ARTYKUŁY

ZAGADNIENIA FILOZOFICZNE
W NAUCE
XXX / 2002, s. 48–88*

Bartosz BROŻEK

Wydział Filozoficzny PAT
Kraków

ROLA PARADOKSU KŁAMCY W KONSTRUKCJI LOGICZNYCH TEORII PRAWDY

1. WSTĘP

Celem niniejszego artykułu jest przegląd logicznych teorii prawdy pod kątem roli, jaką w ich konstrukcji pełni paradoks kłamcy, w szczególności w sytuacji, gdy budowana jest teoria prawdy dla języka naturalnego. Chciałbym bronić tezy, że rola to niepoślednia. Chciałbym też zwrócić uwagę na fakt wykorzystania pewnych niestandardowych technik semantycznych w próbach zdefiniowania prawdy, oraz na przyczyny, zasadność i konsekwencje takiego postępowania.

Rozważania rozpocznę od przywołania różnych wersji paradoksu kłamcy i krótkiej dyskusji nad tym, czy paradoks ten ma jakieś znaczenie filozoficzne. Postawię też pytanie o warunki adekwatności teorii prawdy. Następnie przedstawię trzy — moim zdaniem — paradygmatyczne teorie prawdy: Tarskiego, tzw. teorię „stałych punktów" oraz koncepcję A. Gupty i N. Belnapa, analizując sposób, w jaki teorie te „radzą" sobie z paradoksem kłamcy. W przedostatnim paragrafie porównam semantyki, z których korzystają przedstawione wcześniej koncepcje, by na koniec dokonać krótkiego podsumowania.

*UWAGA: Tekst został zrekonstruowany przy pomocy środków automatycznych; możliwe są więc pewne błędy, których sygnalizacja jest mile widziana (obi@opoka.org). Tekst elektroniczny posiada odrębną numerację stron.

2. PARADOKS KŁAMCY

Paradoks kłamcy znany jest przynajmniej od IV wieku przed Chrystusem. Żyjący w tym czasie Ateńczyk Eubulides pośród swych słynnych siedmiu zagadek umieścił i taką: „jeśli człowiek mówi, że kłamie, to czy mówi prawdę”? (por. Sainsbury, Williamson [1999], Ajdukiewicz [1931], Kotarbiński [1985]). Najsłynniejsze sformułowanie paradoksu kłamcy pochodzi od Savanaroli, który zagadkę tę streścił w lapidarnym zdaniu „ *Hoc est falsum* ”. Taką jednozdaniową formę paradoksu Eubulidesa nazywać będę poniżej **Kłamcą**. Paradoksalność Kłamcy, czyli wyrażenia

(1) To zdanie jest fałszywe

(gdzie zaimek wskazujący odnosi się do (1)) związana jest z faktem, że nie możemy mu jednoznacznie przypisać wartości prawdziwościowej. Jeśli założymy, że (1) jest prawdziwe, to znaczy, że prawdą jest to, co głosi — a głosi, że jest fałszywe. Jeśli zaś założymy jego nieprawdę, to okaże się, że musi być prawdziwe. Tak czy inaczej uzyskujemy sprzeczność.

Obok Kłamcy istnieją jeszcze inne wersje paradoksu. Wyrażenie:

(2) Czerwony Kapturek zjadł wilka i to zdanie jest fałszywe

i zdania analogiczne do (2) nazwać można **Przygodnym Kłamcą** (*Contingent Liar*) . Jak wiadomo, to wilk zjadł Kapturka, więc pierwszy człon (2) jest fałszywy, a zatem i całemu wyrażeniu przypisać możemy wartość „fałsz”[1] . Jednak w alternatywnej bajce to Kapturek mógł zjeść wilka, a wtedy (2) stałoby się paradoksalne. Paradoksalność zdań typu (2) jest więc przygodna, zależy bowiem od tego, że zachodzą takie a nie inne fakty.

---

[1]Wychodząc z założenia, że wszyscy już wiedzą, iż śnieg jest biały, zdecydowałem, że w przytaczanych przeze mnie przykładach pojawiać się będą bohaterowie bajek i baśni. Zdaję sobie sprawę, że status ontologiczny tych postaci może być przedmiotem niekończących się dyskusji. Można mieć też poważne wątpliwości, czy w świecie bajek obowiązują klasyczne zasady logiczne. Nie chcę się przy tych kwestiach zatrzymywać; przyjmijmy tylko, że tam, gdzie pojawiają się baśniowi bohaterowie, światem, o którym mówimy, jest świat baśni, z tym zastrzeżeniem, że nie sprzeciwia się on prawom logiki.

Mianem **Koła Kłamcy** (*Liar Cycle*) określa się nastepujący układ zdań:

(X_1) Zdanie (X_2) jest prawdziwe.

(X_2) Zdanie (X_3) jest prawdziwe.

...

(X_{n-1}) Zdanie (X_n) jest prawdziwe.

(X_n) Zdanie (X_1) jest fałszywe.

Powyższym zdaniom nie sposób jednoznacznie przypisać wartości prawdziwościowych. Obok Kół Kłamcy można też konstruować **Koła Przygodnego Kłamcy**.

Bardzo ciekawe zachowanie wykazuje tzw. **Wzmocniony Kłamca** (*Strenghtened Liar*) . Weźmy następujące zdanie:

(3) To zdanie albo jest paradoksalne albo nie jest prawdziwe.

Jeśli założymy, że (3) jest prawdziwe, to znaczy, że jest albo fałszywe albo paradoksalne. Jeśli przyjmiemy, że (3) jest fałszywe, to okaże się, że jest ono prawdziwe. Jeśli wreszcie założymy, że (3) jest paradoksalne, to także będziemy zmuszeni stwierdzić, że (3) jest prawdziwe[2] .

Wypada jeszcze przywołać **Prawdmównego** (*Truthteller*) . Nie jest to co prawda odmiana paradoksu kłamcy, ale zdanie:

(4) To zdanie jest prawdziwe

gdzie zaimek „to”, tak jak w Kłamcy, odnosząc się do zdania, w którym występuje, zachowuje się bardzo ciekawie. Jak ustalić wartość prawdziwościową (4)? Jeśli założymy, że Prawdomówny jest prawdziwy, czyli że faktycznie jest tak, jak (4) głosi, to okaże się, że Prawdomówny rzeczywiście jest prawdziwy. Jeśli zaś przyjmiemy, że jest fałszywy, to okaże się, że rzeczywiście jest fałszywy. Można się spierać, czy „zachowanie” (4) nazywać paradoksalnym (por. Barwise,

[2]Mianem Wzmocnionego Kłamcy określana bywa też inna postać paradoksu Eubulidesa. Odróżnia się czasem dwa sformułowania: „To zdanie jest fałszywe” oraz „To zdanie nie jest prawdziwe”; drugie z przytoczonych zdań nazywane bywa Wzmocnionym Kłamcą. Argumentuje się, że tak rozumiany Wzmocniony Kłamca pozostaje paradoksalny nawet wtedy, gdy nie przyjmiemy zasady dwuwartościowości (por. Martin [1984], s. 2).

Etchemedy [1987]). Tak czy inaczej w Prawdomównym jest coś zastanawiającego. Cóż bowiem oznaczają użyte przed chwila słowa „faktycznie" i „rzeczywiście"? Odsyłają nas do faktów, do świata. Jednak w przypadku Prawdomównego odesłanie to nie udaje się[3] .

Czy wyliczone powyżej różne odmiany paradoksu kłamcy mają jakiekolwiek filozoficzne znaczenie? Trzeba w tym kontekście zwrócić uwagę na dwa zagadnienia:

(a) Czy należy uważać paradoks kłamcy i – szerzej — paradoksy semantyczne za autentyczne problemy filozoficzne?

(b) Czy podejmując namysł nad paradoksem kłamcy potrzeba zwracać uwagę na wszystkie jego wersje?

Odpowiedź na drugą z przytoczonych wątpliwości wydaje się prosta: skoro trzeba poszukiwać jakiegoś rozwiązania paradoksu kłamcy, to takiego, które poradzi sobie ze wszystkimi jego mutacjami. Jednak wątpliwość (b) można podnieść dopiero wtedy, gdy odpowiemy pozytywnie na pytanie (a) — a o odpowiedź tu niełatwo i to nie tylko ze względu na trudność w zdefiniowaniu „autentycznego problemu". Wielu wybitnych filozofów było zdania, że poważne zajmowanie się konsekwencjami paradoksu kłamcy to wielkie nieporozumienie. Czytamy w Wittgensteinowskich „Uwagach o podstawach matematyki":

> „Cóż szkodzi sprzeczność powstająca wtedy, gdy ktoś mówi: „Kłamię — A więc nie kłamię — A więc kłamię — itd."? Chodzi mi o to: czy nasz język staje się mniej użyteczny przez to, że można w tym wypadku z pewnego zdania wywnioskować wedle zwykłych reguł jego przeciwieństwo, a z tego znów tamto pierwsze zdanie? – Bezużyteczne jest *samo* to zdanie, podobnie jak owo wnioskowanie. Dlaczego jednak nie mielibyśmy go przeprowadzać? – Jest to jałowa sztuka!" (Wittgenstein [1956], s. 95).

Wittgenstein, tak jak i np. nie przepadający za późną filozofią autora „Dociekań" Popper, twierdzi, że sprzeczność, którą pociąga za

[3]Przedstawione paradoksy nie wyczerpują listy paradoksów semantycznych, na której znaleźć można także np. paradoks Grellinga, Berry'ego, czy Löba (por. Gupata, Belnap [1993], Barwise, Etchemedy [1987]).

sobą paradoks kłamcy, nie jest „niebezpieczna", że „żaden most się jeszcze przez nią nie zawalił" (por. Hodges [1998]), a zatem paradoksem tym nie trzeba się przejmować.

W XX wieku padały jednak słowa zgoła przeciwne. Alfred Tarski zauważa:

> „Uważam, że z punktu widzenia nauki byłoby rzeczą błędną i niebezpieczną, gdybyśmy deprecjonowali znaczenie tej i innych antynomii, traktując je jako żarty lub sofizmaty. Jest faktem, że mamy przed sobą absurd, że zostaliśmy zmuszeni do uznania fałszywego zdania [...]. Z tym faktem nie możemy się pogodzić, jeśli swoją pracę traktujemy poważnie. Musimy znaleźć jego przyczynę, tzn. musimy przeanalizować przesłanki, na których opiera się ta antynomia; musimy następnie odrzucić przynajmniej jedną z tych przesłanek i zbadać, jakie są tego konsekwencje dla całej dziedziny naszych badań" (Tarski [1944], s. 241).

W komentarzu polskiego logika rzuca się w oczy nie tyle niezadowolenie z faktu pojawienia się sprzeczności, ale raczej niepokój związany z nieznajomością jej przyczyn. Podobne obawy J. Barwise i J. Etchemedy formułują w sposób następujący:

> „Doniosłość paradoksu nie ogranicza się do niego samego, ale do tego, czego paradoks jest symptomem. Paradoks pokazuje bowiem, że nasze rozumienie pewnych podstawowych pojęć lub związków pojęć jest istotnie wadliwe, że w pewnych sytuacjach pojęcia te się załamują. I pomimo, że owe sytuacje mogą się nam wydać dziwaczne czy wręcz zabawne, sama wadliwość jest cechą pojęcia, a nie sytuacji, które ją ukazują. Jeśli zaś analizowane pojęcia są dla nas istotne, to nie ma tu miejsca na śmiech" (Barwise, Etchemedy [1987], s. 4).

