

# Anna Brożek, Jacek Jadacki

---

## Eksperymenty myślowe w filozofii

---

Filozofia Nauki 20/2, 5-27

---

2012

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej [bazhum.muzhp.pl](http://bazhum.muzhp.pl), gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

Anna Brożek, Jacek Jadacki

## **Eksperymenty myślowe w filozofii**

1. *Quasi-fizyczne* eksperymenty myślowe. 2. Bezludny wszechświat. 3. Wymiana «klepek». 4. Diagnostyczne eksperymenty myślowe. 5. Tor śmierci. 6. Semantyczne eksperymenty myślowe. 7. Statek Tezeusza. 8. Uzasadnione przekonanie. 9. Ziemia Bliźniacza. 10. Chiński pokój. 11. Mózg w naczyniu. 12. Konkluzja.

Bibliografia

### **1. QUASI-FIZYCZNE EKSPERYMENTY MYŚLOWE**

W tekście *Eksperymenty myślowe w nauce* zaproponowaliśmy pewną definicję „eksperymentu myślowego”. Przypomnijmy ją tutaj, gdyż będzie ona tłem dla rozważania funkcji eksperymentów myślowych w filozofii.

Otóż założyliśmy, że pomiędzy pojęciami eksperymentu rzeczywistego i eksperymentu myślowego zachodzi wyraźna analogia. Eksperyment rzeczywisty polega na wywołaniu pewnego stanu rzeczy po to, aby zaobserwować jego efekt.

(1) Osoba *O* eksperymentuje rzeczywiście (przeprowadza eksperyment rzeczywisty) z *X*-em, gdy:

(a) osoba *O* wywołuje *X*-a

i

(b) osoba *O* ustala — za pomocą obserwacji — co jest efektem *X*-a.

W eksperymencie myślowym wywołujemy nie stan rzeczy, a — przedstawienie stanu rzeczy, tj. zakładamy lub wyobrażamy sobie, że dany stan rzeczy zachodzi — po to, aby przewidzieć, za pomocą wnioskowania, jego efekty:

(2) Osoba *O* eksperymentuje myślowo (przeprowadza eksperyment myślowy) z *X*-em, gdy:

(a) osoba *O* przedstawia sobie *X*-a

i

(b) osoba  $O$  ustala — za pomocą rozumowania — co jest efektem  $X$ -a.

Wyobrażony lub założony  $X$  nazwaliśmy „bazą eksperymentu myślowego”, a wywnioskowany efekt  $X$ -a — „rezultatem eksperymentu myślowego”.

Przyjeliśmy też, że osoba przeprowadzająca eksperyment myślowy wywnioskowuje, co jest efektem  $X$ -a, na podstawie pewnych uznawanych już (mniej lub bardziej świadomie) praw lub wysuwanych hipotez. Bywa też, że prawo pozwalające na wywnioskowanie rezultatu z bazy ujawnia się (*scil.* jest uświadamiane) dopiero w momencie «przeprowadzenia» eksperymentu. Owe prawa nazwaliśmy „horyzontem eksperymentu myślowego”.

Eksperymenty — zarówno rzeczywiste, jak i myślowe — miewają cele heurystyczne lub testujące.

W pierwszym wypadku ustalenie (w drodze obserwacji lub rozumowania), że zajście stanu rzeczy  $a$  pociąga za sobą zajście stanu rzeczy  $b$ , staje się pierwszym krokiem do postawienia hipotezy głoszącej, że zajście zjawiska  $\alpha$  pociąga za sobą zajście zjawiska typu  $\beta$  (gdzie zjawisko  $\alpha$  jest typem stanów rzeczy, do którego należy stan rzeczy  $a$ , zjawisko zaś  $\beta$  — typem stanów rzeczy, do którego należy stan rzeczy  $b$ ).

Eksperymenty miewają także cel testujący: jeśli postawiliśmy już hipotezę  $\gamma$ , to ustalenie, że zajście stanu rzeczy  $a$  pociąga za sobą zajście stanu rzeczy  $b$  potwierdza tę hipotezę  $\gamma$ ; gdyby to nie zachodziło — hipoteza  $\gamma$  zostałaby obalona.

Zauważyliśmy, że skoro rezultat eksperymentu myślowego wynika z horyzontu tego eksperymentu, to zakładanie lub wyobrażanie sobie bazy eksperymentu z teoretycznego punktu widzenia jest zbędne. Eksperyment myślowy redukuje się do wywnioskowywania rezultatu z horyzontu i bazy. Wyobrażenie lub założenie bazy i rezultatu — spełniają tylko funkcję praktyczną: ilustrowania lub egzemplifikowania odpowiednich zależności.

Przedstawimy teraz analizę kilku filozoficznych eksperymentów myślowych, traktując powyższe rozróżnienia jako tło analizy.

Pierwsze dwa przykłady — to przykłady eksperymentów *quasi*-fizycznych.

Przypominają one eksperymenty fizyczne — a niezależnie od tego, że nie są rzeczywiste, duże trudności sprawia rekonstrukcja ich horyzontu.

## 2. BEZLUDNY WSZECHŚWIAT

W eseju Wierzbickiego *Boski umysł* (2011)<sup>1</sup> znajdujemy taki oto *passus*:

Proponujemy mały myślowy eksperyment. Spróbujmy wyobrazić sobie [...] świat w sytuacji, gdy nas jeszcze tu nie było i gdy nas znów na tej planecie zabraknie, świat nie oglądany przez

<sup>1</sup> Sam autor zastrzega, że jego esej ma raczej charakter literacki — nie zaś teoretyczny. Zauważmy więc, że przy pewnych założeniach co najmniej niektóre dzieła literatury pięknej (por. np. choćby *Podróże Guliwera* Jonathana Swifta czy *Mikolaja Doświadczyńskiego przypadki* Ignacego Krasickiego) wolno uważać za literackie sprawozdanie z eksperymentów myślowych.

niczyje oczy, nie dotknięty przez niczyją żywą myśl, świat, jak to mawiają, „sam w sobie”, wolny od jakichkolwiek śladów naszych ja, ogołocony z przypadłości (kształtów, kolorów, zapachów), które mu nadają nasze zmysły, pozbawiony widoków, świat, w którym w miejscu obiektów wyodrębnionych i uporządkowanych przez nasz umysł kotłuje się jakaś magma, zaś pajęczynę pojęć utrwalonych przez nasz język zastępuje tępe ani be, ani me, świat, o którym nikt nie słyszał i nie pomyślał. Wyteżamy wszystkie siły duszy, rozkręcamy imaginację, przeczucamy kolejne niewydarzone warianty wizerunku, któremu nie wolno mieć twarzy... Raptem cóż to? Uświadamiamy sobie straszną rzecz: daremne próby, próżny trud, cel nasz jest nieosiągalny z założenia, świata nie oglądanego przez nikogo, świata nie dotkniętego niczym ja, świata, o którym nikt nie słyszał i nie pomyślał, nie zdołamy sobie nigdy wyobrazić, bo on zwyczajnie pod słońcem *nie istnieje*. Jakże by miało być inaczej? Przecież już w pierwszej sekundzie naszych zmagania, ba, jeszcze wcześniej, już wręcz w chwili, w której wpadł nam do głowy sam pomysł, złamaliśmy reguły przedsięwzięcia. Miał to być świat, o którym nie pomyślał nikt. Otóż *pomyślał*. Myśmy sami najpierw o nim *pomyśleli*, potem zaś *myśleli* w najlepsze. Że nie było innego wyjścia... Świeta racja: w tym właśnie sęk, nie było wyjścia, nie ma wyjścia, nasze ja nie odstepuje świata nawet wtedy, gdy zdaje się nam, żeśmy go ogołocili z bagażu naszych zmysłów i umysłu.

Tak wychodzi na jaw najzaszczytniejszy tytuł naszych ja. Są współkreatorami bytu (Wierzbicki 2011: 18-19).

Wywód ten wolno w naszym przekonaniu zrekonstruować następująco.

Zadajmy sobie pytanie:

(1) Jaki byłby świat bezludny, tj. taki, w którym nie byłoby ludzi?

Na pytanie to nie da się odpowiedzieć inaczej niż wyobrażając sobie, jaki byłby ów świat. Zatem wyobraźmy sobie, że:

(2) Świat jest bezludny.

Stan rzeczy stwierdzany przez (2) jest bazą eksperymentu. Okazuje się, że zadanie wyobrażenia sobie, że (2), jest co najmniej dwuznaczne.

Pierwsza interpretacja uwzględniałaby to, że:

(3) Gdyby w świecie nie było ludzi, to nikt by go sobie nigdy m.in. nie wyobrażał.

Można więc (2) zinterpretować jako zadanie wyobrażenia sobie, że:

(4)  $\sim \forall x$  ( $x$  wyobraża sobie świat).

Gdyby to się nam udało, to byłibyśmy kimś, kto wyobraża sobie świat, którego... nikt sobie nie wyobraża. W konsekwencji mielibyśmy sprzeczność:

(5)  $\forall x$  ( $x$  wyobraża sobie świat)  $\wedge \sim \forall x$  ( $x$  wyobraża sobie świat).

A więc zadanie wyobrażenia sobie, że (4), a zatem i zadanie wyobrażenia sobie, że (2) — jeśli do naszego horyzontu należy (3) — wyglądałoby na niewykonalne.

Nie jest to jednak rozumowanie poprawne, gdyż — dodajmy od siebie — zadanie wyobrażenia sobie, że (2), po dokładniejszej analizie, przybiera postać:

(6) Wyobraźmy sobie w chwili  $T$  świat bezludny w chwili  $T^*$  (różnej od  $T$ ).

Wtedy zamiast (5) mamy:

(7)  $\forall x$  ( $x$  wyobraża sobie świat w chwili  $T$ )  $\wedge \sim \forall x$  ( $x$  wyobraża sobie świat w chwili  $T^*$ ).

A tu oczywiście nie ma sprzeczności.

Ponadto — zakłada się, że w formułach zakres zmiennej ‘*x*’ ogranicza się do przedmiotów należących do wyobrazonego świata. Można sobie jednak... wyobrazić, że wyobrażamy sobie świat niejako «spoza niego»: zakres zmiennej ‘*x*’ w pierwszym członie formuły (5) byłby wówczas nadzbiorem zakresu zmiennej ‘*x*’ z drugiego członu.

A oto druga interpretacja formuły (2).

