi.org/10.1080/09515089.2021.2001448

Grellard, C. (2011). Thought Experiments in Late Medieval Debates on Atomism. W: K. Ierodiakonou, S. Roux (red.), *Thought Experiments in Methodological and Historical Contexts* (s. 65–79). Brill.

Gulick, R. van (2001). Reduction, Emergence and Other Recent Options on the Mind/Body Problem. *Journal of Consciousness Studies*, *8*(9–10), 1–34.

Hackett, J. (1997). Roger Bacon on Scientia Experimentalis. W: J. Hackett (red.), *Roger Bacon and the Sciences* (s. 277–315). Brill. https://doi.org/10.1163/9789004444812_015

Hacking, I. (1992). Do Thought Experiments Have a Life of Their Own? Comments on James Brown, Nancy Nersessian and David Gooding. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *1992*(2), 302–308.

Hamdo, M. (2023). *Epistemic Thought Experiments and Intuitions*. Springer.
Hassabis, D., Kumaran, D., Vann, S.D., Maguire, E.A. (2007). Patients with Hippocampal Amnesia Cannot Imagine New Experiences. *Proceedings of the National Academy of Sciences*, *104*(5), 1726–1731. https://doi.org/10.1073/pnas.0610561104
Hepburn, B., Andersen, H. (2021). Scientific Method. W: E.N. Zalta (red.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
Hertogh, C.P. (2018). Logical Analyses of Vipassanā Meditation. *Contemporary Buddhism*, *19*(2), 251–274. https://doi.org/10.1080/14639947.2018.1498604
Hettich, K. (2019). The Literary Thought Experiment as Emotional Experience: Émile Zola's Thérèse Raquin (1867). W: F. Bornmüller, J. Franzen, M. Lessau (red.), *Literature as Thought Experiment? Perspectives from Philosophy and Literary Studies* (s. 109–121). Brill | Fink. https://doi.org/10.30965/9783846764299_010
Hickok, G. (2016). *Mit neuronów lustrzanych*. Copernicus Center Press.
Hill, C.S. (1997). Imaginability, Conceivability, Possibility and the Mind-Body Problem. *Philosophical Studies*, *87*, 61–85.
Hill, C.S. (2016). Conceivability and Possibility. W: H. Cappelen, T.S. Gendler, J. Hawthorne (red.), *The Oxford Handbook of Philosophical Methodology* (s. 326–347). Oxford University Press.
Hill, C.S., Mclaughlin, B.P. (1999). There Are Fewer Things in Reality Than Are Dreamt of in Chalmers's Philosophy. *Philosophy and Phenomenological Research*, *59*(2), 445–454. https://doi.org/10.2307/2653682
Hinwar, R.P., Lambert, A.J. (2021). Anauralia: The Silent Mind and Its Association With Aphantasia. *Frontiers in Psychology*, *12*, 744213. https://doi.org/10.3389/fpsyg.2021.744213
Hoel, E. (2023). *The World Behind the World: Consciousness, Free Will, and the Limits of Science*. Avid Reader Press.
Huffman, C.A. (2005). *Archytas of Tarentum Pythagorean, Philosopher and Mathematician King*. Cambridge University Press.
Hume, D. (2005). *Traktat o naturze ludzkiej*. Aletheia.
Hutto, D.D., Myin, E. (2013). *Radicalizing Enactivism: Basic Minds without Content*. The MIT Press.
Ichikawa, J.J., Steup, M. (2017). The Analysis of Knowledge. W: E.N. Zalta (red.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
Ierodiakonou, K. (2018). The Triple Life of Ancient Thought Experiments. W: M.T. Stuart, Y. Fehige, J.R. Brown (red.), *The Routledge Companion to Thought Experiments* (s. 31–43). Routledge.
Ierodiakonou, K., Roux, S. (2011a). The Emergence Of The Notion Of Thought Experiments. W: K. Ierodiakonou, S. Roux (red.), *Thought Experiments in Methodological and Historical Contexts* (s. 1–33). Brill.
Ierodiakonou, K., Roux, S. (red.). (2011b). *Thought Experiments in Methodological and Historical Contexts*. Brill.
Irish, M., Piolino, P. (2016). Impaired Capacity for Prospection in the Dementias – Theoretical and Clinical Implications. *British Journal of Clinical Psychology*, *55*(1), 49–68. https://doi.org/10.1111/bjc.12090

Irvine, A.D. (1991). Thought Experiments in Scientific Reasoning. W: T. Horowitz, G. Massey (red.), *Thought Experiments in Science and Philosophy* (s. 149–165). Rowman & Littlefield Publishers.

Irvine, E., Sprevak, M. (2020). Eliminativism about Consciousness. W: U. Kriegel (red.), *Oxford Handbook of the Philosophy of Consciousness* (s. 348–370). Oxford University Press.

Jackson, F. (1982). Epiphenomenal Qualia. *The Philosophical Quarterly*, *32*(127), 127–136.

James, W. (1892). The Stream of Consciousness. W: W. James, *Psychology: The Briefer Course* (s. 151–175). Holt.

Jefferies, E., Patterson, K., Jones, R.W., Lambon Ralph, M.A. (2009). Comprehension of Concrete and Abstract Words in Semantic Dementia. *Neuropsychology*, *23*(4), 492–499. https://doi.org/10.1037/a0015452

Johnson-Laird, P.N. (1983). *Mental Models*. The MIT Press.

Johnson-Laird, P.N. (1989). Mental Models. W: M.I. Posner (red.), *Foundations of Cognitive Science* (s. 467–499). The MIT Press.

Johnson-Laird, P.N. (1994). Mental Models and Probabilistic Thinking. *Cognition*, *50*(1), 189–209. https://doi.org/10.1016/0010-0277(94)90028-0

Johnson-Laird, P.N. (2006a). *How We Reason*. Oxford University Press.

Johnson-Laird, P.N. (2006b). Mental Models, Sentential Reasoning, and Illusory Inferences. *Advances in Psychology*, *138*, 27–51. https://doi.org/10.1016/S0166-4115(06)80026-9

Johnson-Laird, P.N. (2008). Mental Models and Deductive Reasoning. W: J.E. Adler, L.J. Rips (red.), *Reasoning: Studies in Human Inference and Its Foundations* (s. 206–222). Cambridge University Press.