Zatem zdaniem Tarskiego, Barwise'a i Etchemedy'ego obok paradoksów semantycznych nie można przejść obojętnie — konieczne staje się tu poszukiwanie jakiegoś rozwiązania.

Nie zamierzam w tym miejscu rozstrzygać, które z przywołanych dwóch stanowisk jest lepiej uzasadnione: zwolenników czy przeciwników „poważnego traktowania paradoksów”. Z całą pewnością sprzeczność jest niedopuszczalna w językach sformalizowanych. Jednak w przypadku języka naturalnego destrukcyjność paradoksu kłamcy nie jest tak oczywista. Nie oznacza to wszakże, iż opinia wyrażona przez Tarskiego jest błędna. Wydaje się, że poszukiwanie wolnych od sprzeczności teorii prawdy — i to nie tylko dla języków sformalizowanych — może jedynie pomóc nam w zrozumieniu, czym jest prawda i czym jest język, nawet jeśli nie jest to sposób działania lepiej uzasadniony niż Wittgensteinowskie lekceważenie Kłamcy.

Poniżej przyjrzymy się trzem próbom rozwiązania paradoksu kłamcy, które opierają się na poważnym potraktowaniu wątpliwości Tarskiego, Barwise'a i Etchemedy'ego. Wspólną cechą tych rozwiązań jest to, że rekonstruują one teorię prawdy w taki sposób, aby uniknąć negatywnych konsekwencji paradoksu. Ale uniknięcie sprzeczności, a więc formalna poprawność nie może stanowić jedynego kryterium adekwatności takiej rekonstrukcji — trzeba też pamiętać o tym, co Tarski nazywa merytoryczną trafnością proponowanego rozwiązania.

Warunek merytorycznej trafności teorii prawdy formułuje polski logik w postaci tzw. Konwencji T (*Vr* oznacza tu zbiór zdań prawdziwych, a *S* to zbiór zdań sensownych):

> „Poprawną formalnie definicję symbolu „ $x \in Vr$ ”, sformułowaną w terminach metajęzyka, nazywać będziemy trafną definicją prawdy, o ile pociąga ona za sobą następujące konsekwencje:
>
> (a) wszystkie zdania dające się uzyskać z wyrażenia „ $x \in Vr$ *wtw* p ” przez zastąpienie symbolu „ x ” nazwą strukturalno-opisową dowolnego zdania rozważanego języka, zaś symbolu „ p ” — wyrażeniem, stanowiącym przekład tego zdania na metajęzyk,
>
> (b) zdanie „dla dowolnego x — jeśli $x \in Vr$ to $x \in S$ ” (lub, innymi słowy, „ Vr zawiera się w S ”) (Tarski [1933], s. 61).

Rodzi się pytanie, czy Konwencja T jest warunkiem wystarczającym merytorycznie trafnej teorii prawdy? Tarski jest przekonany, że w istocie tak jest. Musimy jednak pamiętać, że poszukiwana teoria ma być teorią prawdy dla języka naturalnego. Merytoryczna trafność przeprowadzanej rekonstrukcji zależy więc nie tylko od spełnienia Konwencji T, ale także od zgodości proponowanego rozwiązania z pewnymi intuicjami, *których zbiór nazwiemy I*. Do zbioru tego należą następujące postulaty:

– język jest *semantycznie zamknięty*, tj. zawiera wszelkie pojęcia semantyczne, które go opisują,

– język jest *ekspresywnie kompletny*, tj. zawiera wszystkie możliwe dla danej liczby wartości logicznych funkcje prawdziwościowe,

– Kłamca[4] jest wyrażeniem sensownym,

– Kłamca jest wyrażeniem paradoksalnym[5] .

Dwa pierwsze postulaty, zamkniętość języka i jego ekspresywna kompletność, są pochodną pewnej ogólniejszej tezy podkreślonej przez Tarskiego, zgodnie z którą język naturalny jest językiem uniwersalnym. Przekonanie, że Kłamca jest wyrażeniem sensownym, wynika z konstatacji, że jest to zdanie poprawne gramatycznie (dotyczy to również innych wersji paradoksu kłamcy, a w sposób oczywisty Kół Kłamcy). Ostatni postulat dotyczy paradoksalności Kłamcy. Przez paradoksalność rozumiem tu niestabilność semantyczną tego wyrażenia (przy kolejnych próbach oceny jego wartość prawdziwościowa „oscyluje”).

Nie zamierzam jakoś szczególnie bronić się przed zarzutem, że *zbiór I* został skonstruowany nieco arbitralnie. Wydaje mi się, że niezależnie od możliwych wątpliwości jest to dobry punkt wyjścia do dyskusji nad rekonstrukcją teorii prawdy dla języka naturalnego. Jak można by podsumować ów *point of departure* ? Przywołaliśmy trzy warunki, które uwzględniać powinna dobra teoria prawdy dla języka naturalnego: formalną poprawność, Konwencję T oraz postulaty ze zbioru I. Tarski w swym „Pojęciu prawdy” pokazał, że kumulatywne spełnienie tych warunków jest niemożliwe; w szczególności przyjęcie Konwencji T oraz postulatów zamkniętości języka i sensowności

[4]Piszę tu „Kłamca”, ale mam na myśli wszelkie odmiany paradoksu kłamcy.

[5]Pamiętać należy też o tym, że wspólnym założeniem analizowanych teorii jest teza, iż nośnikami prawdy są zdania.

Kłamcy prowadzi do sprzeczności, a zatem do naruszenia warunku formalnej poprawności przeprowadzanej rekonstrukcji. Zatem budowanie teorii prawdy dla języka naturalnego odbywać się musi w „polu sił" wyznaczonym przez trzy wymienione:

FORMALNA POPRAWNOŚĆ

↗↙ ↖↘

KONWENCJA T ↔ ZBIÓR I

Zauważmy, że decydujące znaczenie ma tu uznanie, że Kłamca, jak inne paradoksy, jest poprawnym, sensownym wyrażeniem. Rodzi to pokusę, by pozbyć się całego problemu, uznając sprawcę całego zamieszania za wyrażenie bezsensowne. Takie postawienie sprawy jest jednak wątpliwe, jeśli jedyną przesłanką dla tak daleko idącej decyzji miałoby być pojawienie się sprzeczności. Przypominałoby to bardzo Wittgensteinowskie „nie przejmowanie się" paradoksami. Można odrzucić tezę o sensowności Kłamcy, ale tylko wtedy, gdy odpowiednio wykaże się jego bezsensowność.

3. TARSKI

Przedstawienie rozwiązania zarysowanego przed chwilą problemu, o jakie pokusił się Alfred Tarski, trzeba zacząć od wyjaśnienia ważnej wątpliwości. Jak powyżej wspomniano, Tarski doszedł do przekonania, że prawda dla języka naturalnego jest niedefiniowalna, głównie w związku z faktem, że język ten jest uniwersalny, a zatem i zamknięty. Pozytywne rozwiązanie Tarskiego — semantyczna definicja prawdy — znajduje zastosowanie tylko do pewnej klasy języków sformalizowanych. Jednak kilka wypowiedzi polskiego logika sugeruje, że sytuacja nie jest tak jasna, jak by się mogło wydawać na pierwszy rzut oka. W jednej ze swych późniejszych prac Taski pisze:

> Trzeba podkreślić, że używając terminu „języki sformalizowane" nie mamy na myśli wyłącznie systemów językowych sformułowanych całkowicie za pomocą symboli i nie mamy też na myśli niczego, co byłoby zasadniczo przeciwstawne językom naturalnym. Przeciwnie, tylko te języki sformalizowane wydają mi się naprawdę interesujące, które są fragmentami ję-

> zyków naturalnych (fragmentami wyposażonymi w pełne słowniki i ścisłe reguły syntaktyczne) lub które dają się przynajmniej adekwatnie przetłumaczyć na języki naturalne (Tarski [1969], s. 314).

Można zatem, jak się wydaje, pokusić się o stwierdzenie, że procedura zaproponowana przez Tarskiego potrafi nam coś powiedzieć o prawdzie w języku naturalnym wziętym jako całość, a nie tylko w odniesieniu do jego fragmentów. Oczywiście, nie da się dla takiego języka zbudować, ze znanych nam już powodów, poprawnej definicji zdania prawdziwego. Można jednak traktować definicje zbudowane dla fragmentów języka potocznego jak Carnapowskie eksplikacje lub, szerzej, parafrazy pojęcia prawdy w języku naturalnym (por. Ajdukiewicz [1934], [1937], Woleński [1980], [1985], [1993]).

Pamiętając o powyższym zastrzeżeniu, przyjrzyjmy się teorii Tarskiego. Wprowadźmy najpierw, na potrzeby niniejszej analizy, pojęcie *modelu bazowego*. Przez model bazowy rozumieć będziemy model klasycznego języka pierwszego rzędu *J*. Wszystkie predykaty języka *J* są w pełni zdefiniowane, przy czym nie ma wśród nich predykatu „prawdziwy" (*T*) . Właśnie odpowiednie wprowadzenie tego predykatu, a więc zdefiniowanie prawdy dla *J* wyposażonego w model bazowy, jest zadaniem, którego realizacji przyjrzymy się poniżej.

Załóżmy teraz, że, oddając sprawiedliwość intuicjom, chcemy potraktować model bazowy jak model języka naturalnego. Język taki jest językiem zamkniętym, zatem musimy umieć w *J* mówić o tym choćby, które zdania *J* są prawdziwe. Należy więc uzupełnić model bazowy o interpretację predykatu *T* oraz kody wszystkich poprawnie zbudowanych zdań języka *J*. Uzupełnienie to ma związek z drugim, obok zamkniętości, założeniem naszej rekonstrukcji: prawda jest orzekana o zdaniach. Pośród poprawnie zbudowanych zdań języka naturalnego znajduje się Kłamca, zatem i jego kod znaleźć się musi w modelu bazowym. To powoduje oczywiście powstanie paradoksu kłamcy, który z kolei prowadzi do konstatacji, że język *J* jest sprzeczny.

Tarski, rezygnując z postulatu zamkniętości języka, zdefiniował prawdę dla języka *J* w metajęzyku *MJ*. Zabieg ten powoduje, że nie musimy nic dodawać do modelu bazowego. Potrzebujemy natomiast drugiego, bogatszego języka (*MJ*) , który zawiera kody wszystkich zdań *J*, predykat „prawdziwy-w- *J* ”, a ponadto da się w nim wyrazić to wszystko, co w *J*.

