Filozofia Nauki
Rok XIX, 2011, Nr 3(75)

Zbigniew Tworak

Paradoks znawcy
(The Knower Paradox)

1. PARADOKS ZNAWCY I JEGO WERSJE

Paradoks znawcy, występujący w literaturze anglojęzycznej pod nazwą *The Knower Paradox*, pochodzi z pracy D. Kaplana i R. Montague analizującej *paradoks kata* (Kaplan, Montague 1960). Stanowi on jego szczególny przypadek — gdy liczba możliwych terminów egzekucji zostaje zredukowana do zera (wtedy postawiony przez sędziego warunek wykonania wyroku przyjmuje postać: skazaniec wie, że niniejszy warunek jest fałszywy).[1] Okazało się — w tym tkwi paradoks — że nie można wyrazić niesprzecznie pojęcia wiedzy, jeśli tylko mamy możliwość budowania zdań na temat wiedzy, które odnoszą się do samych siebie, oraz przyjmujemy następujące trzy zasady (ważne dla analizy wiedzy):

(i) Niezawodność wiedzy: Jeżeli podmiot wie, że α, to α zachodzi (krótko: wiedza implikuje prawdziwość).

(ii) Epistemiczna nieskromność: Podmiot wie, że (i).[2]

(iii) Epistemiczne domknięcie: Jeżeli β jest wyprowadzalne z α i podmiot wie, że α, to wie on też, że β.

---

[1] Warto wspomnieć, że paradoks ten przypomina 13. sofizmat J. Buridana, zamieszczony w VIII rozdziale jego *Sophismata*. Jedyne zdanie napisane na ścianie brzmi: „Sokrates wie, że zdanie napisane na ścianie budzi jego wątpliwości [sprzeciw]". Sokrates czyta owo zdanie i zastanawia się nad jego prawdziwością (Hugens 1982).

[2] Nazywając w ten sposób ową zasadę, nawiązuję do opowieści R. Smullyana o „myślakach" (Smullyan 1987).

Zasada pierwsza ma swe korzenie w starożytności, przede wszystkim w występującym u Platona odróżnieniu *epistéme* i *doxa*. Według Platona, prawdziwość α jest warunkiem koniecznym wiedzy, że α.[3] Wiedzę, w przeciwieństwie do wiary (mniemania), cechuje niezawodność. Powiedzenie przez kogoś „Wiem, że 0 = 1" natychmiast powoduje sprzeciw z powodu fałszywości zdania „0 = 1". Można powiedzieć, że przykłady tego typu testują zasadę (i). Z kolei, zasada druga bezpośrednio dotyczy tego, co podmiot wie, a mianowicie na jej mocy podmiot wie o niezawodności swej własnej wiedzy. W pewnym sensie wyraża ona zdolność podmiotu do refleksji na temat swej własnej wiedzy. Może on dojść do wspomnianej (meta)wiedzy na podstawie introspekcji bądź jakiegoś rodzaju empirycznego uogólnienia. Wreszcie zasada trzecia czyni z wiedzy zbiór zdań zamknięty ze względu na relację wyprowadzalności. Na jej podstawie podmiot automatycznie wie to, co daje się wyprowadzić za pomocą reguł wnioskowania dedukcyjnego (np. reguły *modus ponens*) z tego, co już wie. Tym samym wyraża ona dedukcyjną wszechwiedzę podmiotu (podmiot staje się przez to wyidealizowany).[4]

Paradoks znawcy, podobnie jak antynomia kłamcy, opiera się na możliwości budowania zdań odnoszących się do samych siebie. Generalnie, zdania tego rodzaju mają formę:

S: Zdanie *S* ma własność *P*.

Litera *S* reprezentuje nazwę jednostkową wyrażenia „Zdanie *S* ma własność *P*". Oczywiście, każde takie zdanie musi spełniać następujący warunek: *S* jest prawdziwe wtw „Zdanie *S* ma własność *P*" jest prawdziwe (wtw jest tak, że zdanie *S* ma własność *P*). Przypomnijmy, u podstaw antynomii kłamcy leży zdanie:

L: Zdanie L jest fałszywe.

Natomiast paradoks znawcy związany jest ze zdaniami takimi jak:

(i) K: Zdanie K jest niezgodne z wiedzą podmiotu [= *a* wie, że nie-K].

(ii) K: Zdanie K nie jest wiadome podmiotowi [= *a* nie wie, że K].

Możliwość budowania zdań odnoszących się do samych siebie daje tzw. lemat o diagonalizacji. Stąd właściwą ramą formalizacji rozważanego paradoksu jest arytmetyka liczb naturalnych (czy jakaś teoria ją zawierająca).[5] Przyjmijmy więc, że $\mathbf{\Sigma}$ jest aksjomatyzowalnym rozszerzeniem arytmetyki Robinsona **Q**, dla którego zachodzi lemat o diagonalizacji, a $J(\mathbf{\Sigma})$ językiem owej teorii. Zakładamy, że w języku $J(\mathbf{\Sigma})$ występują dwa predykaty specyficzne $I(x, y)$ oraz $K(x)$. Pierwszy w nich wyraża, że

[3] J. Hintikka takiego rozumienia wiedzy dopatruje się już u Parmenidesa.

[4] Jest to „mocna" wersja zasady epistemicznego domknięcia. Wersja „słabsza" stanowi, iż jeśli podmiot wie, że β daje się dedukcyjnie wyprowadzić z α, to wie on, że β, o ile wie, że α.

[5] Odwołanie się do lematu o diagonalizacji pozwala uniknąć dyskusji na temat dopuszczalności (naturalności) zdań odnoszących się do samych siebie.

zdanie o kodzie y jest wyprowadzalne w teorii $\mathbf{\Sigma}$ ze zdania o kodzie x. Drugi natomiast wyraża, że podmiot wie to, co głosi zdanie o kodzie x. Dla wygody czytelnika przypomnę obecnie wspomniany lemat. Niech $\ulcorner\beta\urcorner$ oznacza liczebnik będący nazwą kodu Gödla formuły β. Natomiast $\alpha[\beta]$ skraca zdanie $\alpha(\ulcorner\beta\urcorner)$.

Lemat o diagonalizacji. Dla dowolnej formuły $\alpha(x)$ języka $J(\mathbf{\Sigma})$ z jedną zmienną wolną x istnieje zdanie β języka $J(\mathbf{\Sigma})$ takie, że $\mathbf{\Sigma} \vdash \beta \equiv \alpha[\beta]$.

Skonstruowane zdanie β „mówi" o sobie samym (*via* numeracja Gödlowska), że spełnia warunek α: Mój kod ma własność α. Oto sformalizowana wersja rozważanego paradoksu:

Twierdzenie 1 (Kaplan, Montague 1960, *The Knower Paradox*). Niech $\mathbf{\Sigma}$ będzie teorią scharakteryzowaną jak wyżej, taką, że dla dowolnych zdań α i β jej języka spełnione są następujące trzy warunki:

K1.	$K[\alpha] \rightarrow \alpha$	(epistemiczna zasada refleksji),
K2.	$\mathrm{K}[\mathrm{K}[\alpha] \rightarrow \alpha]$	(epistemiczna nieskromność)
K3.	$\mathrm{K}[\alpha] \wedge \mathrm{I}[\alpha, \beta] \rightarrow \mathrm{K}[\beta]$	(epistemiczne domknięcie).

Wówczas teoria $\mathbf{\Sigma}$ jest sprzeczna.

Dowód. Na mocy lematu o diagonalizacji, istnieje zdanie δ takie, że

1.	$\mathbf{\Sigma} \vdash \delta \equiv K[\neg\delta]$ [6]	
2.	$\mathbf{\Sigma} \vdash \delta \rightarrow K[\neg\delta]$	1, opuszczanie równoważności
3.	$\mathbf{\Sigma} \vdash K[\neg\delta] \rightarrow \neg\delta$	aksjomat K1; owo uszczegółowienie K1 oznaczmy przez K1d
4.	$\mathbf{\Sigma} \vdash \delta \rightarrow \neg\delta$	2, 3, reguła sylogizmu hipotetycznego
5.	$\mathbf{\Sigma} \vdash \neg\delta$	4, reguła Claviusa
6.	$\mathbf{\Sigma} \vdash I[\mathrm{K1d}, \neg\delta]$	na podstawie kroków 3-5, gdyż $I[\alpha, \beta]$, ilekroć $\alpha \vdash \beta$ (tj. β jest wyprowadzalne w $\mathbf{\Sigma}$ z α)
7.	$\mathbf{\Sigma} \vdash K[\mathrm{K1d}]$	aksjomat K2
8.	$\mathbf{\Sigma} \vdash K[\mathrm{K1d}] \wedge I[\mathrm{K1d}, \neg\delta] \rightarrow K[\neg\delta]$	aksjomat K3
9.	$\mathbf{\Sigma} \vdash K[\neg\delta]$	6, 7, 8, MP (*modus ponens*)
10.	$\mathbf{\Sigma} \vdash \delta$	1, 9, odrywanie dla równoważności; sprzeczność z 5. ■

[6] Zdanie to — tzw. *zdanie znawcy* — jest arytmetycznym analogonem (i). Stanowi ono epistemiczną wersję zdania Jeroslowa stwierdzającego swą własną obalalność w arytmetyce **PA** (lub własną sprzeczność z **PA**).