Nazwijmy (idąc za pewną tradycją) własności, które ludzie przypisują jakiemuś przedmiotowi dzięki temu, że spostrzegają ten przedmiot, „własnościami subiektywnymi”. Z kolei „własnościami obiektywnymi” nazwijmy własności, które dany przedmiot ma niezależnie od tego, czy jest spostrzegany przez ludzi.

Otóż:

(8) Gdyby w świecie nie było ludzi, to świat miałby wyłącznie własności obiektywne.

Załóżmy jednak, że:

(9) Ludzie są w stanie wyobrazić sobie tylko własności subiektywne świata.

(10) Własności obiektywne danego przedmiotu są różne od jego własności subiektywnych.

(11) Nie jesteśmy w stanie wyobrazić sobie niczego, co nie ma własności.

Zatem:

(12) Nie jesteśmy w stanie wyobrazić sobie świata bezludnego.

Podkreślmy: jest to niewykonalne nie dlatego, że w zadaniu tkwi rzekoma sprzeczność (jak w wypadku pierwszej interpretacji), lecz dlatego, że zakładamy (9), (10) i (11).

W tej interpretacji — horyzontem eksperymentu myślowego są zdania (10)-(11). Żadnego rezultatu tego eksperymentu nie da się natomiast — zgodnie z (12) — wskazać.

### 3. WYMIANA «KLEPEK»

O ile poprzedni eksperyment wolno uznać za argument na rzecz ontologicznego idealizmu, o tyle eksperyment Searle’a z wymianą «klepek» jest wykorzystywany w dyskusji nad dualizmem psychofizycznym. Oto jego brzmienie:

Wyobraź sobie, że twój mózg zaczyna psuć się w ten sposób, że stopniowo ślepniesz. Wyobraź sobie dalej, że zdesperowani lekarze, pragnący poprawić twój stan, próbują wszelkich metod przywrócenia ci wzroku. Jako ostatnią deskę ratunku postanawiają wszczepić układy scalone w część twojej kory mózgowej odpowiedzialną za widzenie. Ku twojemu i ich zaskoczeniu układy scalone przywracają ci normalne widzenie. Wyobraź sobie teraz, że twój mózg znów niestety nadal się psuje, a lekarze wszczepiają coraz to nowe układy scalone. Widać już, dokąd zmierza ten eksperyment myślowy: w końcu wyobrażamy sobie, że twój mózg zostaje całkiem zastąpiony przez układy scalone — tak, że gdy potrząsasz głową, możesz słyszeć krzemowe płytki klekoczące w środku (Searle 1992: 67).

Skłonni jesteśmy uznać, że eksperyment z wymianą «klepek» ma następującą strukturę.

Obiektem eksperymentu jest hipoteza:

(1) Jeżeli przedmiot  $P_1$  złożony z części  $\{C_1, \dots, C_j\}$  pełni funkcje  $\{F_1, \dots, F_k\}$ , to przedmiot  $P_2$ , złożony z części  $\{D_1, \dots, D_j\}$ , takich, że  $D_1$  pełni taką samą funkcję jak  $C_1$ ,  $D_2$  taką samą jak  $C_2$  *etc.*, także pełni funkcje  $\{F_1, \dots, F_k\}$ .

Bazę eksperymentu stanowi sytuacja polegająca na tym, że:

(2) Wszystkie części pewnego mózgu zostają wymienione na ich odpowiedniki w postaci układów scalonych.

Z kolei na horyzont eksperymentu składają się dwie tezy:

(3) Mózg pełni co najmniej dwie funkcje: steruje zachowaniem i jest ośrodkiem świadomości.

(4) Pewne układy scalone pełnią te same funkcje, co pewne części mózgu.<sup>2</sup>

Searle nie przesądza żadnego wyniku eksperymentu, a wymienia tylko różne możliwości:

(5) System układów scalonych pełni wszystkie funkcje pełnione przez mózg — w szczególności steruje zachowaniem i jest ośrodkiem świadomości.<sup>3</sup>

(6) System układów scalonych steruje zachowaniem, ale nie jest ośrodkiem świadomości.<sup>4</sup>

(7) System układów scalonych jest ośrodkiem świadomości, ale nie steruje zachowaniem.<sup>5</sup>

Zauważmy, że nie jest to pełna lista «czystych» możliwości; istnieje np. jeszcze możliwość, polegająca na tym, że po wymianie jednego z neuronów na układ scalony zanikają wszystkie funkcje mózgu.

Mamy tutaj w gruncie rzeczy inną wersję starożytnego paradoksu nieostrości. Załóżmy mianowicie, że:

(8) Wymiana jednego neuronu mózgu na odpowiedni układ scalony nie narusza funkcjonowania mózgu.

<sup>2</sup> Nie jest jasne, czy teza (4) należy do horyzontu, czy też jest już częścią wyniku eksperymentu. Chodzi tu o to, że po wszczęciu układu scalonego dochodzi do odzyskania wzroku.

<sup>3</sup> „One logical possibility, not to be excluded on any a priori grounds alone, is surely this: you continue to have all of the sorts of thoughts, experiences, memories, *etc.* that you had that you had previously; the sequence of your mental life remains unaffected. In this case, we are imagining that the silicon chips have the power not only to duplicate the mental phenomena, conscious and otherwise that are normally responsible for your input-output functions.” Co ciekawe, Searle dodaje: „I don't for a moment think that such a thing is even remotely empirically possible. I think it is empirically absurd to suppose that we could duplicate the causal powers of neurons entirely in silicon. But that is an empirical claim of my part. It is not something that we could establish a priori. So the thought experiment remains valid as a statement of logical or conceptual possibility.”

<sup>4</sup> „The second possibility: your conscious experience slowly shrinks to nothing while your externally observable behavior remains the same.”

<sup>5</sup> „Your thoughts, feelings, experiences, memories *etc.* remains intact but your observable external behavior slowly reduces to total paralysis.”

(9) Dla każdego  $k$ : jeżeli wymiana  $k$  neuronów mózgu na odpowiednie układy scalone nie narusza funkcjonowania mózgu, to i wymiana  $k+1$  neuronów mózgu na układy scalone nie narusza funkcjonowania mózgu.

Przy założeniach (8) i (9) wymiana wszystkich neuronów mózgu na odpowiednie układy scalone nie narusza funkcjonowania mózgu; jeśli zatem mózg jest ośrodkiem świadomości, to i powstały w wyniku takiej wymiany obiekt jest ośrodkiem świadomości.

Oczywiste jest, że oba założenia — (8) i (9) wymagają faktycznego sprawdzenia empirycznego; żadne rozumowanie takiego sprawdzenia nie zastąpi. Jasne jest też, że dopiero włączenie do horyzontu hipotezy, zgodnie z którą:

(10) Żadne systemy układów scalonych nie są ośrodkami świadomości.  
świadcząoby o dopuszczalności odrzucenia dualizmu.

#### 4. DIAGNOSTYCZNE EKSPERYMENTY MYŚLOWE

O wiele częściej niż z eksperymentami *quasi*-fizycznymi — spotykamy się w filozofii z eksperymentami myślowymi, które nazwiemy tu „myślowymi eksperymentami diagnostycznymi”. Mówiąc swobodnie: eksperyment diagnostyczny służy do ujawnienia czyichś niewyjawionych dotąd przekonań — o różnym stopniu ogólności.<sup>6</sup> Takie niewyjawione (*resp.* niezwerbalizowane) przekonania nazywane bywają „intuicjami” (chodzi oczywiście o jedno ze znaczeń tego słowa).<sup>7</sup>

Otóż chociaż eksperyment diagnostyczny może być zarówno eksperymentem myślowym, jak i rzeczywistym,<sup>8</sup> to jest on szczególnie częsty w formie eksperymentu myślowego w filozofii. Pokazują to zanalizowane przez nas przykłady filozoficznych eksperymentów myślowych,<sup>9</sup> z których kilka przedstawimy poniżej.

<sup>6</sup> Według Nowaczyka i Żołnowskiego — eksperymenty diagnostyczne są „podejmowane w celu uzyskania odpowiedzi na pytania o cechy swoiste badanych zjawisk, pozwalające ustalić miejsce badanego zjawiska w systemie jakiejś klasyfikacji” (1974: 153). W ujęciu Hajduka — eksperyment diagnostyczny jest to eksperyment „służący rozpoznaniu własności obiektu lub jego zaklasyfikowaniu” (2002: 83). Podobnie według Wojtysiaka — jest to eksperyment „stanowiący podstawę do odpowiedniego zaklasyfikowania danego przedmiotu” (2007: 344). Pozostałe eksperymenty — „podejmowane w celu ustalenia charakteru zależności między zjawiskami” — Nowaczyk i Żołnowski nazywają „eksperymentami wyjaśniającymi” (1974: 153).

<sup>7</sup> Rola, jaką odgrywają intuicje w eksperymentach myślowych, została częściowo zanalizowana w Brendel (2004). Stwierdza się tam, że wszelkie eksperymenty myślowe mogą być zrekonstruowane jako argumentacje, ale ich moc ekspanacyjna płynie właśnie z odwoływania się do intuicji. Zgadza się jedynie po części z tym stwierdzeniem. Prawdą jest, że w takich eksperymentach obecny jest pewien komponent intuicyjny, ale w ostateczności, gdy odpowiadamy na pytanie zawarte w scenariuszu eksperymentu, odwołujemy się do *twierdzeń naukowych*, tj. dobrze uzasadnionych, nie zaś do intuicji.

<sup>8</sup> To, że zamiast eksperymentów rzeczywistych robi się tylko eksperymenty myślowe, ma najczęściej uzasadnienie etyczne (np. dlatego, że eksperymenty rzeczywiste nie są obojętne moralnie).

<sup>9</sup> Diagnostyczne eksperymenty myślowe stosowane są również poza filozofią; zaliczylibyśmy do nich np. Schrödingera eksperyment z kotem i trucizną (1935).

Schemat eksperymentu diagnostycznego wyglądałby następująco. Rozważmy zależność o postaci:

$$(1) \wedge x (Px \Rightarrow Qx).$$

Zależność (1) jest obiektem eksperymentu. Załóżmy teraz, że mamy sprawdzić hipotezę:

$$(2) \text{ Osoba } O \text{ jest przekonana, że } \wedge x (Px \Rightarrow Qx).^{10}$$

Aby to sprawdzić — możemy np. osobę  $O$  zapytać:

$$(3) \text{ Osobo } O, \text{ czy jesteś przekonana, że (1)?}$$

Byłoby to uzasadnienie bezpośrednie. Niekiedy jednak — z różnych powodów — korzystamy ze sprawdzania pośredniego. Rozważmy formułę:

$$(4) P^*a \Rightarrow Q^*a.$$

Predykaty w formule (4) celowo oznaczamy gwiazdkami, wskazując na to, że są w pewnym sensie odpowiednikami predykatów występujących w (1).

