Definicje i teorie prawdy

Cezary Cieśliński

W ostatnich latach w literaturze filozoficznej obserwuje się wzmożone zainteresowanie koncepcjami prawdy. Pojawiają się nowe artykuły; padają coraz to nowe argumenty zwolenników i przeciwników danego stanowiska filozoficznego.[1] Tymczasem już w 1933 roku w pracy „Pojęcie prawdy w językach nauk dedukcyjnych" Alfred Tarski przedstawił słynną definicję pojęcia prawdy, która weszła do kanonu współczesnej logiki. Pojawia się pytanie o relację pomiędzy współczesnymi dyskusjami a klasycznymi rozwiązaniami, wypracowanymi w ramach dwudziestowiecznej logiki. Czy uczestnicy tych dyskusji uważają klasyczne podejście za błędne lub niezadowalające? Jeśli tak, to z jakich powodów? Celem artykułu będzie przedyskutowanie tej właśnie kwestii. Główna obserwacja będzie polegać na tym, że uczestnicy wielu (choć oczywiście nie wszystkich) współczesnych dyskusji na temat prawdy stawiają sobie inne cele teoretyczne niż Tarski w latach trzydziestych. Stąd bierze się różnica podejść: nie poszukują oni definicji prawdy, lecz zadowalającej teorii prawdy. W artykule zamierzam opisać dokładniej tę różnicę podejść i jej konsekwencje.

I. Definicje prawdy

Nie będę tu przedstawiał w szczegółach konstrukcji zawartej w klasycznej pracy Tarskiego. Ograniczę się jedynie do szkicowego zarysu oraz paru uwag, które będą dla nas istotne ze względu na omawiane tutaj zagadnienie.

Rozważamy dany język przedmiotowy L, w którym predykat prawdy dla zdań tego języka nie występuje. Konstrukcję przeprowadzamy w metajęzyku M, zawierającym, dla każdego zdania φ języka przedmiotowego, przekład φ na metajęzyk. Celem Tarskiego było podanie definicji prawdy, spełniającej warunki formalnej poprawności oraz materialnej adekwatności. A zatem:

Formalna poprawność: Definicja prawdy powinna mieć postać:

dla każdego zdania φ języka L, $Tr(\varphi)$ wtedy i tylko wtedy, gdy α,

[1] Te nowe dyskusje odnajdzie czytelnik np. w takich pracach jak Halbach [1999] i [2001], Horwich [1990], Ketland [1999], Sheard [1994] i [2001], Shapiro [1998].

gdzie α to formuła metajęzyka nie zawierająca definiowanego predykatu „*Tr*". (W istocie, Tarski wymagał, by w definicji prawdy nie wystąpiły żadne niezdefiniowane terminy semantyczne).

Materialna adekwatność: Dla dowolnego zdania φ języka przedmiotowego, z aksjomatów metateorii (teorii sformułowanej w metajęzyku, w której pracujemy budując definicję prawdy), powinno wynikać zdanie:

$$Tr(\varphi) \text{ wtedy i tylko wtedy, gdy } \psi,$$

gdzie ψ jest zdaniem metajęzyka, będącym przekładem zdania φ. W warunku materialnej adekwatności chodzi o to, by zdefiniowane przez nas pojęcie było rzeczywiście pojęciem prawdy – by spełniały je dokładnie te zdania języka przedmiotowego, które są (intuicyjnie) prawdziwe.

W ramach konstrukcji Tarskiego określa się najpierw pojęcie spełniania dla formuł języka przedmiotowego (formuły, w odróżnieniu od zdań, mogą zawierać zmienne wolne). W intuicyjnych terminach, formuła $\varphi(x_1 \ldots x_n)$ języka przedmiotowego jest spełniona przez skończony ciąg przedmiotów $a_1 \ldots a_n$, gdy jest ona prawdziwa o tych przedmiotach – tzn. staje się prawdziwa, jeśli zinterpretujemy zmienne $x_1 \ldots x_n$ jako nazwy obiektów $a_1 \ldots a_n$. Tarski zauważa jednak, że możemy wyrazić powyższą intuicję bez posługiwania się pojęciem prawdy – w szczególności, jeśli w naszym języku występuje skończenie wiele pierwotnych predykatów, możemy scharakteryzować rekurencyjnie spełnianie poprzez określenie najpierw spełniania dla formuł atomowych, a potem dla złożonych. W odpowiednio bogatej metateorii, taką rekurencyjną charakterystykę spełniania dla formuł języka przedmiotowego będzie można wyeliminować na rzecz definicji równościowej. Na dalszym etapie konstrukcji możemy zdefiniować pojęcie prawdy dla zdań (formuł bez zmiennych wolnych) – powiemy, że zdanie języka przedmiotowego jest prawdziwe, gdy jest spełnione przez pewien ciąg, albo – jak się okazuje, równoważnie – gdy jest spełnione przez pewien ciąg obiektów.