Co jednak, jeśli chcemy zdefiniować prawdę dla *MJ* ? Oczywiście musimy odwołać się do meta-metajęzyka. Mamy więc w teorii Tarskiego nieskończoną hierarchię języków J_1, J_2, ..., J_n, ..., takich że dla dowolnego języka J_i język J_{i+1} jest tożsamy z MJ_i. Z każdym językiem J_i związana jest definicja prawdy dla J_i, czyli pewien predykat P_i, należący oczywiście do MJ_i (J_{i+1}) i w nim zdefiniowany. Można więc też, w pewnym sensie, mówić o hierarchii predykatów „prawdziwy”, przy czym hierarchiczność jest tu „dziedziczona” od języka, w którym dany predykat może zostać zdefiniowany. Konsekwencją tej sytuacji jest fakt, że nie mamy jednego, uniwersalnego predykatu „prawdziwy”. Mamy jedynie „prawdziwy-w-J_1”, „prawdziwy-w-J_2” itd. To jeden z zarzutów, jakie S. Kripke formułuje przeciw koncepcji Tarskiego (por. Kripke [1975]). Próbując odpowiedzieć Kripkemu, możemy, za J. Woleńskim (por. Woleński [1993]), zwrócić uwagę na fakt, że każdy język *J* jest w pełni przekładalny na swój *MJ*. A więc, powiedzmy, w MJ_{10} można wyrazić wszystko to, co w pozostałych 10 językach znajdujących się niżej w hierarchii, a zbiór zdań prawdziwych zdefiniowany w MJ_{10} (dla J_{10}) będzie obejmował, gdy uwzględnimy proces przekładu, wszystkie zdania (lub ich tłumaczenia) należące do zbiorów zdań prawdziwych języków J_1 — J_9.

Analiza powyższa sugeruje, że w każdym praktycznym przypadku można ograniczyć się do dwóch języków: *J* i *MJ*. Przyjmując dowolny język za *J*, możemy skonstruować dla niego odpowiedni *MJ*, w którym możliwa jest definicja predykatu „zdanie-prawdziwe-w- *J* ”. Stopień języka w koncepcji Tarskiego jest więc zrelatywizowany „sytuacyjnie”, inaczej niż np. podział przedmiotów na typy w tzw. prostej teorii typów Russella (por. np. Woleński [1993]).

Wprowadzona przez Tarskiego hierarchia języków powoduje, że Kłamca, Wzmocniony Kłamca, Przygodny Kłamca i Prawdomówny okazują się być wyrażeniami bezsensownymi, bowiem mieszają poziomy językowe. Kłamca orzeka swą fałszywość, a przecież ów predykat „fałszywy" może pojawić się w języku o stopień wyższym niż język, w którym wypowiedziany jest Kłamca. Nieco gorzej rzecz się ma z Kołami Kłamcy — poszczególnym zdaniom wchodzącym w skład Kół po prostu nie sposób jednoznacznie przypisać poziomów językowych!

Problem przypisywania poziomów językowych to drugi — obok rozwarstwienia predykatu „prawdziwy" i kilku innych wątpliwości (por. Kripke [1975], s. 57-63) — zarzut, który Saul Kripke kieruje pod adresem Tarskiego w swym słynnym artykule z 1975 roku. Przyjmijmy, powiada Kripke, że Jones wypowiada następujące zdanie:

(5) Wszystkie wypowiedzi Nixona na temat afery Watergate są fałszywe.

Zdaniu (5) przypiszemy od razu poziom co najmniej o jeden wyższy niż najwyższy poziom jakiejkolwiek wypowiedzi Nixona o Watergate. Tymczasem Nixon mógł w tej sprawie powiedzieć m.in.:

(6) Wszystkie wypowiedzi Jones'a na temat afery Watergate są fałszywe.

Wypowiedź ta musi być przynajmniej o poziom wyżej niż wypowiedź Jonesa! Oczywiście w przykładzie tym możemy się bronić utrzymując, że (6) wcale nie jest wypowiedzią na temat afery Watergate.

Koncepcja Tarskiego jest formalnie poprawna, spełnia przy tym Konwencję T, nie jest jednak zgodna z postulatami zawartymi w I : język, dla którego Tarski buduje definicję prawdy, nie jest zamknięty, a Kłamca nie jest uważany za wyrażenie sensowne. Jeśli dodamy do tego przywołane powyżej wątpliwości Kripkego, jasnym stanie się, że propozycja polskiego logika nie może być niczym więcej, jak tylko niedoskonałą parafrazą teorii prawdy dla języka naturalnego, z czego Tarski oczywiście całkowicie zdawał sobie sprawę. Rodzi się w tej sytuacji pytanie, czy teoria prawdy dla języka naturalnego — lub choćby parafraza lepsza od koncepcji Tarskiego — w ogóle da się skonstruować?

4. KRIPKE

W tytule niniejszego rozdziału znalazło się nazwisko Saula Kripkego, ale obok niego można było wspomnieć też L.M. Martina i P.W. Woodruffa. Tak się bowiem złożyło, że w 1975 roku Kripke oraz Martin i Woodruff opublikowali niezależnie dwa artykuły (Kripke [1975], Martin, Woodruff [1975]), które prezentowały nowe podejście do definiowania prawdy i rozwiązywania paradoksów semantycznych. Teorie takie jak te, które wspomniani autorzy zaproponowali w 1975 roku, nazywane są teoriami *stałopunktowymi*.

Jednym z głównych zarzutów Kripkego wobec koncepcji Tarskiego było spostrzeżenie, że teoria zaproponowana przez polskiego logika prowadzi do rozwarstwienia predykatu „prawdziwy", co jest sprzeczne ze sposobem, w jaki predykatu tego używamy w języku potocznym. Aby uniknąć tej niefortunnej konsekwencji, trzeba by zbudować język zawierający własny predykat prawdziwy. Kripke to właśnie robi, ale za cenę rezygnacji z logiki dwuwartościowej[6] . Procedura zastosowana przez Amerykanina sprowadza się do skonstruowania hierarchii języków; każdy kolejny język ma pełniej zdefiniowany predykat „prawdziwy". W którymś momencie, w tzw. stałym punkcie, proces „definiowania" rzeczonego predykatu się kończy, tj. kolejne języki nie dodają nic do jego interpretacji. O tych językach można powiedzieć, że zawierają swój własny predykat „prawdziwy". Hierarchia języków nie prowadzi więc tu, jak u Tarskiego, do „rozwarstwienia prawdy". Interesująca i potrzebna jest nam tylko ta ekstensja predykatu „prawdziwy", którą osiąga on w stałym punkcie. Cały proces zaś, jak utrzymuje Kripke, można porównać do sposobu, w jaki uczymy się używać słowa „prawda". Zacytujmy dłuższy fragment z „Outline of a Theory of Truth":

> „Chcemy wychwycić intuicję mniej więcej takiego rodzaju.
> Załóżmy, że wyjaśniamy użycie słowa „prawdziwy" komuś, kto

[6]Stwierdzenie to nie jest do końca precyzyjne. Można bowiem zastosować strategie „stałych punktów" dla nieco osłabionej logiki dwuwartościowej. Z całą pewnością nie jest to jednak logika klasyczna.

> słowa tego nie rozumie. Możemy mu powiedzieć, że jesteśmy uprawnieni do uznania (lub odrzucenia) prawdziwości jakiegoś zdania dokładnie w tych samych okolicznościach, w których możemy uznać (lub odrzucić) samo to zdanie. Nasz rozmówca będzie wtedy rozumiał, co to znaczy przypisać prawdę np. zdaniu „Śnieg jest biały”, ale ciągle nie będzie sobie umiał poradzić z przypisaniem prawdy zdaniom, które same zawierają słowo „prawdziwy”.
>
> Z czasem jednak pojęcie prawdy, nawet stosowane do różnych zdań zawierających słowo „prawdziwy”, stopniowo stawać się będzie jaśniejsze. [...] W ten sposób nasz bohater będzie umiał przypisać prawdziwość coraz większej liczbie zdań zawierających pojęcie prawdy. Nie ma powodu, by przypuszczać, że *wszystkie* zdania zostaną na tej drodze ocenione, ale *większość* z nich — na pewno” (Kripke [1975], s. 65-66).

Przyjrzyjmy się bliżej konstrukcji Kripkego. Rozważmy język pierwszego rzędu L z modelem M zawierającym kody wszystkich zdań L i zbiorem w pełni zdefiniowanych predykatów **R**, wśród których, rzecz jasna, nie ma predykatu „prawdziwy”. Rozszerzmy teraz język L do języka L_1 poprzez dodanie częściowo zdefiniowanego predykatu $T(S_{1\_1}, S_{1\_2})$, gdzie S_{1_1} jest ekstensją, a S_{1_2} antyekstensją T (Kripke stosuje więc logikę trójwartościową; posługuje się tzw. mocnym schematem Kleenego).

Zdefiniujmy teraz relację $L_x \models f$, gdzie $x \in N$:

F	$L_x\| = f$	$L_x\| =\sim f$
$R(d_1, ...d_m)$	$(d_1, ...d_m) \in R$	$(d_1, ...d_m) \notin R$
$g_1 \wedge g_2$	$L_x\| = g_1$ i $L_x\| = g_2$	$L_x\| = (\sim g_1) \vee (\sim g_2)$
$g_1 \vee g_2$	$L_x\| = g_1$ lub $L_x\| = g_2$	$L_x\| = (\sim g_1) \wedge (\sim g_2)$
$\forall x(g)$	$L_x\| = g(x = d)$ dla wszystkich $d \in D$	$L_x\| = \exists x(\sim g)$
$\exists x(g)$	$L_x\| = g(x = d)$ dla jakiegoś $d \in D$	$L_x\| = \forall x(\sim g)$
$\sim\sim g$	$L_x\| = g$	$L_x\| =\sim g$
$T(d)$	$d \in S_{x\_1}$	$d \in S_{x\_2}$

gdzie: $R \in \mathbf{R}$, d, d_1 ,...,d_m $\in D$ (uniwersum modelu M) , a g, g_1, g_2 są formułami języka L. Teraz możemy zdefiniować, za Kripkem, hierarchię języków, poprzez zdefiniowanie hierarchii par zbiorów (S_{x_1}, S_{x_2}) . W szczególnym przypadku, gdy $S_{1\_1} = \emptyset$ i $S_{1\_2} = \emptyset$, definicja tej hierarchii wygląda następująco:

(7)

$S_{1\_1} = \emptyset$; $S_{1\_2} = \emptyset$; $L_1 = L$ rozszerzony o $T(S_{1\_1}, S_{1\_2})$,

$S_{x+1\_1} = \{ f: f$ jest zdaniem L_x i $L_x \models f \}$

$S_{x+1\_2} = \{ f : f$ jest zdaniem L_x i $L_x \models \sim f \} \cup \{ d \in D : d$ nie jest zdaniem $L_x \}$

$L_{x+1} = L$ rozszerzony o $T(S_{x+1\_1}, S_{x+1\_2})$.

Konstrukcja ta stanowi monotoniczną funkcję, która ma stały punkt, tj. taką wartość x, dla której $S_{x\_1} = S_{x+1\_1}$ oraz $S_{x\_2} = S_{x+1\_2}$, a więc $L_x = L_{x+1}$. Biorąc pod uwagę fakt, że — intuicyjnie — język L_{x+1} zawiera T stanowiący predykat „prawdziwy" dla L_x, to skoro $L_x = L_{x+1}$, a zatem $T(S_{x\_1}, S_{x\_2}) = T(S_{x+1\_1}, S_{x+1\_2})$, to język L_x zawiera swój własny predykat „prawdziwy".