Założmy teraz, że:

$$(5) \text{ Formuła (4) jest szczególnym przypadkiem formuły (1).}^{11}$$

Pytamy teraz osobę  $O$ :

$$(6) \text{ Osobo } O, \text{ czy jesteś przekonana, że (4)?}$$

W innej wersji pytanie może brzmieć:

$$(7) \text{ Osobo } O, \text{ co jest pociągane przez to, że } P^*a?$$

albo:

$$(8) \text{ Czy } 'P^*a' \text{ pociąga } 'Q^*a', \text{ czy } '\sim Q^*a'?$$

Jeżeli osoba  $O$  odpowie na pytanie (6) pozytywnie (bądź odpowie na pytanie (7) lub (8), stwierdzając, iż ' $P^*a$ ' pociąga ' $Q^*a$ '), to możemy uznać, że hipotezę (b) potwierdziliśmy. Jeżeli natomiast osoba  $O$  odpowie na pytanie (6) negatywnie (bądź odpowie na pytanie (7) lub (8) stwierdzając, iż ' $P^*a$ ' pociąga ' $\sim Q^*a$ '), to możemy to zinterpretować różnie.

Po pierwsze, może to świadczyć o tym, że osoba  $O$  odrzuca (1) — i tym samym hipoteza (2) jest obalona.

Po drugie, może to świadczyć o tym, że osoba  $O$  odrzuca (4) nie dlatego, że odrzuca (1), lecz dlatego, że odrzuca (5), tj. nie uznaje (4) za szczególny przypadek (1).

Przypuśćmy bowiem, że (1) i (4) są schematami odpowiednio zdań:

$$(9) \text{ Oczernianie innych ludzi jest złem.}$$

$$(10) \text{ Osoba } A, \text{ mówiąca coś złego o osobie } B, \text{ jest godna potępienia.}$$

Osoba  $O$  może deklorować odrzucenie tego, że (10), gdyż uważa, że:

$$(11) \text{ Oczernianie ludzi nie jest złem.}$$

$$(12) \text{ Mówienie o kimś czegoś złego nie (zawsze) jest oczernianiem ludzi.}$$

<sup>10</sup> Sytuację tę można równoważnie opisać mówiąc, że chodzi o ujawnienie przekonań danej osoby co do tego, czy zależność ' $\wedge x (Px \Rightarrow Qx)$ ' ma miejsce, czy nie.

<sup>11</sup> Bycie-szczególnym-przypadkiem w tym wypadku może polegać np., na tym, że predykaty ' $P^*$ ' i ' $Q^*$ ' są po prostu równozakresowe z predykatami ' $P$ ' i ' $Q$ ', bądź na tym, że zakresy predykatów ' $P^*$ ' i ' $Q^*$ ' są podzbiorem zakresów, odpowiednio, predykatów ' $P$ ' i ' $Q$ '.



Opisany eksperyment ujawnić może więc — w zależności od okoliczności — intuicje etyczne (co do tego, czy oczernianie jest złem) lub intuicje semantyczne (co do znaczenia słów „oczernianie”, „zło” *etc.*) uczestniczącej w nim osoby.

Dać jeszcze należy, że osobą *O*, «na której» przeprowadza się eksperyment diagnostyczny, może być w szczególności... sam eksperymentator. Chodzi mu wtedy o ujawnienie własnych intuicji. Jeśli z kolei eksperymentator jest różny od osoby *O*, to może mieć różne cele, przede wszystkim perswazyjne. Chodzić może m.in. o:

- (a) uświadomienie osobie *O*, jakie żywi intuicje;
- (b) uświadomienie osobie *O* potrzeby zrewidowania żywionych intuicji;
- (c) uświadomienie osobie *O*, że jej «ukryte» intuicje nie pokrywają się z jej poglądami głoszonymi *explicite*.

## 5. TOR ŚMIERCI

Ograniczymy się tutaj do zreferowania najbardziej znanego eksperymentu myślowego, należącego do eksperymentów diagnostycznych, a mianowicie do eksperymentu z torem śmierci, mającego zresztą rozmaite warianty. Przytoczmy go w wersji Thomson:

Załóżmy, że jesteś maszynistą tramwaju. Tramwaj zakręca i jedzie wprost na pięciu robotników, którzy remontują tory. Tory prowadzą w tym miejscu przez dolinkę o bardzo stromych zboczach, więc żeby nie najechać na robotników, musisz zatrzymać tramwaj. Naciskasz na hamulec, ale okazuje się, że nie działa. Nagle dostrzegasz jednak boczną prowadzącą w prawo. Możesz w nią skręcić i w ten sposób ocalić pięciu robotników stojących na torze przez tobą. Niestety [...] na bocznym torze pracuje jakiś robotnik. Nie zdąży on zbiec z toru na czas, więc jeśli skierujesz na niego tramwaj, zabijesz go (Thomson 1985: 94).

Opisana wyżej sytuacja — to baza eksperymentu; nazwijmy ją ‘*B*’.

Pytanie diagnostyczne brzmieć może różnie, np.:

- (1) Czy jeżeli znalazłbyś się w sytuacji *B*, to skręciłbyś w prawo?
- (2) Co zrobiłbyś w sytuacji *B*?
- (3) Czy w sytuacji *B* skręciłbyś w prawo, czy nie?
- (4) Co powinno się zrobić w sytuacji *B*?
- (5) Czy w sytuacji *B* powinno się / wolno skręcić w prawo?

Załóżmy, że chodzi o pytanie (5). Zauważmy, że może być ono traktowane jako szczególny wypadek pytań dotyczących wielu różnych norm etycznych, np.:

- (6) Czy jeżeli ktoś może ocalić życie pięciu osób, poświęcając życie jednej osoby, to powinien / może poświęcić życie jednej osoby?
- (7) Czy powinno się / wolno kogoś zamordować, by uratować życie komuś innemu?
- (8) Czy w każdych okolicznościach powinno się / wolno wybierać czyny, które minimalizują ogół cierpienia?

Zauważmy, że czyjeś intuicje językowe mogą być takie, że nie uznałby (5) za szczególnie przypadek (6), (7) lub (8) (może np. nie uznać przedstawienia zwrotnicy w celu skierowania pociągu na jakąś osobę za zamordowanie tej osoby).

Negatywna odpowiedź na (5) może jednak świadczyć o tym, że osoba, z którą się eksperymentuje, odrzuca (6), (7) lub (8).

Dodajmy, że postawienie kogoś przed koniecznością odpowiedzi na (5) pozwala na uświadomienie sobie, że pewne spontanicznie uznawane normy pozostają względem siebie w konflikcie — np. normy (6)-(8) są w konflikcie z normą „Nie powinno się (nikogo) zabijać”.

## 6. SEMANTYCZNE EKSPERYMENTY MYŚLOWE

Bywają eksperymenty myślowe, których bezpośrednim celem jest ujawnienie intuicji semantycznych — dotyczących denotacji lub konotacji poszczególnych terminów.<sup>12</sup>

Załóżmy, że chodzi o ustalenie konotacji terminu  $T$ , przy czym są dwie «kandydatury» na tę konotację —  $K_A$  i  $K_B$  — gdyż:

(1) Każdy znany desygnat terminu  $T$  posiada zarówno własności składające się na  $K_A$ , jak i własności składające się na  $K_B$ .

(2) Żaden znany nie-desygnat terminu  $T$  nie posiada łącznie własności składających się na  $K_A$  ani własności składających się na  $K_B$ .

Inaczej mówiąc, zarówno  $K_A$ , jak i  $K_B$ , dobrze wyznaczają zbiór znanych desygnatów terminu  $T$ .

Pytanie, która z proponowanych konotacji jest istotnie konotacją terminu  $T$ , można niekiedy rozstrzygnąć przez odwołanie się do eksperymentu myślowego. Załóżmy, że chcemy potwierdzić hipotezę semantyczną, zgodnie z którą  $K_A$  jest konotacją  $T$ . Wyobrażona baza takiego eksperymentu wskazywałaby sytuację, w której:

(3) Pewien desygnat terminu  $T$  nie posiada jakiegoś elementu  $K_B$ ;  
bądź:

(4) Pewien przedmiot posiada własność  $K_B$ , ale nie jest desygnatem terminu  $T$ .

Baza byłaby więc pewnym «granicznym» przypadkiem użycia terminu  $T$ .<sup>13</sup>

<sup>12</sup> Poza omówionymi niżej — do eksperymentów semantycznych należą m.in. takie, jak Locke’a z księciem i szewcem (1690: 479-280), Zenona z Elei z ciągłością czasoprzestrzeni — a w szczególności tzw. paradoksów dychotomii, Achillesa i żółwia oraz strzały i stadionu (Ajdukiewicz 1931), Quine’a z uprzywilejowaną logiką języka (1960: 58-59) i Fregego z prawdziwością (1893: 14-15).

<sup>13</sup> Dodajmy od razu, że ktoś, kto chciałby kontrargumentować, mógłby kontrargumentować — i utrzymywać, że wskazany w (a) przedmiot wcale nie byłby desygnatem terminu  $T$ , a w wypadku (b) — że wskazany przedmiot do denotacji  $T$  należy. Swoje stanowisko uzasadniać mógłby jedynie, podobnie zresztą jak  $A$ , swoimi i cudzymi intuicjami pojęciowymi. Pamiętajmy, że reguły semantyczne — wiążące wyrażenia z ich konotacjami — są zawsze konwencjonalne i mogłyby być dowolne, gdyby nie to, że niekiedy jest pożądanym, by odpowiadały intuicjom użytkowników języka. Do żadnej wyższej instancji — poza intuicją tych użytkowników — w opisanym sporze odwołać się nie można.

Eksperymenty myślowe służą w takich wypadkach do uzasadniania pewnych twierdzeń analitycznych i przyjmowanych reguł semantycznych.

