*Filozofia Nauki*
Rok I, 1993, Nr 2-3

# W sprawie obserwacyjnej definiowalności terminów teoretycznych

Jednym z klasycznych problemów empirystycznej filozofii nauki jest problem definiowalności terminów teoretycznych przez terminy obserwacyjne. Szczególnego znaczenia problem ten nabiera na gruncie założeń neopozytywistycznych. Nic też dziwnego, że od samego powstania tego kierunku znajdował się w centrum zainteresowań jego przedstawicieli. Ale choć poświęcono mu sporo uwagi i wysiłku, wysunięto wiele sugestii i propozycji, trudno uznać go za problem definitywnie rozwiązany. W znanym artykule „The Theoretician's Dilemma" z r. 1958, zawierającym rodzaj rekapitulacji dotychczasowych badań nad tym zagadnieniem, Hempel podsumowuje przegląd omawianych wyników sceptycznym wnioskiem stwierdzającym brak konkluzywnych argumentów przemawiających za lub przeciw definiowalności terminów teoretycznych nauk empirycznych za pomocą terminów czysto obserwacyjnych. Co więcej, wyraża wątpliwość, czy jakikolwiek argument może tę kwestię przesądzić w sposób ostateczny[1]. W ciągu piętnastu lat, jakie upłynęły od tego stwierdzenia, sytuacja nie zmieniła się w sposób zasadniczy. W opublikowanym w r. 1973 artykule „On the Different Ingredients of an Empirical Theory" Hintikka powtarza niemal dosłownie konkluzję Hempla. „Filozofowie nauki dyskutowali często nad tym, czy terminy teoretyczne mogą być zdefiniowane przez pojęcia obserwacyjne. Wydaje się, że istnieje zgoda co do tego, że nie mogą, choć trudno znaleźć przekonujące racje, dlaczego by tak miało być"[2]. Diagnoza ta wydaje się trafna. Wyniki zawarte w licznych pracach poświęconych problemo-

1) *Minnesota Studies in the Philosophy of Science*, vol. II, 1958, s. 70.
2) *Logic, Methodology and Philosophy of Science IV, 1973, s. 319*

wi obserwacyjnej definiowalności terminów teoretycznych mają z reguły charakter rozwiązań cząstkowych. W sposób konkluzywny okazują jedynie, że pewne definicje obserwacyjne (ewentualnie — definicje obserwacyjne pewnych typów) nie nadają się na definicje określonych terminów teoretycznych (ewentualnie — terminów teoretycznych określonych typów)[3]. Brak jednak argumentów okazujących w sposób przekonujący, że terminów tych nie można zdefiniować za pomocą żadnych definicji obserwacyjnych — a tym bardziej, że nie można w ten sposób zdefiniować żadnych terminów teoretycznych. Jeśli się wysuwa tego rodzaju przypuszczenia, to popiera je się wywodami mającymi świadczyć nie tyle o tym, że takie procedury definicyjne są niemożliwe, ile raczej o tym, że są to procedury nie wskazane z pewnych względów pragmatycznych[4]. Z drugiej strony, przytacza się niekiedy przykłady terminów teoretycznych zdefiniowanych *explicite* przez terminy obserwacyjne[5]. Jeśli są to przykłady trafne, ogólna teza o obserwacyjnej niedefiniowalności terminów teoretycznych utrzymać się nie daje.

Problem jest więc nadal otwarty. Czy istnieje jakakolwiek możliwość rozstrzygnięcia go w sposób ogólny? Czy można w szczególności podać konkluzywne argumenty przemawiające przeciw obserwacyjnej definiowalności terminów teoretycznych — jeśli nie wszystkich, to przynajmniej niektórych wyróżnionych ich rodzajów? Jedno jest niewątpliwe. Niezbędnym warunkiem sformułowania takich argumentów jest uprzednia precyzacja samego problemu. To, czy terminy teoretyczne są, czy też nie są definiowalne przez terminy obserwacyjne, zależy w sposób oczywisty od tego, co przez pojęcia terminu obserwacyjnego i teoretycznego rozumiemy. Tylko określona, dostatecznie wyraźna interpretacja tych wieloznacznych i niejasnych pojęć może pozwolić na uchwycenie ich wzajemnych związków. Taka też interpretacja leży u podstawy przedstawionej niżej argumentacji. Wskazuję w niej pewne konsekwencje płynące na gruncie owej interpretacji z założenia definiowalności terminów teoretycznych przez terminy obserwacyjne — konsekwencje przemawiające, w moim przekonaniu, przeciwko takiemu założeniu. Dostarczają one tym samym częściowego przynajmniej uzasadnienia dla owego dominującego w literaturze przedmiotu poglądu o obserwacyjnej niedefiniowalności terminów teoretycznych. Argumenty poniższe odwołują się do pewnych środków formalnych i zakładają, co za tym idzie, określoną formalizację rozważanego problemu. Jest to formalizacja dokonana za pomocą pojęć teroriomodelowych i umożliwiająca dzięki temu wykorzystanie pewnych ogólnych teoriomodelowych zależności. Rozważania te stanowią tym samym przykład zastosowania pewnych metod formalnych do rozstrzygania problemów filozoficznych.