Całego procesu nie musi się zaczynać od założenia, że zbiory S_{1_1} i S_{1_2} są puste. Jednak nie wszystkie początkowe hipotezy sprawią, że procedura będzie mogła osiągnąć stały punkt. Np. jeśli założymy, że zdanie $d_k = \sim T(d_k)$ i $d_k \in S_{1\_1}$, a więc że Kłamca należy do ekstensji $T(S_{1\_1}, S_{1\_2})$, to w kolejnych krokach iteracji d_k będzie należeć raz do S_{x_1} (dla nieparzystych x) , innym razem do S_{x_2} (dla parzystych) i nigdy nie osiągniemy stałego punktu. Problemu tego nie ma przy Prawdomównym. Jeśli umieścimy go w S_{1_1}, to na każdym stopniu hierarchii należeć będzie do ekstensji prawdy; jeśli początkowo znajdzie się w S_{1_2}, to pozostanie w antyekstensji.

Stały punkt, który osiągamy zaczynając od $S_{1\_1} = \emptyset$ i $S_{1\_2} = \emptyset$, nazywamy najmniejszym punktem stałym (*least fixed point*) . Zdanie, które ma wartość logiczną w takim punkcie, a więc należy do ekstensji lub antyekstensji predykatu T, Kripke nazywa *przedmiotowym* (ugruntowanym — *grounded*) . Natomiast zdanie paradoksalne to takie, które nie ma wartości logicznej w żadnym punkcie stałym. Oto sposób, w jaki teoria Kripkego radzi sobie z Kłamcą: okazuje się on zdaniem sensownym, ale nie posiadającym wartości ani prawdy ani fałszu. Podobnie teoria ta radzi sobie z Przygodnym Kłamcą oraz z oboma rodzajami Kół Kłamcy. Status Prawdomównego zależy od tego, jaką pierwotnie przyjmiemy hipotezę odnośnie ekstensji i antyekstencji predykatu *T*. W konkretnym punkcie stałym Prawdomówny może być prawdziwy, fałszywy lub nie mieć wartości logicznej. Nigdy jednak, zgodnie z definicją Kripkego, nie nazwiemy go paradoksalnym.

Największy problem dla prezentowanej konstrukcji sprawia Wzmocniony Kłamca (por. Haack [1978]). Przypomnijmy, że brzmi on: „To zdanie jest fałszywe albo paradoksalne”. Problem polega na tym, że semantyczne pojęcie paradoksalności jest zrelatywizowane do wszystkich punktów stałych, nie może więc należeć do hierarchii budowanych przez Kripkego języków. Jedyny sposób na poradzenie sobie ze Wzmocnionym Kłamcą polega na przyznaniu, że predykat „paradoksalny” należy do metajęzyka. Otrzymujemy tym samym takie rozwiązanie dla Wzmocnionego Kłamcy, jakie Tarski zaproponował dla Kłamcy. „Ale to może powodować niesmak; wszak jest to nieco rozczarowujące, że nowatorstwo podejścia Kripkego do Kłamcy pociąga za sobą neo-Tarskowskie odrzucenie Wzmocnionego Kłamcy” (Haack [1978], s. 148).

Od połowy lat siedemdziesiątych pojawiło się wiele prac rekonstruujących teorię prawdy w paradygmacie zaproponowanym przez Kripkego, Martina i Woodruffa. Uzyskano wiele ciekawych rezultatów, a badania prowadzono z wykorzystaniem różnych schematów waluacyjnych, nie tylko mocnego schematu Kleenego. Wszystkie te prace borykają się z podobnymi problemami i można im przedstawić te same zarzuty.

Po pierwsze, strategia „stałych punktów” ucieka się do użycia logik wielowartościowych, co dla wielu autorów jest cechą na tyle niepożądaną, by strategię tę odrzucić. Na temat poprawności logik wielowartościowych wylano już tyle atramentu, a dyskusja jest tak daleka od konkluzji, że, nie wchodząc w rozbudowane argumentacje, oddajmy głos samemu Kripkemu:

„Byłem zadziwiony usłyszawszy, jak użycie przeze mnie waluacji Kleenego porównywano czasami do propozycji tych, którzy są za odrzuceniem logiki klasycznej w sferze «mechaniki kwantowej» [...]. Nie można powiedzieć, że «klasyczna logika» generalnie nie obowiązuje [...]. *Jeśli* pewne zdania wyrażają sądy (tj. są ugruntowane — dopisek B.B.), dowolna tautologiczna funkcja prawdziwościowa mająca je za argumenty wyraża prawdziwy sąd [...]. Konwencje co do sposobu traktowania zdań nie wyrażających sądów nie stanowią w żadnym filozoficznie znaczącym sensie «zmian w logice»" (Kripke [1975], s. 64-65).

Drugi zarzut podnoszony przez krytyków Kripkego *et al.* związany jest z interesującą cechą rozwiązań stałopunktowych. Otóż okazuje się, że dla danego modelu bazowego (por. definicja w §3) istnieje zwykle więcej niż jeden stały punkt, a dodać można, że konstruuje się takie modele, dla których liczba stałych punktów jest równa $\aleph_0$, czy nawet takie, których zbiór stałych punktów ma moc kontinuum (por. Belnap, Gupta [1993]). Ta okoliczność powoduje, że konstruując teorię prawdy trybem stałopunktowym musimy uznać któryś z możliwych stałych punktów za właściwą interpretację prawdy. Problem w tym, że taki wybór zawsze można oskarżyć o arbitralność.

Aby przedstawić trzeci zarzut wobec analizowanej teorii zdefiniujmy:

(8) Schemat waluacyjny p ma własność stałego punktu wtedy i tylko wtedy, gdy dla wszystkich języków L z jednoargumentowym predykatem G i dla wszystkich modeli podstawowych M języka L istnieje stały punkt przy waluacji p.

Okazuje się, że tylko takie schematy waluacyjne mają własność stałego punktu, które, z punktu widzenia tego, co można wyrazić choćby w języku naturalnym, są niekompletne. Np. jeśli dany schemat trójwartościowy zawiera równoważność Łukasiewicza, to nie ma on własności stałego punktu.

„Mówiąc ogólniej, dowolny n -wartościowy schemat waluacyjny, który ma własność stałego punktu, musi być logicznie

> niekompletny w tym sensie, że niektóre n -wartościowe funkcje prawdziwościowe nie są w nim wyrażalne. Z tego punktu widzenia fakt, że schemat klasyczny nie ma własności stałego punktu, można wyjaśnić jego logicznym bogactwem, tj. przez to, że wszystkie dwuwartościowe funkcje prawdziwościowe są w nim wyrażalne" (Gupta, Belnap [1993], s. 94).

Zatem wszystkie języki zawierające własny predykat „prawdziwy", zbudowane za pomocą metody Kripkego muszą być „ekspresywnie niekompletne".

Spójrzmy na koniec naszej prezentacji teorii stałopunktowej na to, jak koncepcja ta ma się do warunków adekwatności teorii prawdy, które wyróżniliśly w §2. Rozwiązanie Kripkego jest formalnie poprawne, choć pamiętać należy, że jest to poprawność w rachunku wielowartościowym. Zmiana logiki pociąga też modyfikację Konwencji T. Nie ma w niej już odróżnienia języka od metajęzyka, natomiast trzeba uwzględnić trójwartościowość: merytorycznie trafna teoria prawdy powinna pociągać zdania w formie ' x jest prawdziwe' albo ' x jest fałszywe' wtedy i tylko wtedy, gdy odpowiednio A jest prawdziwe albo A jest fałszywe (por. definicję w §2).

Język stałopunktowy jest zamknięty, jeżeli chodzi o pojęcie prawdy, ale już np. definicja pojęcia paradoksalności zmusza do wykorzystania pewnej hierarchii języków. Rozwiązanie Kripkego prowadzi przy tym do języka ekspresywnie niekompletnego. Zatem spośród dwóch pierwszych postulatów zbioru I tylko postulat zamkniętości języka jest częściowo spełniony.

Postulat trzeci z I jest spełniony (Kłamca jest wyrażeniem sensownym), natomiast czwarty — nie:

> „Podstawową cechą Kłamcy wydaje się być jego niestabilność w semantycznej ewaluacji: nieważne, jaką hipotezę przyjmiemy co do jego wartości, ewaluacja semantyczna odrzuci tę hipotezę. Teoria prawdy powinna wychwycić *tę* intuicję. Powinna też opisać sposób odróżniania zdań wykazujących się takim niestabilnym zachowaniem od reszty zdań, i powinna wreszcie wyjaśnić, *dlaczego* niektóre zdania tak się zachowują" (Gupta, Belnap [1993], s. 100).

Teoria Kripkego stara się wypełnić jedynie drugie z wymienionych przed chwilą trzech zadań. Jak już wspominaliśmy, autor „Outline of a Theory of Truth" definiuje zdanie paradoksalne w sposób następujący: jest to zdanie, które nie ma wartości logicznej *w żadnym* punkcie stałym. Nie bez kozery podkreśliłem wyrażenie „w żadnym", jako że ujawnia ono dosyć dziwną cechę strategii „stałych punktów". Definicja zdania paradoksalnego kwantyfikuje wszystkie stałe punkty, a przecież tylko jeden z nich uznawany jest za interpretację prawdy!

Wszystkie wymienione zarzuty nie pozostawiają wątpliwości, że konstrukcje stałopunktowe pozostają jedynie parafrazami a nie pełnoprawnymi modelami języka naturalnego. Można się zastanawiać, czy jest to propozycja lepsza od teorii Tarskiego. Prosty „bilans intuicji" — zderzenie obu koncepcji z warunkami adekwatności z §2 — wskazywałby, że strategia stałych punktów ma przewagę. Jednak zastrzeżenia wobec niej są nader poważne.

5. GUPTA I BELNAP[7]

Koncepcja, którą zajmiemy się teraz, rozwijana była od początku lat osiemdziesiątych ubiegłego wieku przez Hansa Herzbergera ([1982]), Anila Guptę ([1982]) i Nuela Belnapa ([1982]), a swój ostateczny kształt znalazła w książce tych dwóch ostantnich zatytułowanej „The Revision Theory of Truth" (Gupta, Belnap [1993]).

Revision Theory wznosi się na dwóch filarach. Pierwszy stanowi sugestia Tarskiego wyrażona w „Semantycznej koncepcji prawdy i podstawach semantyki":

> „Możemy [...] powiedzieć, że każdą równoważność postaci (T)[8] uzyskaną przez zastąpienie ' *p* ' określonym zdaniem, a ' *X* '

[7]Niniejszy paragraf stanowi, przede wszystkim, ekspozycję idei Gupty i Belnapa zawartych w ich książce [1993]. Z wielu przyczyn zdecydowałem, że najlepiej będzie, gdy relacja ta będzie możliwie wierna. Można więc uważać dużą część niniejszego tekstu za streszczenie rozdziału IV pracy Gupty i Belnapa. Podane przeze mnie przykłady stanowią modyfikację przykładów amerykańskich logików.

[8]Chodzi oczywiście o T-równoważność: (T) *X* jest prawdziwe wtedy i tylko wtedy, gdy *p*.

> nazwą tego zdania, uważamy za cząstkową definicję prawdy, wyjaśniającą, na czym polega prawdziwość tego konkretnego zdania” (Tarski [1944] s. 236).

