

# Studia z Kognitywistyki i Filozofii Umysłu

**Tom 8/Nr 1**

**Poznań 2014**

**ISSN 2082-7083**



---

# Studia z Kognitywistyki i Filozofii Umysłu

Tom 8/Nr 1

Poznań 2014

## REDAKCJA

**Redaktor naczelny:**

Andrzej Klawiter

**Z-ca red. naczelnego:**

Józef Bremer

**Sekretarz redakcji:**

Bolesław Czarnecki

**Redaktorzy merytoryczni:**

Andrzej Dąbrowski

Wioletta Dziarnowska

Sebastian Kołodziejczyk

Piotr Przybysz

Mariusz Urbański

Maciej Witek

**Redaktor działu recenzji:**

Tomasz Komendziński

**Redaktor techniczny:**

Maja Kittel

**Redaktor elektroniczny:**

Sebastian Stankiewicz

## ADRES

Instytut Psychologii UAM

ul. Szamarzewskiego 89a

60-568 Poznań

pok. 97

<http://www.skfu.amu.edu.pl>

[skfu@amu.edu.pl](mailto:skfu@amu.edu.pl)

Skład w systemie L<sup>A</sup>T<sub>E</sub>X

K. Paluszkiewicz

---

## Spis treści

ANDRZEJ DĄBROWSKI

*Wprowadzenie* **5**

MAREK LECHNIAK

*Reprezentacjonistyczne ujęcie przekonań* **7**

ANDRZEJ BIŁAT

*W stronę ekstensjonalnej teorii przekonań* **23**

TOMASZ A. PUCZYŁOWSKI

*Czy można być przekonanym, że Sherlock Holmes jest detektywem? O sensowności zdań z nazwami fikcyjnymi* **34**

ADRIANNA SMURZYŃSKA

*Kognitywne teorie urojeń — o racjonalności urojeń* **51**

BARTOSZ JANIK

*Próba oceny Bayesowskiego modelu anomalnego wnioskowania w procesie formowania się urojeń* **67**

---

# Wprowadzenie

ANDRZEJ DĄBROWSKI

Tworzenie przekonań to jedna z najbardziej podstawowych i zarazem najważniejszych umiejętności umysłu ludzkiego. Człowiek myśli, podejmuje decyzje i działa m.in. w oparciu o przekonania. Przez przekonanie rozumie się zwykle postawę, w której pewien podmiot *P* uznaje *x* (sąd lub treść zdania) za subiektywnie pewne lub przynajmniej wysoce prawdopodobne. W zależności od kontekstu nabywania i stosowania przekonań istnieje wiele ich rodzajów: (a) dotyczące samego siebie, (b) zdroworozsądkowe, (c) naukowe, (d) polityczne, (e) moralne, (f) religijne, (g) estetyczne i inne. Kwestia przekonań to interdyscyplinarna i niezwykle bogata problematyka. W obecnym tomie SKFU podejmujemy rudymenarne kwestie związane z naturą, uzasadnieniem i racjonalnością przekonań.

Istnieją różne teoretyczne stanowiska w kwestii przekonań. Należą do nich m.in.: reprezentacjonizm, funkcjonalizm, dyspozycjonalizm, interpretacjonizm. Pierwszy artykuł w prezentowanym tomie, *Reprezentacjonistyczne ujęcie przekonań* Marka Lechniaka, dotyczy pierwszego ze wskazanych rozwiązań. Autor artykułu wychodzi od zrekonstruowania ogólnej struktury przekonania (stan rzeczy, akt i treść przekonania, podmiotowy korelat i językowy odpowiednik), a następnie poddaje analizie reprezentacjonizm, zgodnie z którym przekonania wiążą się z reprezentacjami w umyśle, mającymi swój nośnik i treść. Na tych też aspektach, jako kluczowych dla rozumienia reprezentacji, koncentrują się rozważania Lechniaka. Autor artykułu rozważa argumenty za i przeciw takiemu podejściu.

W artykule *W stronę ekstensjonalnej teorii przekonań* A. Biłat analizuje możliwość dołączenia do ogólnej, ekstensjonalnej teorii przekonań (jako postaw propozycjonalnych) niektórych zasad uznawania sądów, jakie leżą u podstaw intensjonalnej logiki epistemicznej. Wówczas język takiej teorii przekonań nie może być jednak scharakteryzowany jako typowy język przedmiotowy. Biłat, wbrew skrajnemu ekstensjonalizmowi, postuluje przywrócenie bytów intensjonalnych, niegdyś skazanych przez Quine'a na „ontologiczną banicję”. Analizy te pokazują, że łączenie ekstensjonalnych teorii z intensjonalnym dyskursem jest nie tylko możliwe, ale — przynajmniej w niektórych przypadkach — potrzebne i korzystne.

Zdania, zawierając pewne nazwy, mogą stwierdzać określone stan rzeczy lub mówić coś o przekonaniach użytkowników języka. Wszelako istnieją różne nazwy, m.in. takie, które nazywają obiekty istniejące czasoprzestrzennie oraz takie, które niczego i nikogo nie nazywają (nazwy fikcjonalne). Wielu uważa, że zdania z nazwami fikcjonalnymi np. *Święty Mikołaj przynosi prezenty* są pozbawione wartości logicznej i informacyjnej. Nie odmawiają jednak wartości informacyjnej zdaniom o przekonaniach np. *Antek wierzy, że Święty Mikołaj przynosi prezenty*. To ostatnie zdanie jest informatywne w tym sensie, że pozwala przewidywać zachowanie Antka. Czy rzeczywiście tak jest? Między innymi tym problemem w artykule *Czy można być przekonanym, że Sherlock Holmes jest detektywem?* O sensowności zdań z nazwami fikcjonalnymi zajmuje się Tomasz A. Puczyłowski.

Autor broni stanowiska, zgodnie z którym proste zdania, zawierające nazwy fikcyjne, nie wyrażają żadnego sądu i nie mają wartości logicznej, prawdy lub fałszu. Przede wszystkim jednak pokazuje jak wyjaśnić rzekomą sensowność zdań o przekonaniach z nazwami fikcyjnymi.

Dwa ostatnie artykuły dotyczą przekonań urojeniowych. Adrianna Smurzyńska w artykule *Kognitywne teorie urojeń — o racjonalności urojeń* koncentruje się przede wszystkim na formalno-poprawnościowym aspekcie urojeń. Autorka analizuje wybrane kognitywne teorie (lub proto-teorie) przekonań urojeniowych, których podstawowym celem jest wyjaśnienie genezy urojeń. Za przyczynę ich powstawania zwykle uważa się jakąś formę zaburzeń: zaburzenie w rozumieniu i operowaniu twierdzeniami ramowymi lub zaburzenie powszechnie podzielanych przekonań zdroworozsądkowych (zaburzenie Tła), zaburzenie mechanizmów atrybucyjnych (np. zaburzenie mechanizmu monitorowania własnych myśli i działań) lub zaburzenia związane z funkcjonowaniem wyobraźni. Analizowane teorie autorka poddaje ocenie.

Natomiast Bartosz Janik w artykule *Próba oceny Bayesowskiego modelu anomalnego wnioskowania w procesie formowania się urojeń* prezentuje dwuczynnikowy model tworzenia przekonań urojeniowych M. Colhearta, w którym decydującą rolę odgrywają dysfunkcje poznawcze (np. niewłaściwa reakcja emocjonalna na bodźce percepcyjne). Model ten Janik uzupełnia Bayesowskim modelem wnioskowania probabilistycznego (przekonaniom naszym towarzyszy określony poziom prawdopodobieństwa), zaproponowanym przez R. McKay'a. Autor nie przyjmuje modelu tego bezkrytycznie i w końcowej części artykułu wyraźnie wskazuje na pewne trudności, które zasadniczo pokrywają się z klasycznymi już zarzutami wobec wnioskowań Bayesowskich.

---

# Reprezentacjonistyczne ujęcie przekonań

MAREK LECHNIAK  
Wydział Filozofii KUL

**Streszczenie.** *W artykule zostanie przedstawione stanowisko, zgodnie z którym przekonanie stanowi mentalną reprezentację stanu rzeczy. Analizie poddamy zatem reprezentacyjne ujęcie postaw propozycjonalnych i racje za tym podejściem. Zaprezentujemy również uwagi dotyczące minimalnego zasobu wyrażeń języka myśli niezbędnego do zapewnienia logicznego charakteru przekonań oraz spostrzeżenia na temat nabywania i zmiany przekonań.*

**Słowa kluczowe:** *przekonanie, reprezentacjonizm, treść, język myśli*

## 1. Momenty (składniki) przekonania

Zacznijmy od podstawowych ustaleń. Weźmy najprostszy kontekst, w którym używamy zwrotu „jest przekonany, że”, np. założmy, że Jan jest przekonany, że w Lublinie pada deszcz. Co to wyrażenie znaczy? Jakie warunki należy spełnić, żeby sformułować takie twierdzenie? Założmy na wstępie, że faktycznie istnieją przekonania, czy inne tego typu „postawy”, takie jak pragnienia, przypuszczenia, wątplenia, oczekiwania, przypomnienia, obawy, chęci, itp.<sup>1</sup>. Dla porządku będziemy mówić o byciu przekonanym podmiotu  $S$ , że  $p$  (o  $p$ ), albo o żywieniu przez  $S$  przekonania, że  $p$ , ażeby uniknąć mówienia o samych przekonaniach<sup>2</sup>.

Mamy więc następujący stan rzeczy polegający na tym, że w dniu 20 czerwca 2013 roku w Lublinie pada deszcz. Jan spostrzega to i wierzy swojemu spostrzeżeniu, że tak się dzieje lub słyszy od świadka tego zdarzenia zdanie „Teraz (20 czerwca 2013 r.) w Lublinie pada deszcz” ( $p$ ) i wierzy temu świadkowi. W rezultacie do jego umysłu „dociera” sąd, że teraz pada deszcz. Jeśli ten sąd jakoś „zadomowi” się w umyśle Jana (niekoniecznie

---

<sup>1</sup>Piszemy „postawy” w cudzysłowie, żeby podkreślić, że nie zakładamy tu, iż są to postawy sensu stricte — chodzi tylko o jakieś odniesienie do czegoś. Fodor podaje następujące warunki, które powinna spełniać teoria postaw propozycjonalnych:

- a) postawy propozycjonalne powinny być analizowane jako relacje;
- b) teoria powinna wyjaśniać paralelizm pomiędzy czasownikami dotyczącymi postaw a czasownikami mówienia (warunek Vendlera);
- c) teoria powinna ujmować nieprzejrzystość postaw propozycjonalnych (warunek Fregego);
- d) przedmioty postaw propozycjonalnych winny mieć formę logiczną (warunek Arystotelesa);
- e) teoria powinna być zgodna z empirycznym ujęciem procesów umysłowych; (por. Fodor 1978).

Przyjęcie obowiązywalności tych warunków (co nie jest sprawą bezdyskusyjną) prowadzi do odrzucenia niektórych koncepcji, np. teorię stanów monadycznych albo, być może, koneksjonistyczne ujęcia przekonań.

<sup>2</sup>Mówienie o przekonaniach prowadzi do wielu niepotrzebnych trudności związanych z pomieszaniem sądów z postawami, zachowaniami, treściami zdań czy wręcz ze zdaniem.



w jego świadomości), można powiedzieć, że Jan nabył (zaczął żywić) przekonanie, że *p*. Teraz zapytamy, co stało się z Janem w tym momencie (oczywiście, Jan nie musi sobie zdawać z tego sprawy, czasem może też długo dochodzić do bycia przekonanym)? Czy jakiś nowy byt „zadomowił się” w umyśle Jana? Jaka jest jego natura (jeśli faktycznie jest to coś, czego dotąd w umyśle Jana nie było)?

Na podstawie powyższego szkicu można wyróżnić następujące momenty bycia przekonanym<sup>3</sup>:

- a) przedmiot przekonania: stan rzeczy, obiekt, którego zachodzenie, istnienie, podmiot uznaje;
- b) treść przekonania: sąd w sensie logicznym, stwierdzający istnienie przedmiotu przekonania;
- c) akt (moment) uznania istnienia stanu rzeczy będącego przedmiotem przekonania (sądu za prawdziwy) — ów moment decyduje o nabyciu przekonania;
- d) podmiotowy korelat przedmiotu przekonania (coś w umyśle, w podmiocie, co odpowiada przedmiotowi przekonania);
- e) językowy odpowiednik (wyraz) treści przekonania;
- f) uposażenie bytowe podmiotu, które umożliwia mu żywienie przekonania;
- g) typ odniesienia do przedmiotu, „jakość nastawienia (aktu), która decyduje o tym, że dane odniesienie (nastawienie, akt) jest przekonaniem, wątpieniem, przypuszczeniem czy innym nastawieniem intencjonalnym, takim jak pragnienie czy oczekiwanie<sup>4</sup>.

Oczywiście niektóre z powyższych momentów mogą zostać zakwestionowane albo stać się przedmiotem dyskusji (co zaraz będzie widoczne, gdy przejdziemy do omówienia różnych stanowisk), ale jak się zdaje, ich wprowadzenie przyczynia się do uporządkowania dyskursu o przekonaniach. W szczególności trzeba stwierdzić, że przedmiot przekonania nie musi być realnie, transcendentnie względem podmiotu, istniejącym przedmiotem, ale raczej czymś takim, jak przedmiot aktu intencjonalnego (wszak można żywić prawdziwe przekonania dotyczące czegoś nieistniejącego lub fałszywe przekonania o czymś istniejącym). Dlatego trudno uznać, że relacja łącząca podmiot z obiektem przekonania ma charakter realny (przyczynowy)<sup>5</sup>. Trzeba również dodać, że rzeczywistość przekonań jest dynamiczna, tzn. poszczególne wyżej wymienione

<sup>3</sup>Por. Lechniak (2011, s. 87–111).

<sup>4</sup>Tu pozostaniemy przy odniesieniu charakteryzującym się uznaniem, że jest tak a tak, wyrażanym w stwierdzeniu. To nastawienie nazywa się właśnie byciem przekonanym; oczywiście, powyższe punkty pozwalają scharakteryzować odpowiednio inne nastawienia intencjonalne, np. w chceniu będziemy mieli jakiś przedstawiony stan rzeczy, który podmiot chce osiągnąć; w tym nastawieniu nie występuje moment uznania prawdziwości; obecny jest za to moment decyzji, czyli uznania, że sąd stwierdzający istnienie tego stanu rzeczy będzie prawdziwy, etc.

<sup>5</sup>Można, odróżnić dwa podstawowe rozumienia intencjonalności: relacyjne, według którego relacja intencjonalna ma dwa człony, chociaż drugi z nich nie musi być przedmiotem realnie istniejącym (ale subsystemnym i immanentnym); np. akt wyobrażenia sobie złotej góry ma za przedmiot złotą górę (intencjonalny korelat tego aktu). Korelatami aktu intencjonalnego nie muszą być zatem przedmioty realne; akceptując teorię relacyjną musimy rozdzielić każdy radykalnie transcendentny obiekt na wiele deskryptywnie rozdzielnych wystąpień simulacrow obiektu i te wystąpienia mogą grać rolę intencjonalnych korelatów aktów; kierunkowe, według którego główna rola w intencjonalności odgrywana jest przez intencjonalny pośrednik, który determinuje kierunek aktu intencjonalnego oraz jest odpowiedzialny za „jak” aktu

momenty mogą ulegać zmianie; tak więc np. przekształcona może zostać jakoś odniesienia do przedmiotu przekonania, treść przekonania, może też nastąpić wycofanie się podmiotu (na skutek nowych informacji) z żywienia przekonania (uznawania). Z kolei ważne jest również to, że sądy będące treścią przekonania są powiązane logicznie i tylko niektóre z nich są przekonaniem żywionym *explicite*.

## 2. Czym są, o ile istnieją, przekonania?

Problem natury przekonania wpisuje się w szerszą dyskusję o naturze stanów umysłowych. Stanowiska w tej debacie można podzielić na realistyczne i antyrealistyczne. Pierwsze z nich przyjmuje istnienie takich obiektów jak przekonania (choć mogą być różnie rozumiane), drugie zaś twierdzi, że nie warto przyjmować zobowiązań ontologicznych w kwestii istnienia przekonania (czy innych postaw propozycjonalnych), a mówienie o nich jest albo użyteczną fikcją (instrumentalizm), albo może być zastąpione przez teorie wyrażone w języku psychologii funkcjonalistycznej, która redukuje wyjaśnianie posługujące się pojęciem przekonania do wyjaśnień, korzystających jedynie z pojęć funkcjonalnych. Według Fodora, tak rozumiany funkcjonalizm głosi istnienie pośredniego poziomu wyjaśniania między psychologią zdroworozsądkową a wyjaśnianiem neurologicznym, przy czym wyjaśnienia zdroworozsądkowe redukują się do wyjaśnień funkcjonalnych:

przekonania i pragnienia są stanami funkcjonalnymi, a każdemu (prawdziwemu) wyjaśnieniu psychologicznemu odpowiada interpretacja wyrażana w kategoriach „nauki twardej” mówiąca o tym, w jaki sposób stany funkcjonalne, które ona postuluje są „realizowane” w badanym systemie. Wiele różnych interpretacji w kategoriach „nauki twardej” może odpowiadać jednemu i temu samemu wyjaśnieniu funkcjonalnemu, jak widzieliśmy bowiem, kryteria wyodrębniania egzemplarzy stanów funkcjonalnych abstrahują od ich fizycznego charakteru (Fodor 1999, s. 26)<sup>6</sup>.

Z kolei realizm głosi istnienie stanów, które nazywamy przekonaniem oraz przypisuje im cechy, takie jak wchodzenie w przyczynowe relacje z zachowaniami czy innymi stanami umysłowymi oraz posiadanie treści i intencjonalność. Wśród realistów można z grubsza odróżnić dyspozycjonistów i reprezentacjonistów<sup>7</sup>. Według tych pierwszych, umysłowa struktura wewnętrzna (istnienie owego podmiotowego korelatu przedmiotu przekonania) ma niewielki wpływ na to, czym jest przekonanie — żywienie przekonania (czy innej „postawy”), że *p*, polega na posiadaniu określonej dyspozycji względem *p* (choćby aktualny stan mózgu człowieka, który wierzy, że *p*, różni się od stanu mózgu człowieka, który w to nie wierzy, istnienie takich stanów mózgu jest według behawioryzmu (za intensjonalny aspekt intencjonalności). Ten intencjonalny pośrednik nie jest tożsamy z intencjonalnym obiektem, ale też nie jest częścią właściwą aktu intencjonalnego; por. Pańiczek 1996.

<sup>6</sup>Por. także, Ciecierski (2012). N. Block (1980) odróżnia trzy różne rozumienia funkcjonalizmu: metodologiczne, czyli analizę funkcjonalną, funkcjonalizm analizy reprezentacji komputacyjnej oraz funkcjonalizm metafizyczny, według którego stany mentalne mają charakter funkcjonalny, i winny być charakteryzowane właśnie za pomocą ich roli przyczynowej względem bodźców zmysłowych czy rezultatów zachowaniowych.

<sup>7</sup>Według Fodora można odróżnić pięć głównych stanowisk realistycznych: antyfunkcjonalizm, funkcjonalizm monadyczny, zakładający semantykę ról pojęciowych, funkcjonalizm monadyczny, odrzucający semantykę ról pojęciowych, funkcjonalizm relacyjny z semantyką ról pojęciowych i funkcjonalizm relacyjny, który odrzuca taką semantykę; por. Ciecierski (2012, s. 329).

obojętne dla pojęcia przekonania). Mieć przekonanie, że  $p$  to tyle, co podtrzymywać sąd, że  $p$  (rozumieć zdanie  $p$ ) oraz mieć dyspozycję do działania, jak gdyby ten sąd był prawdziwy (Braithwaite 1973)<sup>8</sup>. Dyspozycyjny charakter bycia przekonanym podkreśla, np. Armstrong (1982), który twierdzi, że:

przekonanie jest dyspozycyjnym stanem umysłu, trwającym dłuższy lub krótszy czas i mogącym przejawiać się lub nie w okresie swego trwania a postrzeżenie, stanowiące początek przekonania, jest nabywaniem stanu dyspozycyjnego, które nie jest ani stanem, ani procesem, lecz zdarzeniem<sup>9</sup>.

Dyspozycyjny charakter owego stanu jest niezbędny, ponieważ bycie przekonanym jest względnie trwałe, tzn. nie ustępuje, np. w czasie snu czy nieprzytomności. Jeśli zgodzimy się na takie określenie bycia przekonanym, to trzeba zapytać, co jest źródłem owej dyspozycji, na czym ona się opiera — czy istnieje w umyśle podmiotu jakiś czynnik, który pozwala utrzymać dyspozycyjną własność bycia przekonanym nawet podczas, np. stanu nieprzytomności<sup>10</sup>?

Jeśli więc mówienie o dyspozycjach nie tłumaczy natury przekonań, gdyż odsyła do własności podmiotu przekonań, w których te dyspozycje są ugruntowane, to pozostaje przyjąć, że albo treść przekonania (czy innej postawy propozycjonalnej) jest pewną monadyczną własnością postawy propozycjonalnej, która przynależy do struktury izomorficznej względem jakiejś struktury inferencyjnej (Fodor 1999, s. 29–33), albo istnieje w umyśle podmiotu jakiś element (stan?), umożliwiający podmiotowi znajdowanie się w stanie (dyspozycyjnym), zwanym byciem przekonanym. Taki element można nazwać reprezentacją. Według Fodora, reprezentacjonizm (reprezentacyjna teoria umysłu) głosi, że:

dla każdego organizmu  $O$  i dla jakiejś postawy  $P$  zajmowanej wobec sądu w sensie logicznym  $p$  zachodzi relacja  $R$  i [istnieje] reprezentacja umysłowa  $RM$  taka, że  $RM$  znaczy  $p$  (wyraża sąd  $p$ ) i  $O$  jest w stanie  $P$  (np. jest przekonany, że  $p$ ) wtedy i tylko wtedy, gdy organizm  $O$  jest w relacji  $R$  do  $MP$  (Fodor 1991, s. 16; Fodor 1999, s. 34).

Zwykle za Ch. S. Peirce’em wymienia się cztery podstawowe wskaźniki (aspekty) reprezentacji: nośnik reprezentacji, treść, relacje między przedmiotem a treścią i możliwość interpretacji reprezentacji (von Eckardt 1999; Żegleń 2012). Relacja pomiędzy przedmiotem a treścią powinna to być relacją reprezentowania, czyli relacją

<sup>8</sup>Dyspozycjonizm znalazł szczególnie wyraz w behawioryzmie logicznym Ryle’a; obszerną analizę zalet i wad tego ostatniego stanowiska można znaleźć, np. w: Armstrong (1982, s. 77–103).

<sup>9</sup>Dyspozycjonizm nie musi jednak utożsamiać się z behawioryzmem. Dowodzi tego Armstrong. Choć behawioryzm głosi redukcję wewnętrznych stanów umysłowych do zachowań obserwowalnych, to jednak takie zjawiska, jak obliczanie w głowie czy sen nie mogą być całkowicie adekwatnie opisane za pomocą kategorii behawiorystycznych. Ale samo pojęcie dyspozycji sięga głębiej.

<sup>10</sup>Dyspozycja jest to zdolność, moc czy sposobność przedmiotu do działania w pewien określony sposób w pewnych określonych warunkach. Jeśli stoimy na gruncie dyspozycjonalnego realizmu, głosimy, że „dyspozycyjne określenia odnoszą się do określonych stanów przedmiotu na mocy własności, które te przedmioty posiadają, a nie są tylko stwierdzeniami o przeszłym lub możliwym zachowaniu. Same własności są co najmniej z natury dyspozycyjne i są zdolne do wspierania wielu stanów dyspozycyjnych swoich nosicieli”; stwierdzenie to wydaje się ważne w kontekście rozważań nad przekonaniem (por. Williams 2005). Warto podkreślić za Williamsem, że dyspozycje nie zawsze muszą się manifestować; mogą one trwać jako realne składniki przedmiotu, nawet jeśli ich manifestacja nigdy nie będzie miała miejsca.

wskazywania, „bycia o czymś”. Warunkiem bycia reprezentacją jest odniesienie do czegoś, co jest źródłem treści. Innymi słowy, żeby teoria zasługiwała na miano reprezentacjonistycznej, winna mieć charakter „semantyczny”<sup>11</sup>, przy czym często mówi się tu o kowariancji, która zachodzi pomiędzy systemem poznawczym a otoczeniem, co oznacza, że zmiany zachodzące na zewnątrz systemu są odpowiednio skorelowane ze zmianami zachodzącymi w systemie na skutek związków przyczynowych systemu poznawczego z otoczeniem (Żegleń 2012, s. 220). Interpretacja zaś jest warunkiem koniecznym mówienia o reprezentacji. Informacja ma charakter reprezentacyjny jedynie wtedy, gdy jest zinterpretowana przez system poznawczy, czyli gdy ma znaczenie dla osoby, w której umyśle rezyduje. (Interpretant mentalnej reprezentacji  $R$  dla pewnego podmiotu  $S$  składa się ze zbioru wszystkich możliwych determinujących procesów kontyngentnych ze względu na przyjęcie  $R$  w  $S$ , np. za pomocą pamięci).

Pozostałymi składnikami reprezentacji są nośnik reprezentacji, czyli jakiś materialny substrat, który umożliwi materialne zapodmiotowanie reprezentacji oraz treść, czyli to, o czym jest reprezentacja, mentalne „odwzorowanie” odpowiednio ujętego przedmiotu przekonania (czy innej postawy propozycjonalnej). Ponieważ w perspektywie reprezentacjonizmu te dwa czynniki mają decydujące znaczenie, poświęcimy im nieco więcej miejsca.

### 3. Reprezentacjonistyczny obraz przekonań

Wróćmy na chwilę do koncepcji przekonań jako funkcjonalnych stanów monadycznych.

Stany funkcjonalne są indywidualizowane jako stany należące do określonego typu przez odniesienie do ich (rzeczywistych czy potencjalnych) relacji przyczynowych. (...) Opis wszystkich stanów umysłowych będzie implikował sieć wzajemnych relacji przyczynowych. (...) Rzeczywiste życie umysłowe danego organizmu będzie musiało przedstawiać się jako ścieżka wiodąca poprzez tę sieć. (...) Każdy stan umysłowy może być utożsamiony z węzłem takiej sieci: dla każdego stanu umysłowego istnieje odpowiadająca mu rola przyczynowa oraz dla każdej przyczynowej roli istnieje odpowiadający jej węzeł. Inaczej to ujmując, każdy stan umysłowy może być powiązany z pewną formułą, która w sposób unikatowy określa jego położenie w sieci poprzez wyszczególnienie jego potencjalnych interakcji z każdym innym stanem umysłowym (Fodor 1999, s. 30).

Takie ujęcie nie określa jednak tego, co znaczy, że postawa propozycjonalna ma treść propozycjonalną. Rozwiązaniem tego problemu jest redukcja pojęcia treści propozycjonalnej do pojęcia roli przyczynowej. Innymi słowy, zakłada się, że istotne dla pojęcia sądu są jego relacje inferencyjne do innych sądów, a sieć relacji inferencyjnych jest izomorficzna względem sieci ról przyczynowych, które odgrywają postawy propozycjonalne względem odpowiadających im sądów. Tak więc sądy mogą być przypisane do poszczególnych postaw propozycjonalnych na mocy izomorfizmu

<sup>11</sup>Tę relację von Eckardt (1993, s. 51) nazywa podstawą reprezentacji umysłowej, gdyż „determinuje ona fakt, że reprezentacja ma swój przedmiot. Ta podstawa jest naturalistyczna (niesemantyczna ani nieintencjonalna), może składać się z wewnętrznych lub zewnętrznych składników, jednak dowolny składnik musi spełniać następujące ograniczenie: jeżeli dwa egzemplarze reprezentacji mają różne podstawy, i ta różnica podstawy determinuje różnicę treści, wówczas muszą one także różnić się co do kauzalnej siły wytwarzania istotnych skutków w nomologicznie możliwych kontekstach”.

między siecią relacji semantycznych a siecią relacji przyczynowych pomiędzy stanami umysłowymi (por. Fodor 1999, s. 32). Referowany tu pogląd Fodor nazywa standardowym realizmem. Jest on funkcją dwóch twierdzeń, a mianowicie twierdzenia, że postawy propozycjonalne są monadycznymi stanami funkcjonalnymi oraz tezy, że źródłem semantycznych własności postaw propozycjonalnych jest izomorfizm pomiędzy rolą przyczynową stanów a implikacyjną strukturą sądów.

Ujęcie funkcjonalne niesie jednak trudności, które mogą być usunięte przez koncepcję reprezentacjonistyczną. Przede wszystkim reprezentacjonizm usuwa niemożliwe do przyjęcia twierdzenie, że własności semantyczne postawy propozycjonalnej są determinowane przez wewnętrzną rolę przyczynową stanów umysłowych. Tym, co odróżnia własności semantyczne od przyczynowej roli stanów umysłowych, jest właśnie reprezentacja. W poglądzie reprezentacjonistycznym podstawową rolę pełni bowiem rozróżnienie nośnika reprezentacji i treści reprezentacji.

Dwa najczęściej głoszone stanowiska w kwestii natury nośnika reprezentacji odwołują się do modeli powiązanych z metaforą komputerową (koncepcja obliczeniowa) albo z metaforą „mózgową” (koncepcja koneksjonistyczna). W pierwszym z tych modeli zakłada się, że istnieje poziom obliczeniowy naszych stanów umysłowych, które mogą być ujęte jako struktury danych, które są przekształcane zgodnie z określonymi regułami. Drugie zaś z ujęć odwołuje się do modelu rozproszonego przekształcania równoległego, według którego nośniki reprezentacji stanów umysłowych są stanami aktywności węzłów sieci koneksjonistycznej. Jednak jakkolwiek rozumiałoby się strukturę nośnika reprezentacji, samo uznanie jego istnienia jest nieuniknione<sup>12</sup>.

Następną kluczową dla rozumienia reprezentacji kwestią jest rozumienie jej treści. Dyskusja o treści reprezentacji w odniesieniu do postaw propozycjonalnych jest bezpośrednio związana z debatą, która dotyczy rozumienia sądów. Odróżnia się m. in. następujące koncepcje sądów: modalną, według której sądy są zbiorami wartościowań, takich jak możliwe światy, syntaktyczną, wedle której sądy są syntaktycznymi obiektami, takimi jak zinterpretowane zdania języka naturalnego i teoriomnogościową, według której sądy są strukturami teoriomnogościowymi. Zaletą pierwszej z teorii jest to, że u jej podstaw leży intuicja, iż sądy dotyczą sortowania sposobów, na jakie rzeczy mogą być. Sąd w sensie logicznym to zbiór światów możliwych, w którym zdanie, którego znaczeniem jest sąd, jest prawdziwe. Podstawową zaś wadą tej koncepcji jest własność zwana gruboziarnistością (*coarse-grained*), polegająca na tym, że dowolne dwa zdania, które są prawdziwe w tych samych światach możliwych, są nierozróżnialne, gdyż mają tę samą treść (zdania są synonimiczne, gdy są prawdziwe w tych samych światach możliwych), np. ponieważ wszystkie twierdzenia matematyki są prawdziwe we wszystkich możliwych światach, to powinny mieć tę samą treść.