Johnson-Laird, P.N. (2010). Mental Models and Human Reasoning. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(43), 18243–18250. https://doi.org/10.1073/pnas.1012933107

Johnson-Laird, P.N. (2023). Possibilities and Human Reasoning. *Possibility Studies & Society*, *1*(1–2), 105–112. https://doi.org/10.1177/27538699231152731

Johnson-Laird, P.N., Byrne, R.M.J. (1991). *Deduction*. Lawrence Erlbaum Associates.

Johnson-Laird, P.N., Savary, F. (1999). Illusory Inferences: A Novel Class of Erroneous Deductions. *Cognition*, *71*(3), 191–229. https://doi.org/10.1016/S0010-0277(99)00015-3

Kahneman, D. (2012). *Pułapki myślenia. O myśleniu szybkim i wolnym*. Media Rodzina.

Kahneman, D., Tversky, A. (1982). The Simulation Heuristic. W: D. Kahneman, P. Slovic, A. Tversky (red.), *Judgment under Uncertainty* (s. 201–208). Cambridge University Press. https://doi.org/10.1017/CBO9780511809477.015

Kauppinen, A. (2007). The Rise and Fall of Experimental Philosophy. *Philosophical Explorations*, *10*(2), 95–118. https://doi.org/10.1080/13869790701305871

Kerr, E. (2019). Epistemology without Borders. Epistemological Thought Experiments and Intuitions in Cross-Cultural Contexts. W: M. Fricker, P.J. Graham, D. Henderson, N.J.L.L. Pedersen (red.), *The Routledge Handbook of Social Epistemology* (s. 196–208). Routledge.

Khemlani, S.S., Barbey, A.K., Johnson-Laird, P.N. (2014). Causal Reasoning with Mental Models. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00849

Kim, J. (1990). Supervenience as a Philosophical Concept. *Metaphilosophy*, *21*(1/2), 1–27. https://www.jstor.org/stable/24436754

Kim, M., Yuan, Y. (2015). No Cross-Cultural Differences in the Gettier Car Case Intuition: A Replication Study of Weinberg et al. 2001. *Episteme*, *12*(3), 355–361. https://doi.org/10.1017/epi.2015.17

Kind, A. (2001). Putting the Image Back in Imagination. *Philosophy and Phenomenological Research*, *62*(1), 85–109. https://doi.org/10.2307/2653590

Kind, A. (red.). (2016). *The Routledge Handbook of Philosophy of Imagination*. Routledge.

Kind, A., Kung, P. (red.). (2016). *Knowledge Through Imagination*. Oxford University Press.

King, P. (1991). Mediaeval Thought-Experiments: The Metamethodology of Mediaeval Science. W: T. Horowitz, G. Massey (red.), *Thought Experiments in Science and Philosophy* (s. 43–64). Rowman & Littlefield Publishers.

King, P. (2003). Two Conceptions of Experience. *Medieval Philosophy and Theology*, *11*, 203–226. https://doi.org/10.1017/S1057060803000082

Kirk, R. (1974). Sentience and Behaviour. *Mind*, *83*, 43–60.

Kirk, R., Squires, J.E.R. (1974). Zombies v. Materialists. *Proceedings of the Aristotelian Society*, *48*, 135–152.

Klawiter, A. (2004). Powab i moc wyjaśniająca kognitywistyki. *Nauka*, *3*, 81–100.

Klawiter, A. (2012). Świadomość. Wprowadzenie do współczesnych dyskusji. W: M. Miłkowski, R. Poczobut (red.), *Przewodnik po filozofii umysłu* (s. 353–394). Wydawnictwo WAM.

Kneer, M., Colaço, D., Alexander, J., Machery, E. (2021). On Second Thought: Reflections on the Reflection Defense. W: M. Kneer, D. Colaço, J. Alexander, E. Machery, *Oxford Studies in Experimental Philosophy* (t. 4, s. 257–296). Oxford University Press. https://doi.org/10.1093/oso/9780192856890.003.0010

Knobe, J. (2019). Philosophical Intuitions Are Surprisingly Robust across Demographic Differences. *Epistemology and Philosophy of Science*, *56*(2), 29–36. https://doi.org/10.5840/eps201956225

Koch, C. (2019). *The Feeling of Life Itself. Why Consciousness Is Everywhere But Can't Be Computed*. The MIT Press.

Kornmesser, S., Bauer, A.M., Alfano, M., Allard, A., Baumgartner, L., Cova, F., Engelhardt, P., Fischer, E., Meyer, H., Reuter, K., Sytsma, J., Thompson, K., Wyszynski, M. (2024). *Experimental Philosophy for Beginners: A Gentle Introduction to Methods and Tools*. Springer.

Kriegel, U. (2006). Consciousness, Theories of. *Philosophy Compass*, *1*(1), 58–64. https://doi.org/10.1111/j.1747-9991.2006.00008.x

Kripke, S.A. (1972). *Naming and Necessity*. Harvard University Press.

Kripke, S.A. (2001). *Nazywanie a konieczność*. Aletheia.

Kripke, S.A. (2007). *Wittgenstein o regułach i języku prywatnym*. Aletheia.

Krupenye, C., Call, J. (2019). Theory of Mind in Animals: Current and Future Directions. *WIREs Cognitive Science*, *10*(6), e1503. https://doi.org/10.1002/wcs.1503

Kuhn, Th. S. (1985). Rola eksperymentów naukowych. W: Th.S. Kuhn, *Dwa bieguny. Tradycja i nowatorstwo w badaniach naukowych* (s. 336–369). PIW.

Kuhn, Th.S. (2011). *Struktura rewolucji naukowych*, tłum. H. Ostromęcka, J. Nowotniak. Aletheia.