Drugim filarem *Revision Theory* jest ogólna teoria definicji. Gupta i Belnap rozwijają tę teorię w przekonaniu, że tradycyjne wymogi stawiane definicjom mogą zostać rozluźnione, co sprowadza się głównie do uznania, iż definicje koliste są logicznie dopuszczalne. Zaznaczyć wypada już w tym miejscu, że teoria definicji, której szkic poniżej przedstawimy, jest narzędziem, którego użyteczność nie wyczerpuje się w konstrukcji koncepcji prawdy.

Zacznijmy od obserwacji pewnego podobieństwa w zachowaniu definicji kolistych i paradoksów semantycznych. Najpierw wprowadźmy pojęcie reguł Eliminacji Definiendum (*ED*) i Wprowadzenia Definiendum (*WD*) . Jeśli mamy choćby następującą definicję słynnego Quine’owskiego słowa „ *gavagai* ”:

(9) x jest *gavagi* $=_{Df}$ x jest białe lub x jest jednocześnie niebezpieczne i nie- *gavagai*

to ze zdania

(10) x jest *gavagai*

możemy wywnioskować

(11) x jest białe lub x jest jednocześnie niebezpieczne i nie- *gavagai*

lub odwrotnie: z (11) możemy wywnioskować (10). W pierwszym przypadku zastosowana została reguła *ED*, w drugim — *WD*. Schematycznie reguły te można przedstawić w sposób następujący:

$ED\frac{Definiendum}{Definiens}$ $WD\frac{Definiens}{Definiendum}$

Podobne reguły związane są z użyciem predykatu prawdziwy. Nazwiemy je Wprowadzeniem Prawdy (*WP*) i Eliminacją Prawdy (*EP*):

$EP\frac{'A'\ jest\ prawdziwe}{A}$ $WP\frac{A}{'A'\ jest\ prawdziwe}$

EP i *WP* potraktować można jako reguły ustalania, czy dane zdanie *A* jest prawdziwe czy nie. Aby określić status ’ *A* jest prawdziwe’, zmuszeni jesteśmy ustalić status *A* ; by poznać status ’ *A* nie jest prawdziwe’, odsyłani jesteśmy do nie- *A*. W przypadku Kłamcy otrzymamy w ten sposób niekończący się łańcuch kolejnych „odesłań" (przez *K* oznaczmy Kłamcę):

K jest prawdziwe
↓
K nie jest prawdziwe
↓
K jest prawdziwe
↓
K nie jest prawdziwe
...

Podobnie można potraktować *ED* i *WD*. Jeśli przyjmiemy, że obiekt *a* jest jednocześnie niebezpieczny i nie-biały, problem ustalenia, czy zgodnie z definicją (9) *a* jest *gavagai*, zredukuje się do pytania, czy *a* nie jest *gavagai*, co z kolei z powrotem odeśle do pytania, czy *a* jest *gavagai* itd.:

a jest *gavagai*
↓
a nie jest *gavagai*
↓
a jest *gavagai*
↓
a nie jest *gavagai*
...

Zachowanie definicji (9), przy poczynionym założeniu, jest więc analogiczne do zachowania Kłamcy. Można też łatwo stworzyć taką definicję, która zachowywać się będzie jak Prawdomówny. Wystarczy choćby, przyjmując nadal, że *a* jest jednocześnie niebezpieczne i nie-białe, zmienić definicję (9) w taki sposób, by drugi człon alternatywy stanowiącej *definiens* brzmiał: „ *x* jest niebezpieczne i *gavagai* ". Oczywistym jest, że odpowiedni schemat dla tak zmodyfikowanego (9) będzie wyglądał tak, jak schemat dla Prawdomównego.

Można budować takie definicje, które łączą w sobie cechy zdania nieparadoksalnego, Kłamcy i Prawdomównego. Przyjrzyjmy się następującemu przykładowi:

(12) *x* jest *supergavagai* = Df albo (*x* jest białe i niebezpieczne) albo (*x* jest białe i nie-niebezpieczne i *supergavagai*) albo (*x* jest nie-białe i niebezpieczne i nie- *supegavagai*) .

Jeśli *x* jest białe i niebezpieczne, to możemy orzec, że jest też *supergavagai*. Jeśli *x* jest białe i nie-niebezpieczne, to mamy problem podobny do Kłamcy, jeśli zaś *x* jest nie-białe i niebezpieczne, to otrzymujemy definicyjną wersję Prawdomównego. Można więc powiedzieć, że „każdy rodzaj patologicznego zachowania, które wykazuje pojęcie prawdy, można odwzorować w pojęciach uwikłanych w koliste definicje" (Gupta, Belnap [1993] s. 116-117).

Opracowanie teorii definicji kolistych wymaga, zdaniem Gupty i Belnapa, odpowiedzi na dwa pytania. Definicja ustala znaczenie definiowanego wyrażenia. Z tym związane jest pierwsze z kluczowych pytań: jak należy rozumieć owo znaczenie? W tradycyjnej, a więc unikającej definicji kolistych teorii sprawa jest prosta — znając ekstensje wszystkich predykatów wchodzących w skład *definiensa* jesteśmy w stanie określić ekstensję definiowanego terminu. Gdy w grę wchodzą definicje koliste, ten prosty mechanizm się zacina. Potrzebne jest stworzenie innego mechanizmu. Pierwsze pytanie jest zatem pytaniem o mechanizmy semantyczne rządzące kolistymi pojęciami. Drugie pytanie dotyczy, jak można się domyśleć, reguł logicznych rządzących definicjami. Trudno się spodziewać, by były to czyste reguły logiki klasycznej, skoro semantyka nie będzie klasyczna. Jaki jednak będzie stosunek logiki definicji kolistych do logiki klasycznej? Oto ciekawa kwestia, której poświęcimy poniżej nieco uwagi.

Zacznijmy od pytania pierwszego. „Definicja kolista, mimo że nie określa ekstensji *definiendum*, dostarcza pewnej reguły, która może być użyta do ustalenia („wyliczenia"), jaka powinna być ta ekstensja, przy określonej hipotezie dotyczącej ekstensji *definiendum*. [...] Nie możemy wskazać jakiegoś zbioru i powiedzieć, że to jest szukana ekstensja. Możemy tylko stwierdzić, jaka powinna być ta ekstensja, jeśli taki a taki inny zbiór został założony jako ekstensja" (Gupta, Belnap [1993] s. 119). Autorzy *Revision Theory* proponują więc, by uznać, iż znaczenie, które definicja przypisuje *definiendum*, ma *hipotetyczny* charakter. By zrozumieć dokładnie, o co chodzi, przyjrzyjmy się raz jeszcze definicji (12):

(12) x jest *supergavagai* = Df albo (x jest białe i niebezpieczne) albo (x jest białe i nie-niebezpieczne i *supergavagai*) albo (x jest nie-białe i niebezpieczne i nie- *supergavagai*) .

Załóżmy przy tym, że zachodzą następujące fakty (model M):

(13) Uniwersum D składa się z czterech obiektów **a,b,c,d**, a interpretacje poszczególnych predykatów i denotacje nazw przedstawiają się następująco:

– I (biały) = { **a,b**}

– I (niebezpieczny) = { **a,c**},

– I (a) = **a**,

– I (b) = **b**,

– I (c) = **c,**

– I (d) = **d**.

Nie sposób ustalić jednoznacznie ekstensji *supergavagai*, gdyż w *definiensie* (12) występuje definiowany predykat. Możemy jednak przyjąć np. następującą arbitralną hipotezę: I (*supergavagai*) = $\emptyset$. Teraz, korzystając z (12), możemy ustalić, jaka powinna być ekstensja *supergavagai* przy przyjętej hipotezie. „Podstawiając” pod każde wystąpienie predykatu *supergavagai* w *definiensie* (12) zbiór pusty, stwierdzamy, że przy założeniu, iż I (*supergavagai*) = $\emptyset$, definicja zmusza nas do stwierdzenia, że I (*supergavagai*) = { **a,c**}. Podobnie, jeśli założymy, że I (*supergavagai*) = { **a,b,c,d**}, to okaże się, że I (*supergavagai*) powinno być równe { **a,b**}.

Z naszą definicją (12) związana jest więc pewna funkcja $\delta_{D,M}$ (zrelatywizowana do definicji D i przyjętego modelu M) , której argumentami są hipotetyczne ekstensje *supergavagai*, a wartościami ekstensje *supergavagai* wyliczone z definicji w oparciu o przyjętą hipotezę. Funkcję związaną z (12) określają w pełni następujące reguły:

jeśli $\mathbf{b} \in X$ i $\mathbf{c} \in X$, to $\delta_{D,M}$ $(X) = \{\mathbf{a,b}\}$,

jeśli $\mathbf{b} \in X$ i $\mathbf{c} \notin X$, to $\delta_{D,M}$ $(X) = \{\mathbf{a,b,c}\}$,

jeśli $\mathbf{b} \notin X$ i $\mathbf{c} \in X$, to $\delta_{D,M}$ $(X) = \{\mathbf{a}\}$,

jeśli $\mathbf{b} \notin X$ i $\mathbf{c} \notin X$, to $\delta_{D,M}$ $(X) = \{\mathbf{a,c}\}$.

Jak łatwo zauważyć, **c** $\in X$ wtedy i tylko wtedy, gdy **c** $\notin \delta_{D,M}$ *(X)*, a zatem dla wszystkich argumentów X

$$\delta_{D,M}(X) \neq X.$$

Funkcja $\delta_{D,M}$ nie ma więc żadnego punktu stałego. Oznacza to tyle, że niezależnie, jakie hipotezy przyjmować będziemy, nie ustalimy ostatecznie, jaka jest ekstensja *supergavagai*. Nie jest to jednak równoważne temu, że nic nie wiemy. Mamy przecież funkcję $\delta_{D,M}$, która, jak utrzymują autorzy *Revision Theory*, zawiera podstawowe informacje semantyczne o *supergavagai*.

Mając funkcję $\delta_{D,M}$, nie mamy za dużo. Możemy jednak powiedzieć, że jeśli jakiś przedmiot z uniwersum znajdzie się w $\delta_{D,M}(X)$ dla każdego X, to uprawnia nas to do kategorycznego stwierdzenia, że przedmiot ten znajduje się w ekstensji definiowanego predykatu (w powyższym przykładzie rzecz tak ma się z **a**) . Rodzi się pytanie, jak precyzyjnie ustalić, na czym polega to przejście od stwierdzenia hipotetycznego do kategorycznego.

Aby odpowiedzieć na nie, należy, zdaniem Gupty i Belnapa, potraktować $\delta_{D,M}$ jak regułę rewizji (*rule of revision*) . Każdy zbiór otrzymany jako wartość $\delta_{D,M}$ traktować trzeba jak lepszego (a w każdym razie nie gorszego) kandydata na ekstensję definiowanego predykatu niż zbiór stanowiący argument funkcji. Z kolei uzyskaną ekstensję używamy jako nowy argument, otrzymując jako wartość kolejną, lepszą (nie gorszą) ekstensję. W ten sposób uzyskujemy ciąg zbiorów, które pozwalają na wypowiedzenie pewnych kategorycznych stwierdzeń. Zdefiniujmy rekurencyjnie taki ciąg rewizji:

$$\begin{array}{lll} 1. & \delta^0_{D,M}(X) & = X \\ 2. & \delta^{n+1}_{D,M}(X) & = \delta_{D,M}(\delta^n_{D,M}(X)) \end{array} \tag{14}$$

Zatem np. dla definicji (12) i hipotezy, że X=∅, ciąg rewizji wygląda następująco:

$$\begin{array}{ll} \delta^0_{D,M}(X) & = \emptyset \\ \delta^1_{D,M}(X) & = \{a, c\} \\ \delta^2_{D,M}(X) & = \{a\} \\ \delta^3_{D,M}(X) & = \{a, c\} \\ \delta^4_{D,M}(X) & = \{a\} \\ M & \end{array} \tag{15}$$

Zdefiniujmy:

(16)

(a) M+X to sytuacja M wraz z hipotezą, że ekstensją *supergavagai* jest X.