Główną zaletą teorii syntaktycznej, głoszącej, że treści sądów są istnościami o strukturze zdaniopodobnej, jest to, że umożliwia ona odróżnianie pomiędzy treściami, których teoria możliwych światów nie mogła rozdzielić (w przeciwieństwie do gruboziarnistości teorii modalnej jest to podejście drobnoziarniste (*fine-grained*)). Jednak założenie, że zdania są myślami, które są ujmowane jako byty języka naturalnego, niesie ze sobą pewne koszty, albowiem nie pozwala ono uznać, że dwa zdania głoszące to samo w różnych językach (np. „Snow is white” i „La neige est blanche”) mają tę samą treść (inna trudność polega na tym, że zdanie *p* oraz „*p* i *p*” winny mieć

<sup>12</sup>Czasem reprezentacjami nazywa się zarówno sam nośnik, jak i to, co on niesie (treść); funkcja nośnika jest podobna do funkcji substratu materialnego znaku, który sam nie będąc znakiem, jest niezbędny do tego, by znak pełnił swe funkcje.

różną treść) — ta trudność może być usunięta, jeśli uznamy, że treść sądów jest wyrażona w języku myśli. Z kolei podejście teoriomnogościowe przyjmuje, że sądy są złożone ze składników, zespolonych ze sobą na sposób teoriomnogościowy, takich jak indywidualne obiekty (sposoby prezentowania tych obiektów), które korespondują z imionami własnymi, własności — korespondujące z predykatami oraz funkcje, pozostające w relacji korespondencji z funktorami zdaniowymi. Taka teoria unika problemów koncepcji możliwo-światowej, czyli jest w stanie odróżnić treści sądów logicznie równoważnych, a także unika trudności teorii syntaktycznej — mimo różnych języków, w których zdania są wygłoszone, zdania mają tę samą treść, gdyż korelaty składników tych zdań są tymi samymi obiektami. Przyjmując tę koncepcję, nie można jednak poradzić sobie z sądem „ $p$  i  $p$ ” — obiekt propozycjonalny dotyczy relacji koniunkcji, a nie  $p$  (Oppy 1998).

Gdy idzie o koncepcje treści reprezentacji przekonań wskazuje się na dwa główne podejścia: koncepcję funkcyjną (przyczynowo-pragmatyczną), która odwołuje się do modalnej koncepcji sądu oraz koncepcję lingwistyczną, posługującą się syntaktyczną koncepcją sądu. Zgodnie z pierwszą z nich stan przekonaniowy może być reprezentowany jako zbiór światów możliwych. Przekonania indywidualne są własnościami takiego stanu przekonaniowego:

przekonanie, że  $p$  jest dla tego sądu tym samym, co bycie prawdziwym we wszystkich światach możliwych w tym stanie. Jeśli w ten sposób rozważa się przekonania, jawią się one jako coś negatywnego: „wiara, że  $p$ ” jest tym samym, co przebywanie w stanie przekonaniowym pozbawionym każdego ze światów, w których  $p$  jest fałszywe (Stalnaker 1986, s. 69).

W filozofii umysłu tę koncepcję krytykuje się w szczególności za trudności, które wiążą się z implikowaną przez nią gruboziarnistością, czyli niemożnością odróżnienia między sobą sądów logicznie równoważnych. Natomiast za zaletę tego stanowiska uznaje się to, że abstrahuje się tu od wszelkich struktur innych niż inferencyjna, np. nie nakłada się na treść żadnej formy poza tą, która jest absolutnie konieczna; definicja treści, która korzysta z pojęcia możliwych światów spełnia ten warunek, ponieważ odwołuje się do relacji między systemem poznawczym a sądem, która powinna być niewrażliwa na sposób kodowania informacji<sup>13</sup>.

Wśród filozofów umysłu dominuje jednak koncepcja lingwistyczna. Według tej koncepcji, nośniki treści reprezentacji mają strukturę syntaktyczną (symboliczną), a struktura sądu (treści reprezentacji) jest funkcją struktury nośnika treści reprezentacji i treści elementów atomowych struktury syntaktycznej<sup>14</sup>. Charakterystyczną cechą reprezentacjonizmu jest to, że przyjmuje on, iż nośnik postawy propozycjonalnej jest

<sup>13</sup>Dlatego, zdaniem jej zwolenników, zaletą tej koncepcji jest możliwość wyjaśnienia możliwości posiadania przekonań przez dzieci (czy nawet zwierzęta); mimo niskich kompetencji językowych (dziecko może nie być w stanie zrozumieć stwierdzenia) dziecko jest w stanie wyeliminować pewne, choć nie wszystkie, możliwości, których nie może pogodzić z posiadaną wiedzą. Bardziej szczegółowe omówienie tej koncepcji występuje w pracy Ciecierskiego (por. Ciecierski 2012, s. 331–333).

<sup>14</sup>„Reprezentacje umysłowe mają syntaksę kombinatoryczną i semantykę, w której: a) rozróżnia się reprezentacje o strukturze atomowej i reprezentacje o strukturze molekularnej; b) reprezentacje o strukturze molekularnej mają składowe syntaktyczne, które są albo strukturalnie molekularne albo strukturalnie atomowe; c) semantyczna treść reprezentacji molekularnej jest funkcją treści semantycznej swoich składowych syntaktycznych wziętych wraz z jej strukturą składniową” (Fodor, Pylyshyn 1988, s. 12–13).

(inaczej niż w funkcjonalnym ujęciu umysłu) strukturalnie zinterpretowany. Treść postawy propozycjonalnej jest wyrażona przez „że-zdania”, przy czym struktura treści postawy (sądu) jest izomorficzna względem struktury wyrażającego ją zdania. Relacja pomiędzy strukturą treści a strukturą nośnika jest analogiczna do relacji między strukturą sądu a strukturą zdania języka naturalnego, wyrażającego ten sąd. Inaczej mówiąc, ponieważ nośniki postaw propozycjonalnych są ciągami symboli, które formują zdania języka myśli, to powinniśmy rozumieć relację między nośnikiem a treścią w języku myśli analogicznie do relacji między syntaktyką a semantyką systemu formalnego. W systemie formalnym mamy do czynienia z oddzieleniem syntaktyki od semantyki i, dzięki temu, mamy możliwość spojrzenia na system od dwu stron; na poziomie syntaktycznym system jest zbiorem symboli i reguł operowania nimi, natomiast poziom semantyczny określa związek między systemem formalnym i tym, o czym on jest. Semantyka języka zapewnia jego symbolom interpretację, co z kolei umożliwia przejście od symboli pozbawionych znaczenia do systemu reprezentującego. Z kolei związek między pewnymi ciągami symboli a innymi ciągami może być analizowany na poziomie syntaktycznym za pomocą relacji wyprowadzalności, albo za pomocą wynikania logicznego — poziomem semantycznym. Na poziomie języka myśli uznanie, że nośnikiem postawy propozycjonalnej jest ciąg symboli w języku myśli pozwala rozumieć relację pomiędzy nośnikiem a treścią analogicznie do relacji między syntaktyką a semantyką systemu formalnego. Zdania w języku myśli mogą być rozumiane czysto syntaktycznie, jako fizyczne struktury złożone z symboli bazowych konkatenowanych zgodnie z regułami składania, albo też mogą być rozpatrywane na poziomie semantycznym — wówczas interesuje nas sposób, w jaki reprezentują świat.

Zwraca się uwagę, że wzmiankowany wyżej izomorfizm strukturalny między nośnikiem (zdaniem języka myśli)<sup>15</sup> a treścią postawy (sądem) winien spełniać dwa warunki<sup>16</sup>:

1. Powinno być możliwe zidentyfikowanie w nośniku mojego przekonania (np., że Warszawa jest stolicą Polski) odróżnialnych fizycznych odpowiedników podstawowych elementów zdania, wyrażającego przekonanie. Te podstawowe składniki zdania wskazują na jego semantycznie podstawowe elementy, np. pojęcia, których ono dotyczy. Te odróżnialne elementy fizyczne można ujmować jako symbole, które pełnią rolę surogatów bazowych pojęć stosownego sądu, przy czym ich fizyczną odróżnialność można utożsamić z możliwością pojawienia się ich w innych strukturach fizycznych, pełniących funkcję nośnika innych postaw propozycjonalnych. To zaś jest umożliwione przez drugi wymóg, a mianowicie:
2. Istnieje funkcja odwzorowująca sposób złożenia elementów fizycznych tworzących nośnik postawy propozycjonalnej na sposób złożenia pojęć w sądzie (myśli).

Za ujęciem reprezentacjonistycznym postaw propozycjonalnych (w tym, przede wszystkim, przekonañ) przemawia to, że umożliwia ono wyjaśnienie faktów takich jak<sup>17</sup>:

<sup>15</sup>Hipoteza języka myśli jest formułowana na poziomie algorytmicznym, a nie poziomie implementacji, co umożliwia abstrahowanie od fizycznych uwarunkowań działania mózgu.

<sup>16</sup>Por. rozdział 4 *Causes in the mind* pracy (Bermudez 2005, s. 71–96); ażeby istniał wspomniany izomorfizm wewnętrzny język myśli musi mieć charakter jednoznaczny, tzn. nie może np. zawierać wieloznaczności.

<sup>17</sup>Można podać za G. Reyem (Rey 2002) następującą listę zalet (najczęściej natury empirycznej) rozwiązania reprezentacjonistycznego w wersji lingwistycznej:

a) ujmuje ono propozycjonalną strukturę postaw przez określenie obliczeniowej roli różnych

a) systematyczność przekonań, która polega na tym, że to, iż jakaś osoba może myśleć np. o relacji  $R$  między  $a$  i  $b$ , pociąga jego dyspozycję do myślenia o wszystkich innych możliwych układach owych przedmiotów pozostających w relacji  $R$  (np.  $aRb$ ,  $aRa$ ,  $bRb$ , itp.). Teoria reprezentacjonistyczna wyjaśnia również produktywność przekonań, czyli zdolność podmiotu do tworzenia z prostych myśli nieskończonej liczby myśli złożonych. Produktywność przekonań jest analogiczna do produktywności języka — reguły gramatyczne (składniowe) umożliwiają budowanie z wyrazów składowych coraz to bardziej złożonych wyrażań. Z kolei posiadanie złożonego wyrażenia języka umożliwia jego zamianę na wiele wyrażań sensownych. To dokonuje się poprzez zastępowanie jego składników przez wyrażenia należące do tej samej kategorii składniowej, czego odpowiednikiem jest systematyczność naszego myślenia<sup>18</sup>;

b) mnogość postaw propozycjonalnych o tej samej treści. W ujęciu funkcjonalnym nie jest możliwe wyjaśnienie faktu, że wiele różnych postaw (np. przekonanie i pragnienie) posiada tę samą treść z tego prostego powodu, że funkcjonalizm nie odróżnia samej postawy od jej treści (treść jest utożsamiona z rolą funkcjonalną); na poziomie wyjaśniania funkcjonalnego trzeba zatem przyjąć oddzielne ujęcie tego, że  $x$  jest przekonany, że  $p$ , że,  $x$  pragnie, żeby  $p$ ,  $x$  ma nadzieję, że  $p$ , itp. Trudno bowiem określić, co łączy te postawy: przekonanie, że  $p$  jest na ogół powodowane przez to, że coś (np. to, że pada) jest faktem, podczas gdy pragnienie, żeby  $p$  (np. żeby padało) jest wywoływane przez to, że nie jest faktem, że  $p$  (Bermudez 2005, s. 80). Natomiast w ujęciu reprezentacjonistycznym treść jest zdeterminowana przez nośnik, czyli zdanie w języku myśli. Ze względu na to, że nośnik może pełnić różne role funkcjonalne (w różnych momentach lub w tej samej chwili), dana treść propozycjonalna może występować w różnych postawach propozycjonalnych — np. w przekonaniu podmiot odnosi się do świata jako ujętego takim, jaki jest stwierdzany w zdaniu  $p$ , w pragnieniu zaś odnosi się do świata, takiego, jaki chcemy, żeby był (ten pożądaný stan świata stwierdzony jest przez zdanie  $p$ ), itd.;

c) przyczynowe oddziaływanie postaw propozycjonalnych na zachowania, które opiera się na treści postaw. To zagadnienie wiąże się z powyżej omówioną kwestią. W ujęciu funkcjonalnym proponowana jest analiza problemu przyczynowania przez treść na bazie odniesienia treści postawy propozycjonalnej do roli funkcjonalnej pełnionej przez strukturę fizyczną realizującą tę postawę. Ilekroć stan umysłowy jest włączony w określony związek przyczynowy, ów związek określa w sposób efektywny jego rolę

---

rodzajów symboli (np. predykatów, nazw własnych, zmiennych, etc);

b) jest użyteczne dla wielu teorii psychologicznych (np. teorii widzenia) — teoria lingwistyczna próbuje udzielić odpowiedzi na pytania implikowane przez teorie psychologiczne przez sugestię, że system poznawczy wykonuje obliczenia na reprezentacjach własności, a nie na samych własnościach;

c) ujmuje produktywność i systematyczność postaw propozycjonalnych;

d) ujmuje racjonalne relacje między postawami;

e) ujmuje hiperintensjonalność postaw, czyli fakt, że są one referencjalnie nieprzejrzyste oraz kauzalną skuteczność postaw (np. wyjaśnia fizyczne skutki stanów mentalnych);

f) umożliwia wyjaśnienie w jaki sposób wiele postaw może mieć ten sam przedmiot przypisując różne relacje obliczeniowe do tej samej reprezentacji wewnętrznej.

<sup>18</sup>W dyskusji z koneksjonistami Fodor i McLaughlin (1990) wskazują, że można akceptować systematyczność bez produktywności. Takie rozwiązanie przyjmują teorie koneksjonistyczne, w których jest ona gwarantowana przez cechy architektury koneksjonistycznej.



funkcjonalną. Natomiast dla reprezentacjonisty treść jest izomorficzna ze strukturą syntaktyczną nośnika reprezentacji, co umożliwia jej kauzalne (via fizycznie realizowalny nośnik) oddziaływanie na obserwowalne zachowania podmiotu;

d) mechanika przekonań i związki inferencyjne między przekonaniem. W ujęciu funkcjonalnym relacje między poszczególnymi egzemplarzami postaw propozycjonalnych są wyznaczone przez diachroniczną „mechanikę” stanów, ponieważ poszczególne postawy propozycjonalne są realizowane przez struktury fizyczne, które wchodziły w fizyczne interakcje z innymi strukturami, realizującymi inne postawy. Należy zaznaczyć, że w funkcjonalnym wyjaśnianiu zakłada się, że oddziaływania przyczynowe, w których uczestniczy stan umysłowy na mocy jego roli przyczynowej, idą tropem racjonalnych relacji pomiędzy stanami umysłowymi. Innymi słowy, mówienie o roli funkcjonalnej przekonania ma sens jedynie wówczas, gdy założymy, że częścią roli przyczynowej przekonania, że  $p$ , jest to, że ono powoduje przekonania, że  $q$  i że  $r$ , musimy uznać, że jest w jakimś sensie racjonalne uznawać, że  $q$  i  $r$ , jeśli przyjęto się, że  $p$ . Ujęcie reprezentacyjne unika tej trudności przez zakorzenienie związków między przekonaniem w strukturalnych związkach na poziomie języka myśli. Jest tak, ponieważ syntaktyka języka wyznacza pewną logiczną strukturę inferencyjną, której z kolei prawdziwościowym odpowiednikiem są związki semantyczne między stosownymi sądami. (Bermudez 2005, s. 88).

#### 4. Reprezentacjonistyczny obraz przekonań a kompetencje logiczne

Naszkicowano wyżej reprezentacjonistyczny obraz postaw propozycjonalnych i kilka argumentów na jego rzecz. Teraz warto postawić bardziej szczegółowe pytanie odnośnie do charakteryzowanego modelu przekonań (i innych postaw propozycjonalnych), a mianowicie o to, jakie składniki języka myśli są minimalne dla przyjęcia możliwości przeprowadzania operacji logicznych. To pytanie dotyczy więc w mniejszym stopniu ontologicznej natury przekonań, a bardziej stosowalności omawianej koncepcji. Jednym z podstawowych faktów dotyczących przekonań jest bowiem to, że nie wszystkie przekonania są żywione *explicite* przez podmiot<sup>19</sup>; jesteśmy przekonani nie tylko do tego, co myślimy, ale także uznajemy logiczne konsekwencje tego, co myślimy (przekonania *implicite*). W takim sensie umysł nie jest jakimś „podobnym do strychu nieuporządkowanym składowiskiem” informacji o świecie, ale raczej logicznie uporządkowanym i zdolnym do wyprowadzania wniosków systemem inferencyjnym.

O jednej z ważnych wymaganych tu cech mówił Fodor w *The Language of Thought* (Fodor 1975, s. 32–33). Jest nią „mechanizm do wyrażania własności intensjonalnych”, który będzie umożliwiawał odróżnienie aktualnych od nieaktualnych (możliwych) stanów rzeczy<sup>20</sup>. O ile bowiem język ma umożliwiać odróżnianie przekonań (dotyczących

<sup>19</sup> „Informacja jest reprezentowana *explicite* w systemie wtedy, gdy faktycznie istnieje w funkcjonalnie istotnym miejscu tego systemu fizycznie ustrukturyzowany przedmiot, formuła lub ciąg elementów systemu (języka), dla którego istnieje semantyka lub interpretacja i mechanizm dla odczytywania lub analizy składniowej formuły”; „informacja jest reprezentowana *implicite*, gdy jest implikowana logicznie przez coś, co jest magazynowane *explicite*” (Dennett 1982, s. 216).

<sup>20</sup> „Założenie, że system reprezentacyjny jest zdolny do rozróżniania między oddzielnymi możliwymi stanami rzeczy jest tym samym, co dozwoleń na uznanie, że zachowanie aktualnie produkowane jest wyborem pomiędzy tymi opcjami, które sprawca uznaje za żywe” (Fodor 1975, s. 33).

aktualnych, czyli uznanych za rzeczywiste, stanów rzeczy) od pragnień, dążeń (dotyczących przyszłych, możliwych stanów rzeczy) musi zawierać funktory modalne. Szereg elementów, które powinien zawierać język myśli, aby umożliwić przypisanie podmiotowi podstawowych kompetencji logicznych<sup>21</sup>, wymienia w pracy *Logika i psychologia* J. Macnamara (Macnamara 1993, s. 60–62). Twierdzi on, że podmiot posiada podstawowe kompetencje logiczne, jeśli język myśli zawiera określony zestaw elementów.

Zestaw ten obejmuje następujące elementy:

- a) podstawowe spójniki logiczne („nie”, „i”, „albo”, „jest konieczne”);
- b) zwroty związane z logiką nazw gatunkowych (nazw generalnych), predykatów i kwantyfikatorów („wszystkie”, „niektóre”);
- c) predykaty intencjonalne („wierzy”, „chce”);
- d) logiczne środki do ujęcia imion własnych, zwroty relacyjne (np. rodzice) i zmienne indywidualne (np. „ktoś”).

Jednocześnie kompetencja logiczna zakłada istnienie w systemie poznawczym takich elementów, jak fundamentalne zasady logiczne, a w szczególności zasada niesprzeczności, podstawowe reguły wnioskowania (np. reguły dołączania, opuszczania koniunkcji) czy środki dotyczące presupozycji zdań, rozkazów, pytań (Macnamara 1993, s. 61)<sup>22</sup>.

Składniki kompetencji logicznej mogą być realizowane dzięki dwóm typom mechanizmów, a mianowicie tzw. interpretatorom oraz implikatorom.

Aby dokonać interpretacji [zdania] niezbędny jest mechanizm wprowadzający umysł w intencjonalny kontakt z wartościami semantycznymi składników zdania i ich kombinacji. Niezależnie od trudności logicznych, zdania są zbyt bierne: nie są one zdolne do wykonywania zadań (Macnamara 1993, s. 63).

Mogą je zastąpić procedury, zwane właśnie interpretatorami, na mogące pełnić dwie funkcje: syntaktyczną, odpowiedzialną za zbudowanie wyrażenia oraz funkcję semantyczną, której zadaniem jest interpretacja wyrażenia. Z kolei implikatory są to mechanizmy do wyprowadzania wniosków (np. najprostsze reguły wnioskowania, takie jak reguły dla koniunkcji, alternatywy rozłącznej czy redukcji do absurdu). Fundamentalne zasady logiczne (zasada niesprzeczności) mają działać w ramach kompetencji logicznej za pomocą semantyki związanej z funktorem negacji oraz za pomocą reguły *reductio ad absurdum*. Dlatego też zasada niesprzeczności może być realizowana jako interpretator (funktora negacji) i jako implikator (*reductio ad absurdum*).

<sup>21</sup>Koncepcja kompetencji logicznych Macnamary jest oparta na Chomsky’ego pojęciu kompetencji językowej oraz na koncepcji języka myśli Fodora. Choć toczyła się dyskusja co do adekwatności koncepcji Macnamary (por. np. Olson 1999; Putnam 1994), przytaczam tu kilka uwag jej dotyczących, jako że stanowi ona próbę aplikacji koncepcji języka myśli.

<sup>22</sup>Trzeba tu podkreślić, że np. wybór spójników logicznych nie jest przypadkowy. Jak bowiem pokazuje Macnamara, wiele osobliwych tez Piageta na temat późnego opanowania logiki zdań przez dzieci było oparte na badaniach empirycznych, w których np. używano spójnika alternatywy łącznej, a nie rozłącznej. Jednak inni autorzy (np. Olson) poddają w wątpliwość występowanie, np. zwrotów negacyjnych wśród tych podstawowych elementów języka myśli, podkreślając, że używanie negacji jest tym, co wyróżnia myślenia człowieka od myślenia np. wysoko rozwiniętych zwierząt.

## 5. Nabywanie i zmiana przekonania

Na koniec warto zastanowić się jeszcze nad jednym problemem, mianowicie kwestią modelu nabywania i zmiany przekonań. Jest bowiem podstawowym faktem dotyczącym przekonań, że ich treść nie jest wrodzona, ale zależy od przedmiotu (świata zewnętrznego) oraz to, że przekonania mogą ulegać zmianie. W pierwszej kwestii, najprościej mówiąc, przyjmuje się, że zasadniczymi źródłami przekonań są doświadczenie (percepcja) lub świadectwo innych podmiotów, a w drugiej, że przekonania możemy nabywać, tracić czy też rewidować (zastępować wiarę w  $p$  wiarą w  $\neg p$ )<sup>23</sup>. Obie te sprawy ściśle wiążą się ze sobą — można tu mówić po prostu o dynamice przekonań.

Założmy dla uproszczenia, że jedynym źródłem informacji jest percepcja. Czasem w psychologii kognitywnej mówi się o dwóch modelach nabywania przekonań: kartezjańskim i spinozjańskim (Gilbert 1991). Według Gilberta, posiadanie przekonań na ogół wiąże się z reprezentacją mentalną oraz z pozytywną oceną informacji. Sąd jest uznawany, gdy jego znaczenie jest reprezentowane w systemie mentalnym i gdy podmiot uznaje prawdziwość reprezentacji symbolicznej. Uznawane przez podmiot zdania tworzą więc system umysłowy, ponieważ zawierają komponent reprezentacyjny oraz komponent ocenny, który odnosi się do relacji pomiędzy daną informacją a innymi informacjami już należącymi do systemu.

Według modelu kartezjańskiego w umyśle występują dwie odrębne dziedziny. Pierwszą z nich jest odpowiednik intelektu, czyli bierna, automatyczna dziedzina zrozumienia, druga zaś to przypisywana woli (czyli aktywnej, świadomej i wolnej sile umysłu) zdolność do akceptacji lub odrzucenia. Istnieje zatem zasadnicza różnica między rozumieniem (nabyciem idei) a oceną informacji, opierającą się na rozumieniu. W przeciwieństwie do Kartezjusza, w modelu spinozjańskim nie odróżnia się idei od przekonań, jako że wszystkie idee są przekonaniem. Rozumienie bowiem sądu zakłada jego akceptację, zaś odrzucanie sądu jest wtórnym aktem psychicznym, w którym „wstępna akceptacja, która nieodmiennie towarzyszy rozumieniu, zostaje cofnięta. Odrzucenie sądu jest (...) po prostu deliberatywną rewizją przekonania” (Gilbert 1991, s. 108). Innymi słowy, odrzucanie sądu jest wtórną względem uznawania operacją wycofania (pod wpływem jakichś innych informacji) akceptacji, która normalnie (jakby „z natury”) towarzyszy sądowi.

Standardowo przyjmuje się, że skoro prawdziwość sądu jest oceniana przez podmiot, to umysłowa reprezentacja tego sądu musi być w jakiś sposób „znacznikowana”, tak aby wskazywać na tę prawdziwość. W przeciwnym wypadku, podmiot ilekroć chciałby użyć danej informacji, musiałby na nowo oceniać jej wartość logiczną. Różnica między systemem kartezjańskim a spinozjańskim mogłaby być zatem przedstawiona jako różnica w zastosowanych systemach znacznikowania. Według modelu kartezjańskiego każdy sąd najpierw występuje we władzy przedstawiania, a dopiero później jest poddawany ocenie i w rezultacie znacznikowany. Sądy ocenione jako prawdziwe otrzymują jeden

<sup>23</sup>Najbardziej znanym modelem zmiany przekonań, w którym przyjmuje się wspomniane trzy typy zmiany, a mianowicie tzw. ekspansję (nabywanie nowego przekonania), kontrakcję (usuwanie przekonania) i rewizję, jest tzw. model AGM zmiany przekonań. Model ten, jak i inne modele formalne, jest silną idealizacją zmiany przekonań (np. zakłada, że każda informacja musi wywołać zmianę w systemie przekonań, abstrahuje od językowego ujęcia przekonań, ignoruje inne typy zmian); więcej informacji o teorii AGM i jej założeniach znajduje się w: Lechniak 2011, s. 290–383.

rodzaj znacznika, a sądy ocenione jako fałszywe znacznik innego rodzaju. W modelu spinozjańskim stosuje się zaś inny system znacznikowania — każdy sąd, o ile jest zrozumiany, zostaje uznany za prawdziwy i jako taki pozostaje w systemie bez znacznika. Dopiero, jeśli zostanie rozpoznany jako fałszywy, otrzymuje znacznik<sup>24</sup>.

Ostateczny rezultat obu tych operacji rozpatrywany jako dwa rozłączne zbiory jest taki sam. Rozstrzygnięcia, który model adekwatnie opisuje ludzkie przekonania, dostarcza porównanie ich działania w sytuacji, gdy normalna praca systemu zostaje przerwana. W obu modelach zakłada się modularność systemów umysłowych, czyli twierdzenie, że przetwarzanie informacji przebiega w skończonej sekwencji różnych podsystemów; takie systemy przekazują informację z jednego modułu do następnego, dokonując stosownych jej modyfikacji. Każdy z modeli powinien zatem dać inną prognozę rezultatu procesu przetwarzania informacji przerwanej na danym etapie. Jeśli zatem założymy, że systemy mentalne są modularnymi systemami, które umożliwiają uzyskanie wyniku przetwarzania informacji po każdym etapie tego przetwarzania, to ów wynik informuje o tym, w jaki sposób informacja została przetworzona na etapie poprzednim. A zatem, według modelu kartezjańskiego, system, który ulega zaburzeniu, powinien reprezentować sądy (produkt pierwszego modułu), ale nie powinien być zdolny do oceny ich prawdziwości (ta bowiem jest dopiero dokonywana w następnym module). W modelu spinozjańskim zaburzenie pracy systemu prowadzi do rezultatu w postaci przekonania; zakłada się w nim, że system nie powinien produkować sądów odrzuconych lub potwierdzonych (gdy podmiot upewnił się o ich prawdziwości).

Przegląd badań psychologicznych, w których można znaleźć argumenty za większą adekwatnością jednego z modeli, przemawia za wyższością modelu spinozjańskiego. Badania wskazują na to, że, np. dla dzieci wątpliwość (i w ogólności zaprzeczanie) jest trudniejsze niż bezrefleksyjne stwierdzanie. Młodzi ludzie zachowują się więc jak niedojrzałe systemy spinozjańskie; w systemach kartezjańskich można by oczekiwać, że uznawanie i wątpliwość zachodzą równie łatwo. Jednak wyniki eksperymentów wskazują, że tak nie jest. Badania empiryczne dowodzą również, że odrzucanie sądu (dezakceptacja) jest trudniejsze niż akceptacja (destrukcja wątpliwości). Wszystkie zdania są wstępnie kodowane jako prawdziwe i stan porównywania prowadzi do dalszej pracy umysłowej, jedynie wówczas, gdy zdanie okaże się fałszywe. Podobnie, negacja (jako operacja) jest wtórna względem afirmacji zdania — zaprzeczenie treści zdania wymaga jej wcześniejszej afirmacji (Horn 1989, s. 155-203). Jako jeszcze jeden argument za modelem spinozjańskim podaje się problem tzw. „prostego zrozumienia” (czyli supozycji traktowanej jako rozumienie treści sądu bez jego uznania). Otóż wiele badań wskazywało, że ludzie są zasadniczo niezdolni do reprezentowania zdań w sposób neutralny co do wartości logicznej, nawet wówczas, gdy byli do takiego ujmowania zdań bezpośrednio motywowani.

Niniejsze uwagi stanowiły — z konieczności jedynie szkicową — próbę przedstawienia reprezentacjonistycznego ujęcia przekonania. Za stanowiskiem reprezentacjonistycznym przemawia, zdaniem autora, w szczególności to, że wyjaśnia ono produktywność i strukturalność przekonania, fakt istnienia wielu postaw o różnej jakości, ale o tej samej treści, pośredniczącą (pomiędzy postawami a zachowaniami) rolę treści

<sup>24</sup>Inaczej mówiąc, w modelu spinozjańskim, każda percypowana treść jest uznawana, a więc staje się przekonaniem, nazywanym przez L. Koja, spontanicznym (w terminologii Denneta, przekonaniem należącym do systemu intencjonalnego pierwszego rzędu, por Dennett 1988, s. 243). Odrzucenie sądu w tej perspektywie jest już przekonaniem refleksyjnym (według terminologii Koja), albo należy do systemu intencjonalnego II rzędu (w terminologii Denneta).

przekonań, czy wreszcie zakorzenienie związków między przekonaniem w związkach strukturalnych w języku myśli. Inną zaletą prezentowanego podejścia jest możliwość jego (a w szczególności związanej z nim koncepcji języka myśli) wykorzystania w tłumaczeniu psychologicznych podstaw logiki, czy też mechanizmu nabywania i zmiany przekonań. Oczywiście zaprezentowane tu analizy wymagają rozwinięcia i pogłębienia — w szczególności dotyczy to przyczynowej roli treści czy zakorzenienia relacji między przekonaniem w związkach strukturalnych języka myśli. Należy również zauważyć, że sama koncepcja języka myśli wymaga dalszych pogłębionych analiz.

## Literatura

- Armstrong, D. (1982). *Materialistyczna teoria umysłu*, tłum. H. Krahelska, Warszawa: PWN.
- Bermudez, J. (2005). *Philosophy of Psychology*. New York, London: Routledge.
- Block, N. (1980). What is Functionalism. W: Block, N. (red.) *Reading in Philosophy of Psychology*, v. 1, s. 171–184.
- Braithwaite, R. (1973). The Nature of Believing. W: Griffiths, P. (red.) *Knowledge and Belief*, Oxford: OUP, s. 28–40.
- Ciecierski, T. (2012). Postawy propozycyjne, W: Miłkowski, M., Poczobut, R. (red) *Przewodnik po filozofii umysłu*. Kraków: Wydawnictwo WAM, s. 311–349.
- Dennett, D. (1982). Styles of Mental Representation. *Proc. of the Aristotelian Society*, New Series 83, 213–226.
- Dennett, D. (1988). *The Intentional Stance in the Theory and Practice*. Oxford: OUP.
- Fodor J. (1975). *The Language of Thought*. Hassocks: The Harvester Press.
- Fodor, J. (1978). Propositional Attitudes. *The Monist* 61 (4), 501–523.
- Fodor, J. (1991). Theory of Content, W: Fodor, J. *Representations*. Hassocks: The Harvester Press.
- Fodor, J., McLaughlin, B. (1990). Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work, *Cognition* 35, 183–204.
- Fodor, J. (1999). Jak grać w reprezentacje umysłowe — poradnik Fodora. W: Chlewiński, Z. (red.) *Modele umysłu*. Warszawa: PWN, s. 17–49.
- Fodor, J., Pylyshyn, Z. (1988). Connectionism and Cognitive architecture. A Critical Analysis. *Cognition* 28, 3–71.
- Gilbert, D. (1991). How Mental Systems Believe. *American Psychologist* 46 (2), 107–119.
- Horn, L.A. (1989). *Natural History of Negation*. Chicago: Chicago University Press.
- Lechniak, M. (2011). *Przekonania i zmiana przekonań*. Lublin: Wyd. KUL.
- Olson, D.R. (1999). Truth and its Negation: Macnamara's Analysis of the Place of Logic in Cognitive Psychology, W: Jackendoff, R., Bloom, P. (red.) *Language, Logic and Concepts*. Cambridge Ma., MIT Press, s. 109–117.
- Oppy, G. (1998). Propositional Attitude. W: Craig, E. (red.) *Routledge Encyclopedia of Philosophy*, t. 7, Routledge, s. 719–720.
- Paśniczek, J. (1996). The Relational vs. Directional Conception of Intentionality. W: Zeidler-Janiszewska, A. (red.) *Epistemology and History*, Amsterdam: Rodopi, s. 373–380.
- Putko, A. (1995). *O naturze przekonań jako reprezentacji umysłowych*. Poznań: Wyd. Fund. Humaniora.
- Putnam, H. (1994). Logic and Psychology. W: Macnamara, J., Reyes, E. (red.) *The Logical Foundation of Cognition*. Oxford New York: OUP, s. 35–42.
- Rey, G. (2002). Language of Thought. W: *Encyclopedia of Cognitive Science*. New York: Wiley and Sons.