3) Taki charakter ma np. słynny wywód Carnapa w „Testability and Meaning", *Philosophy of Science*, 1936-1937.

4) Taki rodzaj argumentacji zawiera m. in. *Scientific Explanation* Braithwaite'a, 1953.

5) Przykład taki podaje np. R. Wójcicki w publikacji zbiorowej *Teoria i doświadczenie*, 1966, s. 67

Eksplikacja pojęć terminu obserwacyjnego i teoretycznego, jaką tu zakładam, przedstawiona została szerzej w innych moich publikacjach, w szczególności w monografii *The Logic of Empirical Theories*. Tutaj ograniczę się do podkreślenia tych tylko jej rysów, które mają charakter decydujący dla dyskutowanej przez nas sprawy. Można je ująć najkrócej, jak następuje:

*Terminy obserwacyjne* to terminy wyposażone w określoną interpretację wyłącznie w dziedzinie przedmiotów spostrzegalnych, a *terminy teoretyczne* to terminy odnoszące się (między innymi lub wyłącznie) do przedmiotów niespostrzegalnych.

Charakterystyka ta wskazuje zarazem źródło niedefiniowalności terminów teoretycznych przez terminy obserwacyjne. Jakie intuicje kryją się za powyższymi założeniami? Stwierdzić trzeba przede wszystkim wyraźnie, że przyjęte tu rozumienie terminów obserwacyjnych reprezentuje tylko jedną z możliwych eksplikacji tego pojęcia: eksplikację rygorystyczną i wąską, ale za to dostatecznie wyraźną i wyróżniającą, jak sądzę, ważną metodologicznie kategorię wyrażeń. Terminy obserwacyjne tak rozumiane utożsamić można z terminami zinterpretowanymi w sposób bezpośredni. Najprostszym rodzajem takiej interpretacji jest tzw. *definicja ostensywna*. A terminy zdefiniowane ostensywnie to terminy, których jedynym kryterium stosowalności jest wygląd danego przedmiotu (szeroko rozumiany, a więc nie tylko wizualny). Kryterium takie z natury rzeczy ograniczone jest do przedmiotów bezpośrednio spostrzegalnych (a więc pewnych makroobiektów). W stosunku do przedmiotów bezpośrednio niespostrzegalnych (w szczególności wszelkich mikroobiektów) terminy zdefiniowane ostensywnie pozbawione są jakichkolwiek kryteriów stosowalności. W rezultacie tylko w dziedzinie przedmiotów spostrzegalnych interpretację terminów obserwacyjnych można uważać za ustaloną. We wszelkiej dziedzinie przedmiotów niespostrzegalnych terminy te pozostają nie określone; mogą zatem być interpretowane dowolnie.

Formalnym odpowiednikiem tych założeń może być następująca konstrukcja teoriomodelowa. Językiem obserwacyjnym $J_o$ niech będzie język rachunku predykatów z identycznością, zawierający prócz stałych logicznych (spójników zdaniowych, kwantyfikatorów i predykatu identyczności) oraz zmiennych indywiduowych, predykaty obserwacyjne: $o_1, ..., o_n$. Niech $U_o$ będzie zbiorem przedmiotów spostrzegalnych. W zbiorze tym bezpośrednia interpretacja predykatów $o_1, ..., o_n$ przyporządkowuje im jako denotacje określone relacje $O_1, ..., O_n$, wyznaczając tym samym model właściwy języka $J_o$, $\mathfrak{m}_o = \langle U_o, O_1, ..., O_n \rangle$. Jest to założenie upraszczające stan faktyczny, gdyż w rzeczywistości nawet w zbiorze przedmiotów spostrzegalnych interpretacja predykatów obserwacyjnych nie bywa jednoznaczna. Wyznacza ona w rezultacie nie jeden model właściwy języka obserwacyjnego, lecz pewną ich rodzinę. Przyjmujemy tu jednak owo założenie upraszczające, ponieważ przy założeniach bardziej realistycznych przedstawiona niżej argumentacja zachowuje walor *a fortiori*.