Dygresja 1. W teorii $\mathbf{\Sigma}$ można skonstruować predykat $Pr_{\Sigma}(x)$, który reprezentuje dowodliwość w $\mathbf{\Sigma}$: $Pr_{\Sigma}(x) =_{df} \exists y Prov_{\Sigma}(y, x)$, gdzie $Prov_{\Sigma}(y, x)$ odczytujemy: y jest dowodem x na gruncie $\mathbf{\Sigma}$. Spełnia on warunek: jeśli $\mathbf{\Sigma} \vdash \alpha$, to $\mathbf{\Sigma} \vdash Pr_{\Sigma}[\alpha]$. To wraz z faktem, że dla teorii $\mathbf{\Sigma}$ zachodzi twierdzenie o dedukcji sprawia, iż formuła $I[\alpha, \beta]$ jest wyrażalna za pomocą formuły $Pr_{\Sigma}[\alpha \rightarrow \beta]$. Stąd w założeniu K3 można zastąpić $I[\alpha, \beta]$ przez $Pr_{\Sigma}[\alpha \rightarrow \beta]$. ■

Dygresja 2. Założenie o epistemicznym domknięciu, tj. K3, można wyrazić w alternatywnej (słabszej) formie:

K3* $K[\alpha \rightarrow \beta] \rightarrow (K[\alpha] \rightarrow K[\beta])$,

jeśli dla teorii $\mathbf{\Sigma}$ zachodzi twierdzenie o dedukcji oraz obowiązuje w niej następująca reguła:

RK. Jeżeli $\mathbf{\Sigma} \vdash \alpha$, to $\mathbf{\Sigma} \vdash K[\alpha]$.

Regułę RK możemy potraktować jako postulat wzbogacania wiedzy oraz jako warunek uzasadnienia. Generalnie, wyraża ona logiczną wszechwiedzę podmiotu (tj. wszechwiedzę w zakresie wszystkich tez systemu). ■

W artykule z roku 1963 — w związku z twierdzeniem Tarskiego pokazującym, że naiwna teoria prawdy jest sprzeczna — Montague udowodnił pewne twierdzenie, które odpowiednio zinterpretowane stanowi uproszczoną wersję rozważanego paradoksu.[7] Oto ona:

Twierdzenie 2 (Montague 1963). Niech $\mathbf{\Sigma}$ będzie aksjomatyzowalnym rozszerzeniem arytmetyki Robinsona **Q**, $K(x)$ zaś będzie arytmetyczną formułą języka $J(\mathbf{\Sigma})$ taką, że dla dowolnego zdania α tegoż języka spełnione są następujące dwa warunki:

K1. $K[\alpha] \rightarrow \alpha$,

RK. Jeżeli $\mathbf{\Sigma} \vdash \alpha$, to $\mathbf{\Sigma} \vdash K[\alpha]$,

Wówczas teoria $\mathbf{\Sigma}$ jest sprzeczna.

Dowód. Na mocy lematu o diagonalizacji, istnieje zdanie δ takie, że

1. $\mathbf{\Sigma} \vdash \delta \equiv \neg K[\delta]$ [8]
2. $\mathbf{\Sigma} \vdash \delta \rightarrow \neg K[\delta]$ 1, opuszczanie równoważności
3. $\mathbf{\Sigma} \vdash K[\delta] \rightarrow \delta$ aksjomat K1
4. $\mathbf{\Sigma} \vdash K[\delta] \rightarrow \neg K[\delta]$ 2, 3, reguła sylogizmu hipotetycznego

[7] Twierdzenie owo stanowić miało sceptyczny argument przeciwko predykatywnej koncepcji pojęć modalnych na korzyść koncepcji operatorowej.

[8] Zdanie to jest arytmetycznym analogonem (ii); stanowi ono epistemiczną wersję zdania Gödla stwierdzającego swą własną niedowodliwość w arytmetyce PA.

5.	$\mathbf{\Sigma} \vdash \neg K[\delta]$	4, reguła Claviusa
6.	$\mathbf{\Sigma} \vdash \delta$	1, 5, odrywanie dla równoważności
7.	$\mathbf{\Sigma} \vdash K[\delta]$	6, RK; sprzeczność z 5. ■

Dygresja 3. Jeszcze inna wersja ma formę paradoksu Curry'ego. Niech β będzie dowolnym zdaniem języka $J(\mathbf{\Sigma})$. Na mocy lematu o diagonalizacji, istnieje zdanie δ takie, że (*) $\mathbf{\Sigma} \vdash \delta \equiv (K[\delta] \to \beta)$. W oparciu o K1 otrzymujemy $\mathbf{\Sigma} \vdash K[\delta] \to (K[\delta] \to \beta)$, a stąd po zastosowaniu reguły absorpcji uzyskujemy (**) $\mathbf{\Sigma} \vdash K[\delta] \to \beta$. To wraz z (*) daje $\mathbf{\Sigma} \vdash \delta$. Na mocy RK uzyskujemy $\mathbf{\Sigma} \vdash K[\delta]$, a stąd i (**) dostajemy ostatecznie $\mathbf{\Sigma} \vdash \beta$. ■

Chociaż paradoks znawcy pierwotnie został sformułowany w języku arytmetyki liczb naturalnych, można go też sformułować w języku modalnego rachunku zdań. Predykatowi wiedzy odpowiada wówczas właściwy funktor modalny stosowany bezpośrednio do formuł zdaniowych (jest on typu „koniecznościowego"). Niech **L** będzie modalnym rachunkiem zdań zawierającym aksjomaty klasycznego rachunku zdań, aksjomat K (tj. $L(\alpha \to \beta) \to (L\alpha \to L\beta)$), domkniętym na regułę *modus ponens* MP i regułę ukonieczniania RG (tj. jeżeli $\vdash \alpha$, to $\vdash L\alpha$) oraz spełniającym twierdzenie o dedukcji. Logikę **L** rozszerzamy następnie do logiki $\mathbf{L}^+$. $\mathbf{L}^+$ zawiera dodatkowo schemat T (tj. $L\alpha \to \alpha$) oraz spełnia następujący lemat:

Lemat de Jongha-Sambina o punkcie stałym. Dla dowolnej formuły $\alpha(p, \vec{q}\,)$ języka logiki $\mathbf{L}^+$, w której zmienna zdaniowa p jest zmodalizowana (tj. występuje w zasięgu funktora konieczności L), istnieje formuła $\beta(\vec{q}\,)$ niezawierająca zmiennej p taka, że

$$\mathbf{L}^+ \vdash \beta(\vec{q}\,) \equiv \alpha(\beta(\vec{q}\,),\ \vec{q}\,).$$ [9]

Oto typowe przykłady modalnych formuł i ich punktów stałych w systemie Gödla–Löba (**GL**), zwanym logiką dowodliwości:

Modalna formuła $\alpha(p, q)$:	Jej punkt stały $\beta(q)$:
Lp	$\top$
$\neg Lp$	$\neg L\bot$
$L\neg p$	$L\bot$
$\neg L\neg p$	$\bot$
$Lp \to q$	$Lq \to q$

[9] Symbol $\vec{q}$ oznacza sekwencję innych zmiennych niż p (włączając sekwencję pustą). Dowód tego lematu znajduje się np. w Boolos 1993.

Łatwo zauważyć, że aksjomat T jest modalnym odpowiednikiem K1. Reguła ukoniecznienia RG odpowiada regule RK. Stosując do aksjomatu T regułę RG, otrzymujemy formułę $L(L\alpha \rightarrow \alpha)$ będącą modalnym odpowiednikiem założenia K2 (formułę tę oznaczamy literą U). Wreszcie aksjomat K jest modalnym odpowiednikiem założenia o epistemicznym domknięciu w postaci K3* (zob. Dygresja 2).