- Stalnaker, R. (1986). *Inquiry*. Cambridge, Mass.: The Bradford Book.
- von Eckardt, B. (1993). *What is Cognitive Science?*. Cambridge, Mass: Bradford Books/MIT Press.
- von Eckardt, B. (1999). Mental Representation. W: Wilson, R.A., Keil, F.C., *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, Mass: The Bradford Book, s. 527–529.
- Williams, N. E. (2005). Static and Dynamic Dispositions. *Synthese* 146, 303–324.
- Żegleń, U. (2003). *Filozofia umysłu*, Toruń: Wyd. A. Marszałek.
- Żegleń, U. (2012). Treść i reprezentacje umysłowe, W: Miłkowski, M., Poczobut, R. (red.) *Przewodnik po filozofii umysłu*, Kraków, Wyd. WAM, s. 213–252.

## Representational Conception of Beliefs

MAREK LECHNIAK

*John Paul II Catholic University of Lublin*

**Abstract.** *The paper presents representational conception of belief according which beliefs are mental representations of the states of affairs and gives main reasons for this approach. Additionally are presented some comments on the minimum resource of expressions of the language of thought necessary to ensure logical character of beliefs and some mention of the mechanism of acquisition and change of beliefs.*

**Keywords:** *belief, representationism, content, language of thought*

---

# W stronę ekstensjonalnej teorii przekonań

ANDRZEJ BIŁAT

Wydział Administracji i Nauk Społecznych Politechniki Warszawskiej

**Streszczenie.** W artykule analizowane są niektóre ontologiczne i metodologiczne założenia logiki epistemicznej, w tym: (1) ogólna teza ontologicznego intensjonalizmu, głosząca, że istnieją byty intensjonalne oraz (2) pewne specyficzne tezy dotyczące postaw przekonaniowych odpowiadające niektórym prawom tej logiki. Praca składa się z dwóch części. Celem pierwszej części jest wsparcie zasady (1) oraz ekstensjonalnego podejścia w konstrukcji teorii bytów intensjonalnych. Celem części drugiej jest określenie pewnej metody konstrukcji ekstensjonalnych teorii przekonań, w których wyrażalne są tezy typu (2)<sup>1</sup>.

**Słowa kluczowe:** przekonanie, sąd, logika epistemiczna, teoria ekstensjonalna, byty intensjonalne, segmentacja wyrażań, logika nazw propozycjonalnych

## I.

### 1.1

Zgodnie z pewną popularną w filozofii umysłu i języka koncepcją, *przekonanie* jest postawą uznawania (akceptacji) pewnej abstrakcyjnej treści zdaniowej, zwanej *sądem*. Uznanie prawdziwości/słuszności sądu  $S$  jest jedną z postaw (obok wątpienia w prawdziwość/słuszność sądu  $S$ ; życzenia, aby sąd  $S$  się spełnił; obawy, że  $S$  spełni się itd.), jakie osoby mogą przyjmować wobec  $S$ . Zgodnie z tą ogólną *koncepcją postaw propozycjonalnych*, jeśli  $S$  jest sądem wyrażonym w zdaniu  $Z$ , to prawdziwa jest każda równoważność o postaci:

$x$  jest przekonany, że  $Z$  zawsze i tylko wtedy, gdy  $x$  uznaje sąd  $S^2$ .

Tę koncepcję uznajemy za ogólne tło dla dalszych rozważań.

U podstaw znanych i powszechnie stosowanych schematów wnioskowań, specyficznych dla logiki epistemicznej, leżą pewne ogóle zasady uznawania sądów. Wydaje się, że co najmniej niektóre tego rodzaju zasady mogą być wyrażone w „postaci kanonicznej” zdań ekstensjonalnego języka przedmiotowego.

---

<sup>1</sup>Projekt został sfinansowany ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji nr DEC-2011/03/B/HS1/04586.

<sup>2</sup>*Przekonanie dotyczące sądu  $y$*  można formalnie określić jako parę uporządkowaną  $\langle O, S \rangle$ , gdzie  $O$  jest klasą wszystkich osób, które uznają sąd  $S$ . To określenie jest uszczegółowieniem ogólniejszej definicji: „Postawa propozycjonalna jest, najogólniej rzecz biorąc, parą złożoną z sądu i zbioru systemów intencjonalnych, [...] które przyjmują tę postawę w stosunku do tego sądu.” (Fodor 1999, s. 29).



Rozważmy jeden z prostszych schematów tego typu:

(1)  $\frac{\text{Sądę, że nieprawda, że } p.}{\text{Nieprawda, że sądę, że } p}$

(1) jest schematem powszechnie stosowanym w ludzkim myśleniu. Fakt ten można wyjaśnić, odwołując się do zasady:

(1') Dla każdego  $x$ , jeśli  $x$  jest istotą racjonalną oraz  $x$  uznaje w danej chwili sąd *nieprawda-że- $p$* , to *nieprawda, że  $x$  uznaje w tej lub następnej chwili sąd  $p$* .

Analiza tego rodzaju zasad, które leżą u podstaw logiki epistemicznej (czy też ogólniej: logiki filozoficznej), może być bogatym i wiarygodnym źródłem twierdzeń z zakresu ontologii przekonań (ogólniej: ontologii formalnej). Te zasady dostarczają odpowiedzi na pytanie, dlaczego akurat takie, a nie inne schematy rozumowań, są powszechnie stosowane w ludzkim myśleniu. Czy zatem stoi coś na przeszkodzie przed dołączeniem ich do ogólnej teorii przekonań?

O ile wyrażenie „teoria” jest rozumiane standardowo (w sensie zbliżonym do „teorii względności”, „teorii ewolucji”, „teorii racjonalnego wyboru” itd.), to wydaje się, że z pewną formalną trudnością w istocie mamy tu do czynienia. Sformułowania w rodzaju (1') zawierają — w przeciwieństwie do typowych teorii naukowych — „strukturalno-opisowe” nazwy złożonych sądów: „*nie- $p$* ”, „ *$p$ -oraz- $q$* ” itd. Ponadto, uwzględniając wnioskowania, w których funktor przekonaniowy jest iterowany lub występuje w kombinacji z innymi funktorami, takie sformułowania mogą zawierać nazwy złożonych sądów, których częściami są funktory przekonaniowe. Oto przykład schematu takich wnioskowań:

(2)  $\frac{\text{Jestem przekonany, że nieprawda, że } p.}{\text{Jestem przekonany, że nie jestem przekonany, że } p.}$

Fakt, że schemat (2) jest powszechnie stosowany, można wyjaśnić za pomocą następującej zasady:

(2') Dla każdego  $x$ , jeśli  $x$  uznaje (sąd) *nieprawda-że- $p$*  oraz „ $a$ ” jest nazwą używaną przez  $x$  w celu oznaczenia  $x$ , to  $x$  uznaje (sąd) *nieprawda-że- $a$ -jest-przekonany-że- $p$* .

Jak widać, poprawna formalizacja zasad typu (1') i (2') odbywa się w języku, w którym występują: (a) nazwy sądów, w tym nazwy złożone z części reprezentujących *funktory ekstensjonalne* („nieprawda, że”, „i”, „lub” itd.) i *funktory intensjonalne* („uznaje”, „sądzi”, „jest przekonany” itd.) oraz (b) symbole nazw reprezentujących pewne nazwy używane przez użytkowników języka. Zarówno proste nazwy sądów, jak i syntaktycznie złożone nazwy sądów, które składają się z takich funktorów i symboli nazwowych, nazywać będziemy *nazwami propozycjonalnymi*.

Język teorii, która zawiera tezy takie, jak (1') i (2'), nie jest typowym językiem teorii naukowej. Fakt, że w tych tezach występują złożone nazwy propozycjonalne, decyduje o jego podobieństwie do języka semantyki logicznej. Jednakże nie jest to metajęzyk w standardowym znaczeniu współczesnej semantyki, ponieważ sądów, funktorów i symboli nie ujmuje się w nim jako wyrażen jakiegokolwiek języka. W takiego rodzaju teorii relacji między osobami, sądami i momentami czasowymi, a także między samymi sądami, abstrahuje się od jakichkolwiek systemów językowych.

Skoro język omawianej wersji teorii przekonań nie może być scharakteryzowany ani jako typowy język przedmiotowy, ani jako metajęzyk, to pojawiają się podstawowe pytania o charakterze metodologicznym: Na czym polega specyfika tego języka? Czy teorie w nim formułowane mogą być ujęte — analogicznie do typowych teorii naukowych — jako systemy ekstensjonalne? Jeśli tak, to jakie reguły rządzą składnią tego języka?

Pytania te wyznaczają główną problematykę tej pracy; zostanie ona omówiona w drugiej części. Pozostały fragment pierwszej części zawiera uściślenie stosowanej dotychczas terminologii, dyskusję wokół ekstensjonalnego podejścia w konstrukcji teorii przekonań oraz pewien argument wspierający to podejście.

## 1.2

Przez *teorię* rozumiemy logicznie usystematyzowany zbiór ogólnych twierdzeń, dotyczących danej dziedziny przedmiotowej. Jeśli  $T$  jest teorią i  $\alpha, \beta, \gamma$  są wyrażeniami zdaniowymi jej języka, to symbolem „ $\gamma(\alpha//\beta)$ ” oznaczamy wyrażenie zdaniowe tego języka, które powstaje w wyniku zastąpienia (w jednym lub w wielu miejscach)  $\beta$  za  $\alpha$  w  $\gamma$ . Symbolem „ $\leftrightarrow$ ” oznaczamy spójnik *maturalnej równoważności*, tj. spójnik wyrażający identyczność wartości logicznych swoich argumentów. O dowolnej teorii  $T$  mówimy, że jest *teorią ekstensjonalną*, gdy obowiązuje w niej *reguła ekstensjonalności* (względem materialnej równoważności):

**(RE)** dla dowolnych wyrażeń zdaniowych  $\alpha, \beta, \gamma$  języka teorii  $T$ , jeśli tezami teorii  $T$  są wyrażenia o postaci  $\alpha \leftrightarrow \beta$  i  $\gamma$ , to tezą teorii  $T$  jest wyrażenie o postaci  $\gamma(\alpha//\beta)$ .

W przeciwnym razie mówimy, że  $T$  jest *teorią intensjonalną*.

Jak wiadomo, reguła ekstensjonalności jest jednym z filarów standardowej logiki. Okoliczność ta najwidoczniej nie robi większego wrażenia na widzach „Stawki większej niż życie”, którzy słusznie uznają następujące wnioskowanie za niepoprawne:

- (a) Hans Kloss jest tą samą osobą, co J-23.  
 (\*) (b) Hermann Brunner jest przekonany, że Hans Kloss jest tą samą osobą, co Hans Kloss.  
 (c) Hermann Brunner jest przekonany, że Hans Kloss jest tą samą osobą, co J-23.

Skoro zdanie (a) jest prawdziwe, to ma ono tę samą wartość logiczną, co tautologiczne zdanie „Hans Kloss jest tą samą osobą, co Hans Kloss”<sup>3</sup>. Zgodnie z regułą ekstensjonalności, zdania te powinny być wymienne *salva veritate* w dowolnym zdaniu, w tym również w zdaniu (b). Pomimo to, jak wiadomo, zdania te nie są wymienne: przesłanki (a) i (b) są przez widzów akceptowane, w przeciwieństwie do wniosku (c).

Uwzględniając okoliczność, że w logice przekonań nie obowiązuje reguła ekstensjonalności RE, tego rodzaju przykłady mogłyby nasunąć przypuszczenie, że należy budować teorię przekonań bezpośrednio na bazie tej logiki. Taka „intensjonalna metodologia” mogłaby wydawać się atrakcyjna również z tego względu, że zasady leżące u podstaw rozumowań typu (1) i (2) można by po prostu utożsamić z prawami logiki przekonań:

- (1'')  $\forall x(B_x \neg p \rightarrow \neg B_x p)$ ,  
 (2'')  $\forall x(B_x \neg p \rightarrow B_x \neg B_x p)$ .

<sup>3</sup>Ostatnie z wymienionych zdań jest podstawieniem prawa tożsamości. Na użytek tych rozważań zakładamy, że Brunner zawsze stosuje w swoich rozumowaniach to prawo.

Konstruując teorię przekonań bezpośrednio na gruncie odpowiednio bogatej logiki intensjonalnej, w której obowiązują prawa (1'') i (2'') (i nie obowiązuje RE), zapewne dałoby się też uniknąć komplikacji związanych z konstrukcją złożonych nazw propozycjonalnych. Jeśli uznamy słusność tych racji, to powstaje pytanie: czy taka „intensjonalna metodologia” nie jest warta filozoficznego wsparcia?

Przeciwko tej metodologii przemawia okoliczność, że wszystkie dobrze ugruntowane w nauce koncepcje (np. znane teorie współczesnej fizyki) są teoriami ekstensjonalnymi. A zatem „intensjonalna metodologia” jest niezgodna ze stosowanym w „twardej” nauce postulatem *metateoretycznego ekstensjonalizmu*, który głosi (w przeciwieństwie do *metateoretycznego intensjonalizmu*), że każda dobrze określona teoria jest teorią ekstensjonalną.

Nieco paradoksalnie, stanowisko metateoretycznego ekstensjonalizmu — wciąż niemożliwe wśród współczesnych filozofów analitycznych — może być wsparte za pomocą pewnej ontologicznej wersji tezy intensjonalizmu. Celem następnego punktu jest sformułowanie argumentu, w którym ta teza jest główną przesłanką.

### 1.3

Teza *ontologicznego intensjonalizmu* głosi, że istnieją *byty intensjonalne*, czyli specyficzne obiekty postulowane w celu wyjaśnienia warunków prawdziwości zdań, zawierających funktry intensjonalne: światy możliwe, sądy, własności itd. Za uznaniem istnienia takich bytów przemawia okoliczność, że w ludzkiej mowie „roi się” od zwrotów intensjonalnych, a więc zwrotów typu „jest możliwe, że”, „i z tej przyczyny jest tak, że”, „w naszym przekonaniu jest tak, że”, „zwolennicy teorii *T* są w błędzie sądząc, że” itd.

Dotyczy to nie tylko mowy potocznej, ale również „miękkiego” dyskursu naukowego, w którym wykorzystywane są środki typowe dla języka naturalnego. W naukach o dojrzałej, „okrzepłej” metodologii, na przykład we współczesnej fizyce, widoczna jest różnica między takim dyskursem a formalną konstrukcją teorii empirycznej. Wydaje się, że zasady formalnej poprawności „miękkiego” dyskursu naukowego są określane przez stosowne systemy logiki intensjonalnej (głównie logiki zdaniowej), a zasady poprawnej konstrukcji teorii naukowej – przez systemy ekstensjonalnej logiki kwantyfikatorów<sup>4</sup>.

Skoro taka symbioza intensjonalnego dyskursu z ekstensjonalną teorią jest sytuacją typową w rozwiniętych naukach, to wydaje się, że analogiczna sytuacja powinna mieć miejsce w naukowej teorii przekonań.

Uznanie dyskursu intensjonalnego za istotny składnik tak ważnych procesów poznaw-

<sup>4</sup>W. Heisenberg zwracał uwagę na fakt, że fizycy posługują się w obrębie swojej dyscypliny zarówno językiem matematyki, jak i pewnego rodzaju językiem wyobraźniowym (Heisenberg 1987a, 1987b, 1979). Rozproszone uwagi Heisenberga na ten temat skomentował S. Kiczuk: „[...] fizyka nowożytna i współczesna posługuje się dwoma językami. Jednym z nich jest tzw. język matematyczny, który zwięźle opisuje stosunki zachodzące w przyrodzie i pozwala obliczyć wartości wielkości fizycznych [...]. Logiką takiego języka jest klasyczny rachunek logiczny. Fizykowi potrzebny jest również język wyobraźniowy zbliżony do języka potocznego, za pomocą którego można mówić o eksperymentach i przekazywać zmysłowo uchwytnie obrazy natury. Logika przyczynowości, do której skonstruowania zmierzamy, będzie logiką poszukującą praw rządzących poprawnym użyciem funktrów nieekstensjonalnych, związanych z kluczowymi terminami występującymi w języku wyobraźniowym teorii fizycznych” (Kiczuk 1995, s. 112).

czych, jakim jest dyskurs fizykalny, nadaje analizie jego ontologicznych zobowiązań dość silny status poznawczy. W semantyce wypowiedzi, które zawierają funktry intensjonalne, powszechnie przyjmowane są (poprzez warunki prawdziwości lub warunki spełniania dla zdań zawierających dany funktry intensjonalny) pewne ontologiczne założenia dotyczące bytów intensjonalnych. Na przykład warunki spełniania dla formuł logiki modalnej zakładają kwantyfikację po światach możliwych i/lub innych przedmiotach możliwych, warunki spełniania dla formuł logiki przekonań zakładają kwantyfikację po sędach reprezentowanych przez zbiory formuł lub (zależnie od wersji semantyki) po światach możliwych itd.

Przynajmniej w przypadku podstawowych systemów intensjonalnej logiki filozoficznej tego rodzaju zobowiązania nie mają swoich wyraźnych odpowiedników w języku przedmiotowym (w postaci tez o istnieniu odpowiednich rodzajów przedmiotów będących wartościami kwantyfikowanych zmiennych). Na przykład w języku przedmiotowym normalnych systemów logiki modalnej światy możliwe nie należą do zakresu kwantyfikacji (pomimo, że odpowiednie tezy o istnieniu światów możliwych są zakładane w metajęzyku tych logik). Tę cechę „dziedziczą” teorie budowane w ramach tego rodzaju logiki. Może się więc na przykład zdarzyć, że modalna teoria sędów „milczy” w kwestii istnienia światów możliwych, których istnienie *de facto* przyjmuje w swojej metateorii.

Jednym z podstawowych zadań teorii naukowej jest wyraźna, przedmiotowa charakterystyka wszystkich zakładanych przez nią — również w metajęzyku — rodzajów bytów. Dotyczy to zwłaszcza tak kontrowersyjnych obiektów, jakimi są byty intensjonalne. Brak realizacji tego zadania jest wadą teorii: narusza wymóg jej semantycznej i ontologicznej przejrzystości. Co najmniej w tym sensie teorie intensjonalne nie są dobrze określonymi teoriami.

Uwzględniając różnicę między dwiema wersjami ekstensjonalizmu, metateoretyczną i ontologiczną, można teraz lepiej ocenić podstawy dość szerokiej we współczesnej filozofii analitycznej krytyki (klasycznego już) ekstensjonalizmu Willarda Van Ormana Quine’a. Jest to stanowisko skrajne w tym sensie, że obejmuje obie wspomniane wersje ekstensjonalizmu. Według Quine’a, przedmiotem analizy ontologicznych zobowiązań mogą być wyłącznie twierdzenia ekstensjonalnych teorii naukowych, a więc zdania, których warunki prawdziwości nie wymagają specjalnych założeń, dotyczących istnienia bytów intensjonalnych. Uwzględniwszy dodatkowo trudności związane z określeniem warunków identyczności dla tego rodzaju bytów, Quine przyjął stanowisko ontologicznego ekstensjonalizmu (wyrażając je w jednej ze swoich słynnych sentencji: byty intensjonalne są w filozofii *entia non grata*).

Uwzględniając przedstawioną w tym punkcie argumentację, znosimy w dalszych rozważaniach „ontologiczną banicję” nałożoną na byty intensjonalne i uznajemy jednocześnie ekstensjonalną metodologię konstrukcji teorii tych bytów. Dotyczy to w szczególności teorii sędów, która leży u podstaw teorii przekonań

## II.

### 2.1

Przyjmując dość naturalne założenie, że nazwy własne „Hans Kloss” i „J-23” są logicznie proste (a więc nie są deskrypcjami określonymi), musimy jednocześnie uznać, że jedna z klasycznych metod analizy relacji przekonaniowej, której źródłem są prace Bertranda Russella, nie znajduje zastosowania w analizie wnioskowań typu (\*). Stosując tę metodę, schemat tego wnioskowania wyglądałby z grubsza następująco:

- (a') Kloss = J-23  
 (\*')  $\frac{(b') \text{ Bel (Brunner, Kloss, =, Kloss)}}{(c') \text{ Bel (Brunner, Kloss, =, J-23)}}$

Zgodnie z regułą ekstensjonalności — wbrew naszym oczekiwaniom — wnioskowanie (\*') jest formalnie poprawne.

Powyższa analiza pokazuje, że wartość logiczna konkluzji (c) we wnioskowaniu (\*) zależy nie tyle od odniesienia nazw własnych „Hans Kloss” i „J-23”, ile od odniesienia całej nazwy propozycjonalnej „(to), że Hans Kloss jest tą samą osobą, co J-23”. Jest to konkluzja zbieżna z koncepcją Gottloba Fregego, zgodnie z którą zdania występujące w zasięgu funktorów przekonaniowych, „zachowują się” jak nazwy propozycjonalne (oznaczają sądy wyrażone w tych zdaniach). Jednakże wadą ujęcia Fregego jest brak wyraźnego, a więc syntaktycznego odróżnienia zdań denotujących sądy od nazw sądów. Ten brak prowadzi do kolizji z Fregeańską zasadą jednoznaczności wyrażen („języka idealnego”): wbrew tej zasadzie, każde zdanie  $Z$  ma dwie różne denotacje, zależnie od tego, czy występuje w kontekście ekstensjonalnym (wówczas  $Z$  denotuje wartość logiczną  $Z$ ), czy intensjonalnym (wówczas  $Z$  denotuje sąd wyrażony w  $Z$ ).

W literaturze nie podejmowano dotąd prób (autorowi tych słów nie są one znane) określenia reguł składni i logiki języków ekstensjonalnych z nazwami propozycjonalnymi, które czyniłyby istotny użytek — w postaci konstrukcji stosownego systemu logiki tych nazw — z dwóch warunków:

- W1** Wszystkie nazwy propozycjonalne są wyrażeniami *logicznie prostymi*, tzn. reguły składni wykluczają możliwość stosowania w ich obrębie jakichkolwiek logicznych operacji (w tym operacji zastępowania i podstawiania).  
**W2** Niektóre nazwy propozycjonalne są wyrażeniami *lingwistycznie złożonymi*, tzn. składają się z wielu znaków prostych (tj. elementów danego alfabetu).

Warunek **W1** wyklucza wymienialność *salva veritate* członów identyczności wewnątrz nazw propozycjonalnych, ale nie wyklucza możliwości porównywania (na poziomie składni) struktur sądów oznaczonych za pomocą tych nazw; wręcz przeciwnie, **W2** dopuszcza taką możliwość. Z tego powodu wydaje się, że w sposób czysto składniowy można określić pewną relację równoznaczności między logicznie prostymi nazwami sądów (przykład takiego określenia zostanie podany w ostatnim punkcie artykułu).

Naszym głównym zadaniem jest obecnie objaśnienie specyfiki reguł składni i reguł ekstensjonalnej logiki języka z takimi nazwami. Pozostała część tego artykułu jest poświęcona tej kwestii.

## 2.2

Kluczowa dla postawionego zadania jest kwestia sposobu tworzenia wewnątrz danego języka logicznie prostych i zarazem lingwistycznie złożonych nominalizacji wyrażen tego języka, w szczególności nominalizacji wyrażen zdaniowych, które mogłyby pełnić funkcję nazw lub ogólniej: termów propozycjonalnych. Takie nominalizacje powinny odzwierciedlać „działanie” zwykłych funktorów, a więc powinny być efektywne i odwracalne. Ponadto, zgodnie z warunkiem **W1**, mają być logicznie proste.

Przyjmijmy następujące oznaczenie. Jeśli  $W$  jest *wyrażeniem podstawowym* (tj. zdaniowym lub nazwowym) danego języka, to symbolem  $'[W]'$  oznaczamy pewne wyrażenie nazwowe tego języka. Intuicyjnie,  $'[W]'$  odnosi się do treści wyrażenia  $W$ .

Rodzi się pytanie, jakie — możliwie naturalne i proste — cechy syntaktyczne musi

posiadać język  $L$ , aby dało się zdefiniować odwzorowanie  $f$  zbioru wszystkich wyrażeń podstawowych języka  $L$  w zbiór wyrażeń nazwowych (termów) tego języka, określone wzorem:

$$f(W) = [W]$$

i spełniające warunki:

- (I)  $f$  jest funkcją obliczalną,
- (II)  $f$  jest funkcją jednojednoznaczną (tj. jeśli  $f(W) = f(W')$ , to  $W = W'$ )<sup>5</sup>,
- (III)  $f(W)$  jest termem prostym, dla każdego wyrażenia podstawowego  $W$  języka  $L$ .

W celu lepszego objaśnienia tej kwestii skorzystamy z podstawowej aparatury pojęciowej lingwistyki matematycznej. Przyjmujemy, że *alfabet* jest niepustym i przeliczalnym zbiorem znaków. Dowolny element alfabetu  $V$  nazywamy jego *literą*, a skończony ciąg liter — *wyrażeniem* nad  $V$ . Podobnie jak zwykle wyróżniamy *wyrażenie puste*  $e$ ; symbol „ $V^+$ ” oznacza zbiór wszystkich wyrażeń nad  $V$ .

Jeżeli  $x$  i  $y$  są wyrażeniami nad  $V$ , to ich *konkatenacją* (oznaczaną tu symbolem „ $xy$ ”) nazywamy wyrażenie powstałe przez wypisanie najpierw kolejnych liter ciągu  $x$ , a następnie — bezpośrednio po jego ostatniej literze — kolejnych liter ciągu  $y$ . Formalnie, operacja konkatenacji spełnia ogólne warunki:

- (4)  $xe = ex = x$ , dla każdego  $x \in V^+$ ,
- $(xy)z = x(yz)$ , dla każdego  $x, y, z \in V^+$ .

Tę standardową (w lingwistyce matematycznej) aparaturę pojęciową uzupełniamy o pojęcie SEGMENTACJI WYRAŻENIA. Dla każdego alfabetu  $V$ , dla każdego niepustego znaku  $\delta \in V$ , określamy (indukcyjnie) operację  $[ ]^\delta$ :

- (5)  $[x]^\delta = \delta x$ , dla każdego  $x \in V$ ,
- $[xy]^\delta = [x]^\delta [y]^\delta$ , dla każdego  $x, y \in V^+$ .

*Segmentacja wyrażenia*  $x$  za pomocą znaku  $\delta$  polega więc na dopisaniu tego znaku przed każdą literą występującą w  $x$ . Oto przykłady:

$$\begin{aligned} [ab]^a &= aaab, \\ [Ewa]^\# &= \#E\#w\#a, \\ [Kloss = J-23]^- &= -K-l-o-s-s-- = --J---2-3. \end{aligned}$$

Operacja segmentacji w oczywisty sposób spełnia warunek (I) (obliczalności). Jest też jednojednoznaczna: w rezultacie „wytarcia” nieparzystych wystąpień znaku  $\delta$  w ciągu liter  $[x]^\delta$  (dokładniej, w rezultacie podstawienia litery pustej  $e$  za wszystkie nieparzyste wystąpienia znaku  $\delta$  w w tym ciągu) uzyskujemy ponownie wyrażenie  $x$ . Spełniony jest więc warunek (II):

FAKT

Jeśli  $[x]^\delta = [y]^\delta$ , to  $x = y$ , dla dowolnych  $x, y \in V^+$ .

Co więcej, wyrażenie  $[x]^\delta$  może służyć do „markowania” charakterystycznej cechy funktorów intensjonalnych, jaką jest niewymienialność w ich zasięgu członów materialnej równoważności. Człony te są niewymienialne w nazwach propozycjonalnych ze względów czysto syntaktycznych: poddane segmentacji, tracą swoją logiczną strukturę

<sup>5</sup>Dzięki tej własności (odwracalności), funkcja  $f$  „markuje” podstawową własność syntaktyczną dowolnego (jednoargumentowego) funktora  $F$ : jeśli  $'F\alpha' = 'F\beta'$ , to  $\alpha = \beta$

(w szczególności — zdaniowy charakter)<sup>6</sup>. Możliwa jest więc konstrukcja ekstensjonalnej logiki nazw typu  $[x]^\delta$  jako logicznie prostych nazw propozycjonalnych (zgodnie z warunkiem (III)).

W tego typu logice reguła ekstensjonalności jest „pusto spełniona” w każdym przypadku, w którym człony danej równoważności są poddane segmentacji. Z takim właśnie przypadkiem mamy do czynienia we wnioskowaniu:

- (a'') Kloss jest tą samą osobą, co J-23.  
 (\*\*) (b'') Brunner jest przekonany, że  $[Kloss \text{ jest tą samą osobą, co Kloss}]^\delta$ .  
 (c'') Brunner jest przekonany, że  $[Kloss \text{ jest tą samą osobą, co J-23}]^\delta$ .

Wnioskowanie (\*\*) nie jest logicznie poprawne. Wniosek (c'') nie może być prawidłowo wyprowadzony z (a'') i (b''), gdyż wyrażenie „Kloss” nie występuje w zasięgu funktora przekonaniowego w (b'') (odpowiednikiem tego wyrażenia występującym w nazwie  $[Kloss \text{ jest tą samą osobą, co Kloss}]^\delta$  jest wyrażenie „ $\delta K \delta l \delta o \delta s \delta s$ ”).

Odpowiednie zastąpienie we wnioskowaniu podobnym do (\*\*) byłoby syntaktycznie dopuszczalne na podstawie reguły ekstensjonalności, gdyby następująca równość była prawdziwa:

$$[Kloss \text{ jest tą samą osobą, co J-23}]^\delta = [Kloss \text{ jest tą samą osobą, co Kloss}]^\delta$$

Jednakże tego właśnie nie możemy uczynić, skoro tę równość intuicyjnie oceniamy jako fałszywą (jej człony odnoszą się do różnych sądów).

## 2.3

Wykluczenie możliwości zastępowania i podstawiania wyrażeń wewnątrz nazw propozycjonalnych nie wyklucza możliwości konstrukcji nietrywialnego, ekstensjonalnego systemu logiki takich nazw. Wręcz przeciwnie: technika segmentacji wyrażeń umożliwia taką konstrukcję; w jej obrębie można w dość naturalny sposób określić stosowną relację identyczności między sądami.

Podstawowy system logiki termów propozycjonalnych możemy utożsamić z ekstensjonalną wersją najmniejszego systemu tzw. *logiki równoważności analitycznej*, czyli systemu będącego rozszerzeniem najmniejszego rachunku zdaniowego ze spójnikiem identyczności „ $\equiv$ ” Romana Suszki (SCI, *Sentential Calculus with Identity*) o następującą regułę inferencji (symbol „ $At(\alpha)$ ” oznacza tu zbiór wszystkich zmiennych zdaniowych występujących w formule  $\alpha$ ):

$$\frac{\alpha \leftrightarrow \beta, At(\alpha) = At(\beta)}{\alpha \equiv \beta} \text{ (RQF)}$$

Zgodnie z RQF tezą logiki jest dowolna równość, której członami są inferencyjnie

---

<sup>6</sup>Metafunktor segmentacji ‘[...]’ ma częściowo podobne własności do funktora #, którego wprowadzenie — do języka przedmiotowego standardowej logiki — proponował G. Harman (Harman 1993). Przedstawione powyżej motywy wprowadzenia tego metafunktora zasadniczo nie różnią się od motywów Harmana (tamże, s. 95 i ss. 101–103). Jednakże istnieje istotna różnica formalna między tymi funktorami: (a) należą do języków różnych stopni (funktor # należy do języka przedmiotowego) oraz (b) w nazwach sądów utworzonych za pomocą funktora Harmana — w przeciwieństwie do nazw sądów utworzonych w wyniku segmentacji zdań — mogą wystąpić specyficzne zmienne podlegające „zewewnętrznej” kwantyfikacji (a więc kwantyfikacji *de re*).

równoważne formuły składające się z tych samych zmiennych zdaniowych. Ta reguła jest zgodna z zasadą głoszącą, że zdania wyrażają dokładnie ten sam sąd, gdy są inferencyjnie równoważne i mają tę samą zawartość treściową<sup>7</sup>.