W sytuacji, gdy język obserwacyjny występuje jako fragment języka teoretycznego, jego interpretacja nie może być ograniczona do zbioru przedmiotów spostrzegalnych. Typowe teorie empiryczne (np. teorie fizyczne) mówią z reguły o pewnych przedmiotach niespostrzegalnych (atomach, elektronach itp.) i takie też przedmioty wchodzić muszą w skład universum języka obserwacyjnego stanowiącego część języka teorii. Ale w zbiorze przedmiotów niespostrzegalnych interpretacja predykatów obserwacyjnych pozostaje, jak widzieliśmy, nie określona. Fakt ten znajduje wyraz w założeniu głoszącym, iż modelem właściwym języka $J_o$ (traktowanego jako część języka teoretycznego) jest dowolne rozszerzenie modelu $\mathfrak{m}_o$, a więc wszelki model $\mathfrak{m}'_o = \langle U'_o, O'_1, \ldots, O'_n \rangle$, w którym uniwersum $U'_o$ zawiera uniwersum $U_o$: $U'_o \supset U_o$, a relacje $O'_1, \ldots, O'_n$ ograniczone do universum $U_o$ pokrywają się z relacjami $O_1, \ldots, O_n$: $O'_i | U_o = O_i$ dla $i = 1, \ldots, n$. Rodzinę takich modeli oznaczymy przez $M_o$.

Terminy teoretyczne ograniczyliśmy wyżej do terminów desygnujących (między innymi przynajmniej) pewne przedmioty niespostrzegalne. Jest to ograniczenie istotne dla dalszych rozważań. Wyłącza ono z kategorii terminów teoretycznych pewne terminy tradycyjnie do niej zaliczane, takie chociażby, jak termin *inteligentny*. Typowe jednak terminy teoretyczne występujące w teoriach empirycznych (m. in. wszystkie terminy fizyczne) warunek powyższy spełniają. Możemy go traktować jako warunek wyróżniający kategorię terminów teoretycznych *sensu stricto*. W przeciwieństwie do terminów obserwacyjnych, terminy teoretyczne najszerzej nawet rozumiane mogą być interpretowane jedynie pośrednio. Takim pośrednim sposobem interpretacji danego terminu teoretycznego jest charakterystyka jego denotacji za pomocą zinterpretowanych uprzednio terminów obserwacyjnych. Załóżmy, iż opisany poprzednio język obserwacyjny $J_o$ wzbogacamy o pewien predykat teoretyczny $t$, przechodząc w ten sposób do języka teoretycznego $J$ o predykatach $o_1, \ldots, o_n, t$. Interpretacja pośrednia predykatu $t$ polega na przyjęciu pewnego zdania $\delta$ języka $J$ (zawierającego predykat $t$ i niektóre przynajmniej z predykatów $o_1, \ldots, o_n$) w roli postulatu charakteryzującego denotację predykatu $t$ jako relację spełniającą warunki sformułowane w zdaniu $\delta$ przy założeniu, iż predykaty $o_1, \ldots, o_n$ interpretowane są w sposób wyznaczony przez rodzinę modeli $M_o$. Innymi słowy, modelem właściwym tak zinterpretowanego języka $J$ jest każde przedłużenie któregoś z modeli należących do rodziny $M_o$, będące modelem zdania $\delta$, a więc, każdy model $\mathfrak{m} = \langle U, O_1, \ldots, O_n, T \rangle$, którego fragment odpowiadający językowi $J_o$, $\mathfrak{m}|_o = \langle U, O_1, \ldots, O_n \rangle$, pokrywa się z pewnym modelem rodziny $M_o$, i w którym prawdziwe jest zdanie $\delta$. Zdanie $\delta$ nosi nazwę *postulatu znaczeniowego* dla predykatu $t$. W skrajnym przypadku postulat taki może mieć postać definicji równoważnościowej predykatu $t$ za pomocą predykatów $o_1, \ldots, o_n$. Ten właśnie przypadek będzie przedmiotem dalszych rozważań.