## 7. STATEK TEZEUSZA

Zacznijmy od eksperymentu ze statkiem Tezeusza, znanym w starożytności, a tak opisanym i uzupełnionym przez Hobbesa:

Jakieś ciało można porównywać z nim samym w coraz to innym czasie. Stąd między filozofami zrodził się wielki spór o *zasadę indywidualizacji* [...], a mianowicie o to, w jakim znaczeniu jakieś ciało raz może być uważane za tożsame, a kiedy indziej za różne od tego, jakim było uprzednio. [...] Wedle pierwszego zdania nie byłby tym samym człowiek, który grzeszy i który podlega karze, a to z tej racji, że ciało ludzkie nieustannie podlega zmianom; i nie byłoby tożsame państwo, które w jednym wieku pewne prawo ustanowiło, a w innym je odwołuje [...]. Wedle drugiego zdania dwa ciała istniejące jednocześnie byłyby jednym i tożsamym co do liczby; oto bowiem na przykład ów okręt Tezeusza byłby tożsamy co do liczby z tym, jakim był od początku, po zmianie wszystkich desek (kiedyś sofiści ateńscy dysputowali o tym, czy będzie on tożsamy, czy inny, gdy stopniowo będzie się wyjmowało z niego stare deski i na ich miejsce będzie się wstawiało wciąż nowe). Otóż gdyby ktoś zachowywał owe stare deski w marę, jak się je wyjmuje, i z zachowanych, zestawiając je w tym samym porządku, później na nowo zbudował okręt, to jest rzeczą niewątpliwą, iż ten okręt byłby tożsamy co do liczby z tym, jaki był na początku; i mielibyśmy dwa okręty tożsame co do liczby, co jest rzeczą jak najbardziej niedorzeczną. Wedle trzeciego zdania żadna w ogóle rzecz nie jest tożsama z tym, czym była; tak że nawet człowiek, który stoi nie byłby tożsamy z człowiekiem, które poprzednio siedział, i nie byłaby tożsama woda, która znajduje się w naczyniu i która niebawem zostaje wylana. Tak więc należy uważać, że zasada indywidualizacji nie zawsze spoczywa w samej materii ani też nie zawsze w samej formie. [...] Okręt, o ile tym mianem oznacza się materię w pewien sposób ukształtowaną, będzie ten sam, jeśli cała materia jest ta sama; jeśli zaś żadna część materii nie jest ta sama, to całkiem różny co do liczby będzie też i okręt; jeśli zaś pozostała część materii, część zaś się zmieniła, to okręt będzie w części ten sam, w części zaś różny (Hobbes 1642: 156-158).

Rozważmy teraz następujące (nie wszystkie!) znaczenia „tożsamości”:

(1)  $\wedge x \wedge y [x \text{ jest tożsamym}_M z y \Leftrightarrow \wedge C (C \text{ jest częścią } x \text{-a} \Leftrightarrow C \text{ jest częścią } y \text{-a})]$ .

(2)  $\wedge x \wedge y [x \text{ jest tożsamym}_F z y \Leftrightarrow \forall f (x \text{ ma formę } f \wedge y \text{ ma formę } f)]$ .

(3)  $\wedge x [x \text{ zachowuje tożsamość}_G \text{ w okresie } T \Leftrightarrow \wedge t_1 \wedge t_2 (t_1, t_2 \in T \Rightarrow x \text{ w chwili } t_1 \text{ jest tożsamym}_M \text{ i tożsamym}_F z x \text{ w chwili } t_2)]$ .

Jeżeli teraz założymy, że tożsamość<sub>M</sub> (co do części) i tożsamość<sub>F</sub> (co do formy) są wzajemnie niezależne, to np. zarazem:

(4) Jeżeli  $A$  i  $B$  mają różne części, ale tę samą formę (kształt), to  $A$  i  $B$  są tożsame<sub>F</sub>, ale nietożsame<sub>M</sub>.

(5) Jeżeli  $A$  i  $B$  mają te same części i tę samą formę (kształt), to  $A$  i  $B$  są zarówno tożsame<sub>F</sub>, jak i tożsame<sub>M</sub>.

(6) Jeżeli  $A$  w chwili  $t_1$  jest tożsamym<sub>F</sub> i tożsamym<sub>M</sub> z  $A$  w chwili  $t_3$ , ale nie jest taki w chwili  $t_2$  (późniejszej od  $t_1$  i wcześniejszej od  $t_3$ ), to  $A$  nie zachowuje tożsamości<sub>G</sub> w okresie  $t_1$ - $t_3$ .

Uproszczone przykłady tych sytuacji są różne przypadki «modyfikacji» statku Tezeusza. Przyjmijmy za bazę eksperymentu stan rzeczy, w którym nastąpiła wymiana wszystkich desek — z zachowaniem kształtu całości. W tych okolicznościach statek utraciłby tożsamość<sub>M</sub> (ale zachowałby tożsamość<sub>F</sub>). Statek zestawiony z desek usuniętych przy tamtej wymianie byłby tożsamy<sub>M</sub> (i ewentualnie tożsamy<sub>F</sub>) ze statkiem przed «modyfikacją», ale utraciłby tożsamość<sub>G</sub>.

Zauważmy, że wszelkie znane nam z codziennego doświadczenia przedmioty, które są tożsame<sub>M</sub>, są także tożsame<sub>F</sub>. Dopiero przypadek graniczny, jakim jest «rozebrany» statek Tezeusza uwiadcza potrzebę rozróżnienia dwóch pojęć: tożsamości<sub>M</sub> i tożsamości<sub>F</sub>. Potraktowany jako eksperyment diagnostyczny pozwala natomiast danej osobie uświadomić sobie, które z tych pojęć odpowiada jej intuicjom związanym z terminem „tożsamość”.

## 8. UZASADNIONE PRZEKONANIE

Opis eksperymentu z uzasadnionym przekonaniem pochodzi od Gettier'a. Przyjrzyjmy się następującemu opisowi tego eksperymentu, który — jak się wydaje — jest skądinąd przykładem eksperymentu nieudanego:

W ostatnich latach podejmowano różne próby ustalenia koniecznych i wystarczających warunków tego, że ktoś wie, że jest tak jak mówi dane zdanie. Dotychczasowe próby można przedstawić w następującej formie [...]:

(a) *S* wie, że *p*

wtedy i tylko wtedy, gdy

(i) '*p*' jest prawdziwe

(ii) *S* jest przekonany, że *p*

oraz

(iii) *S*-a przekonanie, że *p*, jest uzasadnione. [...]

Będę argumentował, że (a) jest fałszywe, ponieważ warunki tam ustalone nie są WYSTARCZAJĄCE dla prawdziwości zdania „*S* wie, że *p*”. Argumentując w ten sam sposób okaże, że (b) i (c) nie zachodzą, jeśli wyrażenia: „jest oczywiste dla” lub „ma prawo być pewnym, że” będą zastąpione przez wyrażenie: „przekonanie (że) jest uzasadnione”. Zauważmy najpierw dwie rzeczy. Po pierwsze, w tym sensie „uzasadnienia”, w którym bycie *S*-a w sposób uzasadniony przekonany, że *p*, jest koniecznym warunkiem wiedzy *S*-a, że *p*, istnieje możliwość, że ktoś będzie przekonany, że uzasadnione jest zdanie, które faktycznie jest fałszywe. Po drugie, dla dowolnego zdania '*p*', jeśli przekonanie *S*-a, że *p*, jest uzasadnione i '*p*' pociąga za sobą '*q*' i *S* wyprowadza '*q*' z '*p*' i przyjmuje '*q*' jako wynik tej dedukcji, to przekonanie *S*-a, że *q*, jest również uzasadnione. Pamiętając o tych dwóch rzeczach, przedstawię teraz [...] [przykłady sytuacji], w których warunki ustalone w (a) są spełnione dla pewnego zdania, chociaż jednocześnie fałszem jest, że dana osoba w danej sprawie posiada wiedzę. [...]

Przypuśćmy, że Smith i Jones zaczęli się starać o pewną posesję. Ponadto przyjmijmy, że dla Smitha jest zupełnie oczywista następująca koniunkcja:

(d) Jones jest tym, który dostanie posesję i Jones ma dziesięć monet w swej kieszeni.

Dla Smitha może być oczywiste (d), na przykład dlatego, że prezes spółki zapewnił go, że w końcu zostanie wybrany Jones; jest też dla niego oczywiste, że on, Smith, policzył monety będące w kieszeni Jonesa przed dziesięcioma minutami. Zdanie (d) pociąga zdanie:

(e) Człowiek, który otrzyma posadę, ma dziesięć monet w kieszeni.

Przypuśćmy teraz, że Smith dostrzega wynikanie (e) z (d) i uznaje (e) na podstawie (d), które jest dla niego zupełnie oczywiste. W tym przypadku przekonanie Smitha, że (e) jest prawdziwe, istotnie jest uzasadnione.

Lecz wyobraźmy sobie dalej, że to właśnie Smith (a nie Jones) dostanie posadę, choć nie będzie o tym wiedział, oraz że to on będzie miał dziesięć monet w kieszeni i tego również nie będzie wiedział. Wówczas zdanie (e) jest prawdziwe, chociaż zdanie (d), z którego Smith wyprowadził (e), jest fałszywe. W naszym przykładzie wobec tego, to wszystko jest prawdziwe: (i) (e) jest prawdziwe, (ii) Smith jest przekonany, że (e) jest prawdziwe oraz (iii) przekonanie Smitha, że (e) jest uzasadnione. Jednakże jasne jest też, że Smith *nie wie*, że (e) jest prawdziwe; (e) jest bowiem prawdziwe z racji ilości monet w kieszeni Smitha, gdy tymczasem Smith nie wie, ile monet jest w kieszeni Smitha, i opiera swoje przekonanie, że (e), na obliczeniu monet w kieszeni Jonesa, co do którego jest fałszywie przekonany, że jest tym człowiekiem, który otrzyma posadę (Gettier 1963: 94-95).

A oto nasza rekonstrukcja.

«Klasyczna» definicja „wiedzy” jest następująca:

(1) Osoba  $O$  wie, że  $p \Leftrightarrow$  [zdanie ‘ $p$ ’ jest prawdziwe  $\wedge$  (osoba  $O$  jest przekonana, że  $p \wedge$  przekonanie osoby  $O$ , że  $p$ , jest uzasadnione)].

Eksperyment Gettier’a to taka odmiana eksperymentu semantycznego, w której «obala» się (w intencji) pewną hipotezę semantyczną, nie proponując żadnej w zamian. Bazą eksperymentu jest sytuacja opisywana przez zdania (2)-(3):

(2) Osoba  $O$  jest przekonana, że  $q \wedge$  przekonanie osoby  $O$ , że  $q$ , jest uzasadnione.