Ekstensjonalnym odpowiednikiem RQF jest reguła:

$$\frac{\alpha \leftrightarrow \beta, At(\alpha) = At(\beta)}{[\alpha] = [\beta]} \text{ (EQF)}$$

Definicja języka logiki termów propozycjonalnych z regułą EQF wygląda następująco. Alfabet języka klasycznego rachunku zdań (KRZ) wzbogacamy o dwa specyficzne symbole: predykat identyczności (=) i znak dywizu (-). Zbiór znaków tak wzbogaconego alfabetu oznaczamy symbolem „V”. Zamiast metajęzykowego symbolu „[ ]<sup>-</sup>”, będziemy pisać „[ ]”. Zbiory *termów propozycjonalnych* *Ter* i *formuł* *For* definiujemy łącznie (indukcyjnie):

- (i) Każda zmienna zdaniowa ( $p, q, r$  itd.) należy do zbioru *For*.
- (ii) Jeśli  $w \in For \cup Ter$ , to  $[w] \in Ter$ .
- (iii) Jeśli  $a, b \in Ter$ , to  $'a = b' \in For$ .
- (iv) Jeśli  $\alpha, \beta \in For$ , to  $'\neg\alpha', '(\alpha \wedge \beta)', '(\alpha \vee \beta)', '(\alpha \rightarrow \beta)', '(\alpha \leftrightarrow \beta)' \in For$ .
- (v) Żaden inny ciąg znaków nie należy do zbioru *Ter* ani do zbioru *For*.

Termami są na przykład wyrażenia: „ $\neg p$ ”, „ $\neg\neg q$ ”, „ $\neg\neg(\neg p \wedge \neg q)$ ” i „ $\neg p \leftrightarrow \neg q$ ”. Intuicyjnie, każde z tych wyrażeń odnosi się do sądu o postaci (odpowiednio): „ $p$ ”, „ $\neg q$ ”, „ $\neg(p \wedge q)$ ” oraz „ $[p] = [q]$ ”.

W podstawowym systemie logiki termów propozycjonalnych jedynymi *aksjomatami* są wszystkie podstawienia tautologii KRZ i formuły odpowiadające aksjomatom SCI:

- A1**  $a = a$ ,
- A2**  $[\alpha] = [\beta] \rightarrow [\gamma] = [\gamma(\alpha//\beta)]$ ,
- A3**  $[\alpha] = [\beta] \rightarrow (\alpha \leftrightarrow \beta)$ .

Jedynymi *regułami pierwotnymi* tego systemu są: (1) reguła odrywania, (2) reguła ekstensjonalności (wymienialności członów) dla identyczności oraz (3) reguła EQF.

Na podstawie ostatniej z wymienionych reguł wnosimy, że rozważany system obejmuje wszystkie tezy, będące odpowiednikami równości algebry Boole’a, w których występują te same zmienne po obu stronach równości. Jego tezami są zatem równości:

- T1**  $[\alpha] = [\neg\neg\alpha]$
- T2**  $[\alpha] = [\alpha \wedge \alpha]$
- T3**  $[\alpha] = [\alpha \vee \alpha]$
- T4**  $[(\alpha \wedge \beta)] = [(\beta \wedge \alpha)]$
- T5**  $[(\alpha \vee \beta)] = [(\beta \vee \alpha)]$
- T6**  $[\neg(\alpha \wedge \beta)] = [(\neg\beta \vee \neg\alpha)]$
- T7**  $[\neg(\alpha \vee \beta)] = [(\neg\beta \wedge \neg\alpha)]$
- T8**  $[(\alpha \wedge \beta) \wedge \gamma] = [(\beta \wedge (\alpha \wedge \gamma))]$
- T9**  $[(\alpha \vee \beta) \vee \gamma] = [(\beta \vee (\alpha \vee \gamma))]$
- T10**  $[(\alpha \wedge \beta) \vee \gamma] = [((\alpha \wedge \gamma) \vee (\beta \wedge \gamma))]$
- T11**  $[(\alpha \vee \beta) \wedge \gamma] = [((\alpha \vee \gamma) \wedge (\beta \vee \gamma))]$

<sup>7</sup>Zasada ta była analizowana m.in. w pracach Biłat (1991), (1993). W pracach tych wprowadzony został termin „system analitycznych równoważności”.



Można ponadto wykazać (korzystając m.in. z reguły ekstensjonalności), że tezami są pewne formuły, w których operacja segmentacji jest iterowana. Przykładem takiej tezy jest równość:

$$\mathbf{T12} \quad [‘[\alpha] = [\beta]’] = [‘[\beta] = [\alpha]’]$$

Język podstawowej logiki termów propozycjonalnych może być wzbogacony o *predykat przekonaniowy* „ $B(x, y)$ ” („ $x$  uznaje sąd  $y$ ”) i predykat nazywania „ $N(x, y, z)$ ” („ $x$  jest nazwą używaną przez  $y$  w celu oznaczenia obiektu  $z$ ”). W takim języku można formułować specyficzne tezy ekstensjonalnej teorii przekonań, które leżą u podstaw logiki epistemicznej. Przykładami takich tez są formuły<sup>8</sup>:

$$\mathbf{T13} \quad \forall x (B(x, [\alpha]) \wedge N([a], x, x) \leftrightarrow B(x, [‘B(a, \alpha’)]) \wedge N([a], x, x))$$

(sądzę, że  $\alpha$  zawsze i tylko wtedy, gdy sądzę, że sądzę, że  $\alpha$ )

$$\mathbf{T14} \quad \forall x (\neg B(x, [\alpha]) \wedge N([a], x, x) \leftrightarrow B(x, [‘\neg B(a, \alpha’)]) \wedge N([a], x, x))$$

(nie sądzę, że  $\alpha$  zawsze i tylko wtedy, gdy sądzę, że nie sądzę, że  $\alpha$ )

$$\mathbf{T15} \quad \forall x (B(x, [-\alpha]) \wedge N([a], x, x) \rightarrow \neg B(x, [‘B(a, \alpha’)]) \wedge N([a], x, x))$$

(jeśli sądzę, że nieprawda, że  $\alpha$ , to nie sądzę, że sądzę, że  $\alpha$ )

$$\mathbf{T16} \quad \forall x (B(x, [‘\alpha \rightarrow \beta’]) \rightarrow (B(x, [\alpha]) \rightarrow B(x, [\beta])))$$

(jeżeli sądzę, że jeśli  $\alpha$ , to  $\beta$ , to jeśli sądzę, że  $\alpha$ , to sądzę, że  $\beta$ )

## Literatura

- Biłat, A. (1991). Zasada Wittgensteina a logika nefregowska. W: Omyła, M. (red.) *Szkice z semantyki i ontologii sytuacji*. Warszawa: seria BMS, s. 63–68.
- Biłat, A. (1993). O pewnym kryterium tożsamości sądów. W: Krysztofiak, W., Perkowska, H. (red.), *Szkice z fenomenologii i filozofii analitycznej*. Szczecin: Wyd. Naukowe Uniwersytetu Szczecińskiego, s. 81–85.
- Fodor, J.A. (1999). Jak grać w reprezentacje umysłowe — poradnik Fodora. W: Chlewiński, Z. (red.) *Modele umysłu*. Warszawa: Wydawnictwo Naukowe PWN, s. 17–49.
- Harman, G. (1993). Forma logiczna. W: Stanosz, B. (red.) *Filozofia języka*. Warszawa: Fundacja Aletheia-Wydawnictwo Spacja, s. 87–111.
- Heisenberg, W. (1987a). Dyskusje na temat języka. W: tenże, *Część i całość*, przeł. K. Napiórkowski, Warszawa: Państwowy Instytut Wydawniczy, s. 164–182.
- Heisenberg, W. (1987b). Pozytywizm, metafizyka i religia. W: tenże, *Część i całość*, przeł. K. Napiórkowski, Warszawa: Państwowy Instytut Wydawniczy, s. 259–273.
- Heisenberg, W. (1979). Język a rzeczywistość we współczesnej fizyce. W: tenże, *Ponad granicami*, przeł. K. Wolicki, Warszawa: Państwowy Instytut Wydawniczy, s. 142–166.
- Kiczuk, S. (1995). *Związek przyczynowy a logika przyczynowości*. Lublin: Redakcja Wydawnictw KUL.
- Tokarz, M. (1993). *Elementy pragmatyki logicznej*. Warszawa: Wydawnictwo Naukowe PWN

<sup>8</sup>Tezy te są ekstensjonalnymi odpowiednikami aksjomatów systemu logiki świadomych przekonań Tokarza (Tokarz 1993, s. 168). Pod każdą tezą podany jest jej odpowiednik w postaci intuicyjnie zinterpretowanego prawa logiki epistemicznej.

## Towards an extensional theory of beliefs

ANDRZEJ BŁĄT

*Warsaw University of Technology*

**Abstract.** *In this paper some ontological and methodological assumptions of the epistemic logic are analysed. One of them is (1) the general thesis of ontological intensionalism (states that there are some intensional entities). A special kind of these assumptions (2) concern propositional attitudes corresponding to some laws of the logic. The paper consists of two parts. The aim of the first part is to support the thesis (1) and to validate the extensional approach to the construction of the theories of intensional entities. The aim of the second part is to define a method of construction of the theories of beliefs, in which theses (2) are expressible.*

**Keywords:** *belief, proposition, epistemic logic, extensional theory, byty intensional entities, segmentation of expression, logic of propositional names*

---

# Czy można być przekonanym, że Sherlock Holmes jest detektywem? O sensowności zdań z nazwami fikcjonalnymi

TOMASZ A. PUCZYŁOWSKI  
*Instytut Filozofii UW*

**Streszczenie.** *W pracy zwięźle przedstawiam problem semantycznej sensowności zdań prostych z fikcjonalnymi nazwami własnymi. Bronię w niej poglądu, zgodnie z którym zdania te pozbawione są semantycznego znaczenia a w konsekwencji nie podlegają wartościowaniu logicznemu. Ocena ta jest niezgodna z popularnym przekonaniem, że zastąpienie w pewnym sensownym zdaniu określonej jednostkowej nazwy własnej jakąś fikcjonalną nie zmienia sensowności zdania. Bronione przeze mnie stanowisko natrafia na jeszcze większą trudność w postaci zdań o przekonaniach z nazwami fikcjonalnymi — pogląd, że zdanie „Jan uważa, że Sherlock Holmes był przystojny” wyraża pewien sąd, wydaje się mniej kontrowersyjny od oceny zdania podrzędnego jako semantycznie nonsensownego. W pracy twierdzą jednak, że również zdania o przekonaniach z nazwami fikcjonalnymi są pozbawione semantycznego znaczenia. Proponuję, jak interpretować tego typu zdania oraz przedstawiam, w jaki sposób wyjaśnić intuicje semantyczne niezgodne z bronionym przeze mnie stanowiskiem, nie odwołując się przy tym do koncepcji tzw. dziurawych sądów albo tzw. obrony implikaturowej.*

**Słowa kluczowe:** *nazwa fikcjonalna, nazwa własna, zdania o przekonaniach, sensowność*

Przyjmijmy, że pomiędzy *Aleksander Kwaśniewski* a *Koszalek Opalek* zachodzi dokładnie ta sama semantyczna różnica co między *Mazowsze* a *Mordor*. Pierwsze imiona z tych par nazw nazywają określone obiekty, istniejące czasoprzestrzennie, drugie nie nazywają i nie nazywały nikogo i niczego takiego. Czy widoczna różnica między wskazanymi nazwami jest filozoficznie ważna? Przedstawione zostaną dalej racje uzasadniające twierdzącą odpowiedź na to pytanie. Argumentować będziemy za stanowiskiem, zgodnie z którym zdania zawierające nazwy drugiego rodzaju — w tym sprawozdające czyjeś przekonania — nie wyrażają żadnego sądu. Innymi słowy, jeśli *Święty Mikołaj* jest nazwą fikcjonalną, to w ścisłym tego słowa znaczeniu zdanie *Jasio uważa, że to Święty Mikołaj przynosi mu prezenty gwiazdkowe* nie może być prawdziwe, bez względu na to, do kogo odnosi się podmiot gramatyczny zdania.

Problem sensowności zdań z nazwami fikcjonalnymi jest od dawna znany i szeroko opisywany we współczesnej literaturze (np. Parsons 1980; Zalta 1983; Kroon, Voltolini 2011; Reicher 2014; Pańniczek 1984, 1991; Gurczyński 2013). Równie wnikliwej analizie poddano również argumenty, które odrzucają tezę, głoszącą, że zdania tego rodzaju są pozbawione semantycznego znaczenia. Dlatego zostanie on przedstawiony stosunkowo skrótowo i tylko na tyle, na ile konieczne to dla inteligibilnego wyeksponowania problemu głównego: jak wyjaśnić rzekomą sensowność zdań o przekonaniach z nazwami fikcjonalnymi, przy założeniu nonsensowności zdań prostych z tego rodzaju nazwami?

Nie odwołuję się dalej do niewątpliwie ciekawych konstrukcji formalnych, rozwijanych również w Polsce, w których podejmowane są próby kodyfikacji praw i regul wnioskowań wyrażanych za pomocą dyskutowanych tu zdań. Nie odwołuję się, ponieważ przedmiotem analiz nie są problemy, które wiążą się z ustaleniem odpowiednich związków inferencyjnych pomiędzy zdaniami z nazwami fikcjonalnymi. W tym artykule będziemy bronić poglądu, zgodnie, z którym zarówno zdania proste z nazwami fikcjonalnymi, jak i zdania o przekonaniach z takimi nazwami są pozbawione semantycznego znaczenia. Przedstawiona propozycja będzie zgodna z niektórymi wariantami tzw. teorii bezpośredniego odniesienia<sup>1</sup>. Różni się od niektórych z nich tym, że nie będzie odwoływała się ani do tzw. obrony implikaturowej (Adams et alia (1993); Adams, Stoker 1997), ani do koncepcji sądu niewypełnionego (*gappy proposition*) (Braun 1993, 2002, 2005; Gnatek 2011; Puczyłowski 2014, s. 133–177)<sup>2</sup>.

## 1. Nazwa fikcjonalna

Mianem *nazwy fikcjonalnej* określa się te nazwy własne, które nie mają żadnego istniejącego aktualnie nominatu<sup>3</sup>. Jest to rodzaj nazwy pustej, co znaczy, że nie ma żadnego aktualnie istniejącego przedmiotu, wobec którego ktokolwiek mógłby prawdziwie ją orzec. Jest to zarazem rodzaj nazwy własnej, czyli używanej z intencją jednostkową (to, czy nazwa tego rodzaju ma treść językową lub konotację — pozostawiamy do dalszego badania), która w sensownym zdaniu podmiotowo-orzecznikowym może pełnić wyłącznie funkcję podmiotu. Jako jej przykład podaje się znane — używane na kartach powszechnie znanych dzieł literackich — imiona w rodzaju: *Sherlock Holmes*, *Cezary Baryka*, *Genezy Kapen*; zalicza się do nich również imiona używane w kulturze popularnej, takie jak *Superman* lub *Święty Mikołaj*, i to nie tylko pewnych fikcyjnych osób lub zwierząt, ale i fikcyjnych rzeczy nieożywionych (w tym np. różnych krain i innych jednostek geograficznych jak *Mordor*), utworów, czy zdarzeń. Nazwy te nie mają nominatów, ponieważ nic, co aktualnie istnieje, nie jest prawdziwie żadną

<sup>1</sup>Koncepcję tę można charakteryzować na różne sposoby — chyba najbardziej zwięźle formuluje ją Arvid Båve (2008), twierdząc, że jej naczelnym twierdzeniem jest teza: Prawdziwy jest sąd wyrażany przez dowolne zdanie o schemacie:

⌈*a* jest takie, że *S(on/ona/ono)* zawsze i tylko wtedy, gdy *S(a)*⌋,

gdzie *a* jest imieniem własnym, *S(x)* reprezentuje jakąś funkcję zdaniową języka polskiego, zaś ⌈*S(on/ona/ono)*⌋ i ⌈*S(a)*⌋ różnią się tylko tym, że w miejscu, w którym w ⌈*S(a)*⌋ znajdowała się nazwa własna *a*, w ⌈*S(on/ona/to)*⌋ znajduje się odpowiedni zaimek.

<sup>2</sup>W obronie implikaturowej wysuwa się tezę, zgodnie z którą zdanie z użytą nazwą fikcjonalną nie wyraża żadnego sądu logicznego. Jednak asercja takiego zdania może już generować taki sąd, będący implikaturą konwersacyjną takiej wypowiedzi; w koncepcji niewypełnionego sądu (określanego niekiedy mianem *dziurawego millianizmu*, zob. Glüer-Pagin, Pagin 2014) utrzymuje się zaś, że zdania rozważanego typu, choć nie wyrażają pełnego sądu logicznego, to wyrażają sąd niewypełniony, który podlega ocenie prawdziwościowej. Nie tylko z uwagi na odmienną od potocznej ocenę zdań z użytymi nazwami własnymi, rozwiązania te wydają się kontrowersyjne (Everest 2003; Green 2007; Reimer 2001).

<sup>3</sup>Należy zwrócić uwagę, że zdaniem niektórych badaczy żadna nazwa własna nie może być nazwa pustą. Hans Reichenbach pisał: „Jeżeli wyraz ma być nazwany *imieniem własnym* konieczne jest, aby istniała odpowiadająca mu rzecz. Imiona mitologiczne, jak *Zeus*, nie są imionami własnymi, lecz tylko *imionami*, tzn. wyrazami używanymi na podobieństwo imion własnych, mianowicie jako argumenty fikcji” (Reichenbach 1948, s. 255–256, Reichenbach 1967, s. 91).

z tych nazw nazywane w kontekście, w którym tych imion używa się jako nazw postaci wyłącznie literackich, filmowych, czy też pop-kulturowych. Tego typu konteksty będziemy dalej uznawać za typowe dla tego rodzaju nazw. (Nie będą więc nimi sytuacje, w których nazwy, np. *Sherlock Holmes* używa się jako nazwy serialu filmowego lub czyjegoś zwierzęcego pupila.) Nie są to oczywiście jedyne konteksty, w których spotykamy tego typu nazwy. Poza przypadkami, w których nazwy te zostają w sposób oczywisty wymienione, a nie użyte, można wyróżnić użycie fikcjonalne metatekstowe (jak w zdaniu: *Sherlock Holmes jest fikcyjną postacią wymyśloną przez Artura Conan Doyle'a*), przedmiotowe użycie fikcjonalne intratekstowe (jak np. „*[Sherlock Holmes] dobrze zna anatomię i jest pierwszorzędnym chemikiem. Ale nigdy chyba systematycznie nie zajmował się medycyną. Jego studia są chaotyczne i dziwaczne; zebrał olbrzymi zapas oderwanych wiadomości, którymi mógłby zadziwić swych profesorów.*” (Conan Doyle 2000, s. 6) i fikcjonalne ekstratekstowe (jak w zdaniu *Sherlock Holmes był londyńskim detektywem, żyjącym w XIX w., który przyjaźnił się z dr. Watsonem*), a także mieszane (*Sherlock Holmes jest fikcyjnym detektywem, mieszkającym przy Baker Street z dr. Watsonem*). Można się zgodzić, że w przypadku użycia metatekstowego nazwy fikcjonalnej, zdania z wyłącznie tak użytą nazwą podlegają wartościowaniu logicznemu, ponieważ można twierdzić, że mówią coś o nazwach, a nie o obiektach, do których nazwy się odnoszą. Zdanie *Sherlock Holmes jest postacią fikcyjną wymyślona przez Artura Conan Doyle'a* można rozumieć w następujący sposób: *Artur Conan Doyle wymyślił i użył w swoich dziełach fikcjonalne imię „Sherlock Holmes”*. W przypadku użycia intratekstowego zazwyczaj nie stawia się pytań o prawdziwość zdań z tak użytymi nazwami fikcjonalnymi: zdania te nie pełnią funkcji informacyjnej, lecz innej. Jednak ekstratekstowe oraz mieszane użycie nazw rozważanego rodzaju prowadzi do kwestii filozoficznie interesujących, jeśli nie doniosłych — w szczególności z perspektywy analitycznej filozofii języka.

## 2. Problem nazw fikcjonalnych

Filozoficzna doniosłość nazw fikcjonalnych ujawnia się w następującym zagadnieniu: określając warunek prawdziwości dla prostych zdań podmiotowo-orzecznikowych w następujący — zdawałoby się intuicyjny i prosty — sposób:

(WP) Zdanie podmiotowo-orzecznikowe  $\lceil P(N) \rceil$  jest prawdziwe zawsze i tylko wtedy, gdy nominat nazwy  $\lceil N \rceil$  ma cechę wyrażaną przez orzecznik  $\lceil P \rceil$ ,

wyzwanie, jakie stanowią nazwy fikcjonalne dla semantyki, której konsekwencją jest (WP), staje się natychmiast widoczne. W sytuacji, w której  $\lceil N \rceil$  reprezentuje imię pozbawione nominatu, nie można stwierdzić prawdziwości takiego zdania — bez względu na orzeczaną za pomocą odpowiedniego predykatu cechę. Tymczasem wydawać się może, że istnieją prawdziwe zdania, w którym taka nazwa zostaje użyta, np.:

*Sherlock Holmes jest detektywem.*

W analogiczny sposób wydawać się może, że niektóre tego typu zdania są fałszywe, np.:

*Sherlock Holmes jest kobietą,*

co jest tym razem niezgodne z ogólnym określeniem warunków fałszywości dla prostych zdań podmiotowo-orzecznikowych w równie intuicyjny, co poprzedni sposób:

(WF) Zdanie podmiotowo-orzecznikowe  $\lceil P(N) \rceil$  jest fałszywe zawsze i tylko wtedy, gdy nominat nazwy  $\lceil N \rceil$  nie ma cechy wyrażanej przez orzecznik  $\lceil P \rceil$ .

Tymczasem większość użytkowników języka posługuje się tą nazwą, formułując zdania, w których ona występuje, w słusznym często przekonaniu, że inni uznają owo twierdzenie za sensowne. Jest zatem tak: albo  $(WP)$  i  $(WF)$  nie są prawdziwe — przynajmniej w stosunku do zdań z tego rodzaju nazwami, albo opinia użytkowników języka jest w tej kwestii błędna i nie powinna być rozstrzygająca w przypadku oceny twierdzeń i reguł semantyki, dotyczących zdań z tego typu nazwami użytymi ekstratekstowo lub w sposób mieszany.

### 3. Nazwa fikcjonalna jako deskrypcja

Nasuwa się kilka oczywistych rozwiązań tego problemu. Po pierwsze można utrzymywać, że warunek prawdziwości lub fałszywości zdań tego rodzaju powinien być inaczej określony. W szczególności można bronić poglądu, zgodnie z którym w przypadku takich zdań nazwa fikcjonalna powinna być interpretowana jak pewna deskrypcja — określona lub nie. O tym zaś, czy dana nazwa fikcjonalna  $N$  powinna być interpretowana jak deskrypcja  $D_1, \dots$ , czy  $D_n$ , rozstrzygać się powinno poprzez odwołanie do, na przykład, kontekstu kulturowego lub sytuacyjnego. W związku z tym z nazwą *Sherlock Holmes* niekiedy powinno wiązać się deskrypcję w rodzaju: *najbardziej genialny brytyjski detektyw końca XIX, mieszkający przy Baker Street, grający na skrzypcach* lub *bohater powieści detektywistycznych Artura Conana Doyle’a, niekiedy zaś: genialny brytyjski detektyw rozwiązujący kryminalne zagadki w Nowym Yorku, niegdyś uzależniony od narkotyków* lub: *bohater amerykańskiego serialu detektywistycznego „Elementary”*; w pewnych zaś okolicznościach możliwe, że najlepiej będzie przyjąć, że jest to *jedyna osoba, do której odnosi się „Sherlock Holmes”*.

Jeżeli za Russellem będziemy uznawać deskrypcję za wyrażenie kwantyfikatorowe, a nie za nazwowe, to wartość logiczna zdań z takimi nazwami będzie niejednokrotnie odbiegać od intuicyjnych ocen respondentów. Otóż zdanie, o którym sądzi się, że jest prawdziwe:

$(SH)$  *Sherlock Holmes jest detektywem,*

okazuje się na gruncie tej interpretacji fałszywe, ponieważ fałszem jest, że istnieje dokładnie jeden obiekt taki, że  $\dots$ , który jest detektywem (gdzie wielokropek zastępuje opis wyrażany w skojarzonej z użytą nazwą fikcjonalną; w rozważanym przypadku zdania. taki jak np.: *jest detektywem, mieszka przy Baker Street, gra na skrzypcach* itp). (Z drugiej zaś strony, zauważmy, że gdyby okazało się, że istnieje dokładnie jeden osobnik lub więcej osób spełniających deskrypcję skojarzoną z fikcjonalną nazwą *Sherlock Holmes*, nie chcielibyśmy uznać za prawdziwe zdania, np. *Sherlock Holmes istnieje.*) Można tej nieintuicyjnej konsekwencji zaradzić, jeśli przyjmie się następujący warunek prawdziwości dla zdań tego typu (patrz np. Tiedke 2011):

$(WP^*)$  jeżeli  $\ulcorner N \urcorner$  jest fikcjonalną nazwą, a  $D$  jest skojarzoną (w pewnym kontekście  $k$ ) z nią deskrypcją, to zdanie  $\ulcorner P(N) \urcorner$  jest prawdziwe (w kontekście  $k$ ) zawsze i tylko wtedy, gdy cecha wyrażana przez  $\ulcorner P \urcorner$  jest zawarta w konotacji  $D$ .

Na gruncie  $(WP^*)$  zdanie  $(SH)$  można uznać za prawdziwe, o ile cecha bycia detektywem będzie należała do konotacji deskrypcji skojarzonej w odpowiednim kontekście z nazwą „Sherlock Holmes”.

Przyjmując  $(WP^*)$  (i analogicznie  $(WF^*)$ ), zdania z nazwami fikcjonalnymi okazują się na mocy przyjętych warunków, które określają ich wartość logiczną, albo analitycznie prawdziwe, albo fałszywe. Rozważmy zdanie:

*Sherlock Holmes ma starszego brata.*

Jeżeli posiadanie starszego brata nie jest cechą należącą do konotacji deskrypcji skojarzonej z tą nazwą, to zdanie jest fałszywe; jeśli jednak do niej należy, to jest analitycznie prawdziwe, podobnie jak zdanie: *Matka Jana jest kobietą*, w którym cecha bycia matką należy do konotacji wyrażenia *matka Jana*. Rozstrzygnięcie o prawdziwości (fałszywości) zdania z użytą nazwą tego rodzaju będzie tu zawsze polegać na odwołaniu się wyłącznie do treści deskrypcji skojarzonej z tą nazwą. Dodatkowym zagadnieniem, które koniecznie należy wyjaśnić, jest charakter relacji skojarzenia nazwy z deskrypcją. Można zadać następujące pytanie: czy z jedną nazwą może być w danej chwili lub okresie skojarzona więcej niż jedna deskrypcja? Co lub kto przesądza o tym, że jakaś deskrypcja jest skojarzona z daną nazwą? Czy deskrypcje skojarzone z daną nazwą mogą zmieniać się w czasie (np. jeśli fikcyjny bohater pojawia się w pewnej serii powieściowej, przez co w nowych utworach przypisuje się mu kolejne cechy)?

Przyjęcie warunków prawdziwości dla zdań z nazwami fikcyjnymi takimi, jak to proponowane w (*WP\**), prowadzi do interesujących problemów, znanych przynajmniej od czasów *Nazywania a konieczności* Saula Kripkego. Można bowiem przypuszczać, że osoba, która uznaje za sensowne i prawdziwe zdanie (*SH*), podobnie oceni zdanie modalne:

*Sherlock Holmes mógłby nie być detektywem.*

Ogólnie rzecz biorąc wydaje się, że dla dowolnej deskrypcji *D* skojarzonej w ten lub inny sposób z nazwą fikcyjną Sherlock Holmes, ktoś, kto uznaje za prawdziwe (*SH*), gotowy będzie uznać, że

*Sherlock Holmes mógłby nie być D.*

Niepożądaną, gdyż nieintuicyjną, konsekwencją — uznania zdań z nazwami fikcyjnymi za analitycznie prawdziwe lub fałszywe — można (przynajmniej do pewnego stopnia) uniknąć, jeśli sformułuje się dla zdań z nazwami fikcyjnymi warunek prawdziwości, który przywołuje na myśl koherencyjną definicję prawdziwości, na przykład:

(*WP\*\**) jeżeli  $\ulcorner N \urcorner$  jest fikcyjną nazwą, a *U* jest jakimś utworem autora nazwy  $\ulcorner N \urcorner$ , to zdanie  $\ulcorner P(N) \urcorner$  jest prawdziwe zawsze i tylko wtedy, gdy (i) z *U* syntaktycznie wynika  $\ulcorner P(N) \urcorner$  i (ii) z żadnego utworu tego autora nie wynika żadne zdanie niezgodne z  $\ulcorner P(N) \urcorner$ ,

oraz

(*WF\*\**) jeżeli  $\ulcorner N \urcorner$  jest fikcyjną nazwą, a *U* jest jakimś utworem autora nazwy  $\ulcorner N \urcorner$ , to zdanie  $\ulcorner P(N) \urcorner$  jest fałszywe zawsze i tylko wtedy, gdy z *U* syntaktycznie wynika  $\ulcorner \neg P(N) \urcorner$  i (ii) z żadnego utworu tego autora nie wynika żadne zdanie niezgodne z  $\ulcorner P(N) \urcorner$ .

Takie określenie (oczywiście jedynie przykładowo) warunków prawdziwościowych zdań z nazwami fikcyjnymi nie pozwala rozstrzygnąć wartości logicznej niektórych zdań, w których będą występować nazwy tego typu. Tak określając warunki prawdziwościowe zdań tego typu, można jednak wyjaśnić, dlaczego zdaniom modalnym, takim jak

*Sherlock Holmes mógłby nie być detektywem*

przypisuje się prawdziwość — nawet jeśli takie zdanie nie zostało nigdy przez Conan-Doyle'a napisane, to mógłby istnieć taki utwór jego autorstwa, z którego wynikałoby zdanie *Sherlock Holmes nie jest detektywem*.

Po drugie, można bronić poglądu zgodnie, z którym niektóre nazwy fikcyjne mają nominaty, choć nie są to nominaty, którym przysługuje istnienie czasoprzestrzenne.

Choć w tym sensie one nie istnieją, to można powiedzieć, że jakoś są lub egzystują. Prawdopodobnie można byłoby się tu odwoływać do pewnej analogii z twierdzeniami, np. arytmetyki, twierząc, że podobny status ontyczny mają liczby — istnieją, lecz nie istnieją czasoprzestrzennie. Nikt rozumny nie sądzi, że ich nominatami są rzeczy lub ich stany. Nie próbuje się również ich szukać wśród indywiduów takiego typu. Z drugiej jednak strony o żadnej liczbie naturalnej nie orzeka się własności, która przysługuje przedmiotom czasoprzestrzennym. Nie twierdzi się, np., że najmniejsza liczba naturalna jest detektywem lub że potrafi latać. Twierdzenie takie byłoby uznane za nonsensowne. Takie twierdzenie — potraktowane całkiem poważnie, a przynajmniej literalnie — zmuszałoby do odszukania wśród bytów materialnych takich, które są nominatami liczb naturalnych, i ustalenia, czy mają one cechy wyrażane przez odpowiednie predykaty. Takich jednak obiektów wśród bytów tego rodzaju nie ma, toteż, np. zdanie *2 jest detektywem* należy uznać za pozbawione wartości logicznej (a być może również jakiegokolwiek sensu). Analogicznie, uznać należy zdanie *Sherlock Holmes jest detektywem* za pozbawione wartości logicznej, nawet jeśli nominat tej nazwy istnieje w sensie pozaczasoprzestrzennym, o ile predykat *jest detektywem* ma być interpretowany «normalnie», tj. jako wyrażający określoną własność przysługującą ludziom<sup>4</sup>.