Załóżmy dla uproszczenia, że predykat $t$ jest predykatem jednoargumentowym (uogólnienie na predykaty $k$-argumentowe nie przedstawia większych trudności). Definicja $\delta$ przybiera w tym wypadku postać:

$$t(x) \equiv \alpha(x), \quad (1)$$

gdzie $\alpha(x)$ jest formułą zdaniową języka obserwacyjnego $J_o$ o jednej zmiennej wolnej. Rozważany przez nas problem definiowalności predykatu teoretycznego $t$ w języku obserwacyjnym $J_o$ zakłada pewną zamierzoną interpretację predykatu $t$ i sprowadza się do pytania, czy istnieje taka obserwacyjna formuła $\alpha(x)$, której interpretacja właściwa pokrywa się z ową zamierzoną interpretacją predykatu $t$. Ta ostatnia nie musi być przy tym interpretacją jednoznaczną. Może przyporządkowywać predykatowi $t$ jako jego denotację nie jeden zbiór przedmiotów, lecz rodzinę takich zbiorów. Idzie wówczas o to, czy ta sama rodzina zbiorów przyporządkowana być może pewnej obserwacyjnej formule $\alpha(x)$ zgodnie z jej właściwą interpretacją.

To, jaka jest zamierzona interpretacja danego predykatu teoretycznego, zależy oczywiście od tego, co to jest za predykat. Wydaje się jednak, że istnieją pewne warunki niezbędne, które spełniać musi interpretacja każdego w zasadzie predykatu teoretycznego. Interpretacja nie spełniająca któregoś z tych warunków nie może być uznana za interpretację adekwatną. Otóż argumentacja nasza zmierza do okazania, że żadna definicja obserwacyjna typu (1) nie może wyposażyć predykatu teoretycznego $t$ w interpretację adekwatną. Dotyczy to wszystkich predykatów teoretycznych *sensu stricto*, a więc predykatów, które z zamierzenia odnosić się mają do pewnych przedmiotów niespostrzegalnych, tj. nie należących do universum $U_o$. Okazuje się mianowicie, iż interpretacja takiego predykatu określona przez definicję (1) nie jest w stanie wyznaczyć żadnego ustalonego podziału owych przedmiotów niespostrzegalnych na przedmioty desygnowane przez predykat $t$ i pozostałe. Dopuszcza zawsze traktowanie pod tym względem wszystkich przedmiotów niespostrzegalnych *en bloc*.

Fakt ten jest konsekwencją pewnych ogólnych zależności teoriomodelowych. Bezpośrednie zastosowanie znajduje tu twierdzenie następujące (szkic jego dowodu podaję w załączonym Dodatku):

(T1) *Niech $\alpha(x)$ będzie formułą języka $J_o$, a $\mathfrak{m}_o$ modelem tego języka o universum $U_o$. Jeżeli pewien element universum $U_o$ spełnia formułę $\alpha(x)$ w każdym rozszerzeniu modelu $\mathfrak{m}_o$, to dla dowolnego zbioru $U$ zawierającego zbiór $U_o$ istnieje rozszerzenie $\mathfrak{m}$ modelu $\mathfrak{m}_o$ o universum $U$ takie, iż każdy element universum $U$ nie należący do $U_o$ spełnia formułę $\alpha(x)$ w modelu $\mathfrak{m}$.*

Innym nieco sformułowaniem twierdzenia (T1) jest jego wersja następująca:

(T2) *Jeżeli pewien element universum $U_o$ nie spełnia formuły $\alpha(x)$ w żadnym rozszerzeniu modelu $\mathfrak{m}_o$, to dla dowolnego zbioru $U$ zawierającego zbiór $U_o$ istnieje rozszerzenie $\mathfrak{m}$ modelu $\mathfrak{m}_o$ o universum $U$ takie, iż żaden element universum $U$ nie należący do $U_o$ nie spełnia formuły $\alpha(x)$ w modelu $\mathfrak{m}$.*

Przyjrzyjmy się konsekwencjom tych twierdzeń dla rozważanej przez nas sytuacji. Co mówią nam one o interpretacji predykatu teoretycznego $t$ wyznaczonej przez definicję (1)? Przypuśćmy, że jest to predykat, który prócz pewnych przedmiotów niespostrzegalnych desygnować ma również pewne przedmioty spostrzegalne. Definiująca go formuła $\alpha(x)$ spełniać musi wówczas poprzednik twierdzenia (T1). Tym samym jedną z jej dopuszczalnych interpretacji okazuje się, na mocy tego twierdzenia, interpretacja taka, przy której formułę tę spełnia każdy przedmiot niespostrzegalny — niezależnie od tego, jaki zbiór przedmiotów niespostrzegalnych weźmiemy pod uwagę.