Kulakova, E., Aichhorn, M., Schurz, M., Kronbichler, M., Perner, J. (2013). Processing Counterfactual and Hypothetical Conditionals: An fMRI Investigation. *NeuroImage*, *72*, 265–271. https://doi.org/10.1016/j.neuroimage.2013.01.060

Kushnir, T. (2022). Imagination and Social Cognition in Childhood. *WIREs Cognitive Science*, *13*(4). https://doi.org/10.1002/wcs.1603

Lakatos, I. (1976). *Proofs and Refutations. The Logic of Mathematical Discovery*. Cambridge University Press.

Langland-Hassan, P. (2015). Imaginative Attitudes. *Philosophy and Phenomenological Research*, *90*(3), 664–686. https://doi.org/10.1111/phpr.12115

Langland-Hassan, P. (2020). *Explaining Imagination*. Oxford University Press.

Leeuwen, N. van (2013). The Meanings of „Imagine" Part I: Constructive Imagination. *Philosophy Compass*, *8*(3), 220–230. https://doi.org/10.1111/j.1747-9991.2012.00508.x

Leibniz, G.W. (1969). Zasady filozofii, czyli monadologia. W: S. Cichowicz (red. i tłum.), *Wyznanie wiary filozofa, rozprawa metafizyczna, monadologia, zasady natury i łaski oraz inne pisma filozoficzne* (s. 297–317). PWN.

Lenharo, M. (2023). Decades-Long Bet on Consciousness Ends – and It's Philosopher 1, Neuroscientist 0. *Nature*, *619*, 14–15. https://doi.org/10.1038/d41586-023-02120-8

Levine, J. (1983). Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly*, *64*, 354–361.

Levinson, S.C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. The MIT Press.

Lewis, C.I. (1929). *Mind and the World Order*. Charles Scribner's Sons.

Lewis, D. (1986). *On the Plurality of Worlds*. Wiley-Blackwell.

Li, J., Zhu, X. (2023). Twenty Years of Experimental Philosophy Research. *Metaphilosophy*, *54*(1), 29–53. https://doi.org/10.1111/meta.12602

Lidell, H.G., Scott, R., Jones, H.S. (1968). *A Greek-English Lexicon*. Clarendon Press.

Liu, H. (2022). What It Is Like. *Inquiry*, 1–29. https://doi.org/10.1080/0020174X.2022.2161069

Loar, B. (1997). Phenomenal States. W: N. Block, O.J. Flanagan, G. Güzeldere (red.), *The Nature of Consciousness: Philosophical Debates* (s. 597–616). The MIT Press.

Ludwig, K. (2007). The Epistemology of Thought Experiments: First Person versus Third Person Approaches. *Midwest Studies in Philosophy*, *31*(1), 128–159. https://doi.org/10.1111/j.1475-4975.2007.00160.x

Mach, E. (1973). On Thought Experiments. *Philosophical Forum*, *4*(3), 449–457.

Mach, E. (1976). On Thought Experiments. W: E. Mach, *Knowledge and Error* (s. 134–147). D. Reidel Publishing Company.

Machery, E. (2012). Expertise and Intuitions about Reference. *Theoria*, *27*(1), 37–54. https://doi.org/10.1387/theoria.3482

Machery, E. (2017). *Philosophy Within Its Proper Bounds*. Oxford University Press.

Maćkiewicz, B., Kuś, K., Hensel, W.M. (2023). The Influence of Philosophical Training on the Evaluation of Philosophical Cases: A Controlled Longitudinal Study. *Synthese*, *202*, 113. https://doi.org/10.1007/s11229-023-04316-x

Mallozzi, A., Vaidya, A.J., Wallner, M. (2023). The Epistemology of Modality. W: E.N. Zalta, U. Nodelman (red.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Manktelow, K. (2012). *Thinking and Reasoning: An Introduction to the Psychology of Reason, Judgment and Decision Making*. Psychology Press.

Marcus, E. (2004). Why Zombies Are Inconceivable. *Australasian Journal of Philosophy*, *82*(3), 477–490. https://doi.org/10.1080/713659880

Martin, M.G.F. (2002). The Transparency of Experience. *Mind & Language*, *17*(4), 376–425. https://doi.org/10.1111/1468-0017.00205

Martin-Ordas, G. (2020). It Is about Time: Conceptual and Experimental Evaluation of the Temporal Cognitive Mechanisms in Mental Time Travel. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(6), e1530. https://doi.org/10.1002/wcs.1530

Márton, M. (2019). What Does the Zombie Argument Prove? *Acta Analytica*, *34*(3), 271–280. https://doi.org/10.1007/s12136-018-0373-4

Masland, R. (2020). *Czego oczy nie widzą. Jak wzrok kształtuje nasze myśli*. Wydawnictwo Poznańskie.

Mather, J.A., Dickel, L. (2017). Cephalopod Complex Cognition. *Current Opinion in Behavioral Sciences*, *16*, 131–137. https://doi.org/10.1016/j.cobeha.2017.06.008

Matthews, R., Wasserstein, R., Spiegelhalter, D. (2017). The ASA's p-Value Statement, One Year On. *Significance*, *14*, 38–41. https://doi.org/10.1111/j.1740-9713.2017.01021.x

McAllister, J.W. (2018). Historicism and Cross-Culture Comparison. W: M.T. Stuart, Y. Fehige, J.R. Brown (red.), *The Routledge Companion to Thought Experiments* (s. 425–438). Routledge. https://doi.org/10.4324/9781315175027-24

McGinnis, J. (2018). Experimental Thoughts on Thought Experiments in Medieval Islam. W: M.T. Stuart, Y. Fehige, J.R. Brown (red.), *The Routledge Companion to Thought Experiments* (s. 77–91). Routledge. https://doi.org/10.4324/9781315175027-5

McNamara, P., Durso, R., Brown, A., Lynch, A. (2003). Counterfactual Cognitive Deficit in Persons with Parkinson's Disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *74*(8), 1065–1070. https://doi.org/10.1136/jnnp.74.8.1065

Menzel, C. (2023). Possible Worlds. W: E.N. Zalta, U. Nodelman (red.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Miščević, N. (1992). Mental Models and Thought Experiments. *International Studies in the Philosophy of Science*, *6*(3), 215–226. https://doi.org/10.1080/02698599208573432

Miščević, N. (2007). Modelling Intuitions and Thought Experiments. *Croatian Journal of Philosophy*, *7*(20), 181–214.