Twierdzenie 3 (Modalny wariant paradoksu znawcy). Niech $\mathbf{L}^+$ będzie modalną logiką zdaniową scharakteryzowaną jak wyżej. Wówczas $\mathbf{L}^+$ jest sprzeczna.

Dowód tego twierdzenia w zasadzie powtarza kroki dowodu Twierdzenia 2.

Interesującą wersję paradoksu znawcy przedstawił C. B. Cross (Cross 2001). Spośród zasad charakteryzujących wiedzę najwięcej wątpliwości budzi zasada epistemicznego domknięcia wyrażająca — jak powiedziano wyżej — dedukcyjną wszechwiedzę podmiotu.[10] Na jej mocy w skład wiedzy podmiotu wchodzą wszystkie jej konsekwencje, nawet te najdalsze (co wydaje się zbyt daleko idącą idealizacją). Cross pokazał, że możemy otrzymać sprzeczność również bez owej zasady, a nawet bez reguły RK wyrażającej logiczną wszechwiedzę podmiotu. W tym celu zdefiniował on nowy predykat wiedzy *K'* za pomocą predykatów *K* oraz *I*:

(DK') $K'(x) =_{\mathrm{df}} \exists y(K(y) \wedge I(y, x))$.

Zakres predykatu *K'* stanowi zbiór (kodów) zdań, które są wyprowadzalne z tego, co jest podmiotowi wiadome w sensie reprezentowanym przez predykat *K*. Jeśli przyjmiemy, że predykat *K* reprezentuje wiedzę aktualną (*explicite*) podmiotu, wtedy predykat *K'* będzie reprezentował jego wiedzę potencjalną (*implicite*) (zob. Dygresja 4 poniżej). Rzecz jasna, każde zdanie należące do zakresu predykatu *K* należy też do zakresu predykatu *K'*, gdyż tezą arytmetyki **Q** jest $I[\alpha, \alpha]$, dla dowolnego zdania α. Mówiąc swobodnie, przedstawiona przez Crossa wersja paradoksu znawcy (występująca pod nazwą *The Pardox of the Knowledge-Plus Knower*) stanowi, że jeżeli prawdą jest wszystko, co daje się wyprowadzić (w danej teorii bazowej) z tego, co podmiot aktualnie wie, to podmiot ów nie może znać każdego przypadku tego faktu, niezależnie od tego, czy jego wiedza aktualna spełnia warunek epistemicznego domknięcia, czy też nie.[11]

Dygresja 4. Różnicę między predykatami wiedzy *K* i *K'* można wyjaśnić za pomocą koncepcji przekonań/wiedzy nazwanej przez R Stalnakera *the sentence storage mo-*

[10] L. Carroll (1895) napisał znakomitą bajkę stanowiącą argument sceptyczny w tej kwestii (tj. epistemicznego domknięcia — zarówno mocnego, jak i słabego). Ma ona formę rozmowy między Żółwiem i Achillesem na temat zastosowania piątego postulatu Euklidesa do pewnego konkretnego przypadku. Achilles próbuje sprawić, by na podstawie przesłanek, które Żółw już uznał, uznał on też wynikający z nich wniosek. Jednak Żółw tego nie czyni, nawet po dowiedzeniu się, że wniosek ów wynika z uznanych przez niego wcześniej przesłanek. Ciągle dopytuje się, dlaczego powinien uznać ów wniosek.

[11] Dyskusja pewnych aspektów wersji Crossa znajduje się w Uzquiano 2004 i Cross 2004.

del of belief (Stalnaker 1991). W koncepcji tej zakłada się, że: (1) przekonania/wiedzę wyznacza pewien zbiór zdań przechowywanych przez podmiot w pamięci — „skrzyni przekonań" (ang. *belief box*) oraz (2) obok przekonań/wiedzy *explicite* (aktualnej) podmiot posiada przekonania/wiedzę *implicite* (potencjalną). To właśnie przekonania *explicite* przechowywane są w „skrzyni przekonań". Wszelkie inne przekonania, w szczególności konsekwencje przekonań *explicite*, tworzą zbiór przekonań *implicite*. Według Stalnakera, warunek dedukcyjnego domknięcia zbioru przekonań *explicite* (tj. zmagazynowanych w „skrzyni przekonań") jest niepotrzebny, a nawet nierealny. Następnie zwraca on uwagę na pewną dwuznaczność pojęcia przekonania/wiedzy *implicite*. Związana jest ona z zagadnieniem ograniczeń umysłowych podmiotu: dostępności przez podmiot do swych własnych przekonań (także tych zmagazynowanych w „skrzyni przekonań"), znajomością relewantnych w danym kontekście reguł inferencyjnych itp. Przy szerszym rozumieniu, przekonania *implicite* obejmują wszelkie przekonania, do posiadania których zobowiązują (wszystkie) przekonania zmagazynowane przez podmiot w „skrzyni przekonań". Są one wówczas dedukcyjnie domknięte. Nawiązując do przedstawionej koncepcji, można przyjąć, że K to predykat odnoszący się do wiedzy *explicite*, K' zaś odnosi się do wiedzy *implicite* (rozumianej szeroko). ■

Prezentację wersji Crossa paradoksu znawcy poprzedzę twierdzeniem pokazującym, że jeśli w założeniach K1 i K2 predykat K zastąpimy przez K', to uzyskana w ten sposób teoria będzie sprzeczna. Jest to zgodne z intuicją, gdyż K' reprezentuje wiedzę *implicite*, która — jak zostało wyżej zasygnalizowane — jest dedukcyjnie domknięta.

Twierdzenie 4 (Cross 2001). Niech (jak poprzednio) $\mathbf{\Sigma}$ będzie aksjomatyzowalnym rozszerzeniem arytmetyki **Q** takim, że predykat K' — zdefiniowany przez (DK') — spełnia następujące dwa warunki: dla dowolnych zdań α i β języka $J(\mathbf{\Sigma})$,

K1'. $K'[\alpha] \rightarrow \alpha$

K2'. $K'[K'[\alpha] \rightarrow \alpha]$.

Wówczas teoria $\mathbf{\Sigma}$ jest sprzeczna.

Dygresja 5. Aksjomat K1' głosi, że jeżeli zdanie α jest wyprowadzalne w $\mathbf{\Sigma}$ z czegoś, co wchodzi w skład wiedzy podmiotu, to α jest prawdziwe (jest tak, jak głosi α). Uzasadnieniem przyjęcia K1' jest zasada K1 (która — w tym przypadku — wyraża niezawodność wiedzy *explicite*) oraz to, że relacja wyprowadzalności zachowuje prawdę. Z kolei K2' głosi, że K1' jest wyprowadzalne w $\mathbf{\Sigma}$ z czegoś, co podmiot wie. Między K1' i K2' istnieje ścisły związek. Skoro każde zdanie jest wyprowadzalne z samego siebie, K2' jest prawdziwe, o ile podmiot wie, że K1'. ■

Dowód. Korzystamy z faktu, że relacja wyprowadzalności spełnia warunek przechodniości, czyli $\forall x \forall y \forall z(I(x, y) \wedge I(y, z) \rightarrow I(x, z))$. Wtedy dla dowolnych zdań α i β języka $J(\mathbf{\Sigma})$,

K3'. $\quad \mathbf{\Sigma} \vdash K'[\alpha] \wedge I[\alpha, \beta] \rightarrow K'[\beta]$.

Istotnie:

1. $\mathbf{\Sigma} \vdash K'[\alpha] \equiv \exists y(K(y) \wedge I(y, \ulcorner\alpha\urcorner))$ definicja (DK')
2. $\mathbf{\Sigma} \vdash K'[\alpha] \wedge I[\alpha, \beta] \equiv \exists y(K(y) \wedge I(y, \ulcorner\alpha\urcorner) \wedge I[\alpha, \beta])$ na podstawie 1
3. $\mathbf{\Sigma} \vdash K'[\alpha] \wedge I[\alpha, \beta] \rightarrow \exists y(K(y) \wedge I(y, \ulcorner\beta\urcorner))$ przechodniość I
4. $\mathbf{\Sigma} \vdash K'[\alpha] \wedge I[\alpha, \beta] \rightarrow K'[\beta]$ definicja (DK').

Łatwo zauważyć, że K1', K2', K3' odpowiadają założeniom K1, K2, K3. Sprzeczność otrzymamy, powtarzając kroki dowodu Twierdzenia 1. ■

Na mocy poniższego twierdzenia podmiot nie może znać — pod groźbą sprzeczności — każdego przypadku zasady K1' (chociaż oczywiście może mu się wydawać, że zna).