Przekonanie o jedynie pozornej pustości referencyjnej nazw fikcyjnych wydaje się przekonaniem stosunkowo popularnym, nie tylko wśród profesjonalnych filozofów. Podczas zajęć z logiki prowadzonych na różnych kierunkach o profilu humanistycznym nieraz przekonywałem się, że podczas rozdawanej na początku kursu ankiety znaczący odsetek studentów (zwykle nieco ponad 60%) uznawał za fałszywą wypowiedź (lub rzadziej za niezrozumiałą):

*N* nie istnieje ani w świecie rzeczywistym, ani w wyobrażonym, ani literackim, ani w świecie wyobraźni, ani w żadnym innym świecie,

przy czym *N* reprezentowało jakąś powszechnie znaną nazwę fikcyjną. Jeśli więc uznaje jakaś osoba takie zdanie za fałszywe, to dopuszczalne jest przypisanie jej przekonania o prawdziwości zdania *N istnieje w świecie rzeczywistym, wyobrażeniowym, literackim, w świecie wyobraźni lub jakimś innym*; wolno przypisać jej więc przekonanie o niepustości nazwy *N*. Te intuicje znajdowały i wciąż znajdują filozoficznie znamienitych obrońców (Meinong 1904). Na przykład Kripke uważa, że nazwy tego rodzaju nazywają jakieś przedmioty abstrakcyjne (2011, s. 63). Uznając trafność zastrzeżeń wysuwanych wobec takich koncepcji (patrz np. Cernera 2012; Everest 2007), uważam, że teorie są błędne. Zdania z użytymi nazwami fikcyjnymi są pozbawione treści semantycznej.

Po trzecie, wskazany na początku problem można rozwiązać, utrzymując, że zdania z fikcyjnymi zdaniami — traktowanymi nie jako skróty, lecz w pełni rozwinięte zdania — są pozbawione — intuicyjnie często przypisywanej im — wartości logicznej. Zdania tego rodzaju nie są ani prawdziwe, ani fałszywe, a powszechne opinie innych ludzi są w tym względzie błędne. Bertrand Russell w często przytaczanym fragmencie pisze:

Wprowadzona w błąd przez gramatykę większość logików, którzy zajmowali się wspomnianą kwestią [nierealności], rozwiązywała ją, kierując się błędnymi zasadami.

<sup>4</sup>Jacek Pańniczek pisze (Pańniczek 1993, s. 173): „Gdy mamy do czynienia ze zdaniami o budowie podmiotowo-orzecznikowej [...] [z nazwą fikcyjną w roli podmiotu — przyp. T.P], to dwuznaczność tę wolno wyjaśnić bądź (a) dwuznacznością grupy orzecznikowej (predykatu), bądź (b) dwuznacznością spójnika ‘jest’, bądź wreszcie (c) dwuznacznością nazwy fikcyjnej tworzącej podmiot zdania”. Wolno też jednak przyjąć, że zdanie takie jest pozbawione semantycznej sensowności.



Uważali oni formę gramatyczną za przewodniczkę w analizie bardziej pewną, niż nią jest w rzeczywistości. [...] Wielu logików doszło do wniosku, że istnieją obiekty nierealne. [...] W teoriach takich, wydaje mi się, brak tego poczucia rzeczywistości, które należy zachować w najbardziej abstrakcyjnych rozważaniach. Logice, skłonny byłbym twierdzić, nie wolno w większym stopniu uznawać jednorożca, aniżeli toleruje się to w zoologii; logika bowiem zajmuje się światem realnym tak samo zgodnie z prawdą jak zoologia [...]. Twierdzić, że istnieją jednorożce w heraldyce czy w fantazji, to jak najbardziej żałosny i marny wykręt. [...] Istnieje tylko jeden świat, świat realny: jego częścią jest wyobraźnia Szekspira; także te myśli, które miał on pisząc Hamleta, są realne. [...] Ale to należy właśnie do istoty fikcji, że tylko owe myśli, uczucia itd. Szekspira i jego czytelników są realne, nie istnieje zaś oprócz nich żaden obiektywny Hamlet. Jeśliście wzięli pod uwagę wszystkie uczucia wzbudzone przez Napoleona w piszących i czytających historię, toście nie zetknęli się z nim: natomiast w wypadku Hamleta na tym sprawa się kończy. Gdyby nikt nie myślał o Hamlecie, nic by z niego nie pozostało; gdyby zaś nikt nie myślał o Napoleonie, już on by się zaraz postarał, żeby o nim pomyślano (Russell 1967, 279-280; patrz również Pelc 1967, s. 325-326).

Czy taką tezę można obronić i jakie zarzuty przeciwko niej należy rozważyć?

#### 4. Informatywność zdań z nazwami fikcjonalnymi użytymi ekstratekstowo

Jest oczywiste, że jeśli *Sherlock Holmes* nie ma żadnego nominatu, to zdanie

(*SH*) *Sherlock Holmes jest detektywem*

nie informuje o żadnym stanie rzeczy. Zauważmy również, że jeśli rozumienie danego zdania utożsamimy z wiedzą o okolicznościach, w jakich byłoby ono prawdziwe, to zdanie (*SH*) nie może wydawać się niezrozumiałe — skoro nie ma nominatu *Sherlock Holmes*, nie wiadomo, czy ma on własność wyrażoną przez predykat jest detektywem. Jeśli (*SH*) nie byłoby zrozumiałe, to nie byłoby powodów, aby uznawać je za sensowne. Kiedy bowiem należałoby uznać je za prawdziwe? Jest wątpliwe, czy uznalibyśmy (*SH*) za prawdziwe (w użyciu ekstratekstowym) w sytuacji, w której spotkalibyśmy kogoś, kto nazywałby się *Sherlock Holmes* i byłby zarazem detektywem. Jeżeli na podstawie znajomości odpowiednich dzieł z tym imieniem można wiązać pewny zbiór deskrypcji określonych  $D_1, \dots, D_n$ , to — zauważmy — spostrzeżenie, że istnieje pewna osoba spełniająca jakąś deskrypcję z tego zbioru i zarazem będąca detektywem, nie prowadziła by nikogo do wniosku, głoszącego, że (*SH*) jest prawdziwe. Identycznie byłoby w sytuacji, w której zaobserwowalibyśmy osobę, która spełniałaby nie jedną, lecz wszystkie deskrypcje z podanego zbioru. Trudno byłoby uznać użyte ekstratekstowo (*SH*) za prawdziwe nawet w sytuacji, w której żylibyśmy w czasach współczesnych Conan Doyle'owi i spotkalibyśmy detektywa, który odpowiadałby opisowi zamieszczonemu w *Studium w szkarłacie*. Nie wiadomo, w jaki sposób ustalić to, do kogo miałyby odnosić się podmiot takiego zdania, rozpatrywanego również w jakich wyobrażonych sytuacjach lub światach, skoro to imię faktycznie do nikogo się nie odnosi, ani nie odnosiło. Ustalenie warunków, w których należałoby uznać (*SH*) za prawdziwe wydaje się trudne, jeśli nie niemożliwe.

Analogicznie, nie informuje wtedy o żadnym stanie rzeczy również negacja (*SH*):

(*NEG-SH*) *Nieprawda, że Sherlock Holmes jest detektywem.*

Skoro bowiem jakieś zdanie nie zawiera żadnej informacji, to jego zaprzeczenie również nie informuje o niczym lub informuje o wszystkim. Ta druga możliwość wydaje się jednak

nieprawdopodobna. Podobnie wydaje się nieprawdopodobne, że Sherlock Holmes jest detektywem nie informowało o niczym, podczas gdy jego negacja miała jakąś określoną treść. Co prawda można argumentować, że  $(NEG-SH)$  jest równoznaczne z alternatywą: (i) *Nie-(Sherlock Holmes) jest detektywem*, (ii) *Sherlock Holmes nie jest detektywem*, lub (iii) *Sherlock Holmes jest nie-detektywem*; (i) jest prawdziwe, ponieważ jest ktoś, kto nie jest Sherlockiem Holmsem, a jest detektywem. W takim wypadku można uznać, że — skoro (i) jest prawdziwe, a pozostałe zaś są pozbawione sensu, to  $(NEG-SH)$  jest zdaniem sensownym, tzn. wyraża, że pewna nieokreślona osoba jest detektywem. Takie rozumowanie można przyjąć, jeśli uzna się zarazem, że

(I) dla dowolnych zdań podmiotowo-orzecznikowych, zachodzi: *Nieprawda, że N jest P* jest równoznaczne z: (i) *nie-N jest P*, (ii) *N nie jest P*, lub (iii) *N jest nie-P*,

a także

(II) jeżeli zdanie  $Z$  jest równoznaczne z  $Z^*$  lub  $Z^{**}$ , to jeżeli  $Z^{**}$  jest pozbawione semantycznej treści, to  $Z$  jest równoznaczne z  $Z^*$ .

Gdyby jednak taka argumentacja za informatywnością  $(NEG-SH)$  była zasadna, to należałoby uznać, że zdanie *Nieprawda, że Sherlock Holmes jest P*, gdzie  $P$  reprezentuje jakąś deskrypcję określoną lub predykat spełniany wyłącznie przez jeden obiekt, informuje o jedynym przedmiocie, który spełnia  $P$ . To zaś wydaje się nieintuicyjne, gdyż trudno przyznać, że zdanie *Nieprawda, że Sherlock Holmes jest stolicą Polski* jest zdaniem o Warszawie.

Podsumowując — proste zdania z użytymi (nie metajęzykowo) nazwami fikcjonalnymi nie wnoszą niczego do wiedzy o świecie. Dotyczy to również twierdzeń egzystencjalnych w rodzaju:

*Sherlock Holmes istnieje; Sherlock Holmes nie istnieje.*

Ponieważ pierwsze zdanie jest pozbawione wartości logicznej, to również nie można jej przypisywać drugiemu zdaniu. Gdyby pierwsze było prawdziwe, to wynikałoby z tego, że istnieje nominat nazwy, która występuje w podmiocie, i że posiada on jakieś cechy. Tymczasem nominat tej nazwy nie istnieje. Gdyby zaś było fałszywe, oznaczałoby to, że ten nominat nie ma żadnej własności, co również nie jest prawdą, ponieważ ta nazwa jest pozbawiona odniesienia. Kiedy więc ktoś — świadomy semantycznych własności nazw i zamierzający odnieść się do stolicy Polski oraz postaci detektywa wymyślonej przez Artura Conan Doyle'a — wypowiada zdania

*Istnieje Warszawa,  
Istnieje Sherlock Holmes,*

to tylko za pierwszym razem mówi coś sensownego.

## 5. Interpretacja metajęzykowa

Kuszącym rozwiązaniem jest przyjęcie, że zdania w rodzaju

*Sherlock Holmes nie istnieje*

interpretować należy metajęzykowo jako:

*Nazwa „Sherlock Holmes” niczego nie nazywa*

a nawet że zdania te są równoważne, np. w takim sensie, że pierwsze jest skrótową formą drugiego, podobnie jak — powiedzmy — za skrótową formę zdania „Maria znenawidziła

Jana, a on ją także znienawidził” można uznać zdanie „Nienawiść zapanowała między Marią a Janem”.

Za tym rozwiązaniem może przemawiać to, że ilekroć prawdziwe jest zdanie  $\lceil N \text{ istnieje} \rceil$ , to nazwa  $\lceil N \rceil$  coś lub kogoś nazywa (i *vice versa*: ilekroć  $\lceil N \rceil$  nazywa coś lub kogoś, prawdziwe jest zdanie  $\lceil N \text{ istnieje} \rceil$ )<sup>5</sup>. Jeżeli więc:

(R)  $\lceil N \rceil$  nazywa coś zawsze i tylko wtedy, gdy prawdziwe jest  $\lceil N \text{ istnieje} \rceil$ ,

to:

$\lceil N \rceil$  niczego nie nazywa zawsze i tylko wtedy, gdy nie jest prawdziwe  $\lceil N \text{ istnieje} \rceil$ .

Jednak — zauważmy — to, że  $\lceil N \text{ istnieje} \rceil$  nie jest prawdziwe, nie upoważnia nas jeszcze do wyprowadzenia wniosku, że fałszywe (a więc i sensowne) jest  $\lceil N \text{ istnieje} \rceil$  i — dalej — że jest prawdziwe:  $N$  nie istnieje. Innymi słowy, konsekwencją (R) nie jest zdanie głoszące, że  $\lceil N \rceil$  nie nazywa niczego zawsze i tylko wtedy, gdy prawdziwe jest  $\lceil N \text{ nie istnieje} \rceil$ .

Jaki jest status (R)? Czy pozwala powiedzieć:  $\lceil N \text{ istnieje} \rceil$  znaczy tyle samo, co  $\lceil \lceil N \rceil \text{ nie jest nazwą pustą} \rceil$ ? Można argumentować, że równoważność (R) nie ma charakteru koniecznego, tj. można bronić prawdziwości zdania:

Możliwa (w określonym świecie możliwym  $m$ ) jest następująca sytuacja: Sherlock Holmes istnieje (w  $m$ ), lecz (w  $m$ ) „Sherlock Holmes” niczego nie nazywa.

Czy jednak jest to faktycznie zdanie prawdziwe? Gdyby było prawdziwe, byłoby również sensowne. Ale czy zdaniem sensownym jest zdanie „Sherlock Holmes istnieje (w możliwym świecie  $m$ )”, czy też „Możliwe (w  $m$ ), że Sherlock Holmes istnieje (w  $m$ )”? Wszak w zdaniach tych użyta jest nazwa fikcyjna, a sensowność zdań tego typu jest właśnie przedmiotem dyskusji. Jeżeli jednak byłoby sensowne, to dlaczego pozbawione sensu miałyby być zdanie *Sherlock Holmes istnieje (w świecie rzeczywistym)*? W celu uzasadnienia sensowności tych zdań można posłużyć się analogią, która zachodzi pomiędzy nimi a zdaniami, w których nie występują nazwy pozbawione nominatów, takimi jak: *Możliwe jest, że Aleksander Kwaśniewski istniałby w pewnym świecie możliwym, ale miałby inne imię i nazwisko*. Takie zdanie — podobnie jak inne z nazwami mającymi referencję — wydawać się będzie sensowne, a nawet prawdziwe. Jednak jest to dokładnie taki sam rodzaj uzasadnienia, który mogą przywołać stronnicy sensowności zdań prostych z nazwami fikcyjnymi: zdanie *Sherlock Holmes jest detektywem* jest sensowne, ponieważ sensowne jest zdanie *Rutkowski jest detektywem*. Sądzi się, że skoro dwa zdania mają taką samą strukturę syntaktyczną, to mają taką samą własność semantyczną — a ściślej, że strukturalne podobieństwo dwóch zdań przesądza o ich semantycznej sensowności. Przyjmuje się, że są albo zarazem sensowne, albo nie są. Większości użytkowników można chyba przypisać takie przekonanie. Stąd też można uznać — pochopnie — argumentację przeciw koniecznemu statusowi (R) za niekonkluzywną. Zauważmy bowiem, że skoro prawdziwe wydaje się zdanie *Możliwa jest sytuacja: Aleksander Kwaśniewski istnieje, ale „Aleksander Kwaśniewski” nikogo*

<sup>5</sup>Można, choć błędnie, również powiedzieć (jak Kripke):

Ilekroć ‘ $N$  nie istnieje’ jest prawdą, tylekroć ‘ $N$ ’ niczego nie nazywa (i *vice versa*).

Należy jednak zauważyć, że zgodnie ze stanowiskiem TBO:

1. Zdanie ‘ $N$  nie istnieje’ nie jest zdaniem prawdziwym
2. Jeżeli ‘ $N$ ’ nazywa jakiś przedmiot, to zdanie ‘ $N$  nie istnieje’ jest fałszem.

*nie nazywa, to Aleksander Kwaśniewski istnieje* nie znaczy „*Aleksander Kwaśniewski*” nie jest nazwą pustą (ponieważ kiedy  $X$  znaczy (na mocy reguł języka  $J$ )  $Y$ , to nie może być tak (na mocy reguł j. polskiego), że  $p(X)$  i nieprawda, że  $p(Y)$  (gdzie  $p(Z)$  — przykład na język polski zdania  $Z$ )). A skoro tak, to — ogólnie —  $\lceil N \text{ istnieje} \rceil$  nie znaczy  $\lceil \lceil N \rceil \text{ nie jest nazwą pustą} \rceil$ . Wtedy jednak  $\lceil N \text{ nie istnieje} \rceil$  nie znaczy  $\lceil \lceil N \rceil \text{ nie jest nazwą pustą} \rceil$ .

Kripke zauważa również, że intuicyjnie fałszywe zdanie:

*Gdyby Tomek został nazwany po ulubionej postaci Flaminga, to James Bond by istniał,*

interpretowane zgodnie z metajęzykową interpretacją zdań egzystencjalnych daje:

*Gdyby Tomek został nazwany po ulubionej postaci Flaminga, to nazwa „James Bond” nie byłaby pusta.*

A to wydaje się fałszem. Interpretacja metajęzykowa zdań z nazwami fikcjonalnymi wydaje się więc niemożliwa do utrzymania. Zwolennik tej tezy mógłby jednak odpowiedzieć, że jego interpretacja nie ma na celu podania sposobu przełożenia zdań intuicyjnie prawdziwych (ponieważ te i tak uznaje za pozbawione sensu) na sensowne, lecz jej zadaniem jest przedstawienie procedury interpretacji zdań pozbawionych literalnego sensu na sensowne, niekoniecznie z zachowaniem intuicyjnych ocen użytkowników co do wartości logicznych tych pierwszych.

## 6. Sensowność zdań o przekonaniach z nazwami fikcjonalnymi

Jak pokazaliśmy, interpretacja metajęzykowa zdań z nazwami fikcjonalnymi jest dyskusyjna, przynajmniej w przypadku zdań, w których nie używa się wyrażen, oznaczających postawy epistemiczne. A jak jest w przypadku sensowności zdań w rodzaju:

*Jan uważa, że Sherlock Holmes jest detektywem.*

Czy są to zdania sensowne? Czy w ich przypadku możliwa jest interpretacja metajęzykowa? Czy można wierzyć w Świętego Mikołaja? Czy może należy powiedzieć: czy można wierzyć w to, że ta nazwa ma nominat? Czy można uznać, że zdanie *Jan wierzy, że Święty Mikołaj daje mu prezenty* wydaje się fałszywe, ponieważ skrótowo wyraża pewną informację, którą należałoby precyzyjniej wyrazić za pomocą *Jan uważa, że istnieje nominat „Święty Mikołaj” i czymkolwiek on jest, to daje mu prezenty?*

O ile większość badaczy rozumie — choć nie zawsze podziela lub uznaje za wystarczające — racje, stojące za negacją sensowności (a przynajmniej niemożnością przypisania wartości logicznej) prostych zdań z użytymi nazwami fikcjonalnymi, to konsekwentne uznanie zdań o przekonaniach, w których zdaniach podrzędnych zostały użyte tego rodzaju imiona, wywołuje bardziej stanowczy sprzeciw. Jeśli nawet zdania w rodzaju

- (1) *Koszalek Opalek opowiadał legendy,*
- (2) *Święty Mikołaj przynosi prezenty*

nie informują o żadnym stanie rzeczy (podobnie jak ich zaprzeczenia), i nie wiadomo, w jakich okolicznościach należałoby uznać takie zdania za prawdziwe, to większą trudność sprawia uznanie, że zdania o pewnym małoletnim Antku pozbawione są wartości logicznej, czy tym bardziej sensowności, że nie informują o niczym:

(3) *Antek utrzymuje, że Koszałek Opalek opowiadał legendy*

(4) *Antek wierzy, że Święty Mikołaj przynosi prezenty.*

Wydaje się bowiem, że zdania (3) i (4) są informatywne, przynajmniej w tym sensie, że pozwalają przewidywać np. pewne zachowania Antka wobec niektórych zdań zrozumiałego dlań języka. Jeśli tym językiem jest język polski, to można przypuszczać, że Antek będzie gotowy uznać za prawdziwe zdanie *Święty Mikołaj przynosi prezenty*, że na pytanie *Kto przynosi prezenty na Święta?* odpowiada: *Święty Mikołaj* itd. Wydaje się również, że — przynajmniej w typowym przypadku — prawdziwość (3) i (4) pozwala przewidywać pewne zachowania Antka, które nie muszą być przez niego werbalizowane i mieć takiego charakteru. Antek może w wigilię Bożego Narodzenia niecierpliwie wyglądać przez okno, zaglądać pod choinkę, wypatrywać kogoś, kto spełnia opis tego, kto rzekomo przynosi tego dnia prezenty. Można sobie wyobrazić, że widząc starca z białą brodą, ubranego w czerwony kubrak, Antek myślałby, że za chwilę ziszczą się jego marzenia o wyczekiwanym prezencie. Zachowania te — nie tylko werbalne — mogłyby być odmienne niż wtedy, gdyby prawdziwe było

(5) *Antek wierzy, że Superman przynosi prezenty.*

W tym bowiem wypadku — poza inklinacją do uznawania prawdziwości innego zdania — Antek mógłby wyczekiwać na pojawienie się kogoś stosunkowo młodego i ubranego w obcisły trykot i zamaszystą pelerynę. A zatem wydaje się, że nawet jeśli wartość poznawcza zdań *Święty Mikołaj przynosi prezenty* i *Superman przynosi prezenty* jest taka sama (tj. ściślej — żadna), to zdań (4) i (5) — różna. Na przykład Tadeusz Ciecierski stwierdza (2013, s. 207):

Jeśli w ogóle chcemy na serio mówić o przekonaniach i zdaniach przekonaniowych, musimy mieć narzędzie pozwalające na opisywanie nastawień sądzeniowych dotyczących flogistonu, perpetuum mobile, świętego Mikołaja, Atlantydy lub Eldorado, a zatem nastawień sądzeniowych, które ludzie rzeczywiście często posiadają lub które często sobie przypisują. Istnienie takich przekonań jest faktem i nie ma żadnej analogii między sytuacją, w której przedmiotem analizy jest język fikcji, a sytuacją, w której mówimy o przekonaniach oraz zdaniach przekonaniowych na temat nieistniejących przedmiotów. Niezależnie jak bardzo kontrowersyjny byłby status logiczny zdania „Otello kochał Desdemonę”, to status logiczny zdania „Jan wierzy, że Otello kochał Desdemonę” jest — jeśli tylko Jan uwierzył Szekspirowi — w ogóle niekontrowersyjny.

Ogólnie rzecz biorąc, przekonania często przypisuje się podmiotom na podstawie ich werbalnych deklaracji, zgodnie z tzw. Zasadą Uwłasnosłowienia (patrz np. Puczyłowski 2014, s. 274–276):

Jeżeli osoba  $X$  uznaje za prawdziwe (lub wysoce prawdopodobne) zrozumiałe zdanie języka polskiego  $\ulcorner p \urcorner$  (w którym nie występują wyrażenia okazjonalne, ekspresywizmy), to prawdziwe jest zdanie  $\ulcorner X \text{ uważa, że } p \urcorner$ ,

jak również uważa się, że prawdziwość raportów przekonaniowych pozwala na określone predykcje zachowań podmiotów wobec zrozumiałych dla nich zdań:

Jeżeli o pewnej osobie  $X$  (rozumiejącej język polski) prawdziwe jest zdanie  $\ulcorner X \text{ uważa, że } p \urcorner$ , to  $X$  uznaje za prawdziwe (lub wysoce prawdopodobne)  $\ulcorner p \urcorner$  (w którym nie występują wyrażenia okazjonalne, ekspresywizmy).

Przekonania wyrażone w powyższych dwóch zasadach natrafiają jednak na znane trudności, określane w literaturze mianem problemu Kripkego (Kaplan 1968, Zabłudowski 2009). Należy też zauważyć, że uściślenie kłopotliwej reguły uwłasnosłowienia, przykładowo (Moore 1999):

Jeżeli (i) kompetentny użytkownik języka *A* szczerze i po namyśle uzna w kontekście *k* za prawdziwe zdanie  $\lceil p \rceil$ , (ii)  $\lceil p \rceil$  wyraża w *k* tę samą treść, co  $\lceil q \rceil$  w kontekście atrybucji, to  $\lceil A$  uważa, że *q* jest prawdziwe w kontekście atrybucji

prowadzi w przypadku rozpatrywanych przez nas zdań do następującej trudności. Jeżeli zdanie (1) i (2) wyrażają tę samą treść (ściślej — nie wyrażają żadnej) i jeśli Antek uzna tylko pierwsze, to za prawdziwe będzie należało uznać nie tylko (3), ale i — wbrew intuicjom — (4). Oczywiście — i takie rozwiązanie wydaje się nam właściwe — można uznać, że skoro ani (1) ani (2) nie wyrażają — ściśle rzecz biorąc, użyte ekstratekstowo — żadnej treści, to nie można w ich wypadku (jak w każdym, w którym do czynienia mamy ze zdaniami z nazwami fikcjonalnymi) stwierdzić, że wyrażają tę samą treść, a w związku z tym przypisywać na podstawie uznania ich prawdziwości jakiegokolwiek przekonania treściowo określonego.

Podobnie jak w przypadku zdań prostych z użytymi nazwami dyskutowanego typu, tak w wypadku takich złożonych wypowiedzi nasuwa się kilka rozwiązań problemu sensowności zdań o przekonaniach z nazwami fikcjonalnymi. Jeśli — odwołując się do znanych odróżnień (Kaplan 1968) — *Koszalek Opatek* i *Święty Mikołaj* są w (3) i (4) użyte w sposób zwykły, a nie przypadkowy czy pośredni, to (3) i (4) z punktu widzenia semantyki ocenić należy tak samo, jak (1) i (2) (zakładając, że w (1) i (2) nazwy są użyte w taki sam sposób co w (3) i (4), a predykaty *utrzymuje, że* oraz *wierzy, że* nie generują dodatkowych komplikacji semantycznych). Jeśli — jak utrzymuję — zdania (1) i (2) są pozbawione semantycznej treści, to sensowności nie można również przypisać zdaniom (3)-(5). Jednak przyjęcie takiego twierdzenia zmusza nie tylko do wyjaśnienia pozornej sensowności (3)-(5), ale i tego, w jaki sposób zdania (4) i (5) — w równym stopniu pozbawione semantycznej treści — wydają się pozwalać na różne predykcje dotyczące zachowań Antka.

Pozorną sensowność (3)-(5) można wyjaśnić, zwracając uwagę na to, że takimi zdają się również zdania (1)-(2). Wydaje się, że ktoś, kto uznaje (1)-(2) za sensowne i podlegające wartościowaniu logicznemu, ten za takie uzna również (3)-(5). Zależność odwrotna zapewne nie jest równie jednoznaczna, ale można podejrzewać, że ktoś, kto uznaje za sensowne (3)-(5), ale neguje sensowność takie (1)-(2), odebrał jakieś filozoficzne wykształcenie. W przeciwnym razie — jak sądzę — osoba, która przypisuje sensowność zdaniom (3)-(5), uzna również, że (1)-(2) są sensowne<sup>6</sup>. Jak jednak odpowiedzieć filozofom, którzy — być może twierdzą, że (1)-(2) są wypowiedziami pozbawionymi semantycznej treści — wskazują na możliwość dokonywania różnych przewidywań dotyczących zachowań Antka przy założeniu prawdziwości (4) i (5)?

Rzeczoną różnicę można byłoby wyjaśnić, wskazując na odmienną funkcję semantyczną, którą pełnią nazwy własne w takich zdaniach, jak (1)-(2) oraz (4)-(5). Tylko w tych pierwszych nazwy pełnią swoją podstawową funkcję wskazywania na określone indywiduum. W drugim przypadku te nazwy wskazują na indywiduum, ale — jak by się chciało powiedzieć — nie w sposób bezpośredni, ale przez opis (lub może wyobrażenie) tego indywiduum, skojarzony z tą a nie inną nazwą tego przedmiotu. W przypadku zdań takich jak (3)-(5), o czyje i jakie skojarzenia wywoływane przez nazwy fikcjonalne chodzi? Czy są nimi skojarzenia osoby, do której odnosi się gramatyczny podmiot takiego zdania (tj. Antek, czy też osoby, która formułuje takie zdanie? I dlaczego skojarzone przez tę lub inną osobę deskrypcje, ale skojarzona z nim nazwa pojawia

<sup>6</sup>W każdym razie nie spotkaliśmy dotąd nikogo, kto uznając za sensowne (3)-(5), a za pozbawione treści (1)-(2), nie miałyby filozoficznego przygotowania lub zainteresowań.

się w takich zdaniach? Czy gdyby faktycznie nazwa  $N$  pełniła rolę jakiegoś skrótu dla skojarzonego z nią opisu, nie byłaby możliwa eliminacja zdań o przekonaniach z nazwami fikcyjnymi na rzecz zdań o przekonaniach, w których nazwy fikcyjne zostały zastąpione przez skojarzone z nimi opisy? Można też przyjąć, że  $N$  powinna zostać zastąpiona przez nazwę lub opis tego, do czego podmiot w rzeczywistości się odniósł, nieświadomie błędnie nazywając owo indywiduum tą nazwą fikcyjną (patrz np. Puczyłowski 2010, 2012, s. 312–331).

Wydawać się więc może, że w przypadku tego rodzaju nazw i tego rodzaju zdań (i być może tylko w ich przypadku), zachodzi:

Zdanie  $(P) \ulcorner A \text{ uważa, że } N \text{ jest } P \urcorner$  znaczy, tyle samo co  $(P^*) \ulcorner A \text{ uważa, że istnieje taki } D, \text{ który jest } P \urcorner$ , gdzie  $D$  jest opisem, który w opinii nadawcy osoba  $A$  łączy z nazwą fikcyjną  $N$ .

To, że  $(P)$  powinno być interpretowane jako równoznaczne z  $(P^*)$  okreśmy mianem hipotezy (H).

Dlaczego więc używa się  $(P)$  a nie  $(P^*)$ ? Zwolennik (H) mógłby wskazać dwa powody. Pierwszym byłby ten, że  $(P)$  jest leksykalnie prostsze od  $(P^*)$ , a zatem użycie  $(P)$  jest bardziej ekonomiczne niż  $(P^*)$ . Jest także równie komunikatywne — to drugi powód — ponieważ z użyciem danej nazwy w kontekście użycia  $(P)$  jednoznacznie skojarzony jest opis  $D$  — nie zachodzi więc znaczące ryzyko nieporozumienia.

Zauważmy jednak, że negując sensowność wypowiedzi takich, jak (3)–(5), nie należy przyjmować hipotezy (H), w proponowanym wyżej brzmieniu, ponieważ zdanie, w którym użyta jest nazwa fikcyjna nie posiada znaczenia: o ile więc  $(P^*)$  ma być semantycznie sensowne, nie może być równoznaczne z pozbawionym tej sensowności  $(P)$ . Jednakże można przeformułować (H), przyjmując:

$(P) \ulcorner A \text{ uważa, że } N \text{ jest } P \urcorner$  należy interpretować:  $(P^*) \ulcorner A \text{ uważa, że istnieje taki } D, \text{ który jest } P \urcorner$ , gdzie  $D$  jest opisem, który w opinii nadawcy  $A$  łączy z nazwą fikcyjną  $\ulcorner N \urcorner$ .

Zauważmy i podkreślmy przy tym, że to, że zdanie  $X$  należy interpretować jako  $Y$ , nie znaczy, że  $X$  znaczy tyle samo co  $Y$ .

Nasuwa się jednak następujące pytanie: dlaczego zdanie takie jak  $(P)$  należy jakoś inaczej rozumieć? Albo — chcąc kwestię postawić bardziej ogólnie — dlaczego jedno zdanie należy rozumieć tak a nie inaczej?