W konsekwencji, jedną z dopuszczalnych interpretacji predykatu $t$ okazuje się zawsze interpretacja zaliczająca do jego desygnatów wszelkie przedmioty niespostrzegalne. A to wydaje się z reguły niezgodne z jego zamierzonym sensem. Z drugiej strony, jeśli istnieją przedmioty spostrzegalne, których predykat $t$ nie desygnuje, jedną z jego dopuszczalnych interpretacji będzie, na mocy twierdzenia (T2), interpretacja wykluczająca spośród jego desygnatów wszelkie przedmioty niespostrzegalne — niezależnie od tego, jakie przedmioty tego typu obejmuje universum danego języka. To również wydaje się świadczyć o nieadekwatności takiej interpretacji. Tak więc, w przypadku predykatów teoretycznych, desygnujących prócz przedmiotów niespostrzegalnych niektóre (i tylko niektóre) przedmioty spostrzegalne, mamy do czynienia z sytuacją następującą. Predykat taki wyposażony zostaje przez definicję obserwacyjną w interpretację niezmiernie wieloznaczną. Wśród jego denotacji znajdujemy, z jednej strony, zbiór nie obejmujący żadnych przedmiotów niespostrzegalnych, z drugiej strony, zbiór obejmujący wszystkie takie przedmioty. Sądzę, że żadnej z tych denotacji nie można uznać za należącą do denotacji zamierzonych.

W przypadku predykatów teoretycznych desygnujących wyłącznie przedmioty niespostrzegalne, ich definicje obserwacyjne pociągają konsekwencje bardziej jeszcze paradoksalne. Bezpośrednie zastosowanie znajduje tu zależność będąca bezpośrednim wnioskiem z twierdzenia (T2):

(T3) *Jeżeli żaden element universum $U_0$ nie spełnia formuły $\alpha(x)$ w żadnym rozszerzeniu modelu $\mathfrak{m}_0$, to dla dowolnego zbioru $U$ zawierającego zbiór $U_0$ istnieje rozszerzenie $\mathfrak{m}$ modelu $\mathfrak{m}_0$ o universum $U$ takie, iż żaden element universum $U$ nie spełnia formuły $\alpha(x)$ w modelu $\mathfrak{m}$.*

Poprzednik tego twierdzenia zachodzić musi dla każdej formuły $\alpha(x)$ definiującej predykat teoretyczny omawianego typu. W konsekwencji, formuła ta dopuszcza zawsze interpretację taką, przy której nie spełnia jej żaden w ogóle element universum — niezależnie od tego, jakie universum dany język zakłada. Tak więc, jedną z denotacji predykatu tak zdefiniowanego okazuje się zawsze zbiór pusty. Interpretacja, zgodnie z którą dany predykat teoretyczny nie desygnuje niczego, nie jest na pewno jego interpretacją zamierzoną. A takiej interpretacji żadna definicja obserwacyjna wykluczyć nie jest w stanie.