(b) Zdanie jest prawdziwe (fałszywe) w M+X wtw jest prawdziwe (fałszywe) w M przy hipotezie, że ekstensją *supergavagai* jest X (tak więc zdanie „ *a* jest *supergavagai* " jest fałszywe w M+∅, a prawdziwe w $M + \delta^n_{D,M}(\emptyset)$ dla wszystkich n>0).

(c) Zdanie jest *paradoksalne*, jeśli jego wartość logiczna oscyluje w ciągu rewizji wychodzącym od *każdej* hipotezy.

(d) Zdanie A jest ważne w M (ze względu na definicję D) wtedy i tylko wtedy, gdy istnieje taka liczba naturalna p, że dla wszystkich $q \geq p$ i dla wszystkich podzbiorów X uniwersum, A jest prawdziwe w $M + \delta^q_{D,M}(X)$ Powiemy, że A jest *kategoryczne w M*, wtedy i tylko wtedy, gdy albo A albo negacja A jest ważna w M ; A jest *patologiczne w M* wtedy i tylko wtedy, gdy nie jest kategoryczne w M.

Naszkicowany powyżej system semantyczny autorzy *Revision Theory* nazywają S_0. System ten, jak się okazuje, ma pewne słabości (por. Gupta, Belnap [1993], przykład 5A.17). Problemy te Gupta i Belnap rozwiązują tworząc systemy $S^\#$ i S^*, wykorzystujące pozaskończone aplikacje reguły rewizji. Ze względu na techniczną złożoność nie będziemy ich tu przedstawiać. Nie jest to zresztą konieczne, jako że podstawowe idee semantyczne, a te są dla nas ważne, są takie same dla wszystkich trzech systemów.

Drugie pytanie, na które odpowiedzieć muszą twórcy teorii definicji, dotyczy, jak pamiętamy, reguł logicznych rządzących definicjami. Rzecz jasna, chodzi nam tu o logikę:

(a) zupełną i odpowiednią dla zarysowanej przed chwilą semantyki,

(b) w której definicje koliste nie są kreatywne, jak to ma miejsce w przypadku stosowania zwykłych reguł logicznych.

Semantyka rewizyjna sugeruje, że, próbując zbudować logikę spełniającą warunki (a) i (b), szczególną uwagę poświęcić trzeba tym środkom logicznym, które pozwalają wnosić coś z *definiendum* o *definiensie* i odwrotnie. Innymi słowy, modyfikacji wymagają reguły *ED* i *WD*. Biorąc pod uwagę fakt, że — w badanej semantyce — jeśli *definiens* można orzec o przedmiocie x w i -tym etapie rewizji, to *definiendum* można orzec o x w etapie i +1, poszukiwane reguły wprowadzania i eliminacji definiendum, *ED r* i *WD r*, przedstawić można następująco:

$$ED_r \frac{Definiendum^i}{Definiens^{(i-1)}} \qquad WD_r \frac{Definiens^i}{Definiendum^{(i+1)}}$$

gdzie indeks górny i określa etap rewizji.

Poza tą modyfikacją stosowane są zwykłe reguły logiczne, z pewnym jednak zastrzeżeniem: stosowanie zwykłych reguł dla spójników boolowskich wymaga, by przesłanki i konkluzje rozumowania miały te same indeksy (wszak na poszczególnych etapach rewizji *definiendum* ma klasyczną ekstensję).

Gupta i Belnap wprowadzają jeszcze dodatkową regułę: *Zmiany Indeksu* (*index shift*) . Zgodnie z tą regułą, można dowolnie zmieniać indeks formuły w sytuacji, w której nie występuje w niej *definiendum*. Zdefiniowawszy ED_r, WD_r i Regułę Zmiany Indeksu, mamy już pełny rachunek logiczny C_0 dla tworzonej teorii definicji.

Dwie cechy C_0 warte są szczególnego podkreślenia. Po pierwsze, stosowanie indeksów jest konieczne tylko przy rozumowaniach hipotetycznych. W rozumowaniach kategorycznych można je pominąć. Po drugie, indeksy są zaniedbywalne zawsze, gdy mamy do czynienia z definicją niekolistą. Innymi słowy, we wszystkich tych kontekstach, w których definicje pozbawione są kolistości, można stosować do nich klasyczne reguły logiczne. System C_0 jest konieczny tylko, gdy dopuszczamy kolistość, choć podkreślić trzeba, że może być z powodzeniem stosowany także do definicji tradycyjnie uznawanych za dopuszczalne. Można zatem powiedzieć, że system ten jest bardziej ogólny niż logika klasyczna. Do kwestii tej powrócimy jeszcze w dalszym toku wywodów.

Znając S_0 i C_0, mamy narzędzie potrzebne do odtworzenia teorii prawdy zaproponowanej przez Guptę i Belnapa. Jak pamiętamy, logicy ci przyjęli sugestię Tarskiego, że T-równoważności stanowią cząstkowe definicje prawdy. Skoro tak, to może się zdarzyć, że *definiendum* (predykat „prawdziwy") pojawi się w *definiensie* takiej cząstkowej definicji, jak to ma miejsce choćby w T-równoważności dla Prawdomównego: „'To zdanie jest prawdziwe' jest prawdziwe wtw to zdanie jest prawdziwe". Tradycyjna teoria definicji uznaje takie definicje za niedopuszczalne, ale zarysowana powyżej teoria Belnapa i Gupty świetnie daje sobie z nimi radę.

Załóżmy, że mamy język, którego T-równoważności wyglądają następująco:

z_1 jest prawdziwe $=_{Df}$ A_1,

z_2 jest prawdziwe $=_{Df}$ A_2,

z_3 jest prawdziwe $=_{Df}$ A_3,

...

Definiens A_x może zawierać predykat „prawdziwy". T-równoważności nie ustalają więc od razu jednoznacznie ekstensji prawdy. Musimy zacząć od pewnej hipotezy tyczącej tej ekstensji. Hipoteza stanowi początek ciągu rewizyjnego, gdyż T-równoważności (cząstkowe definicje prawdy) dyktują nam, w „odpowiedzi" na przyjęta hipotezę, nową ekstensję predykatu „prawdziwy", po czym proces się powtarza. Zatem T-równoważności określają regułę rewizji τ_M, którą Gupta i Belnap nazywają *skokiem Tarskiego* (*Tarski jump*) . Zdefiniujmy τ_M:

(17) Niech M będzie całością relewantnych faktów, a U określonym zbiorem. Wtedy $s_i \in M(U)$ wtedy i tylko wtedy, gdy A_i zachodzi w M przy hipotezie, że ekstensją predykatu „prawdziwy" jest U, gdzie A_i to *definiens* w cząstkowej definicji zdania „ s_i jest prawdziwe". Innymi słowy τ_M jest regułą rewizji, określoną przez nieskończoną definicję kolistą:

(18) x jest prawdziwe $=_{Df}$ (x=s_1 i A_1) lub (x=s_2 i A_2) lub (x=s_3 i A_3) ...

Zauważmy, że konsekwencją akceptacji definicji kolistych jest pewna zmiana w odczytywaniu T-równoważności. Tradycyjnie traktowane są one jako równoważności materialne, tymczasem w *Revision Theory* T-równoważności muszą być równoważnościami definicyjnymi (stąd stosowane powyżej symbole $=_{Df}$). Oczywiście bez tej zmiany cały aparat rewizyjnej semantyki nie mógłby zostać zastosowany, a odpowiedź na pytanie, czy zmiana ta jest zasadna, jest zarazem odpowiedzią na pytanie o trafność całej *Revision Theory*.

Spójrzmy teraz na prosty przykład ukazujący sposób działania reguły rewizji τ_M. Weźmy język pierwszego rzędu L (stanowiący pewien fragment języka polskiego), który posiada następujące środki logiczne: *jest* (jako spójka i na oznaczenie identyczności), *nie, i, lub, jeśli... to..., każdy, istnieje taki że...* Załóżmy dalej, że dysponujemy następującym słownikiem pozalogicznym:

NAZWY: *Królowa Śniegu, Kai*, cudzysłowowe nazwy zdań języka L

PREDYKAT JEDNOARGUMENTOWY: ...*wredny*

PREDYKAT DWUARGUMENTOWY: ...*mówi że*...

Rozważmy sytuację M :

D ={ **Królowa śniegu, Kai**} ∪ zbiór zdań języka L

I (*Królowa Śniegu*) = **Królowa Śniegu**

I (*Kai*) = **Kai**

Nazwy cudzysłowowe mają odpowiednie interpretacje.

I (*wredny*) = { **Królowa Śniegu**}

I (*mówi że*) = {< **Kai**, *Królowa Śniegu jest wredna* >, < **Kai**, *'Królowa Śniegu jest wredna' jest prawdziwe* >, < **Kai**, *Coś, co Kai mówi jest prawdziwe* >}

Dla hipotezy, że ekstensja prawdy jest zbiorem pustym, zbiór $\tau_M(\emptyset)$ będzie miał m.in. następujące elementy: *Królowa Śniegu jest wredna; 'Królowa Śniegu jest wredna' nie jest prawdziwe; "Królowa śniegu jest wredna' jest prawdziwe' nie jest prawdziwe; Wszystko jest nieprawdziwe; Wszystko, co Kai mówi, nie jest prawdziwe.* Z kolei $\tau_M(D)$ zawierać będzie m.in.: *Królowa Śniegu jest wredna* ; *' Królowa Śniegu jest wredna' jest prawdziwe; 'Królowa Śniegu nie jest wredna' jest prawdziwe; Kai jest prawdziwy; Wszystko jest prawdziwe; Wszystko, co Kai mówi, jest prawdziwe*.

Można też, oczywiście, obliczyć wartości τ_M dla innych podzbiorów U uniwersum D. Zauważmy, że już pierwsze zastosowanie τ_M eliminuje arbitralność hipotez w odniesieniu do tych zdań, w których nie występuje predykat „prawdziwy", tak więc zarówno $\tau_M(\emptyset)$ jak i $\tau_M(D)$ zawierają zdanie *Królowa Śniegu jest wredna*. $\tau_M(\emptyset)$ i $\tau_M(D)$ nie zgadzają się co do zdania *'Królowa Śniegu jest wredna' jest prawdziwe*, jednak już po drugim etapie rewizji sytuacja ta zmienia się — dla każdego $U \subset D$, jeśli $n > 1$ *'Królowa Śniegu jest wredna' jest prawdziwe* $\in \tau^n_M(U)$. Z kolei wartość zdania *Wszystko, co Kai mówi, jest prawdziwe* stabilizuje się, niezależnie od hipotezy początkowej, po trzecim etapie rewizji. W powyższym przykładzie, w którym nie występuje samoodniesienie w rodzaju Kłamcy czy Prawdomównego, dla każdego zdania istnieje skończona liczba aplikacji reguły rewizji, po której status zdania jest ten sam niezależnie od przyjętej początkowo hipotezy. Powiemy, że funkcja τ_M wykazuje w analizowanym przypadku *usadowienie* (*settledness*) i *konwergencję* (*convergence*) .