Zwykle — wydawać się może — zdanie należy rozumieć zgodnie z określonym przez reguły języka znaczeniem. W zwykłych okolicznościach zakładamy, że jeśli zdanie  $Z$  znaczy:  $p$ , to  $Z$  należy rozumieć jako stwierdzające:  $p$ . Normatywny charakter tego zdania może się wydawać o tyle osobliwy, że w sytuacji, w której zdanie  $Z$  znaczy:  $p$ , zdanie  $Z$  interpretuje się — po prostu — jako głoszące:  $p$ . Jednak w nadzwyczajnych okolicznościach taki postulat nie razi. Niektóre z tych nadzwyczajnych okoliczności mogą być na tyle częste, że można sformułować ogólne zalecenie, które dotyczy poprawnej interpretacji danych zdań w tego typu sytuacjach. W innych zmienne czynniki kontekstowe uniemożliwiają opisanie ogólnych reguł wskazujących, jak daną wypowiedź należy rozumieć. Na przykład nie można stwierdzić, że fraza *Ojraojraojra* powinna być — w każdych okolicznościach — rozumiana tak a tak, chociaż istnieją przecież takie komunikatywne użycia tej frazy. Ktoś, kto ją niespodziewanie wypowiada podczas przyjęcia, przerywając interlokutorowi żywiołowy wywód na temat wad gospodarza, może komunikować, że temat rozmowy powinien natychmiast zostać zmieniony. W innych okolicznościach, te same słowa — nic przecież nieznaczące — mogą

zostać użyte w sposób, dzięki któremu komunikować będą coś zupełnie odmiennego. Jednak w przypadku innych wyrażań lub wypowiedzi można podać regułę, która — przy założeniu określonych parametrach kontekstu — podaje właściwy sposób rozumienia (różny od dosłownego) określonej wypowiedzi. Kiedy np. między współpracownikami *A* i *B* utrwała się zwyczaj odnoszenia się do swoich przełożonych za pomocą słowa *S*, przyjmując można regułę zgodnie, z którą:

jeżeli osoby *A* i *B* podczas komunikacji używają zdania  $Z(S)$ , to wypowiedź *A* lub *B*:  
 $Z(S)$  należy rozumieć:  $Z$  (przełożony *A* lub *B*).

W innych okolicznościach — kiedy ktoś inny używa *S* lub *A* używa go w komunikacji z kimś innym niż *B*, reguła interpretacji zdań z tego typu słowem nie musi obowiązywać. Jeśli np. między pracownikami restauracji wytworzył się zwyczaj, żeby odnosić się do klientów za pomocą nazw potraw, które zamówili, to nonsensowne zdanie wypowiedziane przez jednego z nich: *Stek po wiedeńsku zostawił napiwek* należy interpretować jako komunikujące sąd: osoba, która zamówiła stek po wiedeńsku, zostawiła napiwek. Warto również zauważyć, że nie zawsze osoba dokonująca asercji zdania dostrzega, że powinno być ono inaczej rozumiane. Kiedy np. ktoś notorycznie przekręca imię i nazwisko polityka Marka Jurka tytułując go *Jerzym Markiem*, to zdanie wypowiedziane przez taką osobę: *Zuzanna zagłosowała na Jerzego Marka* powinno być interpretowane jako sąd o następującej treści: *Zuzanna zagłosowała na Marka Jurka*. Chociaż więc zdanie *Zuzanna zagłosowała na Jerzego Marka* nie jest równoznaczne ze zdaniem *Zuzanna zagłosowała na Marka Jurka*, a nawet to pierwsze może być pozbawione znaczenia w sytuacji, w której nie ma nikogo takiego jak Jerzy Marek, to pierwsze — w przedstawionych okolicznościach — powinno być interpretowane tak, jakby głosiło dokładnie to samo, co drugie.

Można więc utrzymywać, że w przypadku ogółu zdań o przekonaniach istnieje taka reguła ich interpretacji, jak postulowana w zmodyfikowanej (*H*), nawet jeśli większość użytkowników języka takiej potrzeby interpretacji takich zdań nie dostrzega. W szczególności w ten sposób interpretując tego typu zdania można równie skutecznie przewidywać zachowania podmiotów, o których przekonaniach się w nich mówi. Jeśli deskrypcja *D* jest treściowo uboga i głosi jedynie *Istnieje taki, który nazywa się*  $\ulcorner N \urcorner$ , to będzie wolno przewidywać przynajmniej tyle, że pewne zdanie z użyta nazwą *N* podmiot uzna za prawdziwe.

## Literatura

- Adams, F., Fuller, G., Stecker, R. (1997). The Semantics of Fictional Terms. *Pacific Philosophical Quarterly* 78, 128–148.
- Adams, F., Stecker, R. (1994). Vacuous Singular Terms. *Mind and Language*, 387–401.
- Bach, K. (2002). Giordano Was So-Called Because of His Name. *Philosophical Perspectives* 116, 73–103.
- Båve, A. (2008). A Pragmatic Defense of Millianism. *Philosophical Studies* 138 (2), 271–289.
- Braun, D. (1993). Empty Names. *Nous* 27, 449–469.
- Braun, D. (2002). Cognitive Significance, Attitude Ascriptions, and Ways of Believing. *Philosophical Studies* 108, 65–81.
- Braun, D. (2005). Empty Names, Mythical Names, Fictional Names. *Nous* 39, 596–631.



- Ciecierski, T. (2013). *Nastawienia sadzeniowe. Wykłady z filozofii psychologii*. Warszawa: Wydawnictwo Naukowe PWN.
- Cernea, M.-V. (2012). Kripke's Account of Fictional Names. *Revista Română de Filosofie Analitică* VI (2), 92–104.
- Conan Doyle, A. (2000). *Studium w szkarłacie*. Gdańsk: Tower Press.
- Everett, A. (2003). Empty Names and 'Gappy' Propositions. *Philosophical Studies* 116, s. 1–36.
- Everett, A. (2007). Pretense, Existence, and Fictional Objects. *Philosophy and Phenomenological Research* vol. LXXXIV, 56–80.
- Glüer-Pagin, K., Pagin, P. (2014). Vulcan Might Have Existed, and Neptune Not: On the Semantics of Empty Names. W: García-Carpintero, M., Martí, G. (eds.) *Empty Representations. Reference and Non-Existence*. Oxford: Oxford University Press, w druku.
- Gnatek, Z. (2011). Nazwy puste w teorii bezpośredniego odniesienia — koncepcja Davida Brauna i jej słabości. *Diametros* 27, 130–149.
- Green, M., S. (2007). Direct Reference, Empty Names and Implicature. *Canadian Journal of Philosophy* 37, 419–447.
- Gurczyński, J. (2013). Przedmiotowe logiki fikcji. Część IV: Logika fikcji J. Paśniczka. *Diametros* 36, 81–97 DOI: <http://dx.doi.org/10.13153/diam.36.2013.520>.
- Kaplan, D. (1968). Quantifying in. *Synthese* XIX (1-2), 178–214.
- Kripke, S. (2011). Vacuous Names and Fictional Entities. [w:] Kripke, S. *Philosophical Troubles*. New York: Oxford University Press, s. 52–74.
- Kroon, F., Voltolini, A. (2011). Fiction. W: Zalta, E., N. (ed.) *The Stanford Encyclopedia of Philosophy* (2011 Edition), <http://plato.stanford.edu/archives/fall2011/entries/fiction/>.
- Meinong, A. (1904). *Untersuchungen zur Gegenstandstheorie und Psychologie*. Leipzig: Verlag von Johann Ambrosius Barth.
- Moore, J., G. (1999). Misdisquotatation and Substitutivity: When Not to Infer Belief from Assent. *Mind* 108 (403), 335–365.
- Paśniczek, J. (1984). *Logika fikcji*. Lublin: Wydawnictwo UMCS.
- Paśniczek, J. (1991). Przedmioty fikcyjne a światy fikcyjne. W: Paśniczek, J. (red.) *Ontologia fikcji*. Warszawa 1991, s. 153–160.
- Paśniczek, J. (1991). Problemy logiki fikcji. W: Muszyński, Z. (red.) *Język, znaczenie, rozumienie, relatywizm*. Warszawa: Polskie Towarzystwo Semiotyczne, s. 171–184.
- Pelc, J. (1967). Wyrażenia imienne a fikcja literacka. *Studia Estetyczne* 4, 318–336.
- Parsons, T. (1980). *Nonexistent Objects*. New Haven and London: Yale University Press.
- Puczyłowski, T., A. (2010). Życzliwa interpretacja wypowiedzi a semantyka zdań o wierzeniach. *Przegląd Filozoficzny — Nowa Seria* 78, 277–295.
- Puczyłowski, T., A. (2014). *Argument z implikatury konwersacyjnej w polemikach filozoficznych*. Warszawa: Wydawnictwo Wydziału Filozofii i Socjologii UW.
- Reichenbach, H. (1948). *Elements of Symbolic Logic*. New York [wyd. polskie (fragmenty): Reichenbach, H. (1967). Elementy logiki formalnej. W: Pelc, J. (red.), *Logika i język*, tłum. J. Pelc, s. 2–222].
- Reicher, M. (2014). Nonexistent Objects. [w:] *The Stanford Encyclopedia of Philosophy* (Summer 2014 Edition), Zalta, E.N. (ed.), forthcoming <http://plato.stanford.edu/archives/sum2014/entries/nonexistent-objects/>.
- Reimer, M. (2001). The Problem of Empty Names. *Australian Journal of Philosophy* 79, 491–506.
- Russel, B. (1967). Deskrypcje. W: Pelc, J. (red.) *Logika i język*. tłum. J. Pelc, Warszawa: Państwowe Wydawnictwo Naukowe, s. 277–293.

- Tiedke, H. (2011). Proper Names and their Fictional Uses. *Australasian Journal of Philosophy* 89 (4), 707–726.
- Zabłudowski, A. (2009). O rozwiązaniu zagadki Kripkego przez Marcus i kilku pokrewnych sprawach. W: Zabłudowski, A. *Polemiki filozoficzne. Prace zebrane*. Warszawa: Aletheia, s. 197-228.
- Zalta, E., N. (1983). *Abstract Objects: An Introduction to Axiomatic Metaphysics*. Dordrecht: D. Reidel.

## Is it possible to believe that Sherlock Holmes is a detective? On the problem of meaning of belief reports

TOMASZ A. PUCZYŁOWSKI

**Abstract.** *In the paper I briefly present the problem of meaningfulness of the sentences with fictional proper names. I argue that sentences of the kind are deprived of any semantic content and therefore they lack truth value. This stand is naturally challenge with common sense intuitions according to which two alike sentences can differ in their contents and their truth-values, if the only difference between them is that a fictional name is used in one but not in the other. Even more challenging problem arises in the case of belief sentences, for sentences like “John believes that Sherlock Holmes is handsome” seem to semantically express a proposition (assuming that the name “John” has an actual referent) and I claim here they do not. I present possible explanation how to deal with such semantics intuitions and how to interpret these kind of sentences without attributing any semantic value to them and without applying to conversational implicature nor gappy propositions.*

**Keywords:** *fictional name, proper name, belief reports, meaningfulness.*

---

# Kognitywne teorie urojeń — o racjonalności urojeń

ADRIANNA SMURZYŃSKA  
*Uniwersytet Jagielloński*

**Streszczenie.** *Przedmiotem artykułu jest problematyka formalnego aspektu urojeń — rozumowania i wnioskowania. Autorka poddaje krytycznej analizie wybrane kognitywne teorie urojeń, skupiając uwagę na teoriach odgórnych (top-down) (dotyczących zmiany twierdzeń ramowych, zaburzenia Tła, mechanizmów atrybucyjnych, funkcjonowania wyobraźni i mechanizmów wnioskowania, braku harmonii między zasadami konserwatyzmu i obserwacyjnej adekwatności, problemami związanymi z odpornością urojeń na rewizje oraz z niespójnością przekonań i działań). Uwzględniając różne koncepcje urojeń, stawia pytanie o racjonalność osób z urojeniami.*

**Słowa kluczowe:** *urojenia, rozumowanie, racjonalność, kognitywne teorie urojeń*

Urojenia i halucynacje są podstawowymi symptomami zaburzeń psychicznych. W odróżnieniu od halucynacji — zaburzeń percepcji — urojenia dotyczą myślenia i sfery przekonań. We współczesnej filozofii umysłu o przekonaniu (P) osoby A możemy mówić, gdy uznaje ona P za prawdę i zgodnie z tym działa (jest to jedna z definicji, którą warto przyjąć w kontekście definicji urojeń). Przekonania można ujmować albo jako postawy albo jako reprezentacje umysłowe (Schwitzgebel 2014). Wydaje się, że definicja urojeń bliższa jest temu drugiemu podejściu. Natomiast w DSM-IV urojenie definiowane jest jako:

Fałszywe przekonanie oparte na niewłaściwym wnioskowaniu o zewnętrznej rzeczywistości, silnie podtrzymywane pomimo tego, w co wierzy prawie każdy i pomimo ustanowienia bezspornych i oczywistych dowodów lub faktów przeciwnych (DSM-IV 2000).

Powstaje pytanie, co odróżnia przekonania urojeniowe od innego rodzaju przekonań — czy różnica tkwi tylko w treści, czy też formie (sposobie rozumowania)? Wybrane kognitywne teorie urojeń posłużą do analizy problematyki związanej z formalnym aspektem urojeń — rozumowania i wnioskowania<sup>1</sup>.

## 1. Problem urojeń

Poniżej zostaną krótko omówione podstawowe aspekty problemu: definicje i rodzaje urojeń, kognitywne teorie urojeń oraz koncepcje racjonalności przekonań.

### 1.1. Rodzaje urojeń

Urojenia nie stanowią jednorodnej grupy zjawisk. Dla lepszego zrozumienia artykułu, pomocne będzie przywołanie ich klasyfikacji. Można pogrupować je na kilka

---

<sup>1</sup>Składam serdecznie podziękowania dr. S. T. Kołodziejczykowi za pomoc w pracach będących podstawą niniejszego artykułu.

sposobów: urojenia pierwotne i wtórne<sup>2</sup>, organiczne i funkcjonalne<sup>3</sup>, monotematyczne i politematyczne, ograniczone (*circumscribed*) i dopracowane (*elaborated*)<sup>4</sup> (Bortolotti 2010). W kontekście poniższych rozważań szczególnie istotną funkcję pełni podział na urojenia monotematyczne i politematyczne.

Podział na urojenia monotematyczne i politematyczne opiera się na wskazaniu ilości obszarów tematycznych, których urojenie dotyczy (Bortolotti 2010). Urojenie monotematyczne dotyczy jednego problemu i jest zazwyczaj oddzielone od innych przekonań. Urojenia politematyczne dotyczą wielu tematów oraz składają się z większej ilości przekonań, powiązanych ze sobą treściowo i formalnie.

Reprezentatywnymi przykładami urojeń monotematycznych są: urojenie Capgrasa i urojenie Cotarda. Pacjent cierpiący na urojenie Capgrasa twierdzi, że bliska mu osoba (małżonek, rodzeństwo) została porwana i podmieniona przez oszusta (Capgras, Reboul-Lachaux 1923, za: Stone, Young 1997). Urojenie to pojawia się w różnych kręgach kulturowych i historycznych. Forma pozostaje podobna, zmienia się tylko treść — kiedyś, gdy wysuwano hipotezy o istnieniu UFO, o porwaniu bliskiego posądzano kosmitów, obecnie spotyka się chorych, którzy twierdzą, że ich bliski został sklonowany (Kapusta 2010).

Takie urojenie może doprowadzić do werbalnej lub fizycznej agresji skierowanej przeciwko rzekomemu oszustowi. Dzieje się tak mniej więcej w co piątym przypadku choroby (Forstl, Almeida, Owen, Burns, Howard 1991, za: Stone, Young 1997). Pozostali akceptują „osobę podstawioną za współmałżonka”, albo wręcz traktują ją po przyjacielsku (co trzeci chory) (Wallis 1986, za: Stone, Young 1997). Interesujący wydaje się stosunek chorego do „porwanego” współmałżonka. Pacjent urojeniowy zazwyczaj nie wykazuje zainteresowania tym, co mogło się stać z „porwanym”, nie próbuje go szukać (Stone, Young 1997).

Pacjent z urojeniem Cotarda twierdzi natomiast, że jest odcięty lub martwy (Cotard 1882, za: Campbell 2001). W lżejszej formie choroby odczuwa on głęboką niechęć wobec samego siebie, w ciężkiej natomiast, neguje istnienie własnej osoby i świata zewnętrznego. Pacjent podtrzymuje te twierdzenia, pomimo świadomości ich racjonalności (Campbell 2001).

Treść urojeń monotematycznych dotyczy głównie osoby zaburzonej (jej ciała, sposobu istnienia), bliskich osób lub środowiska (Bayne, Pacherie 2004; Stone, Young 1997). Ich szczegółowa analiza pokazuje dwie warstwy tego rodzaju przekonań (Grzywa, Gronkowski 2010). Warstwa powierzchniowa zależna jest od czynników społeczno-kulturowych, głęboka natomiast jest związana z uniwersalnymi, egzystencjalnymi problemami człowieka (takimi jak lęki czy poczucie tożsamości).

Urojenia politematyczne pojawiają się zwykle w przypadku schizofrenii. Ich

<sup>2</sup>Urojenia pierwotne są oparte na nowych doznaniach osoby. Są to przekonania o tym, że dzieje się coś niezwykłego i podejrzanego. Urojenie wtórne jest natomiast sądem sformułowanym na bazie tego zmienionego doświadczenia, w odniesieniu do aktualnej sytuacji oraz kontekstu społecznego (Jaspers 1963, za: Kapusta 2010).

<sup>3</sup>Urojenia organiczne są efektem uszkodzeń mózgu, funkcjonalne natomiast nie mają podłoża organicznego i są wyjaśniane w odniesieniu do teorii psychodynamicznych i motywacyjnych. Obecnie podział ten budzi wątpliwości (Johnstone i in. 1988, za: Bortolotti 2010).

<sup>4</sup>Podział ten odnosi się do relacji urojenia do innych przekonań. Urojenia ograniczone (*circumscribed*) oddzielone są od innych przekonań, mogą wchodzić z nimi w konflikt. Urojenia dopracowane (*elaborated*) tworzą natomiast całą sieć przekonań i uwidaczniają się w zachowaniu jednostki (Bortolotti 2010; Davies, Colheart 2000).

przykładem są wszelkiego rodzaju urojenia prześladowcze i urojenia braku kontroli (na przykład nasyłanie myśli) (Bortolotti 2010). Takie przekonania urojeniowe mają wpływ na niemal wszystkie dziedziny życia pacjenta.

## 1.2. Kognitywne teorie urojeń

Kognitywne teorie urojeń można podzielić na teorie oddolne (*bottom-up*) i odgórne (*top-down*) (Bortolotti 2010). Teorie oddolne upatrują przyczyny urojeń w doświadczeniu (percepcji i doznaniach). W teoriach oddolnych uznaje się, że źródłem urojenia są zaburzenia rozumowania i języka.

## 1.3. Urojenia jako przekonania

Urojenia są z definicji traktowane jako przekonania, które z racji swej fałszywości, uporczywości oraz oparcia na niewłaściwym sposobie rozumowania, uznaje się często za irracjonalne. Warto w tym miejscu wspomnieć, że definicja urojeń z DSM-IV, budzi wiele zastrzeżeń, głównie z powodu korzystania z pojęć epistemicznych, które nie są wyraźnie zdefiniowane (Kapusta 2010). Wobec definicji, która występuje w DSM-V, nie można formułować tego rodzaju zarzutów. Urojenie jest w tej klasyfikacji „trwałym przekonaniem, które nie jest podatne na zmiany w świetle sprzecznych dowodów<sup>5</sup>” (tłum. AS) [DSM-V 2013]. Definicja ta wydaje się być jednak zbyt szeroka, ponieważ opisuje również wiele rodzajów przekonań, które nie są urojeniami.

Istnieją teoretycy urojeń, którzy nie zgadzają się z tezą, zgodnie z którą urojenia są rodzajem przekonań. Podkreślają oni doświadczeniowy charakter urojeń (Gold, Hohwy 2000): sądzą, że są one wyobrażeniami (Currie, Jureidini, 2001 za: Bortolotti 2010) lub postawami wobec pewnych reprezentacji rzeczywistości (Stephens, Graham 2006 za: Bortolotti, 2010), które mogą być związane nie tylko ze sferą poznawczą, ale również ze zmianami w odczuwaniu ciała, emocji, relacji społecznych czy środowiska (Gallagher 2009, za: Bortolotti 2010). Niektórzy uznają je wręcz za puste akty mowy (Berrios 1991, za: Bortolotti, 2010). W poniższych rozważaniach konieczne będzie uznanie urojeń za typ przekonań.

## 1.4. Problem racjonalności

Zazwyczaj przyjmuje się, że logiczność, oparcie na dowodach empirycznych lub pragmatyczność są podstawowymi kryteriami racjonalności przekonań (David, Lynn, Ellis 2010). W takim ujęciu przekonanie musi posiadać tylko jeden z tych atrybutów, żeby można je było uznać za racjonalne. Istnieją jednak również inne definicje racjonalności, a główny problem w analizie racjonalności/ irracjonalności urojeń leży w tym, że jej definicja nie jest zwykle w takich sytuacjach przywoływana.

Autorzy kognitywnych teorii urojeń często posługują się pojęciem racjonalności, którego znaczenia dokładnie nie wyjaśniają. Analiza ich artykułów pozwala na wymienienie trzech różnych odmian tej zdolności (Kapusta 2010a). Pierwsza z nich to racjonalność logiczna (proceduralna). Związana jest ona z wnioskowaniem indukcyjnym i dedukcyjnym oraz spójnością przekonań. Drugą z nich jest racjonalność epistemiczna. Dotyczy ona oceny wiarygodności dowodów, spójności systemu przekonań oraz ich rewizji na skutek zapoznania się z nowymi dowodami czy faktami. Trzecia, to racjonalność „doświadczeniowa” (*experiential*). Odnosi się ona do związku osoby ze

<sup>5</sup> *fixed beliefs that are not amenable to change in light of conflicting evidence.*

światem oraz poczucia sprawczości działań i myśli.

## 2. Podejście odgórne

Przeprowadzenie formalnej analizy urojeń wymaga opisu problematyki związanej z podejściem odgórnym (racjonalistycznym) w obrębie kognitywnych teorii urojeń. Urojenie jest w nim modyfikacją systemu przekonań, która wpływa na zmianę doświadczenia i działań (Bortolotti 2010; Campbell 2001).

Owa modyfikacja systemu przekonań, która jest związana ze stroną formalną, może mieć dwojaki wpływ na doświadczenia i działania jednostki — może warunkować całkowicie zmianę doświadczenia (warunek wystarczający) lub być niezbędną, ale niewystarczającą do jej zaistnienia (warunek konieczny).

Dla niniejszych rozważań istotne znaczenie posiada następujące pytanie: jaka zmiana rozumowania powoduje zmianę doświadczenia prowadzącą do urojeń. Może być to zmiana ilościowa (pewien rodzaj wnioskowania przebiega w inny sposób, np. przy wnioskowaniu indukcyjnym osoba dochodzi do wniosku na podstawie niewystarczającej ilości przesłanek) lub jakościowa (rozumowanie przybiera odmienną postać). Można się również zastanowić, czy zmiana ta ma charakter globalny (dotyczy całkowitej zmiany reguł rozumowania), czy lokalny (dotyczy jedynie pewnego rodzaju rozumowania). W pierwszym przypadku zmiana rozumowania byłaby warunkiem wystarczającym dla powstania urojeń. W drugim natomiast, częściowa zmiana rozumowania umożliwiałaby zmianę przekonań, ale jedynie w wyniku dłuższego procesu inferencyjnego (byłby to warunek konieczny dla powstania urojeń).

Innym pytaniem, jakie warto w tym kontekście postawić, jest kwestia świadomości zmiany rozumowania — czy pacjent zmienia sposób myślenia na podstawie jakichś wskazówek, czy też takie przekształcenie dokonuje się bez jego wiedzy. Świadomość zmiany rozumowania, czy też kontroli logicznej, jest związana z dwiema umiejętnościami — przyznawaniem zdaniom wartości logicznej i rekonstrukcją rozumowania. Pierwsza z nich jest prostsza — dotyczy orzekania prawdy i fałszu. Druga cechuje się większym stopniem złożoności i wiąże się ze zdolnością rewizji przekonań. Warto przyrzeć się, która z nich może zostać zaburzona w przypadku urojeń.

Określony rodzaj rozumowania jest zwykle utożsamiany z racjonalnością danej jednostki. Kolejny problem stanowi zatem irracjonalność pacjentów urojeniowych. Analiza tego zagadnienia odbywa się w dwóch obszarach — przy badaniu związku przekonań i doświadczenia oraz przy sprawdzaniu relacji pomiędzy przekonaniem (Davies, Coltheart 2000). Takie relacje można sprawdzać na dwóch płaszczyznach: formalnej i materialnej. Przy pierwszej z nich, bada się formę logicznego rozumowania pacjentów urojeniowych (na przykład przedstawiając im do rozwiązania zagadkę logiczną). Druga natomiast dotyczy treści rozumowania i poprawności przesłanek.

## 3. Zaburzenie rozumowania jako warunek powstania urojenia

Jeśli — zgodnie z teorią racjonalistyczną — urojenie jest skutkiem zaburzenia rozumowania, to takie zaburzenie powinno być albo wystarczającym, albo koniecznym warunkiem powstania urojeń. Globalna zmiana rozumowania mogłaby pełnić rolę warunku wystarczającego. W konkurencyjnym ujęciu zaburzenie rozumowania można by uznać za warunek konieczny. Taką rolę można by przypisywać lokalnemu defektowi rozumowania, który stanowi podstawę zaburzonego myślenia.

### 3.1. Warunek wystarczający

Przedmiotem teorii, które zakładają, że zaburzenie rozumowania jest warunkiem wystarczającym powstania urojeń, są głównie zmiany podstawowych przekonań uznawanych przez jednostkę i relacji między istniejącymi już przekonaniem.

#### 3.1.1. Zmiana twierdzeń ramowych

Wydaje się, że warunkiem wystarczającym powstania urojeń, może być zmiana twierdzeń ramowych (Campbell 2001; Kapusta 2010; Rhodes, Gipps 2008). Twierdzenia ramowe to podstawowe przekonania danej jednostki, które są podstawą wszelkiego rozumowania (Wittgenstein 1993). Przykłady takich twierdzeń to: „świat istnieje od wielu lat”; „obiekty fizyczne nie znikają, gdy przestanie się na nie patrzeć” oraz wiedza zdobywana w trakcie nauki szkolnej (Wittgenstein 1993). Dotyczą one głównie faktów, które są uznawane przez większość społeczeństwa. Te twierdzenia stanowią punkt odniesienia przy każdym badaniu prawdy i fałszu.

Hipotezy urojeniowe zdają się być przeniesione z obszaru wątpienia i uzyskują status przekonań ramowych. Podstawowe przekonania danej jednostki mogą przyjąć następującą postać: „mój mąż został podmieniony przez kosmitów” czy „jestem martwa” (Kapusta 2010). Stają się one niekwestionowanymi oczywistościami oraz stanowią punkt wyjścia dla innych przekonań. Z tego powodu trudno je odrzucić. Tak jak w przypadku innych przekonań ramowych, nie sposób ich uzasadnić, ponieważ stanowią fundament dla innych przeświadczeń (Rhodes, Gipps 2008; Kapusta 2010).

Zmiana twierdzeń ramowych może prowadzić zatem do powstania urojenia. Taki status przekonań urojeniowych sprawia, że nie tylko stają się one niekwestionowalne, ale też wpływają na znaczenie używanych pojęć (Bayne, Pacherie 2004).

#### 3.1.2. Zaburzenie Tła

Inną teorią, która może wyjaśnić proces powstawania urojeń, jest koncepcja Tła. Tło jest zbiorem zdolności czy dyspozycji, mających charakter niepropozycjonalny, które tworzą podstawę dla percepcji, myślenia i języka (Searle 1983). Jest ono takie samo u większości ludzi i stanowi podstawę przekonań zdroworozsądkowych, charakterystycznych dla danej kultury. Staje się warunkiem wszelkiego rozumowania. W odróżnieniu od twierdzeń ramowych nie stanowi podstawy uzasadnienia akceptowanych twierdzeń.

Wedle tej koncepcji, urojenia są rezultatem zawieszenia lub zmiany podstawowych założeń pre-koncepcyjnych, które zwykle ograniczają tworzenie przekonań i pragnień w obrębie określonych zasad. Wpływają one również na zdolność funkcjonowania w świecie (Kapusta 2010; Rhodes, Gipps 2008). Wynikiem zawieszenia lub przekształcenia podstawowych założeń pre-koncepcyjnych jest skłonność do akceptacji bezpodstawnych lub zadziwiających twierdzeń (Rhodes, Gipps 2008). W takim przypadku zanika również różnica między wyobraźnią i rzeczywistością oraz między świadomością siebie i innych ludzi. Prowadzi to także do zmian w języku. Zaburzeniu podlega sposób rozumienia metafor (Rhodes, Gipps 2008). Zatem zaburzenie Tła — podstawy tworzenia wszelkich przekonań — zdaje się być warunkiem wystarczającym akceptacji zaskakujących twierdzeń, czyli przekonań urojeniowych.

#### 3.1.3. Zaburzenie mechanizmów atrybucyjnych

Wydaje się, że zaburzenie atrybucji może wystarczyć do pojawienia się urojeń. Negatywne myśli osoby A na temat bliskich, które nie są przez nią akceptowane,



mogą być źródłem powstania różnego rodzaju urojeń. Niskie poczucie własnej wartości i obrona przed nim przez eksternalizację winy, prowadzi do uznania, że inni ludzie są negatywnie nastawieni do pacjenta (urojenia prześladowcze) (Davies, Coltheart 2000; Garety, Freeman 1999). Silne negatywne emocje wobec bliskich i świadomość, że te osoby powinny być mimo wszystko kochane, prowadzi do wytworzenia mechanizmu obronnego — uznania, że te emocje nie są żywione względem bliskich, ale podobnych do nich osób (urojenie Capgrasa) (Davies, Coltheart 2000). Wydaje się zatem, że w przypadku urojeń prześladowczych oraz urojenia Capgrasa, pacjent szuka przyczyny swoich doznań poza sobą — ich źródłem są inni ludzie<sup>6</sup>.

Innym rodzajem zaburzenia mechanizmów atrybucyjnych jest zaburzenie mechanizmu monitorowania własnych myśli i działań (Frith 1992, za: Kapusta 2010)<sup>7</sup>. Można je podzielić na dwa podtypy — urojenia braku kontroli (*alien control delusion*) i urojenia nasyłania myśli (*thought insertion delusion*) (Davies, Coltheart 2000). Pierwsze z nich polega na przeświadczeniu, że działania chorego są kontrolowane przez inną osobę. Drugie natomiast objawia się wrażeniem nasyłania myśli z zewnątrz. Oba te urojenia charakteryzują się pasywnością, ponieważ pacjent nie uznaje własnego autorstwa myśli i działań, ale lokuje ich przyczynę w zewnętrznym źródle (Bentall 2006).

Doświadczenie myśli, które zdają się nie być egocentryczne, ale pochodzą z zewnątrz, prowadzi do zaburzenia racjonalności (Gold, Hohwy 2000, za: Kapusta 2010). Pacjent przestaje być inicjatorem swoich myśli i działań. Staje się jedynie ich wykonawcą, co pociąga za sobą dalsze zmiany w przekonaniach i rozumowaniu.

### 3.1.4. Zaburzenia funkcjonowania wyobraźni

Umiejętność rozróżniania tego, co rzeczywiste, od tego, co wyobrażone (snu od jawy, wspomnienia od wyobrażenia) jest jedną z podstawowych umiejętności, umożliwiających działanie w świecie. Zaburzenie funkcjonowania tej zdolności może prowadzić do uznania wyobrażeń lub halucynacji za rzeczywistość i przyczynić się do powstania urojeń (Slade, Bentall, 1988, za: Rhodes i Gipps 2008; Davies, Coltheart 2000).