Twierdzenie 5 (Cross 2001; *The Paradox of the Knowledge-Plus Knower*). Niech teraz $\mathbf{\Sigma}$ będzie aksjomatyzowalnym rozszerzeniem arytmetyki **Q** takim, że predykat K' — zdefiniowany przez (DK') — spełnia następujące dwa warunki: dla dowolnych zdań α i β języka $J(\mathbf{\Sigma})$,

K1'. $\quad K'[\alpha] \rightarrow \alpha$

K2⁺. $\quad K[K'[\alpha] \rightarrow \alpha]$.

Wówczas teoria $\mathbf{\Sigma}$ jest sprzeczna.

Dygresja 6. Aksjomat K2^{+} głosi, iż podmiot wie (*explicite*), że K1'. ■

Dowód. Na mocy lematu o diagonalizacji, istnieje zdanie δ takie, że

$$\mathbf{\Sigma} \vdash \delta \equiv K'[\neg\delta].$$

Przyjmijmy skróty:

K1'd dla $\quad K'[\neg\delta] \rightarrow \neg\delta$,

K2'd dla $\quad K'[K'[\neg\delta] \rightarrow \neg\delta]$,

K2^{+}d dla $\quad K[K'[\neg\delta] \rightarrow \neg\delta]$,

Na podstawie Twierdzenia 4 mamy: K1'd, K2'd $\vdash \bot$ (symbol $\bot$ oznacza stałą *falsum*).

1. $\mathbf{\Sigma} \vdash \text{K2}^{+}\text{d} \wedge I[\text{K1'd}, \text{K1'd}] \rightarrow \text{K2'd}$ na podstawie definicji (DK')
2. $\mathbf{\Sigma} \vdash I[\text{K1'd}, \text{K1'd}]$ każde zdanie jest wyprowadzalne z samego siebie

3. $\mathbf{\Sigma} \vdash K2^{+}d \rightarrow K2'd$ na podstawie 1 i 2

4. $\mathbf{\Sigma} \vdash K1'd \wedge K2'd \rightarrow \bot$ na podstawie Twierdzenia 4

5. $\mathbf{\Sigma} \vdash K1'd \wedge K2^{+}d \rightarrow \bot$ na podstawie 3 i 4.

A zatem, $\mathbf{\Sigma}$ jest teorią sprzeczną (jeśli jej twierdzeniami są K1'd, K2'd i $K2^{+}d$). ■

Podsumowanie: Teoria $\mathbf{\Sigma}$, będąca aksjomatyzowalnym rozszerzeniem arytmetyki **Q**, w której predykat wiedzy charakteryzują następujące zasady:

(1) K1-K3 (Kaplan, Montague 1960)

(2) K1, RK (Montague 1963)

(3) DK', K1', K2' (Cross 2001)

(4) DK', K1', $K2^{+}$ (Cross 2001)

jest sprzeczna.

2. DIAGNOZA I ROZWIĄZANIA

Przechodzę obecnie do omówienia pewnych propozycji rozwiązania paradoksu znawcy. Po pierwsze, wiążą one wiedzę z dowodliwością — dotyczą więc specyficznego podmiotu, a mianowicie „matematycznego znawcy". Po drugie, przyjmują, iż źródłem sprzeczności jest założenie K2 przypisujące podmiotowi wiedzę o tym, że jego wiedza jest niezawodna — może nasza wiedza jest niezawodna (w sensie K1), ale nie możemy wiedzieć z matematyczną pewnością, że tak właśnie jest.

Generalnie, w celu rozwiązania paradoksu znawcy można zdystansować się do każdej z zasad K1, K2 i K3, a także możliwości budowania zdań odnoszących się do samych siebie. W obronie założenia K1, czyli epistemicznej zasady refleksji, można powiedzieć, że odzwierciedla ona na poziomie formalnym intuicje wiązane z pojęciem wiedzy. Bez owej zasady predykat *K* nie byłby adekwatnym predykatem wiedzy (w stosunku do Platońskiego ideału wiedzy jako prawdziwego i uzasadnionego przekonania).[12] Z kolei, zasada K3, czyli zasada epistemicznego domknięcia, mimo że na pierwszy rzut oka jest najbardziej wątpliwym założeniem, nie jest niezbędna do wyprowadzenia sprzeczności (z uwagi na rezultat Crossa). Ponadto jej zakwestionowanie rodzi problem logicznej ignorancji: W jaki sposób odróżnić dedukcje,

---

[12] Dodatkowo, istnieje wersja paradoksu znawcy bez owej zasady przedstawiona w (Thomason 1980). Dotyczy ona przekonań, a nie wiedzy podmiotu. Natomiast w (Woleński 2005, s. 271-274) przedstawia się pewne powody ewentualnego odrzucenia K1. Idea sprowadza się — z uwagi na teoriomodelową formalizację pojęcia prawdziwości — do odróżnienia dla danego zbioru zdań (wiedzy) „modelu intencjonalnego", związanego z domniemaniami podmiotu, i „modelu rzeczywistego", reprezentującego „świat realny". Od wiedzy w sensie *episteme* oczekuje się, że zdanie prawdziwe w „modelu intencjonalnym" będzie też prawdziwe w „modelu rzeczywistym".

które charakteryzują wiedzę *explicite* i których brak uczyni z podmiotu logicznego ignoranta, od tych, które świadczą o jego dedukcyjnej wszechwiedzy? Jeszcze bardziej złożony jest problem zdań odnoszących się do samych siebie. Można wprawdzie powiedzieć o nich, że są konstrukcjami sztucznymi (np. nie posiadają naturalnych egzemplifikacji wśród wypowiedzi) i je wyeliminować, ale wydaje się to zabiegiem instrumentalnym (żeby nie rzec fanaberią). Nie każde tego typu zdanie ma destrukcyjne konsekwencje. Są wśród nich zdania samoodnośne w sposób „niewinny" i samoodnośne w sposób „złośliwy". Jak zauważa S. Kripke, nawet całkiem normalne zdania, jeśli fakty empiryczne towarzyszące ich wypowiedzeniu są niesprzyjające, mogą — niejako przez przypadek — odnosić się do samych siebie i generować paradoksy (Kripke 1975). Na terenie arytmetyki zdania samoodnośne są dopuszczalne *via* lemat o diagonalizacji (i numeracja Gödlowska), a nadto pełnią istotną rolę w dowodach wielu ważnych twierdzeń.

Wycofać zatem należy założenie K2. Pod jego adresem wysuwa się różne obiekcje. Jedne z nich mają charakter intuicyjny. Jeżeli wiedza podmiotu spełnia jakiś warunek W, to czy można zasadnie założyć, że ów podmiot wie, iż W? Odpowiedź na tak postawione pytanie — jak to często bywa — brzmi: I tak, i nie. Z jednej strony wydaje się, że poprzez zdolność do introspekcji, w szczególności refleksji na temat swej własnej wiedzy, uzyskujemy dostateczną ewidencję, aby odpowiedzieć na nie pozytywnie, z drugiej zaś strony mamy również argumenty skłaniające do odpowiedzi negatywnej. Po pierwsze, introspekcja nierzadko zwodzi. Po drugie, osiągnięcie przez podmiot wiedzy, że W może zmienić jego wiedzę w ten sposób, że pewne dotyczące jej warunki, nie wykluczając samego warunku W, zostaną sfalsyfikowane. Wyobraźmy sobie podmiot, którego wiedzę tworzą tylko trzy zdania. Niech to będzie ów warunek W. Osiągniecie przez podmiot wiedzy, że W prowadzi do falsyfikacji owego warunku.

Inne zarzuty mają charakter formalny — odwołują się do analogii pomiędzy predykatem wiedzy K a predykatem dowodliwości Pr_Σ (Tymoczko 1984, s. 453-456). Gdy w warunkach K1, K2 i K3* predykat K zastąpimy przez Pr_Σ, przybiorą one wówczas postać:

D1. $Pr_\Sigma[\alpha] \to \alpha$,

D2. $Pr_\Sigma[Pr_\Sigma[\alpha] \to \alpha]$,

D3. $Pr_\Sigma[\alpha \to \beta] \to (Pr_\Sigma[\alpha] \to Pr_\Sigma[\beta])$.