Miščević, N. (2022). *Thought Experiments*. Springer.

Miyamoto, K., Rushworth, M.F.S., Shea, N. (2023). Imagining the Future Self through Thought Experiments. *Trends in Cognitive Sciences*, *27*(5), 446–455. https://doi.org/10.1016/j.tics.2023.01.005

Moody, T.C. (1994). Conversations with Zombies. *Journal of Consciousness Studies*, *1*(2), 196–200.

Mordarski, R. (2015). Paradoksy omnipotencji. *Filo-Sofija*, *15*(30), 199–215.

Moszkowski, A. (1921). *Einstein the Searcher. His Work Explained from Dialogues with Einstein*. Methuen & Co. Ltd.

Mousavi, S., Gigerenzer, G. (2014). Risk, Uncertainty, and Heuristics. *Journal of Business Research*, *67*(8), 1671–1678. https://doi.org/10.1016/j.jbusres.2014.02.013

Næss, A. (1938). *„Truth" as Conceived by Those Who Are Not Professional Philosophers*. Jacob Dybwad.

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, *83*(4), 435–450. https://doi.org/10.2307/2183914

Nagel, T. (2010). The Facts Fetish. *The New Republic*. https://newrepublic.com/article/78546/the-facts-fetish-morality-science

Nanay, B. (2018). Multimodal Mental Imagery. *Cortex*, *105*, 125–134. https://doi.org/10.1016/j.cortex.2017.07.006

Nersessian, N.J. (1991). Why Do Thought Experiments Work. *Proceedings of the Cognitive Science Society*, *13*, 480–486.

Nersessian, N.J. (1992). In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *1992*, 291–301. http://www.jstor.org/stable/192843

Nersessian, N.J. (2007). Thought Experimenting as Mental Modeling: Empiricism without Logic. *Croatian Journal of Philosophy*, *7*(20), 125–160.

Nersessian, N.J. (2008). *Creating Scientific Concepts*. The MIT Press.

Nersessian, N.J. (2018). Cognitive Science, Mental Modeling, and Thought Experiments. W: M.T. Stuart, Y. Fehige, J.R. Brown (red.), *The Routledge Companion to Thought Experiments* (s. 309–326). Routledge.

Nichols, S., Bruno, M. (2010). Intuitions about Personal Identity: An Empirical Study. *Philosophical Psychology*, *23*(3), 293–312. https://doi.org/10.1080/09515089.2010.490939

Nichols, S., Stich, S., Weinberg, J.M. (2003). Metaskepticism: Meditations in Ethnoepistemology. W: S. Luper (red.), *The Skeptics: Contemporary Essays* (s. 227–247). Ashgate Publishing.

Niedenthal, P.M., Tangey, J.P., Gavanski, I. (1994). „If Only I Weren't" versus „If Only I Hadn't": Distinguishing Shame and Guilt in Conterfactual Thinking. *Journal of Personality and Social Psychology*, *67*(4), 585–595. https://doi.org/10.1037/0022-3514.67.4.585

Nigel, J.T.T. (1998). Zombie Killer. W: S. Hameroff, A.W. Kaszniak, A.C. Scott (red.), *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates* (s. 171–177). The MIT Press.

Nitsche, R.S. (2020). A Historical Analysis of Thought Experiments. *Controvérsia*, *16*(3), 99–121.

Norton, J.D. (1991). Thought Experiments in Einstein's Work. W: T. Horowitz, G. Massey (red.), *Thought Experiments in Science and Philosophy* (s. 129–144). Rowman & Littlefield Publishers.

Norton, J.D. (1996). Are Thought Experiments Just What You Thought? *Canadian Journal of Philosophy*, *26*(3), 333–366. https://doi.org/10.1080/00455091.1996.10717457

Norton, J.D. (2004a). On Thought Experiments: Is There More to the Argument? *Philosophy of Science*, *71*(5), 1139–1151.

Norton, J.D. (2004b). Why Thought Experiments Do Not Transcend Empiricism. W: C. Hitchcock (red.), *Contemporary Debates in Philosophy of Science* (s. 44–66). Blackwell Publishing Ltd.

Norton, J.D. (2013). Chasing the Light Einstein' s Most Famous Thought Experiment. W: M. Frappier, L. Meynell, J.R. Brown (red.), *Thought Experiments in Philosophy, Science and the Arts* (s. 123–140). Routledge.

Oatley, K., Johnson-Laird, P.N. (1996). The Communicative Theory of Emotions: Empirical Tests, Mental Models, and Implications for Social Interaction. W: L.L. Martin, A. Tesser (red.), *Striving and Feeling: Interactions Among Goals, Affect and Self-Regulation* (s. 363–393). Lawrence Erlbaum Associates.

Oliver, S. (2004). Robert Grosseteste on Light, Truth and Experimentum. *Vivarium*, *42*(2), 151–180. https://doi.org/10.1163/1568534043084739

OpenAI. (2023). *GPT-4 Technical Report*. https://doi.org/10.48550/arXiv.2303.08774

Open Science Collaboration. (2015). Estimating the Reproducibility of Psychological Science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Ørsted, H.C. (1811). *Første Indledning Til Den Almindelige Naturlære*. J.S. Schultz.

Østergaard, E. (2019). Echoes and Shadows: A Phenomenological Reconsideration of Plato's Cave Allegory. *Phenomenology & Practice*, *13*(1), 20–33. https://doi.org/10.29173/pandpr29372

Owesen, E. (2023). What-It's-like Talk Is Technical Talk. *Synthese*, *201*, 132. https://doi.org/10.1007/s11229-023-04123-4

Ozdemir, E. (2022). *Empirical Evidence Against Phenomenal Theses*. Te Herenga Waka-Victoria University of Wellington. https://doi.org/10.26686/wgtn.19070153

Palmieri, P. (2018). Galileo's Thought Experiments. W: M.T. Stuart, Y. Fehige, J.R. Brown (red.), *The Routledge Companion to Thought Experiments* (s. 92–110). Routledge.

Papineau, D. (2007). Phenomenal and Perceptual Concepts. W: T. Alter, S. Walter (red.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (s. 111–144). Oxford University Press.

Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.

Peirce, C.S. (1982). Lowell Lecture (Ix). W: M.H. Fisch (red.), *Writings of Charles S. Peirce. A Chronological Edition* (s. 471–486). Indiana University Press.

Pessoa, L. (2018). Understanding Emotion with Brain Networks. *Current Opinion in Behavioral Sciences*, *19*, 19–25. https://doi.org/10.1016/j.cobeha.2017.09.005

Peterson, E. (2021). Imagination, Attention, and Epistemic Significance. W: C. Badura, A. Kind (red.), *Epistemic Uses of Imagination* (s. 218–234). Routledge.

Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., Santos, L., Knobe, J. (2021). Knowledge before Belief. *Behavioral and Brain Sciences*, *44*, e140. https://doi.org/10.1017/S0140525X20000618

Pigliucci, M. (2014). *Neil deGrasse Tyson and the Value of Philosophy*. HuffPost. https://www.huffpost.com/entry/neil-degrasse-tyson-and-the-value-of-philosophy_b_5330216

Pigliucci, M., Boudry, M. (red.). (2013). *Philosophy of Pseudoscience: Reconsidering the Demarcation Problem*. The University of Chicago Press.

Pinker, S. (2021). *Racjonalność: Co to jest, dlaczego jej brakuje, dlaczego ma znaczenie*. Zysk i S-ka.

Piotrowski, R. (2011). *Demon Maxwella. Dzieje i filozofia pewnego eksperymentu*. Dialog.

Place, U.T. (1956). Is Consciousness a Brain Process? *British Journal of Psychology*, *47*(1), 44–50. https://doi.org/10.1111/j.2044-8295.1956.tb00560.x

Platon. (2010). *Państwo*, tłum. W. Witwicki. Wydawnictwo Naukowe PWN.

Poczobut, R. (2009). *Między redukcją a emergencją. Spór o miejsce umysłu w świecie fizycznym*. Fundacja Nauki Polskiej.

Poczobut, R. (2012). Relacje psychofizyczne. W: M. Miłkowski, R. Poczobut (red.), *Przewodnik po filozofii umysłu*. Wydawnictwo WAM.

Poerio, G.L., Smallwood, J. (2016). Daydreaming to Navigate the Social World: What We Know, What We Don't Know, and Why It Matters. *Social and Personality Psychology Compass*, *10*(11), 605–618. https://doi.org/10.1111/spc3.12288

Poincaré, H. (1956). Mathematical Creation. W: J. Newman (red.), *The World of Mathematics* (s. 2041–2050). Simon & Schuster.

Polcyn, K. (2022). Pojmowalność, możliwość i materializm. *Diametros*, *19*(73), 20–34. https://doi.org/10.33392/diam.1663

Popper, K. (2002). *Logika odkrycia naukowego*. Aletheia.

Prelević, D. (2015). Zombies Slap Back: Why the Anti-Zombie Parody Does Not Work. *Disputatio*, *7*(40), 25–43. https://doi.org/10.2478/disp-2015-0002

Princess Elisabeth of Bohemia, Descartes, R. (1645/2007). *The Correspondence between Princess Elisabeth of Bohemia and René Descartes*, red. i tłum. L. Shapiro. University of Chicago Press.

Putnam, H. (1960). Mind and Machines. W: S. Hook (red.), *Dimensions of Mind* (s. 148–179). New York University Press.

Putnam, H. (1975). The Meaning of „Meaning". *Language, Mind, and Knowledge. Minnesota Studies in the Philosophy of Science*, *7*, 131–193.

Putnam, H. (1998). Znaczenie wyrazu „znaczenie". W: H. Putnam, *Wiele twarzy realizmu i inne eseje*, tłum. A. Grobler (s. 93–184). Wydawnictwo Naukowe PWN.

Raffaelli, Q., Wilcox, R., Andrews-Hanna, J. (2020). The Neuroscience of Imaginative Thought: An Integrative Framework. W: A. Abraham (red.), *The Cambridge Handbook of the Imagination* (s. 332–353). Cambridge University Press.

Ragni, M., Knauff, M. (2013). A Theory and a Computational Model of Spatial Reasoning with Preferred Mental Models. *Psychological Review*, *120*(3), 561–588. https://doi.org/10.1037/a0032460

Raichle, M.E. (2015). The Brain's Default Mode Network. *Annual Review of Neuroscience*, *38*(1), 433–447. https://doi.org/10.1146/annurev-neuro-071013-014030

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.

Rescher, N. (1991). Thought Experimentation in Presocratic Philosophy. W: T. Horowitz, G. Massey (red.), *Thought Experiments in Science and Philosophy* (s. 31–41). Rowman & Littlefield Publishers.

Rescher, N. (2017). *What If? Thought Experimentation in Philosophy*. Routledge.

Reuter, K. (2024). Experimental Philosophy of Consciousness. W: S. Nichols, J. Knobe (red.), *Oxford Studies in Experimental Philosophy Volume 5* (rozdz. 16). Oxford University Press.

Ridder, J. de, Peels, R., Woudenberg, R. van (red.). (2018). *Scientism: Prospects and Problems*. Oxford University Press.

Rim, S., Summerville, A. (2014). How Far to the Road Not Taken? The Effect of Psychological Distance on Counterfactual Direction. *Personality and Social Psychology Bulletin*, *40*(3), 391–401. https://doi.org/10.1177/0146167213513304

Rips, L.J. (1994). *The Psychology of Proof: Deductive Reasoning in Human Thinking*. The MIT Press.

Roca-Royes, S. (2011). Conceivability and De Re Modal Knowledge. *Noûs*, *45*(1), 22–49. https://doi.org/10.1111/j.1468-0068.2010.00757.x

Rodzeń, J. (2008). „Archytas of Tarentum. Pythagorean, Philosopher and Mathematician King", Carl A. Huffman, Cambridge 2005 [recenzja]. *Kwartalnik Historii Nauki i Techniki*, *53*(1), 240–243.