Konsekwencją chronicznego zamazania różnicy pomiędzy wyobrażeniem a rzeczywistością jest stworzenie świata psychotycznego, który, z biegiem czasu, coraz łatwiej pomylić ze światem realnym. Chory jest jedynie zdolny do wyobrażenia sobie świata urojonego. Pacjent może przezwyciężyć chorobę dopiero wówczas, gdy zakwestionuje jego realność (Kępiński 2001). Zaburzenie funkcjonowania wyobraźni ma bardzo poważne konsekwencje i może z łatwością prowadzić do powstania urojeń.

## 3.2. Warunek wystarczający — kontrargumenty

Żadna z powyższych teorii nie wyjaśnia procesu powstawania urojeń. Poniżej przedstawione zostaną kontrargumenty, które uwidoczniają defekty tych koncepcji.

<sup>6</sup>W przeciwieństwie do urojenia Capgrasa, przy urojeniu Cotarda pacjent szuka przyczyny zmiany afektu w samym sobie. W tym przypadku występuje globalna zmiana afektu (jego spłylenie, które może być skutkiem silnej depresji) (Debruyne, Portzky, Peremans, Audenaert 2011). Ponieważ zmiana doznań pacjenta ma charakter stały — nie dotyczy jedynie reakcji na konkretne osoby — dochodzi on do wniosku, że przyczyną tej zmiany jest on sam. W ten sposób dochodzi do przeświadczenia, że jest martwy.

<sup>7</sup>Urojenia tego typu są przede wszystkim związane ze schizofrenią.

### 3.2.1. Ad. Zmiana twierdzeń ramowych

Pomimo wielu podobieństw, przekonania urojeniowe różnią się bardzo od twierdzeń ramowych. Trzeba przy tym zaznaczyć, że koncepcja twierdzeń ramowych nie doczekała się szczegółowych analiz w żadnym z dzieł Wittgensteina. Uwagi z *O pewności* (Wittgenstein, 1993) pozwalają jednak wyróżnić kilka ich ważnych cech.

Po pierwsze, twierdzenia ramowe są podobne lub takie same u większości ludzi. Na przykład dzięki temu, że ludzie uznają, że krzesła nie znikają, kiedy nie są przedmiotem percepcji, mogą na nich usiąść bez przyglądania się im. W przeciwieństwie do twierdzeń ramowych, przekonania urojeniowe są bardziej zindywidualizowane (Kapusta 2010). Są na przykład związane z poczuciem odrealnienia. Co więcej, pacjenci urojeniowi często uznają swoje przekonania za absurdalne i dziwią się, że ktokolwiek inny mógłby je zaakceptować (Bayne, Pacherie 2004; Davies, Coltheart 2000). Na przykład, człowiek, który twierdzi, że jego żona została podmieniona, ma zwykle świadomość, że zdanie to brzmi absurdalnie. Dowodzi to, że przekonania urojeniowe nie są przyjmowane z taką samą oczywistością, jak podstawowe przekonania na temat świata.

Po drugie, twierdzenia ramowe dotyczą głównie działań i sposobu funkcjonowania w świecie. W odróżnieniu od urojeń nie są konsekwencją rozważań, lecz są przyjmowane mimowolnie (Campbell 2001; Rhodes, Gipps 2008). Co więcej, twierdzenia ramowe są na tyle podstawowymi prawdami, że nie są zwykle wypowiedane, a często nawet nie są uświadamiane (Rhodes, Gipps 2008; Wittgenstein 1993). Nie trzeba na przykład zastanawiać się, czy książki znikają po włożeniu ich do szuflady. Z przekonaniem urojeniowym jest odwrotnie — zwykle są wypowiedane wprost, ale niekoniecznie stają się podstawą działań (Bayne, Pacherie 2004). Na przykład pacjent, który twierdzi, że jego żona została porwana, nie zaczyna jej szukać.

Po trzecie, twierdzenia ramowe stanowią całą sieć zdań (Wittgenstein 1993). Przekonania urojeniowe natomiast, szczególnie w urojeniach monotematycznych, są często odizolowane i nie stanowią, w przeciwieństwie do twierdzeń ramowych, centrum przekonania danej jednostki. Zwykle nie są również fundamentem innych przekonań i działań (Bayne, Pacherie 2004; Kapusta 2010).

Po czwarte, prawdopodobnie twierdzenia ramowe osób urojeniowych są takie same jak u innych osób. Gdyby tak nie było, urojenia stałyby się całkowicie niezrozumiałe, a zwykle tak nie jest (Kapusta 2010). Gdyby pacjent podważył na przykład twierdzenia ramowe dotyczące istnienia świata, innych ludzi czy krzeseł i postępował w sposób zgodny z przekonaniem, nie mógłby usiąść, czy zacząć rozmowy z psychiatrą.

Urojenia, podobnie jak twierdzenia ramowe, nie podlegają uzasadnieniu. Gdyby koncepcja twierdzeń ramowych została doprecyzowana, można by poszukiwać dowodów na to, że powstanie urojeń wiąże się ze zmianą twierdzeń ramowych. Jednak obecna forma koncepcji twierdzeń ramowych nie wyjaśnia procesu powstawania urojeń.

### 3.2.2. Ad. Zaburzenie Tła

Jeśli Tło ma stanowić bazę dla wszelkiego rozumowania, to jego zmiana pociągałaby za sobą dużą zmianę w obszarze konceptualnym. Zaburzenie Tła skutkowałoby całościowym zaburzeniem procesu powstawania przekonań.

Jednak z łatwością można dostrzec, że gdyby przyczyna powstawania przekonań miała charakter globalny, to różnorodność typów i treści urojeń byłaby o wiele większa (Gold, Hohwy 2000, za: Kapusta 2010). Tymczasem urojenia mają często podobną treść. Urojenia monotematyczne dotyczą głównie bliskich osób, zwierząt i obiektów

(z którymi pacjent ma silny emocjonalny związek) (Bayne, Pacherie 2004). Przy silnym zaburzeniu podstawy tworzenia przekonań, prawdopodobnie nie istniałaby możliwość wyszczególnienia grup treści urojeniowych.

Co więcej, w urojeniach monotematycznych, dotychczasowa sieć przekonań pacjenta nie zostaje zmieniona, a przekonanie urojeniowe zostaje od niej odizolowane (Stone, Young 1997). Nie dochodzi tym samym do analizy hipotezy urojeniowej wewnątrz systemu, jak to bywa w przypadku niezaburzonego rozumowania, w którym zmiana ważnego przekonania pociąga za sobą szereg zmian w innych przekonaniach (Campbell 2001; Wittgenstein 1993). Zmiana sposobu rozumowania nie musi zatem dotyczyć całego systemu przekonań, a taka miałaby miejsce przy zaburzeniu Tła.

Zaburzenie rozumowania u osób urojeniowych ma zatem charakter częściowy, a nie globalny (Davies, Coltheart 2000). Koncepcja Tła nie wyjaśnia procesu powstawania urojeń.

Warto również podkreślić, że koncepcja Tła nie jest jasna. Sam Searle próbuje ją wyjaśnić za pomocą przykładu, w którym jedna osoba prosi drugą o przycięcie trawy (Searle 1995, za: Rhodes, Gipps 2008). Naturalną reakcją na taką prośbę jest wzięcie kosiarki. Gdy natomiast ktoś zaczyna ścinać trawę nożyczkami, można stwierdzić, że nie działa wedle tych samych reguł co większość społeczeństwa. Tak samo można zareagować, gdy zacznie on obcinać nie tylko trawnik przed domem, ale i dziko rosnącą trawę na łące. Tło umożliwia umiejscowienie sytuacji w odpowiednim kontekście i właściwą reakcję.

Taka interpretacja tej koncepcji podważa prawdziwość tezy, zgodnie z którą Tło jest strukturą niepropozycjonalną. Stwierdzenie, że trawę kosi się zwykle na swoim terenie za pomocą kosiarki, przypomina raczej twierdzenie ramowe, które jest podstawą określonych działań. Jeśli zatem Tło można związać ze zbiorem twierdzeń (niekoniecznie uświadamianych), to nie różniłoby się niczym, lub różniło niewiele, od twierdzeń ramowych. Tym samym, ta koncepcja byłaby narażona na zarzuty, które zostały przedstawione w poprzednim podrozdziale.

### **3.2.3. Ad. Zaburzenie mechanizmów atrybucyjnych**

Teoria, która odwołuje się do zaburzenia mechanizmów atrybucyjnych może wyjaśniać jedynie niektóre rodzaje urojeń. Co więcej, każde z doświadczeń, wiążących się z zaburzeniem mechanizmów atrybucyjnych, klóci się z codziennym przeżywaniem świata, które powinno pozostać w pamięci pacjenta. Nasyłanie myśli mógłby on uznać za omam słuchowy, a brak emocjonalnego związku z bliską osobą za zaburzenie układu limbicznego. Tak dużej zmiany w przekonaniach, nie da się wyjaśnić jedynie poprzez odniesienie do mechanizmów monitorowania myśli i działań.

### **3.2.4. Ad. Zaburzenia funkcjonowania wyobraźni**

O ile u wielu pacjentów z urojeniami łatwo można zauważyć, że uznają oni swój wyobrażony świat za rzeczywisty, o tyle nie można stwierdzić, że jest to tendencja uniwersalna. Często pacjenci urojeniowi zachowują się bowiem tak, jakby nie traktowali swoich urojeń za realne (Sass 1994, za: Stone, Young 1997). Świadczy o tym fakt, że mówią czasem o nich w sposób ironiczny lub jednocześnie sądzą, że są one czymś ważnym, ale i nierealnym. Na przykład pacjent, który z uporem twierdzi, że lekarz chce go otruć, nie sprzeciwia się zażyciu rzekomej trucizny.

Co więcej, nie wszystkie urojenia są wynikiem halucynacji lub pojedynczego wyobrażenia. Zaburzenie funkcjonowania wyobraźni nie może być zatem uznawane za

uniwersalną przyczynę w procesie powstawania urojeń.

### 3.3. Warunek konieczny

Przedmiotem teorii, które przyjmują, że zaburzenie rozumowania jest warunkiem koniecznym powstania urojeń, są głównie zmiany związku między przekonaniem a doświadczeniem, zaburzenia wnioskowania, które czerpią materiał z zastanej wiedzy i obserwacji, podatność przekonania na rewizję oraz związek przekonania i działań.

#### 3.3.1. Zaburzenie mechanizmu wnioskowania

Pacjenci urojeniowi wykazują tendencję do szybkiego przeskoku z małej ilości przesłanek do konkluzji. Zebranie zbyt małej ilości dowodów, które potwierdzałyby jakąś hipotezę, może prowadzić do pochopnej akceptacji niewłaściwego sądu (Garety, Freeman 1999). Zaburzenie rozumowania probabilistycznego i indukcyjnego u dużej części pacjentów urojeniowych prowadzi wręcz do przyjęcia konkluzji na podstawie jednej przesłanki (która obiektywnie może być informacją neutralną) (Bentall, Kinderman 1998, za: Davies, Coltheart 2000). Na przykład, niewielka zmiana w zachowaniu współmałżonka, może doprowadzić pacjenta do wniosku, że jest on zdrajcą.

Taki proces rozumowania prowadzi do przyjęcia nawet bardzo niezwykle i nieuzasadnionych hipotez urojeniowych.

#### 3.3.2. Brak harmonii między zasadą konserwatyzmu i zasadą obserwacyjnej adekwatności

W procesie powstawania przekonania biorą udział dwie zasady — konserwatyzmu i obserwacyjnej adekwatności (Fodor 1989, za: Stone, Young 1997). Zgodnie z pierwszą z nich rewizja przekonania wymaga zachowania ostrożności. Druga zasada głosi, że w procesie formowania przekonania należy brać pod uwagę dane obserwacyjne. Dzięki harmonii obu tych zasad możliwe jest jednoczesne zachowanie spójności poglądów oraz zgodność przekonania z obserwacjami. Obie te zasady działają przeciwstawnie i niezależnie względem siebie (Fodor 1989, za: Stone, Young 1997).

Wydaje się, że u pacjentów urojeniowych występuje brak równowagi pomiędzy tymi zasadami. Zasada obserwacyjnej adekwatności staje się ważniejsza — dane obserwacyjne i doznania odznaczają się o wiele większą wiarygodnością niż dotychczasowa wiedza oraz przekonania (Stone, Young 1997; Campbell 2001). Za przykład może posłużyć urojenie Capgrasa, w którym doznania afektywne są dla pacjenta ważniejsze niż dotychczasowe przeświadczenia dotyczące bliskiej mu osoby (Stone, Young 1997).

Takie funkcjonowanie pacjentów urojeniowych umożliwia przyjęcie przez nich hipotez niezgodnych z dotychczasowymi przekonaniem i zastaną wiedzą.

#### 3.3.3. Odporność na rewizję

Racjonalność zakłada umiejętność krytycznej analizy swoich przekonania. Brak umiejętności korekty hipotez w świetle zastanej wiedzy i nowych informacji jest cechą charakterystyczną pacjentów urojeniowych (Kapusta 2010; Bortolotti 2010). Krytycyzm wobec własnych objawów chorobowych stanowi wręcz warunek wyjścia z psychozy (Kępiński 2001).

Chory uznaje swoje twierdzenia za prawdziwe, nawet wówczas, gdy przedstawi się mu poważne racje przeciwko jego twierdzeniu (Rhodes, Gipps 2008). Nie przekonują go też żadne racjonalne argumenty (Kapusta 2010). Pacjentowi z urojeniem Capgrasa nie

wystarczy na przykład to, że „intruz”, który przybył po porwaniu jego współmałżonka, posiada te same wspomnienia, co ów współmałżonek.

Przekonania urojeniowe są odporne na rewizje i wykazują dużą stabilność (Kapusta 2010). Można stąd wyciągnąć wniosek, że u pacjentów urojeniowych dochodzi do zaburzenia umiejętności rekonstrukcji przeprowadzonego przez nich rozumowania. Taki sposób myślenia umożliwia podtrzymywanie twierdzeń urojeniowych.

### 3.3.4. Niespójność przekonań i działań

Racjonalnym nazywamy działanie, które jest spójne z przyjmowanymi przekonaniami. U wielu pacjentów urojeniowych można zauważyć brak tej cechy. Wyraźnym przykładem jest urojenie Capgrasa. Pacjent, który twierdzi, że jego bliski został podmieniony, nie stara się zmienić tego stanu. Nie zgłasza zaginięcia na policję i żyje dalej ze swoim „podmienionym współmałżonkiem” (Davies, Colheart 2000; Kapusta 2010; Stone, Young 1997). W przypadku takich osób zdarzają się wybuchy agresji, które są skierowane przeciwko domniemanemu intruzowi, podszuwającemu się za współmałżonka (mniej więcej w co piątym przypadku zaburzenia). Zwykle nie dochodzi jednak do próby pozbycia się „oszusta” z mieszkania oraz odszukania „porwanej osoby”. Taki pacjent nie wykazuje żadnej inicjatywy, która mogłaby doprowadzić do odnalezienia prawdziwego małżonka i zmiany istniejącego stanu rzeczy. Niespójność myśli i działań zdaje się być zatem jedną z cech towarzyszących urojeniom.

## 3.4. Warunek konieczny — kontrargumenty

Przedstawione powyżej teorie nie są w stanie wyjaśnić procesu powstawania urojeń. Żadna z opisanych cech myślenia nie jest bowiem związana jedynie z zaburzonym myśleniem. Przywołane defekty rozumowania można zauważyć również u osób, u których nie występują zaburzenia w myśleniu.

### 3.4.1. Ad. Zaburzenie mechanizmu wnioskowania

Tendencja do szybkiego przeskoku od przesłanek do wniosku nie musi być skorelowana z urojeniowością (Mortimer, in. 1996, za: Rhodes, Gipps 2008). Często nie pojawia się ona u pacjentów urojeniowych (Garety, Hemsley, Wessely 1991, za: Rhodes, Gipps 2008). Co więcej, u pacjentów nie przejawiających żadnych zaburzeń, można zauważyć taką tendencję. Przyczyną zbyt pochopnego wnioskowania może być na przykład niepokój czy zdenerwowanie (Garety, in. 1991, za: Rhodes, Gipps 2008; Stone, Young 1997).

Szybki przeskoczenie z przesłanek do wniosku może być związany z przejściem od rozumowania na poziomie materialnym do rozumowania na poziomie formalnym. Problem na tej płaszczyźnie dotyczy również ludzi zdrowych, którzy przy rozwiązywaniu zagadki logicznej o odpowiednio dobranej treści, zwykle nie rozumują w sposób logicznie poprawny.

Nawet gdyby można było dojść do wniosku, że u większości osób z urojeniami zaburzone jest rozumowanie indukcyjne, to nie można by powiedzieć, że rozumują oni w sposób nielogiczny. Zdolność rozumowania dedukcyjnego jest u nich nienaruszona (Stone, Young 1997). Zaburzenie racjonalności, które zarzuca się pacjentom urojeniowym, nie jest zatem uzasadnione.

### 3.4.2. Ad. Brak harmonii między zasadą konserwatyizmu i zasadą obserwacyjnej adekwatności

Zasada obserwacyjnej adekwatności, nawet jeśli jest traktowana w sposób uprzywilejowany, nie tłumaczy tworzenia zadziwiających hipotez urojeniowych (Stone, Young 1997). Doświadczenie wymaga wyjaśnienia, doznanie afektywne można interpretować na kilka różnych sposobów. Sama zasada nie objaśnia, dlaczego taka, a nie inna hipoteza (na przykład o podmienieniu współmałżonka) jest przyjmowana przez wiele różnych osób.

Co więcej, w życiu codziennym i w teoriach naukowych, występują twierdzenia, które nie przystają do zastanej wiedzy, a jednocześnie uznaje się je za prawdziwe (Stone, Young 1997). Jeśli na przykład jakaś teoria fizyczna wyjaśnia część zjawisk, a nie da się jej pogodzić z innymi teoriami, to jest mimo wszystko przyjmowana. Naukowcy poszukują potem teorii pomostowej, która umożliwi wyjaśnienie związku pomiędzy nimi. Przewaga zasady obserwacyjnej adekwatności nad zasadą konserwatyizmu, nie musi mieć zatem charakteru patologicznego.

### 3.4.3. Ad. Odporność na rewizje

Wydaje się, że często urojenie nie jest efektem rozumowania. Pacjent „po prostu coś wie” (Rhodes, Gipps 2008). W takim przypadku przekonania stają się odporne na wszelkie argumenty. Jeśli bowiem do jakiegoś przeświadczenia nie doszło się na drodze rozumowej, to trudno je obalić za pomocą argumentów. Jednakże, osoby, które nie cierpią na żadne zaburzenia, również akceptują niektóre przekonania nieopierające się na racjonalnych przesłankach. Wiele twierdzeń, które są akceptowane przez ludzi, nie jest skutkiem wnioskowania, ale pojawia się samoistnie (bądź pozornie samoistnie). Autor takiego przeświadczenia nie potrafi go wtedy uzasadnić; często również nikt nie wymaga od niego takiego uzasadnienia.

W przypadku normalnego rozumowania również może występować opór wobec racjonalnych argumentów. Osoba może uznawać za prawdziwe twierdzenie A, ale nie uznawać twierdzenia B, chociaż jednocześnie zdaje sobie sprawę, że A implikuje B (Davies, Coltheart 2000). Taka odporność na rewizje nie może być zatem przypisywana jedynie pacjentom urojeniowym, ale występuje również u ludzi, u których proces myślenia nie jest zaburzony.

### 3.4.4. Ad. Niespójność przekonań i działań

Żadne działanie nie jest jednoznacznie związane z jakimś przekonaniem (Williams 1970, za: Stone, Young 1997). Jeśli dana osoba mówi coś, co jest niezgodne z jej działaniem, istnieje możliwość, że nie ujawnia swoich przekonań lub ma inny plan działania, niż można sądzić. Co więcej, manifestacja przekonań zależy od okoliczności — pacjent może wstrzymać się od określonych działań w obawie przed negatywnymi konsekwencjami ze strony otoczenia (Quine 1990, za: Stone, Young, 1997). Po przeprowadzeniu analizy szerszego kontekstu sytuacji, niespójność przekonań i działań w urojeniu Capgrasa może okazać się zatem pozorna.

## 4. Czy pacjenci urojeniowi są irracjonalni?

Powyższe kontrargumenty pokazują, że zaburzenia rozumowania, jakie można zaobserwować u pacjentów urojeniowych, występują również u osób, których myślenie

nie podlega zaburzeniom. Rodzi się zatem pytanie: Czy osoby z urojeniami są irracjonalne i jakiej racjonalności od nich oczekujemy?

Pierwszym rodzajem racjonalności, jaki warto rozważyć, jest racjonalność logiczna (proceduralna). Wiąże się ona ze spójnością przekonań oraz wnioskowaniem indukcyjnym i dedukcyjnym (Kapusta 2010a). W świetle przywołanych argumentów łatwo zauważyć, że u pacjentów urojeniowych ten rodzaj racjonalności może być zaburzony, ale nie musi. Co więcej, często zauważa się, że pacjenci ze schizofrenią rozumują bardzo logicznie, a przechodzenie testów logicznego wnioskowania wychodzi im nierzadko lepiej niż osobom zdrowym (Garety, Hemsley 1997, za: Kapusta 2010a). Aspekt formalny rozumowania pozostaje zatem często w normie.

Drugim rodzajem racjonalności, jaki warto wziąć pod uwagę, jest racjonalność epistemiczna. Polega ona na właściwej ocenie wiarygodności dowodów oraz spójności przekonań i ich rewizji pod wpływem nowych informacji (Bortolotti 2009, za: Kapusta 2010a). Tych wymogów pacjenci urojeniowi zazwyczaj nie spełniają (Gerrans 2000, za: Bortolotti, 2010). Jest to związane z zaburzeniem rozumowania na poziomie materialnym (treściowym) oraz związkiem między rozumowaniem na poziomie formalnym i materialnym. W analizach urojeń nawiązuje się również do racjonalności „doświadczeniowej” (*experiential*) (Gold, Hohwy 2000, za: Kapusta 2010a). Łączy się ona głównie z poczuciem sprawczości i odniesieniem do świata. Brak tego rodzaju racjonalności można często zauważyć u pacjentów urojeniowych.

Przyjmując określoną definicję racjonalności, można odpowiedzieć na pytanie, czy osoby z urojeniami wykazują ją, czy nie. Nie można jednak stwierdzić, że są oni całkowicie irracjonalni.

## 5. Podsumowanie

W artykule zostały przedstawione wątki związane z formalnym aspektem urojeń — sposobami rozumowania i wnioskowania, jakie występują u pacjentów urojeniowych. Dokonano prezentacji argumentów, które dowodzą, że zmiany w rozumowaniu są warunkiem wystarczającym i koniecznym powstania urojeń oraz zrekonstruowano odnoszące się do nich kontrargumenty. Pod koniec postawione zostało pytanie o racjonalność osób z urojeniami.

Analiza rozumowań osób z urojeniami prowadzi do wniosku, że różnica pomiędzy ich myśleniem a myśleniem zwykłych ludzi ma charakter ilościowy, a nie jakościowy. Defekty rozumowania u osób bez zaburzeń myślenia przyjmują lżejszą formę, nie są tak długotrwałe i mają mniejszy wpływ na sposób funkcjonowania w świecie i społeczeństwie. Z tej perspektywy można stwierdzić, że tym, co powoduje powstanie urojeń, jest zmiana rozumowania, choć pojmowana nie tak drastycznie, jak zwykle się uważać. Zaburzeniu ulega zwykle rozumowanie na poziomie materialnym (semantycznym), a nie formalnym (syntaktycznym). Problemem w tej zmianie jest głównie konflikt treściowy pomiędzy poglądami, które są przyjmowane przez pacjenta.

Z powyższych rozważań można wyciągnąć jednak dwa istotne wnioski, dotyczące charakterystycznych cech rozumowania osób z urojeniami. Pierwszy z nich odnosi się do kontroli logicznej, a dokładnie do umiejętności rekonstrukcji rozumowania. O ile u osób bez zaburzeń myślenia ta umiejętność jest podatna na zniekształcenia (co wiąże się między innymi z zawodnością pamięci), to u osób z urojeniami działa w o wiele bardziej wadliwy sposób (co wiąże się z ograniczoną możliwością rewizji przekonań). Być może właśnie to sprawia, że sprzeczne przekonania, generowane przez niekontrolowane procesy

inferencji, zostają zaakceptowane, co w ostateczności prowadzi do wytworzenia się urojenia. Drugi wniosek dotyczy konfliktu treściowego, jaki występuje w przekonaniach pacjentów. Dość często są zdolni do akceptacji dwóch sprzecznych przekonań.

Przypisanie pacjentom urojeniowym racjonalności bądź nieracjonalności zależy zatem od przyjmowanej koncepcji racjonalności. Nie ulega wątpliwości, że ich rozumowanie przebiega w inny sposób niż u ludzi bez zaburzeń. Jednak wydaje się, że to nie sama forma przekonań stanowi o istocie urojenia. W celu jej odkrycia należy podjąć szersze i wielowątkowe badania.

## Literatura

- American Psychiatric Association (2000). *Diagnostic Statistical Manual of Mental Disorders*. Fourth edition. Text Revision (DSM-IV-TR).
- American Psychiatric Association (2013). *Diagnostic Statistical Manual of Mental Disorders*. Fifth edition (DSM-V).
- Bayne, T., Pacherie, E. (2004). Bottom up or top down? *Philosophy, Psychiatry and Psychology*, 11 (1), 1–11.
- Bentall, R. (2006). Madness Explained: Why We Must Reject the Kraepelinian Paradigm and Replace It with a ‘Complaint-orientated’ Approach to Understanding Mental Illness. *Medical Hypotheses* 66, 220–233.
- Bentall, R.P., Kinderman, P. (1998). Psychological Processes and Delusional Beliefs: Implications for the Treatment of Paranoid States. W: Wykes, T., Tarrier, N., Lewis, S. (red.) *Outcome and Innovation in the Psychological Treatment of Schizophrenia*. Chichester: John Wiley and Sons, s. 119–144.
- Berrios, G. E. (1991). Delusions as ‘Wrong Beliefs’: a Conceptual History. *British Journal of Psychiatry* 159, 6–13.
- Bortolotti, L. (2009). *Delusions and Other Irrational Beliefs*. Oxford: Oxford University Press.
- Bortolotti, L. (2010). Delusion, *The Stanford Encyclopedia of Philosophy*, Zalta, E.N. (red.), <http://plato.stanford.edu/archives/fall2010/entries/delusion/> (dostęp z dnia 2.09.2012).
- Campbell, J. (2001). Rationality, Meaning and the Analysis of Delusion. *Philosophy, Psychiatry and Psychology* 8, 89–100.
- Capgras, J., Reboul-Lachaw, J. (1923). L’illusion des ‘sosies’ dans un délire systématisé chronique. *Bulletin de la Société Clinique et Médecine Mentale* 11, 6–16.
- Cotard, J. (1882). Du délire des négations aux idées d’énormité. *Archives de Neurologie* 4, 152–170.
- Currie, G., Jureidini, J. (2001). Delusions, Rationality, Empathy. *Philosophy, Psychiatry and Psychology* 8, 159–162.
- David, D., Lynn S.J., Ellis, A. (2010). *Rational and Irrational Beliefs*. Oxford: Oxford University Press.
- Davies, M., Coltheart, M. (2000). Introduction. Pathologies of Belief. *Mind and Language* 15, 1–46.
- Debruyne, H., Portzky, M., Peremans, K., Audenaert, K. (2011). Cotard’s Syndrome. Mind and Brain. *The Journal of Psychiatry* 2, 67–72.
- Fodor, J. (1989). *The Theory of Content and Other Essays*. Cambridge, Massachusetts: MIT Press.
- Forstl, H., Almeida, O.P., Owen, A.M., Burns, A., Howard, R. (1991). Psychiatric, Neurological and Medical Aspects of Misidentification Syndromes: A Review of 260 Cases. *Psychological Medicine* 21, 905–910.



- Frith, C. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Hove-Hillsdale: Lawrence Erlbaum.
- Gallagher, S. (2009). Delusional Realities. W: Broome, M. R., Bortolotti, L. (red.) *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*. Oxford: Oxford University Press, s. 245–268.
- Garety, P., Freeman, D. (1999). Cognitive Approaches to Delusions: A Critical Review of Theories and Evidence. *British Journal of Clinical Psychology* 38, 113–54.
- Garety, P., Hemsley, D. (1997). *Delusions: Investigations into the Psychology of Delusional Reasoning*. Hove: Psychology Press.
- Garety, P., Hemsley, D., Wessely, S. (1991). Reasoning in Deluded Schizophrenic and Paranoid Patients: Biases in Performance on a Probabilistic Inference Task. *Journal of Nervous and Mental Disease* 179, 194–201.
- Gerrans, P. (2000). Refining the Explanation of the Cotard Delusion. *Mind and Language* 15, 111–122.
- Gold, I., Hohwy, J. (2000). Rationality and Schizophrenic Delusion. *Mind and Language* 15, 146–167.
- Grzywa, A., Gronkowski, M. (2010). Urojenia — zaburzenia treści myślenia. *Current Problems of Psychiatry* 11, 67–76.
- Jaspers, K. (1963). *General Psychopathology*, trans. J. Hoenig, M. W. Hamilton, Chicago: University of Chicago Press.
- Johnstone, E., Cooling, N., Frith, C., Crow, T., Owens, D. (1988). Phenomenology of Organic and Functional Psychoses and the Overlap between Them. *British Journal of Psychiatry* 153, 770–776.
- Kapusta, A. (2010). *Szaleństwo i metoda. Granice rozumienia w filozofii i psychiatrii*. Lublin: Wydawnictwo UMCS.
- Kapusta, A. (2010a). Racjonalność szaleństwa. Filozoficzne dyskusje wokół współczesnych koncepcji urojeń. W: Drozdowicz, Z. (red.) *O racjonalności w nauce i życiu społecznym*. Poznań: Wydawnictwo Naukowe UAM, s. 35–44.
- Kępiński, A. (2001). *Schizofrenia*. Kraków: Wydawnictwo Literackie.
- Mortimer, A., M., Bentham, P., McKay, A.P., Quemada, I., Clare, L., Eastwood, N., McKenna, P. (1996). Delusions in Schizophrenia: A Phenomenological and Psychological Exploration. *Cognitive Neuropsychiatry* 1, 289–303.
- Quine, W.V.O. (1990). *The Pursuit of Truth*. Cambridge, MA: Harvard University Press.
- Rhodes, J., Gipps, R. (2008). Delusions, Certainty, and the Background. *Philosophy, Psychiatry, and Psychology* 15 (4), 295–310.
- Sass, L.A. (1994). *The Paradoxes of Delusion: Wittgenstein, Schreber, and the Schizophrenic Mind*. Ithaca, New York: Cornell University Press.
- Schwitzgebel, E. (2014). Belief, *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Zalta, E.N (red.), <http://plato.stanford.edu/archives/spr2014/entries/belief/>.
- Searle, J. (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Searle, J. (1995). *The Construction of Social Reality*. Harmondsworth: Penguin Books Ltd.
- Slade, P.D., Bentall, R.P. (1988). *Sensory Deception: Towards a Scientific Analysis of Hallucinations*. London: Croom Helm.
- Stephens, G.L., Graham, G. (2006). The Delusional Stance. W: Cheung Chung, M., Fulford, W., Graham, G. (red.) *Reconceiving Schizophrenia*. Oxford: Oxford University Press, s. 193–216.
- Stone, T., Young, A.W. (1997). Delusions and Brain Injury: the Philosophy and Psychology of Belief. *Mind and Language* 12, 327–364.

- Wallis, G. (1986). Nature of the Misidentified in the Capgras Syndrome. *Bibliotheca Psychiatrica* 164, 40–48.
- Williams, B. (1970). *Problems of the Self*. Cambridge: Cambridge University Press.
- Wittgenstein, L. (1993). *O pewności*, tłum. M. Sady, W. Sady, Warszawa: Aletheia.