Dygresja 7. D1 to tzw. *lokalna zasada refleksji*. Postuluje ona prawdziwość każdego zdania dowodliwego na gruncie danej teorii. Poszczególne jej przypadki wzięte łącznie wyrażają więc trafność rozważanej teorii. Rola jej — w przeciwieństwie do zasady K1 — jest przedmiotem różnych polemik. ■

W oparciu o D2 otrzymujemy zdanie $Pr_\Sigma[Pr_\Sigma[\bot] \to \bot]$, które jest równoważne zdaniu $Pr_\Sigma[\neg Pr_\Sigma[\bot]]$ (z uwagi na definicję $\neg\alpha$ jako $\alpha \to \bot$). Wyraża ono, że stwierdzenie niesprzeczności rozważanej teorii — tj. zdanie $\neg Pr_{\Sigma}[\bot]$ — jest dowodliwe w owej teorii, co z uwagi na drugie twierdzenie Gödla o niezupełności jest wykluczone (o ile rozważana teoria jest niesprzeczna; w teorii sprzecznej wszystko daje się udowodnić). Epistemicznym odpowiednikiem zdania $Pr_{\Sigma}[\neg Pr_{\Sigma}[\bot]]$ jest zdanie $K[\neg K[\bot]]$. Głosi ono, że podmiot wie o swej własnej niesprzeczności (co można uznać za objaw nieskromności). Gdyby więc podmiot był typu K2 (czyli nieskromny), wtedy wiedziałby o swej własnej niesprzeczności. Analogia między predykatami K i Pr_{Σ} sugeruje, że jest to niemożliwe, o ile podmiot jest niesprzeczny (czyli gdy jego wiedza, że $\neg\alpha$ implikuje niewiedzę, że α, dla dowolnego zdania α).[13]

2.1. Epistemiczna interpretacja logiki dowodliwości

Logika dowodliwości jest modalnym systemem zdaniowym, motywowanym metamatematycznymi wynikami K. Gödla. Jej powstanie wiąże się z modalną interpretacją warunków charakteryzujących predykat dowodliwości $Pr_{\mathrm{PA}}(x)$ (tj. w logice tej funktor konieczności L odpowiada predykatowi dowodliwości w arytmetyce **PA**).[14] Podstawowy system **GL** powstaje z systemu **K4** przez dodanie jako aksjomatu modalnego odpowiednika twierdzenia Löba, a mianowicie formuły:

G. $\qquad L(L\alpha \to \alpha) \to L\alpha.$

Intuicyjny związek pomiędzy **GL** a arytmetyką **PA** daje się uchwycić formalnie przez pokazanie, że **GL** jest zarazem arytmetycznie adekwatna i pełna, czyli wszystko, co w arytmetyce **PA** można udowodnić na temat jej własnego predykatu dowodliwości można też udowodnić w **GL** i na odwrót. Wymaga to wprowadzenia pewnej funkcji t, zwanej *realizacją* (**PA**-*realizacją*), przyporządkowującej formułom danego języka modalnego zdania języka arytmetyki **PA** zgodnie z następującymi warunkami:

$t(p_i)$ = zdanie J(**PA**), gdzie p_i jest zmienną zdaniową,

$t(\bot) = \bot$,

$t(\alpha \otimes \beta) = t(\alpha) \otimes t(\beta)$, gdzie $\otimes$ jest jakimś dwuargumentowym spójnikiem,

$t(L\alpha) = Pr_{\mathrm{PA}}[t(\alpha)]$.

---

[13] Oczywiście podmiot zapytany o własną niesprzeczność może odpowiedzieć twierdząco (i zwykle tak czyni), ale to nie oznacza jeszcze, że wie on o swej własnej niesprzeczności — udziela jej bowiem na podstawie jakiegoś rodzaju samoobserwacji (choćby wczucia się w swój umysł), lecz przecież może się mylić.

[14] Logiki modalne są wygodnym narzędziem analizy pewnych pojęć występujących w językach pierwszego rzędu, choćby z uwagi na to, że są rozstrzygalne. Tworzą one jednak istną wieżę Babel, co stanowi pewien kłopot w takim ich wykorzystaniu.

Arytmetyczną adekwatność i pełność logiki **GL** wyraża następujące twierdzenie udowodnione przez R. Solovaya: Dla dowolnej formuły α języka logiki modalnej, **GL** $\vdash \alpha$ wtw dla każdej realizacji t, **PA** $\vdash t(\alpha)$.[15]

Gdybyśmy chcieli interpretować wyrażenie $L\alpha$ jako odpowiadające zwrotowi „Wiadomo, że α" (a nie „α jest dowodliwe w **PA**"), powinniśmy przyjąć aksjomat T, czyli schemat $L\alpha \rightarrow \alpha$. Okazuje się jednak, że aksjomat ów nie jest twierdzeniem logiki **GL**, a nawet jego dodanie usprzecznia ów system.[16] Logiką dowodliwości, która lepiej niż **GL** nadaje się na logikę wiedzy jest system **GLS**. **GLS**, w przeciwieństwie do **GL**, nie jest systemem *normalnym* (nie jest on bowiem zamknięty na regułę ukonieczniania). Otrzymuje się go przez dodanie schematu T do wszystkich tez systemu **GL** i domknięcie jedynie na regułę *modus ponens*.[17] Związek między nim a arytmetyką **PA** określa tzw. drugie twierdzenie Solovaya, w myśl którego system **GLS** ujmuje wszystko, co jest prawdą o predykacie Pr_{PA} w standardowym modelu **N** arytmetyki **PA**. Dokładniej: Dla dowolnej formuły α języka logiki modalnej, **GLS** $\vdash \alpha$ wtw dla każdej realizacji t, **N** $\models t(\alpha)$. Ważnym faktem na temat tego systemu — z interesującego tu nas punktu widzenia — jest też to, że spełnia on lemat o punkcie stałym.

Spróbujmy teraz nadać systemowi **GLS** interpretację epistemiczną, tj. uznać go za logikę wiedzy jakiegoś wyidealizowanego matematyka. Z punktu widzenia **GLS** wiedza owego matematyka ma następujące własności. Po pierwsze, zawiera ona (modulo realizacja t) tylko prawdy arytmetyczne. Po drugie, obowiązywanie schematu T przesądza o jej niezawodności. Po trzecie, występowanie wśród tez **GLS** wszystkich zdań powstających ze schematu K gwarantuje jej domknięcie na regułę *modus ponens*. Po czwarte, spełniona jest zasada pozytywnej introspekcji, w myśl której jeżeli podmiot coś wie, to także wie, że owo coś wie. Jest tak, gdyż tezami **GLS** są wszystkie zdania powstające ze schematu 4, tj. $L\alpha \rightarrow LL\alpha$. Wreszcie po piąte, z uwagi na występowanie wśród tez **GLS** wszystkich zdań powstających ze schematu G, podmiot jest *skromny* względem każdego zdania α, czyli wie on, że α, o ile wie on, iż jego wiedza, że α implikuje α.[18] W szczególności, **GLS** $\vdash L\neg L\bot \rightarrow L\bot$. Gdyby więc podmiot wiedział o swej własnej niesprzeczności, to popadłby w sprzeczność. Przez kontrapozycję uzyskujemy **GLS**$\vdash \neg L\bot \rightarrow \neg L\neg L\bot$. Oznacza to, że jeśli podmiot jest niesprzeczny, to nie może wiedzieć o swej własnej nie-

[15] W kwestii dowodu tego twierdzenia zob. np. Boolos 1993.

[16] W Smullyan 1987 przedstawiono doksastyczną (związaną z pojęciem wiary) interpretację systemu **GL**. Inne ważne zasady, które charakteryzują wiedzę, lecz nie są twierdzeniami **GL** to: $L\neg L\alpha \rightarrow \neg L\alpha$ (zawsze, gdy podmiot wyklucza, iż wie, że α, nie wie, że α) oraz $L\alpha \rightarrow \neg L\neg\alpha$ (niesprzeczność wiedzy).

[17] Dodajmy, że logika **GLS**, inaczej niż **GL**, jest zamknięta na regułę Arystotelesa, tj. uznawania za możliwe tego, co uznane.

[18] Również to określenie zapożyczyłem od Smullyana (Smullyan 1987). Zauważmy przy okazji, że zinterpretowany epistemicznie schemat G — w przeciwieństwie do zasady K2 — nie dotyczy bezpośrednio tego, co podmiot wie. Ma on bowiem postać warunkową.

sprzeczności. Ponieważ **GLS** $\vdash L\bot \rightarrow \bot$, czyli **GLS** $\vdash \neg L\bot$, więc ostatecznie **GLS** $\vdash \neg L\neg L\bot$, a także **GLS** $\nvdash L\neg L\bot$.[19] To daje, że podmiot nie wie o swej własnej niesprzeczności (mówiąc inaczej, nie wyklucza on swej własnej sprzeczności).