Kluczowa obserwacja Tarskiego polega na tym, że predykat prawdy jest istotnie metajęzykowy – zdefiniowana formuła „$Tr(x)$" nie jest zdaniem języka przedmiotowego. Przyjmijmy dla uproszczenia, że metajęzyk zawiera w sobie język przedmiotowy. Metateoria MT spełniałaby wtedy warunek:

$$\text{Dla każdego } \varphi \in L \text{ MT} \vdash Tr(\varphi) \equiv \varphi$$

Jednakże gdyby „$Tr(x)$" należał do języka przedmiotowego, to zastosowanie znanej, gödlowskiej konstrukcji przekątniowej dałoby nam zdanie ψ języka L, takie że (dla odpowiednio silnej MT):

$$\mathrm{MT} \vdash \psi \equiv \neg Tr(\psi)$$

Dla takiego ψ, mielibyśmy wtedy: $\mathrm{MT} \vdash Tr(\psi) \equiv \neg Tr(\psi)$, a zatem MT byłaby sprzeczna.

Dla naszych celów, istotne jest w tym momencie następujące pytanie: jaka metateoria wystarcza nam do zbudowania materialnie adekwatnej i formalnie poprawnej definicji prawdy dla języka przedmiotowego? Poniżej rozważę przykład, w którym za język przedmiotowy weźmiemy język arytmetyki pierwszego rzędu z dodawaniem i mnożeniem. Okazuje się, że pewne fragmenty arytmetyki drugiego rzędu wystarczą do tych celów – tzn. dla pewnych fragmentów A arytmetyki drugiego rzędu, istnieje formuła drugiego rzędu $Tr(x)$, która (dowodliwie w A) jest predykatem prawdy dla zbioru zdań języka arytmetyki pierwszego rzędu.

Za przykład posłuży tu teoria oznaczana zwykle jako ACA_0 – arytmetyka drugiego rzędu z arytmetycznym wyróżnianiem i arytmetyczną indukcją.[2] Aksjomaty tej teorii można podzielić na trzy grupy. Pierwszą z nich utworzą aksjomaty arytmetyki Peano (PA) – to będą aksjomaty pierwszego rzędu. Drugą grupę utworzą wybrane arytmetyczne podstawienia schematu wyróżniania:

$$\forall x\, \forall X \ldots \exists Z\, \forall n\, [n \in Z \equiv \varphi(n)].$$

„Arytmetycznym podstawieniem" nazywamy tu rezultat wstawienia w miejsce φ dowolnej formuły języka arytmetyki drugiego rzędu, w której żadna zmienna drugiego rzędu nie jest związana kwantyfikatorem. W miejsce trzech kropek należy z kolei wstawić ciąg kwantyfikatorów ogólnych – konkretny przypadek aksjomatu wyróżniania utworzy zatem dowolne uniwersalne domknięcie arytmetycznego podstawienia formuły drugiego rzędu „$\exists Z\, \forall n\, [n \in Z \equiv \varphi(n)]$". Obowiązuje przy tym następujące ograniczenie: w miejsce φ wolno nam wstawić wyłącznie formułę, w której nie występuje zmienna Z (w przeciwnym wypadku dostalibyśmy: $\exists Z\, \forall n\, [n \in Z \equiv n \notin Z]$, a zatem nasza teoria okazałaby się sprzeczna).

Trzecią grupę aksjomatów teorii ACA_0 tworzą podstawienia schematu indukcji. Trzeba tu sformułować podobne zastrzeżenie jak w przypadku schematu wyróżniania – przyjmiemy,

[2] Zob. Simpson [1999].

że w schemacie indukcji wolno nam podstawiać wyłącznie te formuły, w których żadna zmienna drugiego rzędu nie jest związana. W ten sposób określona teoria ACA_0 jest konserwatywnym rozszerzeniem PA – nie dowodzi żadnego twierdzenia języka arytmetyki pierwszego rzędu, które nie byłoby dowodliwe już w samej PA. Ustala to poniższy fakt.

Fakt 1. ACA_0 jest konserwatywnym rozszerzeniem PA.

Idea dowodu. Wystarczy tu zauważyć, że każdy model M dla PA da się rozszerzyć do modelu dla ACA_0. Niech mianowicie D będzie rodziną wszystkich podzbiorów uniwersum modelu M definiowalnych z parametrami w M za pomocą formuł języka PA. Rozważamy następnie model (M, D), w którym interpretujemy zmienne drugiego rzędu jako przebiegające zbiór D. Okazuje się wówczas, że (M, D) spełnia ACA_0. Wnioskujemy z tego, że ACA_0 jest konserwatywna nad PA, gdyż M jest dowolny.