## Cognitive Theories of Delusions about Rationality of Delusions

ADRIANNA SMURZYŃSKA  
*Jagiellonian University*

**Abstract.** *The subject of this article is the formal aspect of delusions — reasoning and inference. The Author describes cognitive theories of delusions and focuses on top-down theories. She presents their pros and cons. She asks also for rationality of deluded persons, concerning different concepts of delusions.*

**Keywords:** *delusions, reasoning, rationality, cognitive theories of delusions*

---

# Próba oceny Bayesowskiego modelu anomalnego wnioskowania w procesie formowania się urojeń

BARTOSZ JANIK

*Katedra Filozofii Prawa i Etyki Prawniczej UJ*

*Zakład Kognitywistyki UJ*

**Streszczenie.** *Głównym celem artykułu jest analiza dwuczynnikowego modelu tworzenia się przekonań urojeniowych, wzbogaconego przez Bayesowski model wnioskowania w procesie tworzenia się przekonań z określonych danych. Analiza rozpocznie się od przedstawienia założeń klasycznych modeli tworzenia się urojeń i nakierowana będzie na pokazanie, że proces wnioskowania od danych anomalnych do przekonania urojeniowego może być modelowany przez anomalny model wnioskowania Bayesowskiego. Szczególnie istotne będzie pokazanie, na czym polega anomalność takiego wnioskowania. Oceniając taki model tworzenia się przekonań urojeniowych, można krytykować go, przywołując klasyczne zarzuty przeciwko analizie Bayesowskiej oraz argumenty związane z neuropsychologią urojeń.*

**Słowa kluczowe:** *wnioskowanie Bayesowskie, urojenia, neuropsychiatria*

## Wstęp

Celem niniejszego artykułu jest opis oraz ewaluacja Bayesowskiego modelu anomalnego wnioskowania proponowanego w literaturze neuropsychologicznej<sup>1</sup>. Po dokonaniu ewaluacji zaproponowane zostaną pewne ogólne modyfikacje istniejących modeli tworzenia urojeń, uwzględniające rolę anomalnego wnioskowania w tworzeniu urojeń i zbliżające owe modele do ogólnych teorii tworzenia nowych przekonań.

Urojeniami przyjęto nazywać fałszywe przekonania, które są podtrzymywane pomimo powszechnych dowodów świadczących przeciwko nim (DSM-5)<sup>2</sup>. Pobieżna analiza poszczególnych elementów definicji prowadzi do konkluzji, że definicja ta jest

---

<sup>1</sup>W tym miejscu chciałbym serdecznie podziękować Michalinie Kaczmarczyk i Pawłowi Banasiowi za pomoc w przygotowaniu niniejszego tekstu.

<sup>2</sup>W nowej definicji pominięte zostały elementy negatywne, tj. wymóg, aby przekonania nie były powszechne w systemie kulturowym oraz aby nie były to przekonania religijne, pojawiające się w klasyfikacji DSM-IV-TR. Dodatkowo zrezygnowano z określenia „fałszywe przekonania” oraz wprost nie wyrażono, że przekonania urojeniowe powstają wskutek niepoprawnego wnioskowania dotyczącego zewnętrznej rzeczywistości. Okoliczności te nie będą istotne z punktu widzenia omawianej tematyki. Należy przyjąć, że przekonania (zarówno urojeniowe jak i zwykle) tworzą się w drodze wnioskowania dotyczącego jakiejś części rzeczywistości i ten aspekt, mimo iż nie wyeksplikowany wyraźnie w definicji urojeń, będzie kluczowy z punktu widzenia niniejszych rozważań. Definicji urojeń brak w klasyfikacji ICD, formułowane w literaturze definicje co do zasady zgadzają się z definicją związaną z klasyfikacją DSM.

wysoce problematyczna. Pozostawiając na boku spory dotyczące przekonaniowego (doksastycznego) charakteru urojeń<sup>3</sup>, niniejszy artykuł poświęcony jest wyeksplikowaniu tego, co to znaczy, że przekonanie urojeniowe powstaje jako efekt wnioskowania o rzeczywistości i w jaki sposób modelować można taki proces inferencyjny. W związku z tym nie będziemy też bliżej zajmowali się ogólnymi modelami tworzenia się przekonań urojeniowych (na ten temat zob. Bell, Halligan, Ellis 2006).

## 1. Dwuczynnikowy model tworzenia się urojeń

Z punktu widzenia problematyki artykułu najważniejsze będzie przedstawienie modelu dwuczynnikowego tworzenia się urojeń, którego autorami są M. Coltheart oraz R. Langdon (Coltheart 2007)<sup>4</sup>. Model ten ma zastosowanie do urojeń monotematycznych, a więc takich, w których treść urojeniowa jest skoncentrowana na jednym temacie. Założeniem dodatkowym modelu, jest ograniczenie się do analizy tych przekonań, które powstają na drodze poznawczej tj. jako odpowiedź na uzyskane w odpowiedni sposób dane wejściowe. W szczególności model ten nie zajmuje się wnioskowaniem z przekonań, o przekonaniach czy tworzeniem się nowych przekonań na bazie przekonań już posiadanych. W literaturze proponowane są ogólne modele tworzenia się przekonań (tzw. modele neurobiologiczne), które generują wiele różnego rodzaju trudności (od których wolne są modele neuropsychiatryczne, por. Corlett 2010). W ramach omawianego modelu problem wyjaśnienia powstawania przekonań urojeniowych pociąga za sobą konieczność odpowiedzi na dwa ściśle ze sobą połączone pytania.

Pytanie pierwsze: Co odpowiada za treść urojeń? Kieruje nas ono ku koncepcji neuropsychologicznego deficytu inicjującego urojenie, tj. takiego upośledzenia mechanizmów poznawczych, które generują treści mogące być uznane za anomalne (np. niewłaściwa reakcja emocjonalna związana z percepcją wzrokową). Pacjent posiada taki deficyt, który może być źródłem myśli inicjującej powstanie przekonania urojeniowego. Warto podkreślić, że w procesie formowania się urojenia, w ramach którego zaangażowane są władze poznawcze (tj. jakiś rodzaj wnioskowania), to deficyt inicjujący dostarcza danych wejściowych (anomalnych danych inicjujących).

Pytanie drugie: Co odpowiada za uporczywy charakter urojeń? Pytanie to kieruje nas wprost do koncepcji deficytu podtrzymującego, tj. takiego upośledzenia systemu ewaluacji przekonań, które uniemożliwia monitorowanie własnych przekonań pacjenta i w efekcie przyjęcie oraz utrzymywanie przekonania urojeniowego powstałego na bazie danych anomalnych, za które odpowiedzialny jest pierwszy deficyt (Coltheart 2007, s. 1044). Obecnie zilustrujemy to, w jaki sposób wyjaśnia się tworzenie określonego syndromu urojeniowego.

Syndrom Capgrasa polega na posiadaniu przekonania, że bliscy pacjenta zostali zastąpieni identycznie wyglądającymi sobowtórami (Berson 1983). Punktem wyjścia dla tworzenia się urojenia tego typu, jest obecność spłaszczonej odpowiedzi afektywnej

---

<sup>3</sup>Odrzucenie przekonaniowej koncepcji urojeń pociągałoby za sobą zmianę definicji urojeń z uwagi na okoliczność, że alternatywne ujęcia (urojenia jako stany beztreściowe, urojenia jako stany metaforyczne, urojenia jako wyobrażenia) nie wpisują się bezpośrednio w klasyczny model wnioskowania prowadzącego do sformułowania przekonań określonej treści. Więcej informacji można znaleźć w Bayne, Pacherie (2005) oraz w Frankish (2009).

<sup>4</sup>Proponowane są różne modyfikacje tego modelu, przede wszystkim zob. Braun, Suffren (2011).

(utruty lub osłabienia odpowiedzi emocjonalnej), powodującej wrażenie, że „coś jest nie tak” podczas obserwacji twarzy bliskich pacjentowi osób<sup>5</sup>. W ramach omawianego modelu stanowi to neuropsychologiczny deficyt inicjujący. Aby powstało przekonanie urojeniowe, wymagana jest obecność deficytu podtrzymującego, tj. mechanizmu, który uniemożliwi „skorygowanie” anomalnej treści wejściowej. W przypadku syndromu Capgrasa upośledzenie to manifestuje się zmniejszonym lub upośledzonym przetwarzaniem afektywnym. Deficyt związany z podtrzymywaniem urojeń zazwyczaj ma związek z uszkodzeniem prawej półkuli — obszaru odpowiedzialnego za ewaluację przekonań. Trudności w precyzyjnym określeniu tego, na czym polega drugi czynnik, zaowocowały istotnymi badaniami dotyczącymi problematyki anomalnego wnioskowania.

## 2. Anomalność wnioskowania czy anomalność treści?

Często podkreśla się, że same dane anomalne nie są wystarczające do powstania przekonania urojeniowego (Coltheart 2010; McKay, Coltheart, Langdon 2007; Garety, Freeman 1999). W licznych publikacjach wskazuje się na przypadki, w których występuje deficyt neuropsychologiczny, polegający na generowaniu anomalnych danych wejściowych, w oparciu o które nie tworzą się przekonania urojeniowe (zestawienie literatury dotyczącej takich przypadków można znaleźć w McKay 2010)<sup>6</sup>.

Czynnik drugi, którego obecność przewiduje model dwuczynnikowy, sam w sobie wymaga pewnego dookreślenia, aby spełnić wszystkie elementy klasycznej definicji urojeń, jak i wyjaśnić sytuację, w której urojenia nie powstają mimo anomalności danych percepcyjnych. Wyjaśnienie takiej sytuacji skłania do przyjęcia, że elementem wpływającym na tworzenie się urojeń jest sposób, w jaki wychodząc od danych anomalnych wnioskuje się i ustala, jakie przekonania powinny zostać przyjęte.

Model wnioskowania ma kilka istotnych założeń, które warto tu wyeksplikować. Przede wszystkim trzeba podkreślić, iż możliwość tego, że anomalne dane wejściowe w pełni przesądzą o anomalności (urojeniowości) przekonań, w sytuacji gdy nie dochodzi do zaburzenia procesu inferencyjnego (wnioskowanie jest poprawne), wydaje się być poważnie osłabiona (jeśli nie wyeliminowana całkowicie) przez zaprezentowaną powyżej argumentację. W związku z tym, wypada przyjąć, że wnioskowanie na danych anomalnych również musi mieć anomalny charakter. Uchwycenie anomalnego charakteru takiego wnioskowania jest zadaniem niezwykle trudnym z kilku powodów. Tutaj podajemy dwa z nich.

Po pierwsze, nie mamy bezpośredniego dostępu do tego procesu. Co więcej, nawet pośredni dostęp (przez werbalne świadectwa pacjenta) wydaje się być zapośredniczony przez obecne w momencie werbalizacji procesu urojenie. Po drugie, w literaturze

<sup>5</sup>Mówiąc nieco bardziej precyzyjnie, system odpowiedzialny za rozpoznawanie twarzy zostaje odłączony od systemu autonomicznej odpowiedzi na znajome twarze. Uczucie, że „jest coś nie tak” stanowi brak odpowiedzi afektywnej w sytuacji, gdy w zwyczajnych warunkach reakcja taka powinna się pojawić (McKay 2012).

<sup>6</sup>Nie porusza się w tym miejscu kilku istotnych problemów: 1) zagadnienia, czy dane na wejściu stanowią „surowe dane anomalne”, czy są one w jakikolwiek sposób kategoryzowane lub czy mają strukturę propozycjonalną? 2) czy abdukcja jest właściwym sposobem modelowania procesu przyjmowania przekonań, czy może zachodzi inny rodzaj wnioskowania na danych, które nie muszą mieć charakteru konceptualnego oraz 3) w jaki sposób dochodzi do wyboru hipotez do ewaluacji — przed dokonaniem wnioskowania? Częściowe odpowiedzi na te pytania, można znaleźć w (Davies, Egan 2013; McKay 2010).

zaobserwować można brak zgody, co do natury takiego procesu wnioskowania. Można sugerować, że będzie to proces wnioskowania abdukcyjnego, ale taka teza musi zostać odpowiednio uzasadniona (Corlett 2010).

### 3. Wnioskowanie Bayesowskie

W modelach wyjaśniających powstawanie urojeń zakłada się, że tworząc przekonania jednostka postępuje zgodnie z Bayesowską metodą abdukcyjną<sup>7</sup>. Abdukcję należy rozumieć tutaj bardzo ogólnie, jako wnioskowanie do najlepszego wyjaśnienia. Nasuwa się pytanie, dlaczego wnioskowanie Bayesowskie zostało wybrane jako model tworzenia się przekonań.

Po pierwsze, wydaje się, że taki sposób tworzenia się przekonań uzasadniony może być probabilistycznym charakterem przekonań, który sprawia, że nasze przekonania mogą mieć pewną siłę (możemy posługiwać się stopniami prawdopodobieństwa w odniesieniu do przekonań lub przyjmować hipotezy o określonej sile jako nasze przekonania).

Po drugie, wykorzystanie reguły Bayesa do dołączania nowych przekonań do zbioru przekonań, które posiadamy oraz do wyboru przekonania, które powinniśmy dołączyć do naszego zbioru przekonań w sytuacji rywalizowania różnych hipotez (w literaturze przedmiotu: modelowanie racjonalności). Anomalne wnioskowanie Bayesowskie mogą więc być adekwatnym modelem dla wnioskowań, które *prima facie* jawią się jako wnioskowania nieracjonalne.

Po trzecie, wiele modeli prezentujących to, w jaki sposób na poziomie poznawczym przebiegają procesy powiązane z powstawaniem urojeń, zakłada obecność modelu Bayesowskiego, lub jakiejś formy Bayesowskich połączeń pomiędzy poszczególnymi układami poznawczymi, co może sugerować zasadność formułowania tezy o mózgu jako wielkiej maszynie probabilistycznej (Corlett 2010, Helmholtz 1977). Niniejsza analiza przyjmuje taki punkt widzenia.

Wracając do opisu metody, rozumie się przez nią strategię połączenia: 1) metody abdukcyjnej, tj. rozumowania zawodnego, zmierzającego do wyboru najlepszego możliwego wniosku oraz 2) wnioskowania Bayesowskiego. Zakłada się również, że ujęcie oraz akceptacja treści jest aktem jednoczesnym, który poprzedza i umożliwia ewaluację przekonań (Davies, Egan 2013). Co więcej należy przyjąć, iż przekonania lokowane są modułowo w mózgu. O ile całość systemu przekonań dopuszczać może sprzeczności, to lokalnie (w ramach modułu) treści takie nie mogą pozostawać ze sobą sprzeczne. Zmusza nas to do wyboru pomiędzy rywalizującymi hipotezami, które po uzyskaniu określonej siły, mogą być dołączone do zbioru przekonań.

Rozumowanie Bayesowskie jest rodzajem rozumowania probabilistycznego, które wychodzi od założenia, że dane, które uzyskujemy w percepcji oraz nasze przekonania mają liczbowo wyrażane poziomy prawdopodobieństwa (jesteśmy o nich przekonani w określonym stopniu), a zmiany w tych elementach dokonują się za pomocą procedury prostego warunkowania, które na podstawie nowych danych i nowych przekonań wymusza na nas zmiany w poziomach zaufania do danych i hipotez posiadanych (Langdon 2011; Langdon, McKay, Coltheart 2008; Langdon, Coltheart 2000). Wzór Bayesa jest regułą dołączania nowych przekonań, na podstawie uzyskanych danych.

<sup>7</sup>Modelowanie przekonań z zastosowaniem rachunku prawdopodobieństwa, może być interpretowane normatywnie oraz deskryptywnie. Dziękuję anonimowemu recenzentowi za zwrócenie uwagi na możliwość nie-probabilistycznego podejścia do natury tworzenia się przekonań. Zagadnienie to jednak wykracza poważnie poza ramy niniejszego artykułu.

Przedstawić go można w postaci:

$$p(H|D) = \frac{p(H)p(D|H)}{p(D)}$$

którą opisać można następująco: prawdopodobieństwo hipotezy H, pod warunkiem posiadania danych D jest równe iloczynowi prawdopodobieństwa samej hipotezy oraz uzyskania posiadanych danych w sytuacji, gdy zakładamy, że jest ona prawdziwa (innymi słowy, jak dobrze hipoteza przewiduje pojawienie się takich danych), podzielonych przez „surowe” prawdopodobieństwo uzyskania danych anomalnych. Z punktu widzenia rozważań dotyczących wnioskowania prowadzącego do formowania się treści urojonych, użyteczny będzie następujący wzór:

$$\frac{p(H_i|D)}{p(H_j|D)} = \frac{p(H_i)}{p(H_j)} \times \frac{p(D|H_i)}{p(D|H_j)}$$

który opisuje sytuację, gdy mamy dwie konkurujące ze sobą hipotezy. Najważniejsze elementy tej formuły to stosunek prawdopodobieństwa obu hipotez oraz stosunek prawdopodobieństwa uzyskania określonych danych, przy założeniu, że hipotezy te są prawdziwe.

Warto w tym miejscu prześledzić, w jaki sposób działać powinien taki Bayesowski system wnioskowania. Najłatwiej będzie rozpatrzyć przykład syndromu Capgrasa, o którym była wcześniej mowa. Dane uzyskane na wejściu to dane anomalne dotyczące niewłaściwej reakcji afektywnej stowarzyszonej z obserwacją twarzy znanej nam osoby. Czy we właściwie działającym systemie wnioskowania hipoteza, że osoba którą widzimy jest podstawionym impostorem, ma szansę stać się przekonaniem (zostać włączona do systemu przekonań)? Wydaje się, że zgodnie z racjonalnie przyporządkowanymi wartościami prawdopodobieństwa taka sytuacja nie jest możliwa, a komentowana hipoteza zostanie z łatwością odrzucona na rzecz hipotezy, że to jednak jest bliska osoba. Dzieje się tak dlatego, że pierwszy człon we wzorze opisującym konkurencję hipotez, w związku z bardzo małym prawdopodobieństwem hipotezy, że obserwowana osoba jest impostorem, każdorazowo spowoduje, iż faworyzowana będzie hipoteza, że to jednak jest bliska nam osoba. W literaturze spotkać można taki dobór wartości parametrów, który dodatkowo wydaje się być racjonalnym wyborem ich wartości (niskie prawdopodobieństwo hipotezy, że napotkana osoba jest sobowtórem, wysokie że jest kimś bliskim, wysokie prawdopodobieństwo uzyskania danych anomalnych przy założonej prawdziwości pierwszej hipotezy oraz niskie prawdopodobieństwo uzyskania danych anomalnych przy prawdziwości hipotezy, że osoba jest bliską nam osobą) który prowadzi do wytworzenia się przekonania urojeniowego<sup>8</sup>.

Aby wyjaśnić, w jaki sposób wnioskowanie staje się anomalne, proponowane są rozmaite zabiegi. Można utrzymywać, że to wartość drugiego członu, a więc stosunku tego, w jaki obie hipotezy przewidują pojawienie się anomalnych danych, musi przysłonić stosunek prawdopodobieństwa obu hipotez (Coltheart). Można też utrzymywać, że

<sup>8</sup>Racjonalna estymacja tych wartości wg. McKay'a to:  $p(Impostor) = 0.00027$ ,  $p(BliskaOsoba) = 0.99973$ ,  $p(D|Impostor) = 0.999$ ,  $p(D|BliskaOsoba) = 0.001$  i w rezultacie wartości  $p(Impostor|D) = 0.21$  oraz  $p(BliskaOsoba|D) = 0.79$ . W efekcie przyjęte zostaje przekonanie, że jest to bliska osoba. Syndrom urojeniowy nie powstaje (McKay 2010). Niznaczna zmiana wartości parametrów doprowadza do przyjęcia powstania takiego syndromu.



pojawia się nowe prawdopodobieństwo początkowe, związane z uzyskanymi danymi, które warunkują przyjęcie pierwszej hipotezy (Corlett). Ostatnią drogą jest pokazanie, w jaki sposób do określonych elementów tego wzoru przypisywana jest określona waga, w związku z czym wnioskowanie staje się „anomalne” (w znaczeniu stanu odbiegającego od założonej, Bayesowskiej racjonalności) oraz możliwe jest jego modelowanie formalne (McKay). Ostatnie ze stanowisk przedstawione zostanie dokładniej.

#### 4. Anomalny model Bayesowski (McKay)

Aby opisać, w jaki sposób podejść należy do problemu anomalnego wnioskowania, należy wrócić do dwuczynnikowego modelu formowania się urojeń i zwrócić uwagę na cechy modelu jako całości, które mogą nam coś powiedzieć o procesie formowania się przekonań.

McKay omawiając ten model wychodzi od obserwacji, że przyjmowanie i utrzymywanie przekonań opiera się na wzajemnym oddziaływaniu na siebie dwóch zdroworoządkowych zasad: (1) doksastycznego konserwatyizmu oraz (2) adekwatności obserwacyjnej (McKay 2010). Pierwsza z tych zasad mówi, że przekonania, które staną się elementem systemu przekonań jakiejś jednostki są utrzymywane z określoną siłą. Dokonywanie zmian w systemie przekonań zachodzi w drodze rewizji przekonań, ale aby doszło do takiej sytuacji, „nowe” przekonania powinny charakteryzować się przynajmniej taką samą siłą jak już posiadane oraz opierać się na adekwatnych danych obserwacyjnych. Druga zasada mówi, że tworząc system przekonań, podmiot stara się w wysokim stopniu uzgadniać przyjmowane przekonania z danymi obserwacyjnymi, które otrzymuje. Perfekcyjnie racjonalny agent, w pełnym stopniu realizuje obie zasady zapewniając równowagę pomiędzy nimi. McKay, przechodząc do procesu tworzenia się treści urojeniowych, akcentuje fakt, że tradycyjna koncepcja anomalnego procesu prowadzącego do formowania się przekonań urojeniowych (model dwuczynnikowy Colthearta), skłania się ku faworyzowaniu doksastycznego konserwatyizmu systemu przekonań (Coltheart 2007; McKay 2012). System taki w chwili „zamknięcia” będzie się opierał próbom rewizji przekonań, w tym wypadku próbom rewizji przekonania urojeniowego.

Rezultatem opisywanego procesu wnioskowania jest wytworzenie (przyjęcie) przekonania urojeniowego, którego siłę i nierewidowalność zapewnia ogólna cecha (anomalność) systemu.

Podejście takie krytykowane jest w rozmaity sposób. Najsilniejszym i zarazem najbardziej przekonywującym argumentem przeciwko takiemu podejściu (do anomalności wnioskowania) jest obserwacja, że proces wnioskowania jest bardzo czuły na dobór wartości prawdopodobieństw „wejściowych”, co z kolei prowadzi do wniosku, iż nie jest obiektywnym narzędziem rozstrzygania o przyjmowaniu (bądź nie) pewnych przekonań. Dodatkowo, jeżeli anomalność wnioskowania dotyczy już sformułowanego przekonania urojeniowego, to wydaje się, że samo przekonanie może być sformułowane bez obecności tego czynnika (choćby tylko czasowo), co blokowane jest przez argumenty z opisu spraw, w których urojenia mimo obecności danych anomalnych nie występują.

Z drugiej strony, w ujęciu McKay’a, kierowanie się ku adekwatności obserwacyjnej może być modelowane formalnie i w większym stopniu może wyjaśniać, w jaki sposób przebiega proces formowania się urojeń. Ponadto dodanie czynnika, kierującego wnioskowanie w kierunku, w którym większą rolę odgrywa prawdopodobieństwo uzyskania danych przy założeniu prawdziwości hipotezy, zmniejsza czułość całego

modelu na warunki początkowe. W literaturze proponuje się również nazywać taki sposób ewaluacji hipotez nakierowaniem na adekwatność wyjaśnienia (Davies, Davies 2009).

Naturalnie model ten nie jest wolny od problemów. Wątpliwości nastrocza sposób, w jaki dobrane powinny zostać parametry powodujące zwrócenie się ku adekwatności wyjaśnienia, co może przypominać problem z wyborem prawdopodobieństw początkowych w modelu klasycznym.

## 5. Ewaluacja i ocena

Racjonalny model tworzenia się przekonań przyjmuje następującą sytuację. Na podstawie danych, tworzy się pewną hipotezę, która w momencie gdy staje się wystarczająco prawdopodobna, przyjmuje się jako przekonanie. Rolą wnioskowania Bayesowskiego w takiej sytuacji jest ewaluacja przekonania wobec posiadanych danych w sposób, który zapewnia nam racjonalność takiego procesu. W ramach modeli tworzenia się przekonań urojeniowych największym problemem jest pokazanie, w jaki sposób dochodzi do przyjęcia przekonania urojeniowego, oraz w jaki sposób staje się ono odporne na rewizję. Bayesowskie modele tworzenia się przekonań urojeniowych mają nam pokazać, w jaki sposób proces przyjęcia przekonania jest możliwy do formalnego wyrażenia, oraz w jaki sposób dokonuje się usztywnienia przekonania tak, aby nie było ono podatne na rewizję. Pojawia się jednak kilka zasadniczych problemów, które chciałbym w tym miejscu omówić. Przyjmuję, że tworzenie takich modeli jest możliwe i chciałbym swoją krytykę skierować raczej ku pewnym elementom takich modeli, niż w ogóle przeciwko koncepcji modelowania procesu formowania się treści urojeniowych.

Po pierwsze, wracając do zarzutu, który już się pojawił, problematyczna wydaje się procedura doboru, zarówno początkowych wartości prawdopodobieństwa, jak i wartości czynnika skłaniającego procedurę wnioskowania do dryfu ku adekwatności obserwacyjnej. Zarzut podobny do powyższego znany jest w literaturze jako problem „pustości” analizy Bayesowskiej. Dotyczy rzekomego braku możliwości obiektywnego decydowania o doborze początkowych wartości (Albert, 2001; Albert, 2003).

Po drugie, problematyczna, z punktu widzenia tworzenia takich modeli, wydaje się sytuacja tylko partykularnej anomalności wnioskowań Bayesowskich (brak możliwości następczej ewaluacji przekonania urojeniowego nie wpływa na niemożliwość re-ewaluacji innych przekonań). Zastanawiającą jest okoliczność, że upośledzenie wnioskowania zachodzi jedynie w określonej klasie sytuacji. Nie możemy jednoznacznie stwierdzić, iż upośledzenie takie z konieczności powodowane jest przez uszkodzenie strukturalne prawej półkuli mózgu (Braun, Suffren 2011). Uszkodzenie takie jest warunkiem koniecznym, ale nie jest warunkiem wystarczającym dla pojawienia się takiego upośledzenia wnioskowania. Z drugiej strony, gdy powiemy że warunkiem koniecznym i wystarczającym jest jednoczesne wystąpienie danych anomalnych i omawianego uszkodzenia, to albo zmuszeni jesteśmy wskazać, w jaki sposób na poziomie funkcjonalnym dochodzi do połączenia obu czynników albo, wobec braku takiego połączenia, należy badać przypadki, w których mimo uszkodzenia mózgu i obecności deficytów neuropoznawczych urojenia się nie pojawiają.

Po trzecie, należy się zastanowić, czy model ewaluacji przekonań, w którym o „miano” przekonania rywalizują dwie hipotezy, które się wzajemnie dopełniają, jest właściwą procedurą. Przede wszystkim zastanawia sposób sformułowania sprzecznych ze sobą hipotez — nie jako zdania i jego zaprzeczenia, lecz jako dwóch zdań niezależnych dopuszczających obecność innych w przestrzeni ewaluacji. Nawet gdy

uznamy, że okoliczność ta nie stanowi trudności, pozostaje problem, w jaki sposób powinniśmy dobierać zdania, tak aby reprezentowały hipotezy, spośród których faktycznie dokonywany jest wybór. Czy model takiej rywalizacji nie powinien zostać zastąpiony jednohipotezowym modelem, w którym hipoteza staje się przekonaniem w momencie osiągnięcia pewnego określonego progu? W tej ostatniej sytuacji od razu otwiera się całe spektrum nowych problemów i możliwości uzupełnienia tego procesu (Leitgeb 2013).

## Literatura

- Albert, M. (2001). *Bayesian Learning and Expectations Formation: Anything Goes*. W: Corfield D., Williamson, J. (red.) *Foundations of Bayesianism*. Dordrecht: Kluwer, s. 351–372.
- Albert, M. (2003). Bayesian Rationality and Decision Making: A Critical Review. *Analyse and Kritik* 25 (1), s. 101–117.
- Bell, V., Halligan, P., Ellis, H. (2006). Explaining Delusions: a Cognitive Perspective. *Trends in Cognitive Science*, 10, s. 219–226.
- Bayne, T., Pacherie, E. (2005). In Defence of the Doxastic Conception of Delusions. *Mind and Language* 20(2), s. 163–188.
- Berson, R. J. (1983). Capgras' Syndrome. *The American Journal of Psychiatry* 140 (8), s. 969–978.
- Braun, C. M., Suffren, S. (2011). A General Neuropsychological Model of Delusion. *Cognitive Neuropsychiatry* 16 (1), s. 1–39.
- Coltheart, M. (2007). Cognitive Neuropsychiatry and Delusional Belief. *The Quarterly Journal of Experimental Psychology* 60, s. 1041–1062.
- Corlett, P.R., Taylor, J.R., Wang, X.J., Fletcher, P.C., Krystal, J.H. (2010). Toward a Neurobiology of Delusions. *Progress in neurobiology* (92 (3), s. 345–369.
- Davies, A.M.A., Davies, M. (2009). Explaining Pathologies of Belief. W: Broome, M.R., Bortolotti, L. (red.) *Psychiatry as Cognitive Neuroscience. Philosophical Perspectives*. Oxford: OUP, s. 285–323.
- Davies, M., Egan, A. (2013). Delusion, Cognitive Approaches: Bayesian Inference and Compartmentalization. W: Fulford, K.W.M., Davies, M., Gipps, R.G.T., Graham, G., Sadler, J., Stanghellini, G., Thornton, T. (red.) *The Oxford Handbook of Philosophy and Psychiatry*. Oxford: OUP, s. 689–727.
- Frankish, K. (2009). Delusions: A Two-level Framework. Neuroscience. W: Broome, M., Bortolotti, L. (red.) *Psychiatry as Cognitive Neuroscience. Philosophical Perspectives*. Oxford: OUP, s. 269–284.
- Garety, P.A., Freeman, D. (1999). Cognitive Approaches to Delusions: a Critical Review of Theories and Evidence. *British Journal of Clinical Psychology* 38 (2), s. 113–154.
- von Helmholtz, H. (1977). The Facts in Perception. W: tenze, *Epistemological Writings*. Dordrecht: Springer Netherlands, s. 115–185.
- Langdon, R. (2011). The Cognitive Neuropsychiatry of Delusional Belief. *WIREs Cognitive Science* 2, s. 449–460.
- Langdon, R., Coltheart, M. (2000). The Cognitive Neuropsychology of Delusions. *Mind and Language* 15 (1), s. 184–218.
- Langdon, R., McKay, R., Coltheart, M. (2008). The Cognitive Neuropsychological Understanding of Persecutory Delusions. W: Freeman, D., Bentall, R., Garety, P. (red.) *Persecutory Delusions: Assessment, Theory and treatment*. UK: Oxford University Press, s. 221–236.

- Leitgeb, H. (2013). Reducing Belief simpliciter to Degrees of Belief. *Annals of Pure and Applied Logic* 164 (12), s. 1338–1389.
- McKay, R. (2012). Delusional Inference. *Mind and Language* 27(3), s. 330–355.
- McKay, R., Langdon, R., Coltheart, M. (2007). Models of Misbelief: Integrating Motivational and Deficit Theories of Delusions. *Consciousness and Cognition* 16, s. 932–41.

## Evaluation of an Anomalous Bayesian Inference Model of Delusions Formation

BARTOSZ JANIK

*Department for the Philosophy of Law and Legal Ethics, Jagiellonian University  
Cognitive Science Unit, Jagiellonian University*

**Abstract.** *The main aim of this article is to analyze the two-factor model of the delusional beliefs formation, enriched by a Bayesian inference model. The analysis begins with a presentation of the assumptions of classical models of the formation of delusions and, as such, aims at showing that the process of inference from anomalous data to delusional beliefs in an appropriate manner can be modeled by an anomalous Bayesian inference model. In assessing such a model of delusional beliefs formation, the author differentiates between the classical arguments against Bayesian analysis and arguments from the neuropsychiatry of delusions.*

**Keywords:** *bayesian reasoning, delusions, neuropsychiatry*