(3) Osoba  $O$  wywnioskowuje ‘ $p$ ’ z ‘ $q$ ’.

Gettier zdaje się ponadto zakładać, że:

(4)  $q \Rightarrow p$ .

(5) [(Osoba  $O$  jest przekonana, że  $q \wedge$  przekonanie osoby  $O$ , że  $q$ , jest uzasadnione)  $\wedge$  (osoba  $O$  wywnioskowuje ‘ $p$ ’ z ‘ $q \wedge q \Rightarrow p$ ’)]  $\Rightarrow$  (osoba  $O$  jest przekonana, że  $p \wedge$  przekonanie osoby  $O$ , że  $p$ , jest uzasadnione).

Czyli: jeżeli coś wywnioskujemy w sposób niezawodny z zasadnie uznanej racji, to nasz wniosek jest uznany zasadnie.

Z (2)-(5) wynika, że:

(6) Osoba  $O$  jest przekonana, że  $p \wedge$  przekonanie osoby  $O$ , że  $p$ , jest uzasadnione.

Załóżmy teraz, że:

(7) Zdanie ‘ $p$ ’ jest prawdziwe.

Z (1), (6) i (7) wynika, że:

(8) Osoba  $O$  wie, że  $p$ .

Rozumowanie Gettier’a miało — w jego intencji — pokazać, że to, że  $p$ , może zachodzić i można mieć uzasadnione przekonanie, że  $p$ , a mimo to nie wiedzieć, że  $p$ . Otóż Gettier uważa, że zdanie (d) pociąga zdanie (e). Zdanie (d) ma formę:

(9)  $Pa \wedge Qa$ ,

gdzie ‘ $P$ ’ to tyle, co „jest tym, który dostanie posadę”, ‘ $Q$ ’ — „ma dziesięć monet w kieszeni”, a ‘ $a$ ’ oznacza Jonesa.

Zdanie (e) można zaś interpretować na kilka sposobów:

(10)  $\wedge x (Px \Rightarrow Qx)$ .

(11)  $\wedge x (Qx \Rightarrow Px)$ .

(12)  $\wedge x (Px \Leftrightarrow Qx)$ .

Żadna z formuł (10)-(12) nie wynika jednak z formuły (9) — bez dodatkowych założeń, a zatem o osobie  $O$  (i o nikim w ogóle) nie można powiedzieć, że ma uzasadnione przekonanie co do którejkolwiek z formuł (10)-(12), a więc nie dotyczy jej formuła (5), czyli nie żywi ona uzasadnionego przekonania, że  $p$ .

Załóżmy więc, że dokładnie jedna osoba dostanie posadę:

(13)  $\wedge x \wedge y [(Px \wedge Py) \Rightarrow x = y]$ .

Odpadają dzięki temu interpretacje (11) i (12) (gdyż co najmniej dwie osoby mają w kieszeni dziesięć monet: Jones i Smith) — pozostaje tylko interpretacja (10). Aby jednak uznać (10) na podstawie uznania formuły (9), trzeba jeszcze uznać, że:

(14)  $x = a$ .

Tymczasem osoba  $O$  nie ma dostatecznych podstaw do uznania formuły (14),<sup>14</sup> która zresztą jest fałszem, gdyż:

(15)  $x = b$ ,

gdzie ‘ $b$ ’ oznacza Smitha.

W tej sytuacji osoba  $O$  myliłaby się sądząc, że jej przekonanie, że (10) jest uzasadnione. Opisana sytuacja — wbrew intencji Gettier’a — nie obala więc formuły (1), tj. krytykowanej definicji „wiedzy”.<sup>15</sup>

## 9. ZIEMIA BLIŹNIACZA

Eksperyment z Ziemią Bliźniaczą opisany jest przez swego autora, Putnama, w sposób następujący:

W celu wykazania, że stan psychiczny nie wyznacza ekstensji, posłużę się odrobiną science fiction. Na użytek przykładów z dziedziny science fiction, które za chwilę wystąpią, przypuścimy, że gdzieś w naszej galaktyce znajduje się planeta zwana Ziemią Bliźniaczą. Ziemia Bliźniacza jest bardzo podobna do Ziemi; ludzie na Ziemi Bliźniaczej nawet mówią *po polsku*. Pomijając różnice, które zostaną wymienione w naszych przykładach z dziedziny science fiction, czytelnik może założyć, że Ziemia Bliźniacza jest *dokładnie* taka sama, jak Ziemia. Może nawet założyć, jeżeli zechce, że ma tam swojego *Doppelgänger* — identyczną kopię siebie samego — aczkolwiek dla moich bajeczek jest to obojętne.

<sup>14</sup> Smith nie może za taką podstawę uważać tego, że „prezes spółki zapewnił go, że w końcu zostanie wybrany Jones”.

<sup>15</sup> Warto dodać, że inne wersje tego eksperymentu, w których występują przykłady zdań z alternatywą logiczną, są raczej argumentacją przeciw zwyczajowo przyjętej w logice interpretacji naturalnojęzykowej alternatywy.

Wprawdzie niektórzy ludzie na Ziemi Bliźniaczej mówią po polsku, istnieją jednak, co nie jest niespodzianką, drobne różnice między polskim dialektem występującym na Ziemi Bliźniaczej a standardowym językiem polskim. Owe różnice wynikają z pewnych osobliwości Ziemi Bliźniaczej.

Jedną z nich jest okoliczność, że płyn zwany na Ziemi Bliźniaczej „wodą” nie jest  $H_2O$ , ale jakimś innym płynem o składzie chemicznym wyrażającym się bardzo długim i skomplikowanym wzorem. Będziemy w skrócie zapisywać go jako XYZ. Założmy, że XYZ jest nieodróżnialny od wody w zwykłym zakresie temperatury i ciśnienia. W szczególności, smakuje jak woda i gasi pragnienie jak woda. Założmy również, że jeziora, morza i oceany Ziemi Bliźniaczej są wypełnione XYZ, a nie wodą, że tamtejsze deszcze składają się z kropel XYZ, a nie wody, itd.

Kiedy statek kosmiczny z Ziemi odwiedzi Ziemię Bliźniaczą, przybyśże z początku będą mniemać, że wyraz „woda” ma takie samo znaczenie na Ziemi i na Ziemi Bliźniaczej. Owo przypuszczenie zostanie skorygowane z chwilą, gdy odkryją, że „woda” na Ziemi Bliźniaczej jest XYZ, co Ziemianie stwierdzą mniej więcej za pomocą słów:

„Na Ziemi Bliźniaczej wyraz «woda» znaczy XYZ”.

(Na marginesie: powyższy sposób użycia słowa „znaczy” jest zgodny z doktryną, w myśl której ekstensja jest jednym z sensów wyrazu „znaczenie”. Zauważ jednak, że choć „znaczy” znaczy w przytoczonym przykładzie coś w rodzaju *ma ekstensję*, nikt nie twierdziłby:

„Na Ziemi Bliźniaczej znaczeniem wyrazu «woda» jest XYZ”.

chyba że — ewentualnie — fakt, iż „woda jest XYZ”, byłby znany wszystkim mówiącym po polsku dorosłym mieszkańcom Ziemi Bliźniaczej. Rzec wyjaśni się na gruncie teorii znaczenia rozwijanej poniżej; na razie zaznaczmy tylko, że choć czasownik „znaczy” niekiedy znaczy: „ma ekstensję”, rzeczownik odsłowny „znaczenie” *nigdy* nie znaczy: „ekstensja”).

Na odwrót, kiedy statek kosmiczny z Ziemi Bliźniaczej odwiedzi Ziemię, przybyśże z początku będą mniemać, że wyraz „woda” ma takie samo znaczenie na Ziemi Bliźniaczej i na Ziemi. Owo przypuszczenie zostanie skorygowane z chwilą, gdy odkryją, że „woda” na Ziemi jest  $H_2O$ , co Bliźniaczanie stwierdzą mniej więcej takimi słowami:

„Na Ziemi<sup>16</sup> wyraz «woda» znaczy  $H_2O$ ”.

Zauważ, że nie ma żadnego problemu z ekstensją nazwy „woda”. Wyraz ten ma po prostu dwa różne znaczenia (jak twierdzą): w sensie używanym na Ziemi Bliźniaczej, w sensie wody<sub>ZB</sub>, to co *my* nazywamy „wodą” po prostu nie jest wodą; podczas gdy w sensie używanym na Ziemi, w sensie wody<sub>Z</sub>, to co Bliźniaczanie nazywają „wodą” po prostu nie jest wodą. Ekstensją „wody” w sensie wody<sub>Z</sub> jest zbiór wszystkich całości złożonych z cząsteczek  $H_2O$ , lub coś w tym rodzaju; ekstensją wody w sensie woda<sub>Z</sub> jest zbiór wszystkich całości złożonych z cząsteczek XYZ, lub coś w tym rodzaju.

Cofnijmy teraz czas mniej więcej do roku 1750. Chemia nie była wówczas jeszcze rozwinięta ani na Ziemi, ani na Ziemi Bliźniaczej. Typowy ziemski Polak nie wiedział, że woda składa się z wodoru i tlenu, a typowy Polak bliźniaczański nie wiedział, że „woda” składa się z XYZ. Niech Oskar<sub>1</sub> będzie takim typowym ziemskim Polakiem, a Oskar<sub>2</sub> jego odpowiednikiem na Ziemi Bliźniaczej. Można założyć, że wszystkie przekonania Oskara<sub>1</sub> na temat wody Oskar<sub>2</sub> żywi w stosunku do „wody”. Można nawet założyć, jeżeli sobie życzysz, że Oskar<sub>1</sub> i Oskar<sub>2</sub> są dokładnie jednaki pod względem wyglądu zewnętrznego, uczuć, myśli, monologów wewnętrznych itd. Niemniej ekstensją nazwy „woda” na Ziemi w 1750 roku była, tak sa-

<sup>16</sup> Lub stwierdzą: „Na Ziemi Bliźniaczej (*Bliźniaczańska nazwa planety Terra* — H.P.) wyraz „woda” znaczy  $H_2O$ ”.

mo jak w roku 1950,  $H_2O$ ; a ekstensją nazwy „woda” na Ziemi Bliźniaczej w 1750 roku była, tak samo jak w roku 1950,  $XYZ$ . Oskar<sub>1</sub> i Oskar<sub>2</sub> różnie rozumieli nazwę „woda” w 1750 roku, *mimo że znajdowali się w jednakowym stanie psychicznym*, i mimo że, biorąc pod uwagę ówczesny stan nauki, ich społeczności uczonych dopiero pięćdziesiąt lat później odkryją, że rozumieli oni nazwę „woda” różnie. Zatem ekstensja nazwy „woda” (i jej „znaczenie”, w intuicyjnym, przedanalizycznym użyciu tego słowa) *nie* jest funkcją stanu psychicznego.