(19) Reguła rewizji wykazuje *usadowienie*, jeśli w każdym ciągu rewizji status każdego zdania stabilizuje się na jednej wartości, tj. nic nie oscyluje.

(20) Reguła rewizji wykazuje *konwergencję*, jeśli status każdego zdania jest zawsze taki sam, niezależnie od hipotezy przyjętej na początku ciągu rewizji.

Wprowadźmy teraz do naszego przykładu Prawdomównego. Załóżmy, że jedyne zdanie, które Kai wypowiada, to zdanie *Coś, co Kai mówi, jest prawdą*. W tym przypadku τ_M wykazuje usadowienie, ale nie konwergencję. Zależnie od przyjętej hipotezy zdanie wypowiedziane przez Kaia ustabilizuje się (usadowi) w zbiorze $\tau_M(U)$ lub poza nim. Jeśli jedynym zdaniem wypowiedzianym przez naszego bohatera będzie *Coś, co Kai mówi, nie jest prawdziwe*, a więc do analizowanej sytuacji wprowadzimy Kłamcę, τ_M wykaże konwergencję, ale nie usadowienie (zdanie wypowiedziane przez Kaia będzie oscylować w ten sam sposób we wszystkich ciągach rewizji). Wreszcie w sytuacjach, w których występują oba rodzaje paradoksalności — Prawdomówny i Kłamca — nie będziemy mieli ani usadowienia ani konwergencji. Ciągi rewizyjne dla tego typu przypadku można jednak badać i wykazać, że lokalnie — w odniesieniu do zdań, które intuicyjnie określimy jako nieparadoksalne — występuje tak konwergencja jak i usadowienie. Pozwala to Gupcie i Belnapowi napisać:

> „Jeśli T-równoważności potraktujemy jak cząstkowe definicje prawdy, to (i) prawda okaże się pojęciem kolistym, (ii) jej interpretację stanowi reguła rewizji, która jest całkowicie określona przez T-równoważności, i wreszcie (iii) spora część zachowania pojęcia prawdy, tak zwykłego jak i patologicznego, może zostać wytłumaczona.” (Gupta, Belnap [1993] s. 137).

Przyjrzyjmy się teraz nieco uważniej, co *Revision Theory* ma do powiedzenia o paradoksach semantycznych. Oczywiście nie są one eliminowane — Gupta i Belnap uznają je za sensowne, pełnoprawne zdania i starają się opisać i wytłumaczyć ich zachowanie. Kłamca charakteryzuje się tym, że oscyluje we wszystkich ciągach rewizyjnych. Prawdomówny, podobnie jak Kłamca, wykazuje zachowanie patologiczne, tj. nie jest zdaniem kategorycznie stwierdzalnym. Jego zachowanie tym różni się od Kłamcy, że ten pierwszy „toleruje” usadowienie τ_M, a uniemożliwia konwergencję, ten drugi zaś — odwrotnie.

Przygodny Kłamca i obie wersje Kół mogą być analizowane w *Revision Theory* podobnie jak Kłamca i Prawdomówny. Pewien problem dla *Revision Theory* stanowi jednak następująca wersja Wzmocnionego Kłamcy:

(21) Albo to zdanie nie jest kategoryczne, albo nie jest prawdziwe.

Łatwo zauważyć, że jeśli (21) nie jest kategoryczne, to jest prawdziwe (bo pierwszy człon alternatywy jest prawdziwy). Ale jeśli jest prawdziwe, to musi być kategoryczne. Jeśli zaś jest kategoryczne, to zachowywać się będzie jak Kłamca (bo pierwszy człon alternatywy jest fałszywy), a więc trzeba go będzie uznać za niekategoryczne. Tak czy owak mamy sprzeczność. Nie można jednak zdania (21) uznać po prostu za paradoksalne. Użyte jest w nim bowiem pojęcie kategoryczności, które nabiera sensu dopiero wtedy, gdy mamy dane wszystkie ciągi rewizji i możemy je ze sobą porównać. Samo pojęcie kategoryczności uważane może być z tej perspektywy za koliste. Zdanie (21) będzie oscylować w ciągu rewizji, ale w ciągu określonym przez definicję kategoryczności a nie prawdy. Można więc powiedzieć, że (21) jest niekategoryczne (paradoksalne), ale użyte tu pojęcie kategoryczności jest pojęciem wyższego poziomu. O Kłamcy można powiedzieć, że jest paradoksalny 1, ale (21) jest już paradoksalne 2. Paradoksalność 1 badamy analizując przebieg ciągów rewizji dla pojęcia prawdy. Paradoksalność 2 — analizując ciągi rewizji dla paradoksalności 1. Oczywiście można konstruować Coraz-Bardziej-Wzmocnionych Kłamców, tworząc nieskończoną hierarchię predykatów „paradoksalny i ”.

Wobec teorii Belnapa i Gupty można podnieść kilka zarzutów. Po pierwsze, stosują oni nieklasyczny rachunek logiczny. Nie jest to jednak zarzut tak poważny, jak w przypadku strategii stałopunktowej. Wszędzie bowiem tam, gdzie w związku z brakiem kolistości można zaniedbać indeksy, *C 0* jest po prostu rachunkiem klasycznym. W obliczu kolistości trzeba jednak uwzględnić „dodatkowy parametr”, czego nie da się zrobić z logiką klasyczną. Można więc w pewnym sensie powiedzieć, że *C 0* stanowi swego rodzaju uogólnienie klasycznego rachunku pierwszego rzędu. *C 0* operuje bowiem na takiej semantyce, która jest całkowicie niedostępna rachunkowi klasycznemu. W szczególnych przypadkach semantyka ta jest równoważna z semantyką tradycyjną i wtedy *C 0* „działa” jak logika klasyczna.

Pod adresem *Revision Theory* wysuwa się jeszcze dwa zarzuty (por. Löwe, D., Welch [0000], Welch [2000], Orilia, Varzi [0000], Hammer [1996]). Po pierwsze, teoria ta wykorzystuje niezwykle wyrafinowane struktury matematyczne. Zbiory, które wchodzą w grę, osiągają skomplikowanie o poziomie lub wyższym. Odnosząc się do tego zarzutu, Gupta i Belnap zauważają, że podstawowe idee stojące za *Revision Theory* wcale nie są tak skomplikowane; co więcej, wydają się one dość intuicyjne. Złożoność, zdaniem twórców *Revision Theory*, związana jest z tym, że ich teoria uwzględnia pewne pozaskończone sytuacje, których na co dzień się nie spotyka.

Ostatni zarzut, o którym trzeba wspomnieć, dotyczy podziału zdań na kategoryczne i patologiczne. Wydaje się bowiem, że przyporządkowanie niektórych zdań do pierwszej z klas jest, jeśli nie kontrintuicyjne, to problematyczne. W szczególności zdanie:

(22) Kłamca jest prawdziwy lub Kłamca nie jest prawdziwy

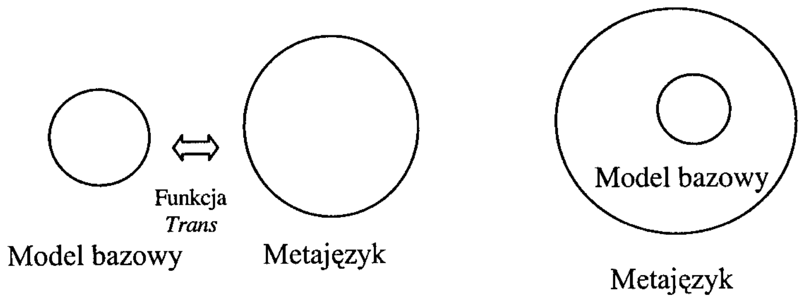
okazuje się być kategorycznie stwierdzalne. Gupta i Belnap argumentują, że jest to konsekwencją zachowania logiki dwuwartościowej. Konsekwencje takie jak (22) nie stanowią istotnej cechy samej *Revision Theory* a jedynie decyzji, by pracować z klasycznym językiem podstawowym. *Revision Theory* równie dobrze sprawdziłaby się, gdyby język podstawowy był np. wielowartościowy.

Podsumowując należy zwrócić uwagę na następujące fakty. Po pierwsze, teoria prawdy zaproponowana przez Belnapa i Guptę jest formalnie poprawna. Po drugie, pewnej modyfikacji ulega w tym podejściu Konwencja T. Autorzy *Revision Theory* zastępują ją Tezą o Sygnifikacji: T-równoważności ustalają sygnifikację prawdy w każdym świecie. Przez sygnifikację Gupta i Belnap rozumieją funkcję, która dostarcza pełnej informacji o interpretacji danego predykatu w świecie *w*. Trzeba też pamiętać o tym, że równoważność w T-równoważnościach jest definicyjna, a nie materialna.

Język, który budują Gupta i Belnap, jest „zamknięty na prawdę", jednak inne pojęcia semantyczne (np. paradoksalność) wymagają użycia hierarchii językowej. W przeciwieństwie do strategii stałopunktowej, rozwiązanie Gupty i Belnapa daje język ekspresywnie kompletny, a Kłamca jest tu nie tylko wyrażeniem sensownym, ale także zachowana zostaje intuicja jego „paradoksalności" (oscylacja wartości logicznej w ciągu rewizyjnym). Wydaje się więc, że spośród trzech opisanych koncepcji *Revision Theory* jest najbliżej tego, by nazwać ją adekwatną teorią prawdy dla języka naturalnego, choć można oczywiście utrzymywać, że wynika to z arbitralności zbioru *I*.