Roese, N.J., Park, S., Gibson, C., Smallman, R. (2008). Schizophrenia Involves Impairment in the Activation of Intentions by Counterfactual Thinking. *Schizophrenia Research*, *103*(1–3), 343–344. https://doi.org/10.1016/j.schres.2007.05.006

Rose, J. (2013). Is Chalmers's Zombie Argument Self-Refuting? And How. *Binghamton Journal of Philosophy*, *1*(1), 105–132. https://doi.org/10.5840/binghamton2013117

Ross, S. (1962). Scientist: The Story of a Word. *Annals of Science*, *18*(2), 65–85. https://doi.org/10.1080/00033796200202722

Roux, S. (2011). Introduction: The Emergence of the Notion of Thought Experiments. W: K. Ierodiakonou, S. Roux (red.), *Thought Experiments in Methodological and Historical Contexts* (s. 1–33). Brill.

Ryle, G. (1970). *Czym jest umysł?*, tłum. W. Marciszewski. PWN.

Schaafsma, S.M., Pfaff, D.W., Spunt, R.P., Adolphs, R. (2015). Deconstructing and Reconstructing Theory of Mind. *Trends in Cognitive Sciences*, *19*(2), 65–72. https://doi.org/10.1016/j.tics.2014.11.007

Schacter, D.L., Addis, D.R., Hassabis, D., Martin, V.C., Spreng, R.N., Szpunar, K.K. (2012). The Future of Memory: Remembering, Imagining, and the Brain. *Neuron*, *76*(4), 677–694. https://doi.org/10.1016/j.neuron.2012.11.001

Schaeken, W., Johnson-Laird, P.N., d'Ydewalle, G. (1996). Mental Models and Temporal Reasoning. *Cognition*, *60*(3), 205–234. https://doi.org/10.1016/0010-0277(96)00708-1

Schaffer, J. (2009). On What Grounds What. W: D.J. Chalmers, D. Manley, R. Wasserman (red.), *Metametaphysics. New Essays on the Foundations of Ontology* (s. 347–383). Oxford University Press.

Schlaepfer, G., Weber, M. (2018). Thought Experiments in Biology. W: M.T. Stuart, Y. Fehige, J.R. Brown (red.), *The Routledge Companion to Thought Experiments* (s. 243–256). Routledge. https://doi.org/10.4324/9781315175027-14

Schroeter, L. (2021). Two-Dimensional Semantics. W: E.N. Zalta (red.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Schrödinger, E. (1935). Die Gegenwärtige Situation in der Quantenmechanik. *Die Naturwissenschaften*, *23*(48), 807–812.

Schwenkler, J., Byrd, N., Lambert, E., Taylor, M. (2022). One: But Not the Same. *Philosophical Studies*, *179*(6), 1939–1951. https://doi.org/10.1007/s11098-021-01739-5

Schwitzgebel, E. (2008). The Unreliability of Naive Introspection. *Philosophical Review*, *117*(2), 245–273. https://doi.org/10.1215/00318108-2007-037

Schwitzgebel, E., Cushman, F. (2012). Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers. *Mind & Language*, *27*(2), 135–153. https://doi.org/10.1111/j.1468-0017.2012.01438.x

Schwitzgebel, E., Cushman, F. (2015). Philosophers' Biased Judgments Persist despite Training, Expertise and Reflection. *Cognition*, *141*, 127–137. https://doi.org/10.1016/j.cognition.2015.04.015

Seager, W. (2020). *The Routledge Handbook of Panpsychism*. Routledge.

Searle, J. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, *3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Searle, J. (1995). Umysły, mózgi i programy. W: B. Chwedeńczuk (red.), *Fragmenty filozofii analitycznej. Tom II: Filozofia umysłu*, tłum. J. Bobryk (s. 301–324). Spacja.

Searle, J. (2010). *Umysł – krótkie wprowadzenie*. Rebis.

Segre, M. (1989). Galileo, Viviani and the Tower of Pisa. *Studies in History and Philosophy of Science Part A*, *20*(4), 435–451. https://doi.org/10.1016/0039-3681(89)90018-6

Shamay-Tsoory, S.G., Mendelsohn, A. (2019). Real-Life Neuroscience: An Ecological Approach to Brain and Behavior Research. *Perspectives on Psychological Science*, *14*(5), 841–859. https://doi.org/10.1177/1745691619856350

Shapin, S. (1996). *The Scientific Revolution*. The University of Chicago Press.

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

Shoemaker, S. (1999). On David Chalmers's the Conscious Mind. *Philosophy and Phenomenological Research*, *59*(2), 439–444. https://doi.org/10.2307/2653681

Smart, J.J.C. (1959). Sensations and Brain Processes. *The Philosophical Review*, *68*(2), 141–156. https://doi.org/10.2307/2182164

Smart, J.J.C. (1995). Doznania a procesy mózgowe. W: B. Chwedeńczuk (red.), *Fragmenty filozofii analitycznej. Tom II: Filozofia umysłu*, tłum. M. Szczubiałka (s. 247–262). Spacja.

Snooks, R.J. (2006). Another Scientific Practice Separating Chemistry from Physics: Thought Experiments. *Foundations of Chemistry*, *8*(3), 255–270. https://doi.org/10.1007/s10698-006-9019-5

Sobczyńska, D. (2001). Eksperymenty myślowe w rozważaniach teoretycznych i praktyce eksperymentalnej przyrodoznawstwa. W: P. Leśniewski, Z. Tworak (red.), *Logos – rozum i logika* (s. 141–168). Wydawnictwo Naukowe Instytutu Filozofii UAM.

Sorensen, R.A. (1998). *Thought Experiments*. Oxford University Press.

Spaulding, S. (2019). What Is Mindreading? *WIREs Cognitive Science*, *11*, e1523. https://doi.org/10.1002/wcs.1523

Stalnaker, R.C. (1984). *Inquiry*. The MIT Press.

Stalnaker, R.C. (2008). Odmiany superweniencji. W: M. Miłkowski, R. Poczobut (red.), *Analityczna metafizyka umysłu. Najnowsze kontrowersje*, tłum. T. Ciecierski. Wydawnictwo IFiS PAN.