$\square$

Choć ACA_0 jest konserwatywnym rozszerzeniem PA, istnieje formuła języka arytmetyki drugiego rzędu, która (dowodliwie w ACA_0) jest predykatem prawdy dla zdań języka arytmetyki pierwszego rzędu. Fakt ten ustala następujące twierdzenie.

Twierdzenie 2. Istnieje formuła $Tr(x)$ języka arytmetyki drugiego rzędu, taka że:

Dla dowolnego zdania φ języka arytmetyki pierwszego rzędu, $ACA_0 \vdash \mathrm{Tr}(\varphi) \equiv \varphi$.

Dowód: Dla zdania φ języka pierwszego rzędu, oznaczmy przez $rk(\varphi)$ rangę φ. Ranga φ to informacja, ile kroków konstrukcyjnych potrzebujemy do zbudowania φ z elementów składowych. I tak: ranga formuły atomowej (czyli równości między termami) wynosi 0, natomiast ranga koniunkcji, negacji i kwantyfikacji jest zawsze większa od rangi formuł składowych.

Dysponując pojęciem rangi, możemy określić teraz dwuargumentowy predykat $tr(n, X)$ o intuicyjnym odczytaniu „X jest zbiorem prawdziwych zdań języka PA o randze $\leq n$". Formuła „$tr(n, X)$" ma następujący kształt:

$$\forall y \in X\ [Sent(y) \wedge rk(y) \leq n] \wedge$$

$$\wedge\ \forall y\ [rk(y) = 0 \Rightarrow (y \in X \equiv \exists t,s \in Tm^{S}\ (y = \ulcorner t = s \urcorner \wedge val(t) = val(s)))] \wedge$$

$$\wedge\ \forall \varphi, \psi \in Sent\ \ [(rk(\varphi) < n \wedge rk(\psi) < n) \Rightarrow$$

$$\Rightarrow [(\ulcorner\neg\varphi\urcorner \in X \equiv \varphi \notin X) \wedge (\ulcorner\varphi \wedge \psi\urcorner \in X \equiv \varphi \in X \wedge \psi \in X)]] \wedge$$

$$\wedge\ \forall\varphi(x) \in Fm\ [rk(\varphi(x)) < n \Rightarrow$$

$$\Rightarrow (\ulcorner\forall x\varphi(x)\urcorner \in X \equiv \forall z\ sub(\varphi(x),\ S^z(0)/x) \in X))].$$

Wyrażenie „Tm^S" to arytmetyczny predykat oznaczający zbiór stałych termów języka PA, *Sent* to zbiór zdań języka PA, zaś przez $sub(\varphi(x),\ S^z(0)/x)$ rozumiemy rezultat podstawienia za zmienną x w formule φ termu o postaci „$S \ldots S(0)$", w którym symbol następnika powtarza się z razy.

Następnie określamy formułę „$Tr(x)$" jako:

$$\exists n\ \exists X\ [tr(n, X) \wedge x \in X].$$

Formuła $Tr(x)$ mówi: dla pewnego n, x należy do zbioru prawdziwych zdań rangi mniejszej lub równej n.
Dalsze rozumowanie przedstawia się następująco:

Fakt 3. $ACA_0 \vdash \forall x, y\ \forall X, Y\ [(x \leq y \wedge tr(x, X) \wedge tr(y, Y)) \Rightarrow Y|x = X]$,

gdzie $Y|x$ to ograniczenie zbioru Y do zdań rangi co najwyżej x.

Fakt 4. Dla dowolnej liczby naturalnej n, ACA_0 dowodzi:

$$\forall X\ [tr(n, X) \equiv X = Tr_n],$$

gdzie warunek „$X = Tr_n$" znaczy: „$\forall\psi\ [\psi \in X \equiv Tr_n(\psi)]$". Wyrażenie „$Tr_n(\psi)$" to arytmetyczny predykat prawdy dla zdań rangi nie większej niż n.

Uzyskujemy więc wniosek:

$$\forall n \in N\ \ ACA_0 \vdash \exists X\ tr(n, X),$$

biorąc bowiem $n \in N$, uzyskamy przez wyróżnianie istnienie zbioru X takiego że $X = Tr_n$.