Brak wśród tez systemu **GLS** schematu U, będącego modalnym odpowiednikiem założenia K2, blokuje powstanie paradoksu znawcy.

2.2. Rozwiązanie C. A. Andersona

Propozycję Andersona rozwiązania paradoksu znawcy można określić mianem *hierarchicznej*, gdyż odwołuje się do odróżnienia poziomów poznania/wiedzy — paradoks znawcy ujawnia ukrytą hierarchiczną strukturę wiedzy (Anderson 1983). Opiera się ona na założeniu, że potoczny zwrot „wiem, że" jest systematycznie wieloznaczny w tym sensie, że jego „stopień" w konkretnym wystąpieniu jest wyznaczony przez kontekst wypowiedzi, a przez to ustalony.[20] Zdaniem Andersona predykat wiedzy K występujący na zewnątrz w schemacie $K[K[\alpha] \rightarrow \alpha]$ reprezentuje wyższy poziom poznania niż predykat K występujący wewnątrz owego schematu.

Niech $J(\mathbf{Q})_\omega$ będzie językiem powstałym z języka arytmetyki Robinsona $J(\mathbf{Q})$ przez dodanie do niego nieskończenie wielu predykatów wiedzy i wyprowadzalności indeksowanych liczbami naturalnymi: $K_0, K_1, K_2, \ldots, I_0, I_1, I_2, \ldots$. Obejmuje on jako swe części wszystkie języki $J(\mathbf{Q}), J(\mathbf{Q})_0, J(\mathbf{Q})_1, J(\mathbf{Q})_2, \ldots$, gdzie $J(\mathbf{Q})_n = J(\mathbf{Q}) \cup \{K_0, K_1, \ldots, K_n, I_0, I_1, \ldots, I_n,\}$.[21] Idea Andersona jest następująca. Niech Alfa będzie pewnym szczególnym matematykiem, takim że jego wiedzę cechuje niezawodność, epistemiczne domknięcie oraz etapowość. Przyjmijmy, że K_0 jest rekurencyjnie przeliczalnym zbiorem zdań języka $J(\mathbf{Q})_\omega$ reprezentującym wyjściową wiedzę owego matematyka. Buduje on teraz teorię **Q'** (rozszerzającą arytmetykę **Q**), zawierającą jako tezy wszystkie zdania wyprowadzalne z aksjomatów arytmetyki **Q** i zdań ze zbioru K_0. Oczywiście, teoria **Q'** jest niesprzeczna, gdyż wszystkie aksjomaty **Q** są prawdziwe w standardowym modelu, zdania ze zbioru K_0, jako tworzące wiedzę Alfy, też są prawdziwe oraz reguły inferencyjne zachowują prawdziwość. Niech teraz Beta będzie drugim matematykiem, który zamierza scharakteryzować wiedzę Alfy. Korzystając z lematu o diagonalizacji, konstruuje on zdanie Gödla dla **Q'**, tj. zdanie δ takie, że **Q'** $\vdash \delta \equiv \neg Pr_{Q'}[\delta]$. Przeprowadzając stosowne rozumowanie, dochodzi następnie do wniosku, że zdanie δ jest prawdziwe (w modelu dla **Q'**) wtw nie jest ono

---

[19] Zdanie $t(L\neg L\bot)$ (= $Pr_{PA}[\neg Pr_{PA}[\bot]]$) wyraża dowodliwość (stwierdzenia) niesprzeczności **PA** w **PA**. Gdyby **GLS** $\vdash L\neg L\bot$, wtedy zdanie $t(L\neg L\bot)$ byłoby prawdziwe w standardowym modelu arytmetyki **PA** (na mocy drugiego twierdzenia Solovaya). Z drugiej strony, drugie twierdzenie Gödla o niezupełności wyklucza dowodliwość niesprzeczności **PA** w niej samej.

[20] Istnieje analogia pomiędzy tą propozycją a Quine'a modyfikacją teorii prawdy Tarskiego, uznającej predykat prawdy jako systematycznie wieloznaczny (Quine 1992, 136-139).

[21] Należy zauważyć, że nie zakazuje się budowania zdań postaci $K_n[\alpha(K_n)]$, np. $K_0[\forall x(K_0(x) \vee \neg K_0(x))]$.

dowodliwe w **Q'**. Znaczy to, że bądź (1) zdanie δ jest prawdziwe, ale nie jest dowodliwe, bądź (2) zdanie δ nie jest prawdziwe, ale jest dowodliwe. Zakładając, że teoria **Q'** jest poprawna (tj. każde zdanie w niej dowodliwe jest prawdziwe oraz żadne zdanie obalane nie jest prawdziwe), ewentualność (2) należy wykluczyć. Zdanie δ, chociaż prawdziwe, nie należy do zbioru K_0, gdyż w przeciwnym razie byłoby dowodliwe w **Q'**. Oznacza to, że Alfa nie wie, że δ. Z drugiej strony, skoro Beta ustalił, że δ jest niedowodliwe w **Q'**, więc wie on, że δ. Znaczy to, że δ można włączyć do zbioru K_1 reprezentującego wiedzę Bety.[22] Przyjmijmy teraz, że Alfa i Beta nie są różnymi osobami. Wówczas — argumentuje Anderson — należy odróżnić to, co Beta wie na poziomie K_0, od tego, co wie on na poziomie K_1. W analogiczny sposób można wyróżnić dalsze poziomy wiedzy K_2, K_3, … .

Przedstawioną wyżej ideę Anderson formalizuje następująco. Etapowość wiedzy uzyskuje reprezentację w postaci hierarchii teorii $\{\mathbf{\Sigma}_n\colon n \in \omega\}$ niesprzecznie rozszerzających arytmetykę **Q**:

$$\mathbf{\Sigma}_0 = \mathbf{Q} \cup \{K_0[\alpha] \to \alpha,\ K_0[\alpha] \wedge I_0[\alpha, \beta] \to K_0[\beta];\ \alpha, \beta \in J(\mathbf{Q})_\omega\},$$

$$\mathbf{\Sigma}_{n+1} = \mathbf{\Sigma}_n \cup \{K_{n+1}[\alpha] \to \alpha,\ K_{n+1}[K_n[\alpha] \to \alpha],\ K_{n+1}[\alpha] \wedge I_{n+1}[\alpha, \beta] \to K_{n+1}[\beta];\ \alpha, \beta \in J(\mathbf{Q})_\omega\}.$$

Semantyczny opis tych teorii przedstawia się następująco. Niech $J(\mathbf{Q})_\omega = \bigcup\{J(\mathbf{Q})_n\colon n \in \omega\}$, gdzie $J(\mathbf{Q})_0 = J(\mathbf{Q}) \cup \{K_0, I_0\}$, $J(\mathbf{Q})_{n+1} = J(\mathbf{Q})_n \cup \{K_{n+1}, I_{n+1}\}$. Przez $g(J(\mathbf{Q})_\omega)$ oznaczamy zbiór kodów zdań języka $J(\mathbf{Q})_\omega$. Niech dalej V_q będzie standardową interpretacją języka $J(\mathbf{Q})$. Definiujemy hierarchię interpretacji $\{V_n\colon n \in \omega\}$ dla języków tworzących hierarchię $\{J(\mathbf{Q})_n\colon n \in \omega\}$:

(1) V_0 jest interpretacją języka $J(\mathbf{Q})_0$ rozszerzającą interpretację V_q.

(2) V_{n+1} jest interpretacją języka $J(\mathbf{Q})_{n+1}$ rozszerzającą interpretację V_n.

(3) Dla dowolnego $n \in \omega$, $V_n(K_n) \subseteq g(J(\mathbf{Q})_\omega)$, zaś $V_n(I_n) \subseteq g(J(\mathbf{Q})_\omega) \times g(J(\mathbf{Q})_\omega)$.

(4) $V = \bigcup\{V_n\colon n \in \omega\}$ jest interpretacją języka $J(\mathbf{Q})_\omega$.

Hierarchia interpretacji $\{V_n\colon n \in \omega\}$ jest *spójna*, jeśli spełnia poniższe warunki:

(5) $V_n(K_n) \subseteq V_{n+1}(K_{n+1})$.