Starikova, I., Giaquinto, M. (2018). Thought Experiments in Mathematics. W: M.T. Stuart, Y. Fehige, J.R. Brown (red.), *The Routledge Companion to Thought Experiments* (s. 257–278). Routledge.

Stenning, K., Lambalgen, M. van (2008). *Human Reasoning and Cognitive Science*. The MIT Press.

Stephens, A.L., Clement, J.J. (2012). The Role of Thought Experiments in Science and Science Learning. W: B.J. Fraser, K. Tobin, C.J. McRobbie (red.), *Second International Handbook of Science Education* (s. 157–175). Springer Netherlands.

Stewart, I., Tall, D. (2015). *The Foundations of Mathematics*. Oxford University Press.

Stich, S.P., Machery, E. (2023). Demographic Differences in Philosophical Intuition: A Reply to Joshua Knobe. *Review of Philosophy and Psychology*, *14*(2), 401–434. https://doi.org/10.1007/s13164-021-00609-7

Stoljar, D. (2007). II – Two Conceivability Arguments Compared. *Proceedings of the Aristotelian Society*, *107*, 27–44. https://doi.org/10.1111/j.1467-9264.2007.00210.x

Stoljar, D. (2023). Physicalism. W: E.N. Zalta, U. Nodelman (red.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Stout, G.F. (1931). *Mind and Matter*. Cambridge University Press.

Strawson, G. (2019). A Hundred Years of Consciousness: „A Long Training in Absurdity". *Estudios de Filosofía*, *59*, 9–43. https://doi.org/10.17533/udea.ef.n59a02

Strawson, P.F. (1980). *Indywidua. Próba metafizyki opisowej*, tłum. B. Chwedeńczuk. Instytut Wydawniczy PAX.

Stuart, M.T. (2016a). Norton and the Logic of Thought Experiments. *Axiomathes*, *26*(4), 451–466. https://doi.org/10.1007/s10516-016-9306-2

Stuart, M.T. (2016b). Taming Theory with Thought Experiments: Understanding and Scientific Progress. *Studies in History and Philosophy of Science Part A*, *58*, 24–33. https://doi.org/10.1016/j.shpsa.2016.04.002

Stuart, M.T. (2021a). Telling Stories in Science: Feyerabend and Thought Experiments. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, *11*(1), 262–281. https://doi.org/10.1086/712946

Stuart, M.T. (2021b). Towards a Dual Process Epistemology of Imagination. *Synthese*, *198*(2), 1329–1350. https://doi.org/10.1007/s11229-019-02116-w

Stuart, M.T., Fehige, Y., Brown, J.R. (2018a). *The Routledge Companion to Thought Experiments*. Routledge.

Stuart, M.T., Fehige, Y., Brown, J.R. (2018b). Thought Experiments. State of the Art. W: M.T. Stuart, Y. Fehige, J.R. Brown (red.), *The Routledge Companion to Thought Experiments* (s. 1–28). Routledge.

Suddendorf, T., Bulley, A., Miloyan, B. (2018). Prospection and Natural Selection. *Current Opinion in Behavioral Sciences*, *24*, 26–31. https://doi.org/10.1016/j.cobeha.2018.01.019

Suhler, C. (2019). Why the Method of Cases Doesn't Work. *Review of Philosophy and Psychology*, *10*, 825–847. https://doi.org/10.1007/s13164-019-00433-0

Sytsma, J. (2012). Revisiting the Valence Account. *Philosophical Topics*, *40*(2), 179–198.

Sytsma, J., Buckwalter, W. (red.). (2016). *A Companion to Experimental Philosophy*. John Wiley & Sons Ltd.

Sytsma, J., Fischer, E. (2023). „Experience", Ordinary and Philosophical: A Corpus Study. *Synthese*, *201*, 210. https://doi.org/10.1007/s11229-023-04190-7

Sytsma, J., Livengood, J. (2015). *The Theory and Practice of Experimental Philosophy*. Broadview Press.

Sytsma, J., Machery, E. (2010). Two Conceptions of Subjective Experience. *Philosophical Studies*, *151*(2), 299–327. https://doi.org/10.1007/s11098-009-9439-x

Sytsma, J., Ozdemir, E. (2019). No Problem: Evidence That the Concept of Phenomenal Consciousness Is Not Widespread. *Journal of Consciousness Studies*, *26*(9–10), 241–256.

Szabó, Á. (1978). *The Beginnings of Greek Mathematics*. D. Reidel Publishing Company.

Szpunar, K.K., Spreng, R.N., Schacter, D.L. (2014). A Taxonomy of Prospection: Introducing an Organizational Framework for Future-Oriented Cognition. *Proceedings of the National Academy of Sciences*, *111*(52), 18414–18421. https://doi.org/10.1073/pnas.1417144111

Świrydowicz, K. (2004). *Podstawy logiki modalnej*. Wydawnictwo Naukowe UAM.

Thagard, P. (2014). Thought Experiments Considered Harmful. *Perspectives on Science*, *22*(2), 288–305. https://doi.org/10.1162/POSC_a_00131

Thomas, K. (2010). Haitian Zombie, Myth, and Modern Identity. *CLCWeb: Comparative Literature and Culture*, *12*(2), 12.

Thomson, J.J. (1971). A Defense of Abortion. *Philosophy & Public Affairs*, *1*(1), 47–66.

Thomson, J.J. (1976). Killing, Letting Die, and the Trolley Problem. *Monist: An International Quarterly Journal of General Philosophical Inquiry*, *59*, 204–217.

Tittle, P. (2005). *What If... Collected Thought Experiments in Philosophy*. Pearson Education.

Tobia, K. (2015). Personal Identity and the Phineas Gage Effect. *Analysis*, *75*(3), 396–405. https://doi.org/10.1093/analys/anv041

Tobia, K., Buckwalter, W., Stich, S. (2013). Moral Intuitions: Are Philosophers Experts? *Philosophical Psychology*, *26*(5), 629–638. https://doi.org/10.1080/09515089.2012.696327

Todd, C. (2023). Affective Memory, Imagined Emotion, and Bodily Imagery. *Synthese*, *202*, 152. https://doi.org/10.1007/s11229-023-04372-3

Tomasello, M. (2015). *Historia naturalna ludzkiego myślenia*. Copernicus Center Press.