Można wszakże zapytać, dlaczego mielibyśmy zgodzić się z twierdzeniem, że nazwa „woda” ma jednakową ekstensję w roku 1750 i 1950 (na obu Ziemiach). Logika nazw oznaczających rodzaje naturalne, jak „woda”, stanowi skomplikowane zagadnienie, niemniej poniżej przedstawię zarys odpowiedzi. Przypuśćmy, że pokazuję palcem na szklankę wody i mówię: „ten płyn nazywa się «woda»” (lub „to się nazywa «woda»”, jeżeli charakteryzator „płyn” jasno wynika z kontekstu). Moja „definicja ostensywna” wody zakłada następującą presupozycję empiryczną: porcja płynu, na którą pokazuję, pozostaje w pewnej relacji jednakowości (*sameness relation*, *x jest tym samym płynem, co y*, lub *x i y są jednakowe*<sub>L</sub>) z większością twórców, które ja i inni członkowie mojej społeczności językowej przy innych okazjach nazywali „wodą”. Jeżeli ta presupozycja jest fałszywa, dajmy na to, jeżeli szklanka, o czym nie wiem, jest napełniona ginem, a nie wodą, wówczas nie oczekuję akceptacji mojej definicji ostensywnej. Zatem definicja ostensywna przynosi coś, co można by nazwać odwoływalnym warunkiem koniecznym i wystarczającym: warunkiem koniecznym i wystarczającym tego, aby być wodą, jest znajdowanie się w relacji jednakowości z płynem w tej oto szklance; co wszakże jest warunkiem koniecznym i wystarczającym tylko wtedy, gdy wspomniana presupozycja empiryczna jest spełniona. W przeciwnym razie zostaje uruchomiony któryś z szeregu, że tak powiem, warunków „zapasowych”.

Kluczowa sprawa polega na tym, że relacja jednakowości<sub>L</sub> jest relacją *teoretyczną*: rozstrzygnięcie, czy dwa płyny są jednakowe, czy nie, może być kwestią nie kończących się badań naukowych. Co więcej, nawet jeśli, dzięki badaniu naukowemu lub na skutek zastosowania jakiegoś „zdroworozsądkowego” testu, uzyska się „ostateczną” odpowiedź, będzie ona *odwoływalna*: dalsze badanie może wywrócić na nice najbardziej „pewny” przykład. Toteż fakt, że Polacy w 1750 roku mogliby nazywać  $XYZ$  „wodą”, a w 1800 czy 1850 już tak by jej nie nazywali, nie dowodzi, że przez ten czas dla przeciętnego Polaka zmieniło się „znaczenie” słowa „woda”. W 1750 roku, lub w 1850 czy w 1950, można było wskazać na, powiedzmy, płyn w dolnym biegu Wisły jako na przykład „wody”. Zmieniłoby się tylko tyle, że w 1750 roku fałszywie sądziłoby się, że  $XYZ$  znajduje się w relacji jednakowości<sub>L</sub> z płynem w dolnym biegu Wisły, podczas gdy w 1800 czy w 1850 roku wiedzielibyśmy, że tak nie jest (pomijam fakt, że w drugiej połowie XX wieku płyn w dolnym biegu Wisły jest tylko mniej więcej wodą<sup>17</sup>) (Putnam 1975: 105-109).

W naszym przekonaniu za całą tą długą historyjką kryje się rozumowanie następujące.

Załóżmy, że:

(1) *A* i *B* używają wyrażenia *W* w odniesieniu do przedmiotu *x* po to, aby przedmiotowi *x* przypisywać konotację *K*.

<sup>17</sup> W oryginale zamiast o mówiących po polsku jest mowa o „English speakers” (mówiących po angielsku), a zamiast o dolnym biegu Wisły, o Jeziorze Michigan. Dlatego nie należy tego fragmentu taktować jako wyraz poglądu Putnama na temat stanu czystości wód w polskich rzekach. — Przep. tłum.



(2) *A* używa wyrażenia *W* po to, aby wskazywać przedmioty należące do denotacji  $D_A$ .

(3) *B* używa wyrażenia *W* po to, aby wskazywać przedmioty należące do denotacji  $D_B$ .

(4)  $D_A \neq D_B$ .<sup>18</sup>

Założmy jeszcze, że:

(5) Znaczenie wyrażenia *W* = konotacja wyrażenia *W*.

Przy tym założeniu z (1) wynika, że:

(6) *A* i *B* używają wyrażenia *W* w tym samym znaczeniu.

Założmy teraz, że:

(7) Znaczenie wyrażenia *W* = denotacja wyrażenia *W*.

Przy tym założeniu z (2)-(4) wynika, że:

(8) *A* i *B* używają wyrażenia *W* w różnych znaczeniach.

Po pierwsze więc, okazuje się, że wyrażenie *W* — przy różnych znaczeniach słowa „znaczenie” — może być używane «zarazem» w tym samym znaczeniu (przy jednym znaczeniu „znaczenia”) i w różnych znaczeniach (przy różnych znaczeniach „znaczenia”).

Po drugie, jest jasne, że rozstrzygnięcie, czy rzeczywiście jest tak, jak głosi założenie (4) (nawiasem mówiąc to samo dotyczy pozostałych założeń), co najmniej niekiedy nie jest proste, i wobec tego trzeba się liczyć z pomyłką.

Tutaj — znowu — do samego przeprowadzenia tego (zgódźmy się dość banalnego) rozumowania nie potrzebne są żadne „bajeczki” o Ziemi Bliźniaczej, które pełnią wyłącznie funkcję ilustracyjną.

## 10. CHIŃSKI POKÓJ

Eksperyment z chińskim pokojem jest autorstwa Searle’a, który pisze następująco:

Mieć umysł, to coś więcej niż realizować formalne czy syntaktyczne operacje. Nasze stany umysłowe, na mocy definicji, mają zawsze jakąś treść. [...] Jeżeli myśl jest zawsze myślą o CZYMŚ, przeto dany ciąg symboli musi mieć jakieś ZNACZENIE, by stać się myślą. [...] Powód, dla którego komputerowy program nie może być umysłem, jest prosty; komputerowy program ma cechy syntaktyczne, umysły mają coś więcej niż syntaktyka. [...]

By zilustrować tę tezę, proponuję rozważenie następującego obmyślonego przeze mnie eksperymentu myślowego. Wyobraźmy sobie, że zespół programistów napisał program, który jest w stanie symulować rozumienie języka chińskiego. Zatem, na przykład, jeśli komputerowi zadaje się pytanie w tym języku, może porównać je ze swoją pamięcią lub bazą danych i wyprodukować w języku chińskim odpowiedź na zadane pytanie. Założmy, dla celów dyskusji, że odpowiedzi te są tak dobre jak odpowiedź osoby, której ojczystym językiem jest język chiński. Czy w takiej sytuacji komputer rozumie język chiński, czy rozumie go dokładnie tak, jak użyt-

<sup>18</sup> Podkreślmy, że jeśli na konotację danego wyrażenia nakłada się tylko warunek, aby była to własność istotna przedmiotów należących do denotacji tego wyrażenia, to wyrażenie o danej denotacji może mieć więcej niż jedną konotację.

kownicy języka chińskiego rozumieją swój ojczysty język? Wyobraźmy sobie teraz, że ktoś z nas jest zamknięty w pokoju, i że w pokoju tym jest szereg koszy wypełnionych znakami z języka chińskiego. Załóżmy, że osoba ta, podobnie jak autor myślowego eksperymentu, nie zna języka chińskiego, otrzymała jednak napisaną w jej ojczystym języku książkę reguł manipulowania znakami języka chińskiego. Reguły te opisują używanie symboli w sposób czysto formalny, opisują manipulowanie nimi w sposób syntaktyczny, nie semantyczny. Mogą mieć postać: „Wybierz ten znak z podwójnym zakrętasem z kosza numer jeden i połóż go za znakiem z dwoma zawijasami z kosza numer dwa”. Przyjmijmy teraz, że w pokoju pojawiają się jakieś nowe symbole, a osoba w nim siedząca otrzymuje instrukcje, jakie chińskie symbole ma wysłać z pokoju w odpowiedzi na te, które się pojawiły. Załóżmy, że siedzący w pokoju nie wie, iż wysyłane przez ludzi z zewnątrz do pokoju symbole nazywane są przez nich pytaniami, zaś symbole, które siedzący w pokoju wysyła na zewnątrz, nazywane są odpowiedziami na pytania. Przyjmijmy poza tym, że programiści napisali na tyle dobry program, zaś siedząca w pokoju osoba jest do tego stopnia dobra w manipulowaniu symbolami, że jej odpowiedzi są nieodróżnialne od odpowiedzi osoby faktycznie znającej język chiński. Zatem jakaś osoba zamknięta jest w pokoju, w którym wybiera symbole chińskie i wysyła je w odpowiedzi na inne pojawiające się w pokoju chińskie symbole. W sytuacji, jaką tu opisałem, nie ma możliwości, by w wyniku takiej manipulacji formalnie zdefiniowanymi symbolami nauczyć się języka chińskiego.