[22] Należy wyraźnie powiedzieć, że wniosek Bety o prawdziwości zdania δ nie pochodzi z formalnych wnioskowań przeprowadzonych wewnątrz teorii **Q'**, której owo zdanie dotyczy, lecz jest „z zewnątrz systemu" (chociaż fakty potrzebne do wyciągnięcia tego wniosku o δ dają się wyrazić i udowodnić w **Q'**). Zdanie δ jest nierozstrzygalne w **Q'**. Oznacza to, że teoria, którą posługuje się Beta, rozstrzygając zdanie δ dla **Q'**, musi być silniejsza od teorii **Q'**. Korzysta on bowiem z przesłanki, że **Q'** jest poprawna (ewentualnie, że **Q'** jest niesprzeczna — poprawność implikuje niesprzeczność). Na fakt, że kwestia wiedzy na temat prawdziwości zdania Gödla dla danej teorii sprowadza się do kwestii, czy wiemy, że rozważana teoria jest niesprzeczna, zwrócił uwagę H. Putnam (Putnam 1960; zob. też Krajewski 2003, s. 119-120).

(6) $V_n(I_n) \subseteq V_{n+1}(I_{n+1})$.

(7) Jeżeli $g(\alpha) \in V_n(K_n)$, to istnieje $m \geq n$ takie, że $V_m(\alpha) = 1$.

(8) Jeżeli $\langle g(\alpha), g(\beta)\rangle \in V_n(I_n)$, to istnieje $m \geq n$ takie, że $V_m(\alpha \rightarrow \beta) = 1$.

(9) Jeżeli $g(\alpha) \in V_n(K_n)$ i $\langle g(\alpha), g(\beta)\rangle \in V_n(I_n)$, to $g(\beta) \in V_n(K_n)$.

Warunki (5) i (6) gwarantują transmisję wiedzy i wyprowadzalności z poziomu niższego na poziom wyższy. Warunek (7) gwarantuje niezawodność wiedzy: to, co podmiot wie na danym poziomie, musi być prawdziwe na tym samym lub jakimś następnym poziomie. Warunek (8) jest semantycznym analogonem twierdzenia o dedukcji (dla każdego I_n). Wreszcie warunek (9) jest semantycznym odpowiednikiem epistemicznego domknięcia.

Interpretacje V_n są wartościowaniami logicznymi formuł języka $J(\mathbf{Q})_\omega$, przy czym nakłada się na nie następujące warunki:

$$V_0(K_0[\alpha]) = 1 \text{ wtw } \mathbf{Q} \vdash \alpha,$$

$$V_{n+1}(K_{n+1}[\alpha]) = 1 \text{ wtw } \mathbf{\Sigma}_n \vdash \alpha,$$

$$V_0(I_0[\alpha, \beta]) = 1 \text{ wtw } \mathbf{Q} \vdash \alpha \rightarrow \beta.$$

$$V_{n+1}(I_{n+1}[\alpha, \beta]) = 1 \text{ wtw } \mathbf{\Sigma}_n \vdash \alpha \rightarrow \beta.$$

Warunki owe gwarantują spójność rozważanej hierarchii interpretacji. Przez V^{koh} oznaczmy interpretację zbudowaną nad spójną hierarchią interpretacji $V_q, V_0, V_1, \ldots$. Zdaniami języka $J(\mathbf{Q})_\omega$ prawdziwymi w każdej interpretacji V^{koh} są m.in. zdania powstające z następujących schematów ($n \in \omega$):

$$K_n[\alpha] \rightarrow \alpha,$$

$$K_n[\alpha] \wedge I_n[\alpha, \beta] \rightarrow K_n[\beta],$$

$$K_{n+1}[K_n[\alpha] \rightarrow \alpha].$$

Zablokowanie paradoksu znawcy opiera się na fakcie, że schemat: $K_n[K_n[\alpha] \rightarrow \alpha]$ posiada fałszywe konkretyzacje; otrzymujemy z niego fałsz w przypadku zdania δ takiego, że $\mathbf{\Sigma}_n \vdash \delta \equiv K_n[\neg\delta]$. W ten sposób zablokowany zostaje krok 7 w dowodzie Twierdzenia 1 i ostatni krok w dowodzie Twierdzenia 2. Przeciwko tej propozycji można wysunąć zarzut, że zdanie „Podmiot *a* (np. Bóg) wie, że na każdym poziomie jego wiedzę cechuje niezawodność” nie jest prawdziwe na żadnym poziomie przedstawionej hierarchii, gdyż dotyczy ono każdego poziomu owej hierarchii.

Dygresja 7. Interpretacja V reprezentuje n-doskonały podmiot epistemiczny, jeśli $V(K_n[\alpha]) = 1$, dla każdego zdania α języka $J(\mathbf{Q})_\omega$, które jest prawdziwe w każdej interpretacji V^{koh}. Zdanie α jest n-transcendentne, jeśli $\neg K_n[\alpha]$ jest prawdziwe w każdej interpretacji V^{koh}, zdanie $\neg K_{n+1}[\alpha]$ zaś nie jest. Anderson dowodzi, że: (1) żadna interpretacja nie reprezentuje n-doskonałego podmiotu epistemicznego, dla dowolnego $n \in \omega$; (2) dla każdego $n \in \omega$ istnieje zdanie, które jest n-trancendentne. ■

2.3. Rozwiązanie P. Égré'ego

Wiąże ono dwa poprzednie rozwiązania. Zdaniem Égré'ego, istnieje odpowiedniość pomiędzy systemem **GLS** a teorią $\mathbf{\Sigma}_0$ hierarchii Andersona. Tezę swą opiera na dwóch przesłankach. Pierwsza postuluje odpowiedniość **GLS** i teorii $\mathbf{PA}^+ = \mathbf{PA} \cup \{Pr_{\mathrm{PA}}[\alpha] \to \alpha\colon \alpha \in J(\mathbf{PA})\}$, tj. teorii powstałej z arytmetyki Peano przez dodanie lokalnej zasady refleksji. Druga postuluje odpowiedniość $\mathbf{PA}^+$ i teorii $\mathbf{\Sigma}_0$. Wybór teorii $\mathbf{PA}^+$ jako łącznika pomiędzy **GLS** i $\mathbf{\Sigma}_0$ uzasadniony jest tym, że daje się w niej udowodnić niesprzeczność arytmetyki **PA**. Teza Égré'ego, choć interesująca, nie ma jednak żadnego formalnego potwierdzenia. Nie przedstawił on bowiem żadnej realizacji przyporządkowującej formułom języka logiki **GLS** zdania języka $J(\mathbf{PA}^+)$ oraz jakiegoś analogonu twierdzeń Solovaya, ustalającego rodzaj odpowiedniości pomiędzy **GLS** i $\mathbf{PA}^+$. Wątpliwości budzi też odpowiedniość teorii $\mathbf{PA}^+$ i $\mathbf{\Sigma}_0$. $\mathbf{PA}^+$ zawiera (modulo realizacja t) arytmetyczne odpowiedniki schematów K, 4 i G. Tymczasem $\mathbf{\Sigma}_0$ nie zawiera odpowiedników owych schematów.

3. ZAKOŃCZENIE

Główną słabością przedstawionych powyżej rozwiązań jest odwołanie się do podobieństwa pomiędzy predykatem (funktorem) wiedzy a predykatem (funktorem) dowodliwości. Argument ten ma wprawdzie dużą moc perswazyjną, można natomiast powątpiewać w jego moc uzasadniającą odrzucenie założenia K2. Właściwie w odwołaniu się do owego podobieństwa zawarta jest sugestia, że umysł ludzki odpowiada systemowi formalnemu, a dochodzenie przez podmiot do wiedzy koresponduje z produkowaniem twierdzeń w danym systemie formalnym.

Istnieją jeszcze inne propozycje rozwiązania paradoksu znawcy. Jedne z nich nawiązują — w sposób mniej lub bardziej wyrafinowany — do teorii S. Kripkego modelującej proces stopniowego poznawania ekstensji i antyekstensji predykatu prawdy (zob. np. Horsten 1998). U ich podstaw leży założenie, że posiadanie przez podmiot wiedzy nie ma charakteru statycznego, lecz dynamiczny, oraz że istnieje tylko jeden predykat wiedzy, ale — podobnie jak predykat prawdy — jest on cząstkowy (tzn. jego ekstensja i antyekstensja nie wyczerpują zbioru wszystkich zdań rozważanego języka). Wiedza podmiotu na żadnym etapie poznawania nie jest zupełna, czyli zawsze oprócz tego, co podmiot wie (czyli należy do ekstensji predykatu wiedzy), i tego, co wyklucza (czyli należy do antyekstensji predykatu wiedzy), istnieje jeszcze to, co nierozstrzygnięte. Zdaniem „absolutnie” nierozstrzygalnym (nieokreślonym w każdym punkcie stałym dookreślania predykatu wiedzy) — i dlatego paradoksalnym — jest zdanie znawcy.