Tononi, G. (2012). *Phi. A Voyage from the Brain to the Soul*. Pantheon Books.
Turing, A.M. (1950). Computing Machinery and Intelligence. *Mind*, *59*(236), 433–360.
Tversky, A., Kahneman, D. (1983). Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review*, *90*(4), 293–315. https://doi.org/10.1037/0033-295X.90.4.293
Tye, M. (2003). A Theory of Phenomenal Concepts. *Royal Institute of Philosophy Supplement*, *53*, 91–105. https://doi.org/10.1017/S1358246100008286
Tye, M. (2006). Absent Qualia and the Mind-Body Problem. *The Philosophical Review*, *115*(2), 139–168. http://www.jstor.org/stable/20446896
Uddin, L.Q., Yeo, B.T.T., Spreng, R.N. (2019). Towards a Universal Taxonomy of Macroscale Functional Human Brain Networks. *Brain Topography*, *32*(6), 926–942. https://doi.org/10.1007/s10548-019-00744-6
Vandierendonck, A., Dierckx, V., De Vooght, G. (2004). Mental Model Construction in Linear Reasoning: Evidence for the Construction of Initial Annotated Models. *The Quarterly Journal of Experimental Psychology Section A*, *57*(8), 1369–1391. https://doi.org/10.1080/02724980343000800
Ven, N. van de, Zeelenberg, M. (2015). On the Counterfactual Nature of Envy: „It Could Have Been Me". *Cognition and Emotion*, *29*(6), 954–971. https://doi.org/10.1080/02699931.2014.957657
Viviani, V. (2019). Historical Account of the Life of Galileo Galilei. W: S. Gattei (red.), *On the Life of Galileo: Vincenzo Viviani's Historical Account and Other Early Biographies* (s. 1–93). Princeton University Press.
Wang, P. (2013). *Non-Axiomatic Logic: A Model of Intelligent Reasoning*. World Scientific.
Weatherson, B. (2004). Morality, Fiction, and Possibility. *Philosophers' Imprint*, *4*(3), 1–27.
Weinberg, J.M., Nichols, S., Stich, S.P. (2001). Normativity and Epistemic Intuitions. *Philosophical Topics*, *29*(1/2), 429–460. https://doi.org/10.5840/philtopics2001291/217
Williamson, T. (2007). *The Philosophy of Philosophy*. Blackwell Publishing.
Williamson, T. (2016). Knowing by Imagining. W: A. Kind, P. Kung (red.), *Knowledge Through Imagination* (s. 113–123). Oxford University Press.
Williamson, T. (2019). *O co chodzi w filozofii? Od zdziwienia do myślenia*. Wydawnictwo Naukowe PWN.
Wilson, J. (2005). Supervenience-based Formulations of Physicalism. *Noûs*, *39*(3), 426–459. https://doi.org/10.1111/j.0029-4624.2005.00508.x
Windt, J.M. (2020). Dreaming: Beyond Imagination and Perception. W: A. Abraham (red.), *The Cambridge Handbook of the Imagination* (s. 659–675). Cambridge University Press. https://doi.org/10.1017/9781108580298.040
Wittgenstein, L. (1953). *Philosophical Investigations*. Macmillan Publishing Company.
Wittgenstein, L. (2000). *Dociekania filozoficzne*, tłum. B. Wolniewicz. Wydawnictwo Naukowe PWN.
Yablo, S. (1993). Is Conceivability a Guide to Possibility? *Philosophy and Phenomenological Research*, *53*(1), 1–42. https://doi.org/10.2307/2108052

Yuan, Y., Kim, M. (2021). Cross-Cultural Convergence of Knowledge Attribution in East Asia and the US. *Review of Philosophy and Psychology*, *14*, 267–294. https://doi.org/10.1007/s13164-021-00523-y

Zawadzki, P. (2017). Poznawczy status eksperymentów myślowych. Platonizm, empiryzm, modele mentalne i analogia. *Filozofia Nauki*, *2*(98), 121–135.

Zeman, A., Dewar, M., Della Sala, S. (2015). Lives without Imagery – Congenital Aphantasia. *Cortex*, *73*, 378–380. https://doi.org/10.1016/j.cortex.2015.05.019

Zielińska, R. (1981). *Abstrakcja, idealizacja, generalizacja. Próba analizy metodologicznej*. Wydawnictwo Naukowe UAM.



Michał Wyrwa — doktor nauk społecznych, kognitywista i filozof. Starszy wykładowca na Wydziale Psychologii i Kognitywistyki Uniwersytetu im. Adama Mickiewicza w Poznaniu. Jego zainteresowania badawcze koncentrują się na filozoficznych i kognitywistycznych aspektach świadomości, w tym na fenomenalnym charakterze życia umysłowego, metaproblemie świadomości i sposobach, w jaki intuicyjne rozumienie świadomości kształtuje nasze sądy moralne i społeczne. Zajmuje się także etyczno-poznawczymi konsekwencjami upowszechniania technologii sztucznej inteligencji w edukacji.

Czy możemy odkrywać nowe prawdy o świecie, nie wychodząc poza własny umysł? Czy myślenie rozpoczynające się od „Jeśli" lub „Co by było, gdyby" może dostarczać nam wiedzy równie wiarygodnej jak empiryczne badania?

Eksperymenty myślowe od wieków pomagają filozofom i naukowcom w ich pracy badawczej – od Platona i Galileusza po współczesną fizykę i naukę o świadomości. Niektóre z nich zmieniły oblicze nauki, inne zaś obnażyły granice ludzkiej intuicji.

Niniejsza książka to przewodnik po świecie eksperymentowania w myślach. Autor bada mechanizmy poznawcze stojące za tą praktyką, analizuje jej historię oraz znaczenie dla naszego rozumienia rzeczywistości. To lektura nie tylko dla tych, którzy chcą zrozumieć, jak myślimy o świecie, ale także dla tych, którzy nie chcą pobłądzić we własnym umyśle.

ISBN 978-83-66983-48-9 (numer tomu 24)
ISBN 978-83-66983-22-9 (numer kolekcji)