Meritum mojego myślowego eksperymentu jest następujące: realizując taki formalny komputerowy program, z punktu widzenia obserwatora z zewnątrz, zachowujemy się dokładnie tak, jak byśmy rozumieli język chiński, jednocześnie jednak nie znamy ani jednego słowa z tego języka naturalnego. Jeśli wykonanie komputerowego programu symulującego rozumienie języka chińskiego nie jest wystarczające dla nas, byśmy ten język rozumieli, NIE MOŻE BYĆ WYSTARCZAJĄCE TAKŻE DLA MASZYNY liczącej. Wyjaśnienie tego stanu rzeczy jest całkiem proste. Jeśli w takiej sytuacji człowiek nie rozumie chińskiego, to inne „komputery” go nie rozumieją, gdyż żaden komputer cyfrowy nie może, dzięki temu, że zrealizuje program, uzyskać zdolności, której nie uzyskuje człowiek. To wszystko, co ma komputer, i co my mamy siedząc w chińskim pokoju, jest jedynie formalnym programem, pozwalającym na manipulowanie niezinterpretowanymi chińskimi symbolami. Powtórzmy to — komputer opanowuje syntaktykę, nie zaś semantykę. Cały morał naszej historii o chińskim pokoju przypomina nam to, co już wiemy. Rozumienie języka — a z pewnością i posiadanie wszelkich innych stanów umysłowych — wymaga czegoś więcej niż garści symboli formalnych. Konieczna tu jest jeszcze interpretacja lub znaczenie związane z tymi symbolami. Komputer cyfrowy zaś, na mocy definicji, nie może mieć niczego więcej poza formalnymi symbolami, gdyż operacje komputerowe, jak to już powiedziano, definiuje się w terminach maszynowej zdolności do wykonywania programów. Programy natomiast mają czysto formalną charakterystykę, to znaczy nie mają semantycznych treści.

Mozemy się przekonać o sile tego argumentu, jeśli porównamy sytuację, w której otrzymujemy pytania i udzielamy odpowiedzi w naszym ojczystym języku, z sytuacją, w której pytają nas i odpowiadamy w języku, w którym nie znamy znaczenia ani jednego słowa. Wyobraźmy sobie, że jesteśmy w chińskim pokoju, gdzie zadają nam w naszym ojczystym języku pytania dotyczące takich spraw, jak nasz wiek czy fakty z naszej biografii, i że odpowiadamy na te pytania. Jaka jest różnica pomiędzy posługiwaniem się językiem chińskim a posługiwaniem się językiem polskim? Otóż, jeśli ktoś nie zna języka chińskiego, a zna język polski, różnica jest oczywista. Rozumiemy pytania zadane po polsku, bo wyrażone są one symbolami, których znaczenie jest nam znane. Analogicznie, odpowiadając po polsku, produkujemy symbole, które mają dla nas znaczenie. W przypadku języka chińskiego zaś nic takiego nie występuje. W przy-

padku chińskiego posługujemy się po prostu formalnymi symbolami, zgodnie z komputerowym programem, jednak nie potrafimy przypisać tym elementom żadnego znaczenia.

Zarówno specjaliści od sztucznej inteligencji, jak i psychologowie oraz filozofowie rozmaicie odpowiadali na moje argumenty. Wszystkie te odpowiedzi miały wspólną cechę: wszystkie były chybione. Stanie się oczywiste, dlaczego były one takie, jeśli oprzemy nasze rozumowanie na prostej logicznej prawdzie, stwierdzającej, że syntaktyka jako taka nie wystarcza dla semantyki, i że komputery w związku z tym, że są komputerami, z definicji, mają tylko syntaktykę (Searle 1984: 28-30).

Mamy tu do czynienia z ilustracją następującego rozumowania.

Załóżmy, że:

(1) *A* rozumie język *J*, gdy zna znaczenie wyrażen języka *J*.

(2) Jeśli *A* dysponuje zestawem wyrażen języka *J* i zna reguły syntaktyczne budowania odpowiedzi właściwych na pytania zadawane w języku *J*, to *A* umie zbudować odpowiedzi właściwe na pytania zadawane w języku *J*.<sup>19</sup>

Zapytajmy:

(3) Czy jeśli *A* umie zbudować odpowiedź właściwą na pytania zadawane w języku *J*, to *A* rozumie język *J*?

Otóż z (3) i (1) wynika, że:

(4) Jeśli *A* umie zbudować odpowiedzi właściwe na pytania zadawane w języku *J*, to *A* zna znaczenie wyrażen języka *J*.

Formuła (4) jest wersją krytykowanej w eksperymencie kryterialnej definicji „rozumienia”.

Z (2) i (4) wynika, że:

(5) Jeśli *A* dysponuje zestawem wyrażen języka *J* i zna reguły syntaktyczne budowania odpowiedzi właściwych na pytania zadawane w języku *J*, to *A* zna znaczenie wyrażen języka *J*.

Załóżmy, że zbudowaliśmy *A*, które dysponuje zestawem wyrażen języka *J* i zna reguły syntaktyczne budowania odpowiedzi właściwych na pytania zadawane w języku *J*, a nie zna znaczenia wyrażen języka *J*.

Zatem:

(6) Istnieje *A*, które dysponuje zestawem wyrażen języka *J* i zna reguły syntaktyczne budowania odpowiedzi właściwych na pytania zadawane w języku *J*, a nie zna znaczenia wyrażen języka *J*.

Formuła (6) to — ujęta w zacytowanym opisie w atrakcyjną formę — baza eksperymentu myślowego.

Z (6), (2) i (1) zaś wynika, że:

(7) Istnieje *A*, które umie zbudować odpowiedzi właściwe na pytania zadawane w języku *J* i nie rozumie języka *J*.

Zatem odpowiedź na pytanie (3) jest negatywna.

<sup>19</sup> Odpowiedź właściwa na dane pytanie to dowolne zdanie będące podstawieniem danej pytania, przy czym dane pytania o postaci „Dla którego *x* jest tak, że *Px*?” to funkcja „*Px*”. Nie każda odpowiedź właściwa jest, rzecz jasna, prawdziwa.

Oczywiste jest jednak przy tym, że nie samo założenie, że (6), tylko zbudowanie *A*, o którym mowa w (6) daje gwarancję tego, że odpowiedź na pytanie (3) jest negatywna.

## 11. MÓZG W NACZYNIU

Szczególną odmianą eksperymentu semantycznego jest eksperyment Putnama z mózgiem w naczyniu, który miałby stanowić argument przeciwko epistemologicznemu pan-iluzjonizmowi — przez wykazanie, że teza ta nie da się w żaden sposób uzasadnić.

Przytoczmy istotne fragmenty opisu tego eksperymentu:

Oto dyskutowana przez filozofów możliwość rodem z *science fiction*: wyobraźmy sobie, że pewien człowiek (Czytelnik może wyobrazić sobie siebie w tej roli) poddał się operacji wykonanej przez niegodziwego uczonego. Jego (twój) mózg został usunięty z ciała i umieszczony w naczyniu wypełnionym pożywką, która podtrzymuje mózg przy życiu. Zakończenia nerwowe zostały podłączone do superkomputera, który powoduje, że osoba, której mózg wyjęto, doświadcza iluzji, iż wszystko jest w najlepszym porządku. Ma ona złudzenie istnienia osób, przedmiotów, niebosłonu itd.; natomiast w rzeczywistości wszystko, czego ów człowiek doznaje (ty doznajesz), jest następstwem impulsów elektronicznych, płynących od komputera do zakończeń nerwowych. Komputer jest tak sprytnie zaprogramowany, że kiedy ofiara eksperymentu usiłuje podnieść rękę w górę, dzięki sprzężeniu zwrotnemu „widzi” i „czuje”, że ręka podnosi się w górę. Odpowiednio modyfikując program, niegodziwy uczoney może spowodować, że jego ofiara „doświadczy” (lub dozna złudzenia) uczestnictwa w dowolnie zaaranżowanej przez niegodziwca sytuacji lub obecności w dowolnie zaprojektowanym otoczeniu. Może on również wymazywać pamięć mózgu zamkniętego w naczyniu tak, że nieszczęsnej ofierze będzie się wydawało, że zawsze przebywała w otoczeniu przeznaczonym jej przez eksperymentatora. Ofierze może nawet wydawać się, że siedzi i czyta te słowa o zabawnym, lecz zgoła absurdalnym domniemaniu, iż pewien niegodziwy uczoney usuwa ludziom mózgi i umieszcza je w naczyniu z pożywką, która podtrzymuje je przy życiu. Zakończenia nerwowe są jakoby podłączone do superkomputera, który powoduje, że osoba, której mózg wyjęto, doświadcza iluzji, iż...

Celem takich opowiadań w toku wykładu z teorii poznania jest oczywiście postawienie w nowoczesnej formie klasycznego problemu sceptycyzmu wobec świata zewnętrznego. (*Skąd wiesz, że nie znajdujesz się w takim żalonym położeniu?*) Niemniej opisane położenie nadaje się również doskonale do zilustrowania kwestii stosunku umysłu do rzeczywistości.

Zamiast rozważać tylko jeden mózg w naczyniu, moglibyśmy wyobrazić sobie, że wszyscy ludzie (a może wszystkie istoty zmysłowe) są mózgami w naczyniu (lub układami nerwowymi w naczyniu, w przypadku istot o najprostszym układzie nerwowym, jaki pozwala zaliczyć je do istot „zmysłowych”). Niegodziwy uczoney musi oczywiście znajdować się na zewnątrz — ale czy aby na pewno? Może nie ma żadnego niegodziwego uczonego, może (choć to absurd) kosmos po prostu przypadkiem składa się z urządzeń automatycznych służących zapatrzeniu naczynia wypełnionego mózgami i układami nerwowymi.

Tym razem przypuścimy, że owe automatyczne urządzenia są zaprogramowane tak, aby wywoływać u nas pewną *zbiorową* halucynację, a nie odrębne, pozbawione wzajemnych związków, rozmaite halucynacje. Kiedy więc wydaje mi się, że rozmawiam z tobą, tobie wy-

daje się, że słyszysz moje słowa. Nie znaczy to, rzecz jasna, że moje słowa faktycznie docierają do twoich uszu — ponieważ (w rzeczywistości) nie masz uszu, a ja w rzeczywistości nie mam ust ani języka. Zamiast tego, kiedy wypowiadam me słowa, bodźce wychodzące z mojego mózgu płyną do komputera, który powoduje, że „słyszę”, jak mój własny głos je wymawia, i „czuję”, jak porusza się mój język, itd., oraz powoduje, że ty „słyszysz” moje słowa, „widzisz” mnie, jak mówię itd. Skoro tak, to w pewnym sensie autentycznie ze sobą rozmawiamy. Nie myślę się co do twojego rzeczywistego istnienia (tylko co do istnienia twojego ciała i „świata zewnętrznego” poza mózgami). Z pewnego punktu widzenia fakt, że „cały świat” jest zbiorową halucynacją, nie jest nawet szczególnie istotny; w końcu słyszysz przecież moje słowa, kiedy mówię do ciebie, mimo że mechanizm tego zjawiska jest inny, niż przypuszczamy. (Gdybyśmy byli parą kochanków oddających się miłosnym uściskom, a nie tylko dwójkiem ludzi zajętych rozmową, wówczas, rzecz jasna, sugestia, że jesteśmy tylko mózgami w naczyniu, mogłaby być nieco kłopotliwa.)