Fakty te przemawiają, w moim przekonaniu, za niedefiniowalnością terminów teoretycznych przez terminy obserwacyjne, rozstrzygając w ten sposób ów tradycyjny problem. Trzeba jednak podkreślić raz jeszcze, że wynik ten jest konsekwencją ściśle określonych założeń. Opiera się w szczególności na swoistym, dość rygorystycznym rozumieniu obu podstawowych pojęć: terminu obserwacyjnego i teoretycznego. To dlatego tylko, że interpretacja terminów obserwacyjnych jest w stosunku do przedmiotów niespostrzegalnych całkowicie nie określona, a terminy teoretyczne takie właśnie przedmioty mają desygnować — nie sposób za pomocą tych pierwszych scharakteryzować wystarczająco ściśle interpretacji tych drugich. Przy innym, bardziej liberalnym rozumieniu obu pojęć można dojść do konkluzji odmiennej. To tłumaczy fakt powoływania się na przykłady definiowalności niektórych terminów teoretycznych przez terminy obserwacyjne. Przyjrzyjmy się jednemu z takich przykładów, wspomnianemu na wstępie naszych rozważań. Terminy obserwacyjne, jakie w tym przypadku wchodzą w grę — to dwuargumentowy predykat „jest dwa razy krótszy od" i nazwa (a raczej deskrypcja) „wzorzec metra z Sèvres". Łatwo można okazać, że za pomocą owego predykatu obserwacyjnego zdefiniować się daje (przez odpowiedni ciąg definicji równoważnościowych) dwuargumentowy predykat „jest dziesięć miliardów razy krótszy od", a za pomocą tego ostatniego i wspomnianej nazwy obserwacyjnej jednoargumentowy predykat „jest dziesięć miliardów razy krótszy od wzorca metra z Sèvres" (inaczej — „ma długość 1 Å"), a więc predykat teoretyczny w ścisłym słowa tego znaczeniu, desygnujący wyłącznie przedmioty niespostrzegalne (mikroobiekty o rozmiarach atomu). Czy definicja nasza nadaje w istocie predykatowi temu interpretację zamierzoną? Odpowiedź zależy najwyraźniej od tego, jak pojmujemy ów wyjściowy predykat „jest dwa razy krótszy od". Jeśli traktujemy go jako predykat obserwacyjny w przyjętym przez nas sensie, a więc zdefiniowany ostensywnie i pozbawiony wobec tego określonej interpretacji w dziedzinie przedmiotów niespostrzegalnych, nie możemy uznać interpretacji zdefiniowanego za jego pomocą predykatu teoretycznego za interpretację adekwatną, choćby dlatego że dopuszcza ona wówczas jako jedną z jego denotacji zbiór pusty. Może to być interpretacja adekwatna tylko przy założeniu, że predykat „jest dwa razy krótszy od" ma interpretację ściśle określoną również w zbiorze przedmiotów niespostrzegalnych. Ale wówczas nie uznamy go, zgodnie z naszą terminologią, za predykat obserwacyjny. Będzie to po prostu jeden z predykatów teoretycznych — i to predykatów teoretycznych *sensu stricto*. Przykład powyższy nie stoi zatem w sprzeczności z tezą o obserwacyjnej niedefiniowalności terminów teoretycznych, rozumianą w przyjęty w tej pracy sposób.

Teza to, jak widzieliśmy, nienowa, ale uzasadniana zwykle w sposób diametralnie różny od przedstawionego powyżej. Tradycyjne argumenty przeciwko obserwacyjnej definiowalności terminów teoretycznych zwracają uwagę na fakt, że definiowanie takie jest procedurą zbyt rygorystyczną. Ma ono wyznaczać interpretację terminu teoretycznego w sposób zbyt sztywny, wiążąc ją zbyt ściśle ze specyficznymi kryteriami jego stosowalności. Argumentacja nasza podkreśla fakt wręcz przeciwny. Definiowanie terminów teoretycznych przez terminy obserwacyjne jest procedurą w pewnym sensie zbyt liberalną. Wyznacza interpretację danego terminu w sposób zbyt luźny, dopuszczając, z konieczności, interpretacje jawnie niezamierzone. Sądzę, że mimo tych przeciwieństw oba rodzaje argumentacji zawierają spostrzeżenia słuszne, tylko dotyczące sytuacji różnych. Nasze wywody odnosiły się do terminów teoretycznych w ścisłym tego słowa znaczeniu, a więc terminów desygnujących pewne przedmioty niespostrzegalne. Argumentacja tradycyjna natomiast zachowuje walor w stosunku do terminów teoretycznych desygnujących wyłącznie przedmioty spostrzegalne. Skoro przy wprowadzaniu tego rodzju terminów do języka obserwacyjnego nie rozszerzamy jego universum poza dziedzinę przedmiotów spostrzegalnych, definicja obserwacyjna takiego terminu determinuje jego interpretację w sposób jednoznaczny, a to w wielu przypadkach wydaje się istotnie ograniczeniem niepożądanym[6]. W ten sposób pogląd głoszący obserwacyjną niedefiniowalność terminów teoretycznych znajduje potwierdzenie w różnych stron i za pomocą różnych argumentów — w zależności od takiej czy innej eksplikacji jego zakresu i treści.

## DODATEK

Szkic dowodu przytaczanego w artykule twierdzenia (T1) ograniczymy do przypadku możliwie najprostszego, ale zarazem dostatecznie ogólnego. Niech $J_o$ będzie językiem rachunku predykatów z identycznością, zawierającym jeden dwuargumentowy predykat pozalogiczny $o$. *Modelami* języka $J_0$ nazywać będziemy wyłącznie modele z absolutnym pojęciem identyczności. Jeśli nie ograniczamy się do modeli tego typu, dowód twierdzenia (T1) jest natychmiastowy. Przy ograniczeniu się do modeli z absolutnym pojęciem identyczności dowodzimy twierdzenia (T1), opierając się na następującym lemacie:

(L) *Niech $\alpha(x)$ będzie formułą języka $J_o$ o jednej zmiennej wolnej, a $\mathfrak{m}_0 = \langle U_o, O_o\rangle$ modelem tego języka. Jeżeli pewien element universum $U_o$ spełnia formułę $\alpha(x)$ w każdym rozszerzeniu modelu $\mathfrak{m}_o$, to istnieje formuła czysto egzystencjalna $\beta(x)$ spełniona przez pewien element universum $U_o$ w modelu $\mathfrak{m}_0$, taka iż*

$$\Lambda x\,(\beta(x) \rightarrow \alpha(x)) \in Cn(O).$$

Lemat ten stanowi pewną odmianę znanych zależności teoriomodelowych i jego dowód przebiega w analogiczny sposób. Niech $J_o'$ będzie językiem $J_o$ wzbogaconym o nazwy wszystkich elementów universum $U_o$, $\{s_u\}_{u \in U_o}$, $Cn'$ konsekwencją w $J_o'$, a $\mathfrak{m}_o' = \langle U_o, O_o, \{a_u\}_{u \in U_o}\rangle$, gdzie $a_u = u$, modelem $J_o'$, będącym przedłużeniem modelu $\mathfrak{m}_o$. Załóżmy, że spełniony jest poprzednik lematu (L). Wówczas, dla pewnej nazwy $s_u$, zdanie $\alpha(s_u)$ jest prawdziwe w każdym rozszerzeniu modelu $\mathfrak{m}_o'$. Okażemy, że zdanie to musi być konsekwencją diagramu modelu $\mathfrak{m}_o$, $D(\mathfrak{m}_o)$. Przypuśćmy, że tak nie jest: $\alpha(s_u) \notin Cn'(D(\mathfrak{m}_o))$. Istnieje wobec tego model $\mathfrak{m}'$ języka $J_o'$, który jest modelem diagramu $D(\mathfrak{m}_o)$, a nie jest modelem zdania $\alpha(s_u)$. Na mocy znanych twierdzeń teorii modeli[7], model $\mathfrak{m}'$ jest rozszerzeniem *sensu largo* modelu $\mathfrak{m}_o'$, czyli jest rozszerzeniem zwykłym (*sensu stricto*) pewnego modelu izomorficznego z $\mathfrak{m}_o'$. Zdanie $\alpha(s_u)$ jest fałszywe w modelu $\mathfrak{m}'$. Ponieważ każde rozszerzenie *sensu largo* modelu $\mathfrak{m}_o'$ jest izomorficzne z pewnym rozszerzeniem zwykłym tegoż modelu, zdanie $\alpha(s_u)$ okazuje się fałszywe w pewnym rozszerzeniu zwykłym modelu $\mathfrak{m}_o'$, a to jest sprzeczne z przyjętym założeniem. Wnosimy stąd, iż $\alpha(s_u) \in Cn'(D_o(\mathfrak{m}_o))$. Istnieje wobec tego skończony podzbiór $D_o(\mathfrak{m}_o)$ diagramu $D(\mathfrak{m}_o)$ taki, iż $\alpha(s_u) \in Cn'(D(\mathfrak{m}_o))$. Niech $\gamma(s_u, s_{u_1}, ..., s_{u_n})$ będzie koniunkcją

6) Ujemne konsekwencje takiego ograniczenia starałem się przedstawić m. in. w artykule „Pojęcia teoretyczne a doświadczenie", *Studia Logica*, 1961.

7) Por. np. J. Łoś, „On the Extending of Models I", *Fundamenta Mathematicae*, 1955.

zdań zbioru $D_o(\mathfrak{m}_o)$, a $s_u$, $s_{u_1}$, ..., $s_{u_n}$ wszystkimi nazwami występującymi w tych zdaniach. Skoro $\alpha(s_u)$ jest konsekwencją $\gamma(s_u, s_{u_1}, ..., s_{u_n})$, a nazwy $s_{u_1}, ..., s_{u_n}$ nie występują w $\alpha(s_u)$, $\alpha(s_u) \in Cn'(\bigvee x_1 \ldots \bigvee x_n\, \gamma(s_u, x_1, ..., x_n))$. To z kolei pociąga zależność:

$$\bigwedge x\, (\bigvee x_1 \ldots \bigvee x_n\, \gamma(x, x_1, ..., x_n) \rightarrow \alpha(x)) \in Cn(O).$$

Ponieważ $\gamma(s_u, s_{u_1}, ..., s_{u_n})$ jest koniunkcją zdań z diagramu modelu $\mathfrak{m}_o$, formuła $\beta(x) = \bigvee x_1 \ldots \bigvee x_n\, \gamma(x, x_1, ..., x_n)$ jest formułą czysto egzystencjalną, spełnioną przez pewien element universum $U_o$ w modelu $\mathfrak{m}_o$. Tym samym spełnia warunki wymienione w następniku lematu (L).