Z kolei, inne rozwiązania nawiązują do rewizyjnej teorii definicji A. Gupty i N. Belnapa. Opierają się one na założeniu, że pojęcie wiedzy jest koliste, tzn. definicja predykatu określającego wiedzę jako uzasadnione przekonanie prawdziwe, jest istotnie kolista. Istnieją bowiem zdania (np. zdanie znawcy), których warunków przynależności do ekstensji predykatu wiedzy nie daje się określić bez posłużenia się pojęciem wiedzy.

Nawiążmy do tej propozycji i przyjmijmy, że pojęcie wiedzy jest definiowane za pomocą nieskończonego zbioru definicji cząstkowych, które podpadają pod następujący schemat:

(TB) $K[\alpha] =_{df} \alpha \wedge B[\alpha]$,

gdzie α reprezentuje zdania rozważanego języka wzbogaconego o predykat wiedzy $K(x)$ i predykat uzasadnionego przekonania $B(x)$. Schemat (TB), można powiedzieć, stanowi regułę użycia predykatu wiedzy. Jak każda definicja, ma on na celu ustalenie ekstensji definiowanego wyrażenia. Zdania będące jego uszczegółowieniami tworzą łącznie (intensjonalnie) adekwatną definicję pojęcia wiedzy. Schemat ów zastosowany do zdania zawierającego predykat wiedzy daje definicję kolistą (tj. w której predykat definiowany występuje w *definiensie*). W przeciwieństwie do teorii opartych na propozycji Kripkego żadna korekta ukierunkowana na ustalenie ekstensji predykatu wiedzy nie ma charakteru kumulatywnego. Jest tak, ponieważ ekstensja i antyekstensja predykatu wiedzy nie tylko są zbiorami rozłącznymi, ale również się dopełniają. Powoduje to, że żadna procedura rewizyjna, mająca na celu ustalenie ekstensji owego predykatu, nie zaczyna się od „pustej" interpretacji oraz pewne zdania zachowują się patologicznie, czyli w wyniku kolejnych rewizji zmieniają swą lokalizację. Innymi słowy, ich przynależność do ekstensji predykatu wiedzy nie stabilizuje się. Zdaniem takim jest m.in. zdanie znawcy δ. Przez κ_M oznaczmy regułę rewizyjną dla predykatu wiedzy (gdzie M jest modelem wyjściowym, tj. modelem *K*-wolnego fragmentu rozważanego języka). Jej działanie opisuje następujący warunek: dla dowolnego zbioru E (kodów zdań rozważanego języka) jako hipotezy dotyczącej ekstensji predykatu *K*,

$$\kappa_M(E) = \{g(\alpha): M + E \models \alpha \wedge B[\alpha]\}.$$

Tak więc, zdanie będące uzasadnionym przekonaniem prawdziwym na jednym etapie rewidowania determinuje ekstensję predykatu wiedzy *K* na etapie następnym. W związku ze zdaniem δ proste rozumowanie pokazuje, że: (1) jeżeli $g(\delta) \in E$, to $g(\delta) \notin \kappa_M(E)$; (2) jeżeli $g(\delta) \notin E$, to $g(\delta) \in \kappa_M(E)$. Oznaczając przez E_α ekstensję predykatu wiedzy na α etapie rewidowania (gdzie α jest liczbą porządkową) oraz przyjmując hipotezę, że $g(\delta) \in E_0$, otrzymujemy następującą sekwencję rewizyjną:

$$g(\delta) \in E_0, g(\delta) \notin E_1, g(\delta) \in E_2, g(\delta) \notin E_3, \ldots .$$

Analogicznie, gdy na początku przyjmiemy, że $g(\delta) \notin E_0$. Zdanie δ jest więc niestabilne dla dowolnej hipotezy wyjściowej w dowolnej sekwencji rewizyjnej, co jest źródłem jego paradoksalności.

Również te rozwiązania mają swoje wady. Obiekcje dotyczą m.in. indukcyjnego charakteru procesu dochodzenia do wiedzy. Modeluje się go za pomocą indukcji pozaskończonej. Stąd można wątpić w empiryczną adekwatność dostarczonych modeli.

BIBLIOGRAFIA

Anderson, C. A. 1983, *The Paradox of the Knower*, „The Journal of Philosophy", 80, s. 338-355.

Boolos, G. 1979, *The Unprovability of Cosistency*, Cambridge University Press.

Boolos, G. 1993, *The Logic of Provability*, Cambridge University Press.

Carroll, L, 1895, *What the tortoise said to Achilles*, „Mind", 4 (new series), s. 278-280.

Cross, C. B. 2001, *The Paradox of the Knower without Epistemic Closure*, „Mind", 110, s. 319-333.

Cross, C. B. 2004, *More on the Paradox of the Knower without Epistemic Closure*, „Mind", 113, s. 109-114.

Égré, P. 2005, *The Knower Paradox in the Light of Provability Interpretations of Modal Logic*, „Journal of Logic, Language and Information", 14, s. 13-48.

Horsten, L. 1998, *A Kripkean Approach to Unknowability and Truth*, „Notre Dame Journal of Formal Logic", 39, 3, s. 389-405.

Hughes, G. E. 1982, *John Buridan on Self-Reference. Chapter Eight of Buridan's 'Sophismata'. An Edition and a Translation with an Introduction and a Philosophical Commentary*. Cambridge: Cambridge University Press.

Kaplan, D., Montague, R. 1960, *A Paradox Regained*, „Notre Dame Journal of Formal Logic", 1, s. 79-90; przedruk w: Montague 1974, s. 271-285.

Krajewski, S. 2003, *Twierdzenie Gödla i jego interpretacje filozoficzne. Od mechanicyzmu do postmodernizmu*, Wydawnictwo IFiS PAN, Warszawa.

Kripke, S. 1975, *Outline of a Theory of Truth*, „Journal of Philosophical Logic", 72, s. 690-716; tłum. P. Garbacz, *Zarys pewnej teorii prawdy*, „Kwartalnik Filozoficzny", XXIX, 4 (2001), s. 97-131.

McGee, V. 1991, *Truth*, *Vagueness and Paradox*. *An Essay on the Logic of Truth*, Hackett Publishing Company, Indianapolis/Cambridge.

Montague, R., 1963, *Syntatical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability*, „Acta Philosophica Fenica", 16, s. 153-167; przedruk w: Montague 1974, s. 286-2302.

Montague, R. 1974, *Formal Philosophy*. *Selected Papers of Richard Montague*, edited and with an introduction by R. H. Thomason, New Haven, Yale University Press.

Murawski, R. 1991, *Funkcje rekurencyjne I elementy metamatematyki. Problemy zupełności, rozstrzygalności, twierdzenia Gödla*, Wydawnictwo Naukowe UAM, Poznań.

Priest, G. 1991, *Intensional Paradoxes*, „Notre Dame Journal of Formal Logic", 2, s. 193-211.

Putnam, H. 1960, *Minds and Machines*, [w:] S. Hook (ed.) *Dimensions of Mind*: *A Symposium*, New Yort Univ. Press, s. 128-164.

Quine, W. V. O. 1992, *Pursuit of Truth*, Harvard University Press, Cambridge, Mass.; tłum. B. Stanosz, *Na tropach prawdy*, Wydawnictwo Spacja, Warszawa 1997.

Smullyan, R. 1987, *Forever Undecided. A Puzzle Guide to Gödel*, Oxford University Press; tłum. J. Pogonowski, *Na zawsze nierozstrzygnięte. Zagadkowy przewodnik po twierdzeniach Gödla*, Książka i Wiedza, Warszawa 2007.

Stalnaker, R. 1991, *The Problem of Logical Omniscence I*, „Synthese", 89, s. 425-440.

Thomason, R. 1980, *A note on syntactical treatments of modality*, „Synthese", 44, 391-395.

Tworak, Z. 2009, *Współczesne teorie prawdy*, Wydawnictwo Naukowe UAM, Poznań 2009.

Tymoczko, T. 1984, *An Unsolved Puzzle about Knowledge*, „Philosophical Quarterly", 34, s. 437-458.

Uzquiano, G. 2004, *The Paradox of the Knower without Epistemic Closure*, „Mind", 113, s. 95-107.

Woleński, J. 2005, *Epistemologia. Poznanie, prawda, wiedza, realizm*, Wydawnictwo Naukowe PWN, Warszawa.