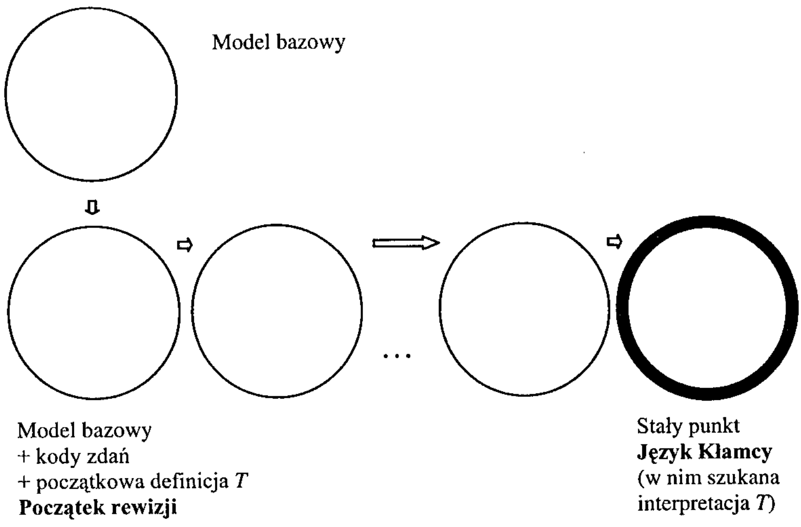
6. SEMANTYKI DLA KŁAMCY

W koncepcji Tarskiego prawda dla języka *J* definiowana jest w bogatszym metajęzyku. Potrzeba tu zatem, obok dwóch modeli semantycznych (bazowego i dla *MJ*) , znajomości pewnej funkcji (nazwijmy ją *Trans*) , umożliwiającej przekład wszystkich zdań *J* na *MJ*. Funkcja ta nie jest potrzebna tylko w tym szczególnym przypadku, w którym *J* jest częścią *MJ*. Można to zilustrować w sposób nastepujący:



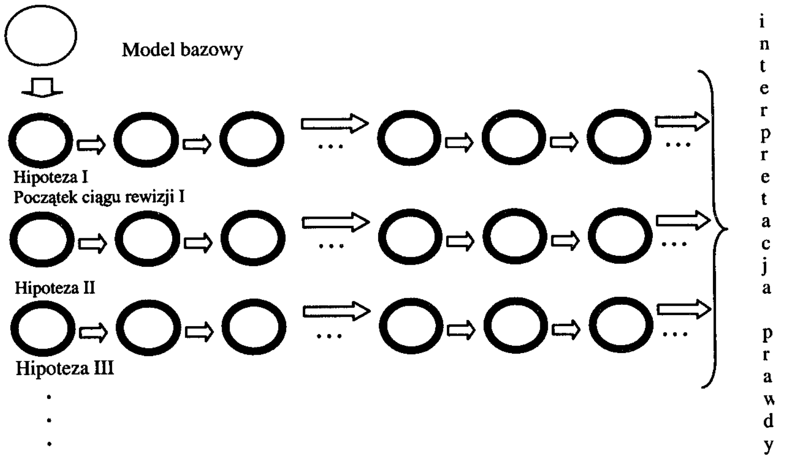
Rozwiązanie Tarskiego eliminuje Kłamcę jako zdanie bezsensowne, tak więc zarysowana powyżej semantyka jest „semantyką dla Kłamcy bez Kłamcy". Nieco inaczej rzecz się ma ze strategią „stałych punktów". Przypomnijmy, że w podejściu tym konstruuje się język, który zawiera własny predykat *T*. Czyni się to kosztem porzucenia dwuwartościowości — predykat *T* jest tylko częściowo zdefiniowany. Jak zatem zwolennicy strategii „stałych punktów" konstruują poszukiwany język, a więc definiują prawdziwość dla *J* i modelu bazowego? Zaczynają od tegoż modelu wzbogaconego o kody wszystkich zdań języka

J, oraz o częściowo (arbitralnie) zdefiniowany predykat *T* (można zacząć od sytuacji, w której ekstensja i antyekstensja *T* są zbiorami pustymi). Definiują następnie pewną funkcję, która w kolejnych etapach („rewizjach") rozszerza interpretację predykatu *T*. Funkcja ta ma stałe punkty, a zatem możliwe jest osiągnięcie takiej interpretacji (ekstensji i antyekstensji) *T*, która nie zmienia się w kolejnych „rewizjach". Język *J* z modelem bazowym, wzbogacony o predykat *T* zinterpretowany w sposób, w jaki to ma miejsce w stałym punkcie, jest językiem zawierającym własny predykat „prawdziwy".



Zauważmy, że cała informacja semantyczna o predykacie *T* znajduje się w stałym punkcie, który na powyższym schemacie zaznaczyliśmy pogrubioną kreską. Nie są więc dla nas interesujące te interpretacje *T*, które pojawiały się w trakcie dochodzenia do stałego punktu, jakkolwiek cała stałopunktowa konstrukcja, symulująca w jakiś sposób proces zdobywania wiedzy, może stanowić ciekawy filozoficznie argument za strategią Kripkego i jego kontynuatorów.

W *Revision Theory* wychodzi się poza klasycznie rozumianą semantykę. Informacja semantyczna na temat prawdy (w terminologii Gupty i Belnapa: sygnifikacja prawdy) jest według *Revision Theory* zawarta w zbiorze ciągów rewizji, same zaś ciągi rewizji są uporządkowanymi zbiorami pewnych (klasycznie rozumianych) modeli semantycznych. Można to zilustrować w sposób następujący:



Warto zaznaczyć, że w teorii przedstawionej przez Guptę i Belnapa tylko interpretacja predykatu „prawdziwy" jest nieklasyczna (na powyższym rysunku reprezentują ją wszystkie pogrubione okręgi). Interpretacje pozostałych predykatów, a więc i model bazowy, nie są modyfikowane. Nie znaczy to, że nie istnieją inne obok prawdy pojęcia, których interpretacja jest nieklasyczna. Rzecz ma się tak ze wszystkimi pojęciami kolistymi, w tym z semantycznymi — np. z „kategorycznością". Zmiana semantyki w odniesieniu do pojęć kolistych jest oczywista. Zamiast jednego — w przypadku logiki dwuwartościowej — lub kilku — przy logikach wielowartościowych — zbiorów indywiduów, do interpretacji kolistych predykatów potrzebne są bardziej złożone obiekty matematyczne. Tym samym pojawia się nowe rozumienie interpretacji, a zatem i języka zinterpretowanego.

Podsumowując przeprowadzone w niniejszym paragrafie rozważania, zauważyć trzeba, że „semantyka dla Kłamcy" wcale nie musi być rozumiana klasycznie. Można budować nowe semantyki w oparciu o klasyczną, jak to czynią Gupta i Belnap. Podejmuje się także próby wykorzystania do tego celu niestandardowej teorii mnogości (por. Barwise, Etchemedy [1987]). Wszystko to prowadzi do bardzo ciekawych rezultatów wzbogacających nasze rozumienie sposobu funkcjonowania języka.

7. ZAKOŃCZENIE

Trzy przedstawione koncepcje prawdy stanowią dowód na to, jak ważną rolę w budowie tego typu teorii odgrywają paradoksy semantyczne. Paradoks kłamcy stanowi swego rodzaju „papierek lakmusowy" konstruktora teorii prawdy, potrafi wskazać „słabe punkty" stworzonej konstrukcji; jego zachowanie powiedzieć może wiele o intuicyjnej trafności wybranego formalizmu. Warto jednak zaznaczyć, że taka — więcej niż heurystyczna — rola paradoksów w analizowanym kontekście związana jest z przyjmowaniem pewnych założeń rekonstrukcyjnych, które składają się na zbiór I ; w szczególności chodzi to o postulat zamkniętości języka.

Poważne potraktowanie paradoksu kłamcy stanowi zatem motyw poszukiwania nowych logicznych teorii prawdy. To ważne, by pamiętać, że obok „panującej" strategii Tarskiego istnieją też inne sposoby definiowania prawdy. Dużo racji jest bowiem w słowach Vanna McGee:

> „Typowe nieszczęście, jakie przytrafia się teorii filozoficznej, to odrzucenie i zapomnienie. Inne nieszczęście dotknęło Tarskiego koncepcję walki z antynomiami: została ona zbyt powszechnie przyjęta. Logicy biorą to za pewnik, że ich prawem i obowiązkiem w rozważaniach nad semantyką danego języka jest używanie istotnie bogatszego metajęzyka. Nie zauważają przy tym, że stosując taktykę Tarskiego związują się z potężną

filozoficzną doktryną o daleko idących filozoficznych konsekwencjach. [...] Doktryna Tarskiego została zaakceptowana tak powszechnie, że stała się niewidzialna" (McGee [1991], s. 147).

Bibliografia

Ajdukiewicz K.

[1931] *Paradoksy starożytnych*, [w:] Ajdukiewicz [1985], ss. 137-144.

[1934] *O stosowalności czystej logiki do zagadnień filozoficznych*, [w:] Ajdukiewicz [1985], ss. 211-214.

[1937] *Problemat transcendentalnego idealizmu w sformułowaniu semantycznym*, [w:] Ajdukiewicz [1985], ss. 264-277.

[1985] *Język i poznanie*, t. I, Państwowe Wydawnictwo Naukowe, Warszawa.

Barwise J., Etchemedy J.

[1987] *The Liar. An Essay on Truth and Circularity*, Oxford University Press, New York — Oxford.

Belnap N.

[1982] *Gupta's Rule of Revision Theory of Truth*, Journal of Philosophical Logic 11, s. 103-116.

Gupta A.

[1982] *Truth and Paradox*, Journal of Philosophical Logic 11, s. 1-60.

Gupta A., Belnap N.

[1993] *The Revision Theory of Truth*, A Bradford Book, The MIT Press, Cambridge (Massachusetts), London (England).

Haack S.

[1978] *Philosophy of Logics*, Cambridge University Press, Cambridge — New York — New Rochelle — Melbourne — Sydney.

Hale B., Wright C.

(red.) [1999] *A Companion to the Philosophy of Language*, Blackwell Publishers.

Hammer E.M.

[1996] *The Revision Theory of Truth*, [w:] Stanford Encyclopedia of Philosophy, edycja internetowa.

Herzberger H.G.

[1982] *Notes on Naive Semantics*, Journal of Philosophical Logic 11, s. 61-102.

Hodges A.

[1998] *Turing*, Amber, Warszawa.

Kotarbiński T.

[1985] *Wykłady z dziejów logiki*, Państwowe Wydawnictwo Naukowe, Warszawa.

Kripke S.

[1975] *Outline o f a Theory of Truth*, Journal of Philosophy 72, ss. 690-716; cytowane za Martin [1984], s. 53-82.

Löwe, D., Welch, P.D.

[0000] *Set-Theoretic Absoluteness and the Revision Theory of Truth*, edycja internetowa

Martin R.L.

(red.) [1984] *Recent Essays on Truth and the Liar Paradox*, Clarendon Press, Oxford University Press, Oxford — New York.

Martin R.L., Woodruff P.W.

[1975] *On Representing 'True-in-L' in L*, Philosophia 5, s. 213-217; cytowane za Martin [1984], s. 47-51.

McGee V.

[1991] *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*, Hackett.

Orilia F., Varzi A.C.

[0000] *Truth and Circular Definitions*, edycja internetowa.

Sainsbury R.M., Williamson T.

[1999] *Sorities*, [w:] Hale, Wright [1999], ss. 458-484.

Tarski A.

[1933] *Pojęcie prawdy w językach nauk dedukcyjnych*, [w:] Tarski [1995], ss. 13-172.

[1944] *Semantyczna koncepcja prawdy i podstawy semantyki*, [w:] Tarski [1995], ss. 228-282.

[1969] *Prawda i dowód*, [w:] Tarski [1995], ss. 292-332.

[1995] *Pisma logiczno-filozoficzne. Tom I: Prawda*, J. Zygmunt (red.), Wydawnictwo Naukowe PWN, Warszawa.

Welch P.D.

[2000] *On Gupta-Belnap Revision Theories of Truth, Kripkean fixed points, and the next stable set*, edycja internetowa.

Wittgenstein L.

[1956] *Uwagi o podstawach matematyki*, wydanie polskie: M. Poręba (tłum.), Wydawnictwo KR, Warszawa 2000.

Woleński J.

[1980] *Z zagadnień analitycznej filozofii prawa*, PWN, Kraków — Warszawa.

[1985] *Filozoficzna szkoła lwowsko-warszawska*, Państwowe Wydawnictwo Naukowe, Warszawa.

[1993] *Metamatematyka a epistemologia*, Wydawnictwo Naukowe PWN, Warszawa.