Lemat ten posłuży nam do dowodu twierdzenia T1:

(T1) *Jeżeli pewien element universum $U_o$ spełnia formułę $\alpha(x)$ w każdym rozszerzeniu modelu $\mathfrak{m}_o$, to dla dowolnego zbioru $U$ zawierającego zbiór $U_o$ istnieje rozszerzenie $\mathfrak{m}$ modelu $\mathfrak{m}_o$ o universum $U$ takie, iż każdy element universum $U$ nie należący do $U_o$ spełnia formułę $\alpha(x)$ w modelu $\mathfrak{m}$.*

Zakładamy prawdziwość poprzednika twierdzenia (T1) i korzystamy z lematu (L). Niech $\beta(x)$ będzie formułą, której istnienie gwarantuje, przy powyższym założeniu, lemat (L). Aby okazać prawdziwość następnika twierdzenia (T1), wystarczy okazać, że warunki sformułowane w tym następniku spełnia formuła $\beta(x)$. Formuła ta ma, jak wiemy, postać $\bigvee x_1 \ldots \bigvee x_n\, \gamma(x, x_1, ..., x_n)$, gdzie $\gamma(x, x_1, ..., x_n)$ jest koniunkcją pewnych formuł atomowych (m. in. identycznościowych) oraz negacji takich formuł. Jest to formuła spełniona przez pewien element universum $U_o$ — niech będzie nim $u$ — w modelu $\mathfrak{m}_o$, a w konsekwencji (jako formuła czysto egzystencjalna) i w każdym rozszerzeniu modelu $\mathfrak{m}_o$. Niech $U$ będzie nadzbiorem $U_o$. Musimy zdefiniować takie rozszerzenie $\mathfrak{m}$ modelu $\mathfrak{m}_o$ o universum $U$, w którym każdy element zbioru $U—U_o$ spełnia formułę $\beta(x)$. Rozszerzenie to konstruujemy definiując relację $O$ w sposób następujący: dla dowolnych $u_i', u_j' \in U$, $O(u_i', u_j') \equiv O_o(u_i, u_j)$, gdzie $u_k = u_k'$, gdy $u_k' \in U_o$, oraz $u_k = u$, gdy $u_k' \in U—U_o$. Tak określona relacja $O$ pokrywa się z relacją $O_o$ w zbiorze $U_o$. Natomiast wszystkie elementy zbioru $U—U_o$ zachowują się ze względu na relację $O$ tak, jak element $u$ ze względu na relację $O_o$. Gdyby formuła $\beta(x)$ nie zawierała formuł identycznościowych, widoczne byłoby natychmiast, że w tak zdefiniowanym rozszerzeniu $\mathfrak{m}$ każdy element zbioru $U—U_o$ spełnia formułę $\beta(x)$ (skoro spełnia ją element $u$). Pozostaje nam wobec tego jedynie rozpatrzenie ewentualnych formuł identycznościowych występujących w formule $\beta(x) = \bigvee x_1 \ldots \bigvee x_n\, \gamma(x, x_1, ..., x_n)$. W grę wchodzą tylko formuły zawierające zmienną $x$. Mogą być one, ogólnie biorąc, czterech rodzajów: (1) $x = x$, (2) $x \neq x$, (3) $x = x_i$, (4) $x \neq x_i$. Formuła (1) jest spełniona przez każdy przedmiot, a formuła (2) przez żaden, a więc w $\beta(x)$ występować nie może. Formuła (3) występuje w sposób nieistotny, bo każde wyrażenie typu $\bigvee x_i\, (x = x_i \wedge \phi(x_i))$ jest równoważne logicznie wyrażeniu $\phi(x)$. Pozostaje formuła (4), a ta spełniona jest przez każdy element zbioru $U—U_o$, skoro w formule $\beta(x)$ spełnionej przez $u$ w modelu $\mathfrak{m}_o$ $x_i$ reprezentuje pewien element zbioru $U_o$. W rezultacie każdy element zbioru $U—U_o$

spełnia formułę $\beta(x)$, a w konsekwencji formułę $\alpha(x)$, w rozszerzeniu $\mathfrak{m}$, co należało okazać.

Przytaczane w artykule twierdzenia (T2) i (T3) są natychmiastowymi konsekwencjami twierdzenia (T1).