Pragnę teraz zadać pytanie, na pozór głupkowane i oczywiste (przynajmniej dla niektórych, w tym również paru nader wyrafinowanych filozofów), które jednak dość szybko doprowadzi nas do autentycznych głębi myśli filozoficznej. Przypuśćmy, że cała nasza historyjka jest rzeczywiście prawdziwa. Czy moglibyśmy, będąc takimi mózgami w naczyniu, *powiedzieć* lub *pomyśleć*, że nimi jesteśmy?

Zamierzam udowodnić, że odpowiedź brzmi: „Nie, nie moglibyśmy”. A właściwie zamierzam udowodnić, że domniemanie, iż naprawdę jesteśmy mózgami w naczyniu, choć nie gwałci żadnego prawa fizyki i w żaden sposób nie kłóci się z całym naszym doświadczeniem, nie może być prawdziwe. *Nie może być prawdziwe*, ponieważ w pewnym sensie obala samo siebie. [...]

Domniemanie samo siebie obala, jeżeli z założenia o jego prawdziwości wynika, że jest fałszywe. [...]

W dalszym ciągu wywodu wykazę, iż domniemanie, że jesteśmy mózgami w naczyniu, ma tę właśnie właściwość. Jeżeli mamy możliwość zastanawiać się nad tym, czy jest prawdziwe, czy fałszywe, wówczas nie jest prawdziwe (co jest do wykazania). A zatem nie jest prawdziwe. [...]

Mieszkańcy możliwego świata, którzy naprawdę są mózgami w naczyniu, [...] mogą wprawdzie pomyśleć i „wypowiedzieć” każde słowo, które my potrafimy pomyśleć lub wypowiedzieć, lecz ich słowa nie mogą (jak twierdzą) *odnosić się* do tego samego, co nasze słowa. W szczególności nie mogą oni pomyśleć ani powiedzieć, że są mózgami w naczyniu (*nawet myśląc: „jesteśmy mózgami w naczyniu”*) (Putnam 1983: 302-306).

Oto nasza rekonstrukcja tego wywodu.

Bazę eksperymentu (w wersji mocnej: „Wszyscy jesteśmy mózgami w naczyniach”) sformułujemy w sposób następujący:

(1)  $\wedge x/\wedge y$  ( $x$  ma dostęp poznawczy do  $y \Rightarrow y$  jest przedmiotem pozornym).

Czyli: przedmioty jawiące się w świadomościach «nabudowanych» na mózgach w naczyniach są jedynie pozorne.<sup>20</sup>

A oto nasza rekonstrukcja horyzontu tego eksperymentu.

Po pierwsze, zgódźmy się, że jeśli twierdzimy coś zasadnie o jakimś przedmiocie, to mamy do tego przedmiotu dostęp poznawczy:

<sup>20</sup> Pomijamy tu trudność dotyczącą statusu ontycznego, który przyznaje się w analizowanym wywodzie „niegodziwemu uczonemu”, którego doznania nie są chyba pozorne.

(2)  $\wedge x \wedge y$  ( $x$  twierdzi zasadnie, że  $Py \Rightarrow x$  ma dostęp poznawczy do  $y$ ).

Po drugie, przedmiot pozorny to przedmiot nierzeczywisty:

(3)  $\wedge y$  ( $y$  jest przedmiotem pozornym  $\Leftrightarrow y$  nie jest przedmiotem rzeczywistym).

Po trzecie, tylko rzeczywiste byty posiadają zasadne przekonania:

(4)  $\wedge x \wedge y$  ( $x$  twierdzi zasadnie, że  $Py \Rightarrow x$  jest przedmiotem rzeczywistym).

Zgódźmy się, że w (4) twierdzi się coś o  $x$ -ach, tj. o przedmiotach posiadających zasadne przekonania.

Zapytajmy teraz: czy jest ktoś, kto twierdzi zasadnie, że (4):

(5)  $\forall z$  ( $z$  twierdzi zasadnie, że (4))?

Zauważmy, że z (1)-(3) wynika, że:

(6)  $\wedge x \wedge y$  ( $x$  twierdzi zasadnie, że  $Py \Rightarrow y$  nie jest przedmiotem rzeczywistym).

Załóżmy teraz, że odpowiedź na pytanie (5) jest pozytywna, czyli:

(7)  $\forall z$  ( $z$  twierdzi zasadnie, że (4)).

Przypuśćmy, że  $A$  byłoby takim  $z$ -em. Na mocy (7) i (6) byłoby, że:

(8)  $A$  nie jest przedmiotem rzeczywistym.

Natomiast na mocy (7) i (4) byłoby z kolei, że:

(9)  $A$  jest przedmiotem rzeczywistym.

Okazuje się, że założenie (7) prowadzi na gruncie poprzednich założeń do sprzeczności. Wynika z tego, że:

(10)  $\sim \forall z$  ( $z$  twierdzi zasadnie, że (4)).

Inaczej mówiąc — na gruncie przyjętych założeń — nikt nie może twierdzić zasadnie, że (4). W konsekwencji — nie może też twierdzić zasadnie, że (6). Jeśli zaś uznamy (6) za adekwatną interpretację tezy, że świat jest iluzją, to trzeba się zgodzić, że nikt tej tezy — nazwijmy ją dla uniknięcia historycznych obciążeń innych terminów „pan-iluzjonizmem” — nie może jej zasadnie stwierdzić (oczywiście przy wyszczególnionych założeniach).

„Opowiadka” o mózgu w naczyniu jest — w naszym przekonaniu — niczym innym jak sugestywną ilustracją tego typu rozumowania.

## 12. KONKLUZJA

Nie dziwi nas, że eksperymenty myślowe są tak bardzo popularne wśród filozofów, a zwłaszcza wśród czytelników niektórych pism filozoficznych.

Jak pokazaliśmy — są to bądź «niedoszłe» eksperymenty rzeczywiste, bądź stimulatory do ujawniania poglądów osób tym eksperymentom «poddanych», w szczególności do ujawniania intuicji semantycznych takich osób. Poza funkcją ilustracyjną i *a fortiori* perswazyjną — filozoficzne eksperymenty myślowe nie pełnią żadnej doniosłej funkcji teoretycznej. Stanowią jednak egzemplifikacje — nieraz pomysłowe i dowcipne — odpowiednich «czystych» rozumowań, a że już od najmłodszych lat lubimy czytać książki z... obrazkami — stąd ich wspomniane wzięcie. Nie byłoby skądinąd niczego złego w posługiwaniu się nimi, gdyby teksty, w których są opi-

sywane, zawierały ponadto jasne wysłowienie bronionych twierdzeń i przejrzyste przedstawioną argumentację. Niestety — znanym nam tekstom filozoficznym, ubarwionym opisami eksperymentów myślowych, jasności i przejrzystości — nie dostaje.

Być może zresztą stąd właśnie bierze się wielość prac, w których się te teksty komentuje. Jak wiadomo, im tekst mniej jasny i przejrzysty, tym więcej ma możliwych interpretacji. W owych komentarzach przy tym zamiast (a nie — obok!) jasno sformułowanych twierdzeń i przejrzystych argumentacji pojawiają się na ogół — nowe eksperymenty myślowe...

## BIBLIOGRAFIA

- Ajdukiewicz, Kazimierz  
(1931) Paradoxy starożytnych, [w:] *Język i poznanie*. Tom I, Warszawa 1960, PWN, s. 137-144.
- Brendel, Elke  
(2004) Intuition Pumps and the Proper Use of Thought Experiments, *Dialectica* vol. 58, No. 1, s. 89-108.
- Brożek, Anna, Jacek Jadacki  
(2012) Eksperymenty myślowe w nauce, *Filozofia Nauki*, nr 1, s. 5-32.
- Frege, Gottlob  
(1893) *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*. Band I, Jena, Verlag Hermann Pohle.
- Gettier, Edmund L.  
(1963) Is Justified True Belief Knowledge?, *Analysis* vol. 23, No. 6, s. 121-123.
- Hajduk, Zygmunt  
(2002) Eksperyment, [w:] Andrzej Maryniarczyk *et al.* (red.), *Powszechna encyklopedia filozofii*. Tom 3, Lublin, Polskie Towarzystwo Tomasza z Akwinu, s. 83-84.
- Hobbes, Thomas  
(1642) *Elementy filozofii*, Warszawa 1956, PWN.
- Locke, John  
(1690) O tożsamości i różnicy, [w:] *Rozważania dotyczące rozumu ludzkiego*. Tom I, Warszawa 1955, PWN, s. 460-493.
- Nowaczyk, Adam & Żołądowski, Zenobiusz  
(1974) Eksperyment, [w:] *Logika i metodologia badań naukowych dla lekarzy*, Warszawa, PZWL, s. 152-153.
- Putnam, Hilary  
(1975) Znaczenie wyrazu „znaczenie”, [w:] Putnam 1998, s. 93-184.  
(1983) Mózgi w naczyniu, [w:] Putnam 1998, s. 295-324.  
(1998) „Wiele twarzy realizmu” i inne eseje, Warszawa 1998, Wydawnictwo Naukowe PWN
- Quine, Willard van Orman  
(1960) *Word and Object*, Cambridge (Massachusetts), The MIT Press.

Schrödinger, Erwin

(1935) Die gegenwärtige Situation in der Quantenmechanik, *Naturwissenschaften* B. XXIII, s. 807-812, 823-828, 844-849.

Searle, John S.

(1984) Czy komputery mogą myśleć?, [w:] *Umysł, mózg i nauka*, Warszawa 1995, Wydawnictwo Naukowe PWN, s. 25-37.

(1992) *The Rediscovery of Mind*, Cambridge (Massachusetts) 1994, The MIT Press.

Thomson, Judith Jarvis

(1985) The Trolley Problem, [w:] *Right, Restitution, and Risk. Essays in Moral Theory*, Cambridge (Massachusetts) & London (England) 1986, Harvard University Press, s. 94-116.

Wierzbicki, Piotr

(2011) *Boski umysł. Esej o tropieniu świata*, Warszawa, Wydawnictwo Sic!.

Wojtysiak, Jacek

(2007) Słownik wybranych terminów filozoficznych, [w:] Antoni B. Stępień, *Wstęp do filozofii*, Lublin, Towarzystwo Naukowe KUL, s. 327-438.