Dwa powyższe fakty wystarczą już do wykazania, że dla dowolnego zdania φ języka arytmetyki pierwszego rzędu, ACA_0 dowodzi: $Tr(\varphi) \equiv \varphi$ (czyli: nasze pojęcie prawdy jest materialnie adekwatne, dowodliwie w ACA_0). Niech mianowicie φ będzie zdaniem języka arytmetyki pierwszego rzędu o randze n. Pracując w ACA_0, załóżmy φ. Skoro φ jest rangi n, uzyskamy $Tr_n(\varphi)$. Korzystając z Faktu 4, ustalmy Y równe Tr_n. Mamy wówczas: $tr(n, Y)$, a przy tym $\varphi \in Y$, zatem $Tr(\varphi)$. W ten sposób otrzymujemy dowód pierwszej implikacji. Dowodząc implikacji odwrotnej, znów pracujemy w ACA_0 i zakładamy $Tr(\varphi)$. Definicja predykatu Tr pozwala nam wybrać X oraz s takie że $tr(s, X) \wedge \varphi \in X$. Okazuje się wtedy, że $n \leq s$, gdyż n jest rangą formuły φ należącej do X. Na mocy Faktu 4 możemy wybrać zbiór Y taki że $tr(n, Y)$. Wówczas z Faktu 3, $Y = X|n$, więc $\varphi \in Y$. Skoro na mocy Faktu 4 $Y = Tr_n$, uzyskujemy: $Tr_n(\varphi)$. Ale φ jest formułą rangi n, uzyskujemy więc φ, co kończy dowód.

□

Mamy zatem pożądany rezultat: dysponujemy formułą „$Tr(x)$" języka arytmetyki drugiego rzędu oraz metateorią, w której potrafimy udowodnić materialną adekwatność naszego pojęcia prawdy, wyrażonego za pomocą wspomnianej formuły. Na zakończenie chciałbym sformułować dwie obserwacje.

Obserwacja 5. Choć arytmetyka drugiego rzędu, w której tu pracujemy, jest dość słabą teorią, to pozostaje jednak faktem, że w celu scharakteryzowania pojęcia prawdy arytmetycznej musieliśmy rozszerzyć zasób środków logicznych – przeszliśmy mianowicie do systemu dowodzenia drugiego rzędu. Inny możliwy ruch polegałby na zachowaniu logiki pierwszego rzędu i przejściu do odpowiedniego fragmentu teorii zbiorów. Można by jednak twierdzić, że w obu przypadkach wykraczamy poza same tylko intuicje związane z pojęciem liczby naturalnej, opisywanym w ramach PA.

Obserwacja 6. Choć ACA_0 dowodzi warunku materialnej adekwatności zdefiniowanego w jej obrębie pojęcia prawdy, to jednak jako teoria prawdy pozostawia sporo do życzenia. Pewne intuicyjne własności pojęcia prawdy nie są w niej dowodliwe. W szczególności, kompozycjonalność tego pojęcia pozostaje poza naszym zasięgiem. Rozważmy własności:

(a) $\forall t_1, t_2 \in Tm^S\, [Tr(\ulcorner t_1 = t_2 \urcorner) \equiv val(t_1) = val(t_2)]$

(b) $\forall \varphi \in Sent\, [Tr(\ulcorner \neg\varphi \urcorner) \equiv \neg Tr(\varphi)]$

(c) $\forall \varphi, \psi \in Sent\, [Tr(\ulcorner \varphi \vee \psi \urcorner) \equiv Tr(\varphi) \vee Tr(\psi)]$

(d) $\forall \varphi \in Fm\, \forall a \in Var\, [Tr(\ulcorner \exists a\varphi \urcorner) \equiv \exists v\, Tr(\ulcorner \varphi(\dot{v}) \urcorner)]$.

Pierwsza z tych własności to adekwatność pojęcia prawdy dla zdań atomowych – prawdziwe są te i tylko te zdania atomowe, dla których wartości odpowiednich termów są identyczne. Pozostałe warunki składają się na tzw. kompozycjonalność – chodzi w nich o to, że predykat prawdziwości zachowuje się w pożądany sposób względem operatorów logicznych (prawdziwość zdania złożonego zależy od prawdziwości odpowiednich zdań prostszych).

Rzecz jednak w tym, że ACA_0 nie dowodzi tych warunków. Jest to wniosek z następującego faktu:

Fakt 7. Istnieje model *M* arytmetyki Peano, w którym nie ma interpretacji predykatu prawdy, czyniącej (a)-(d) prawdziwymi.[3]

Wniosek 8. ACA_0 nie dowodzi warunków (a)-(d).

Dowód Wniosku 8: Załóżmy przeciwnie. Niech *M* będzie dowolnym modelem PA. Na mocy dowodu Faktu 1, można rozszerzyć *M* do modelu dla ACA_0. Ale z założenia, ACA_0 dowodzi warunków (a)-(d). Zatem istnieje w *M* interpretacja predykatu prawdy, czyniąca (a)-(d) prawdziwymi, co jest sprzeczne z Faktem 7, gdyż *M* jest dowolny.

□

Uzyskanie warunków (a)-(d) nie jest zatem możliwe w ACA_0. Istnieją jednak fragmenty arytmetyki drugiego rzędu, w których uzyskamy także i te warunki. Taką arytmetyką jest teoria oznaczana zwykle jako $\Pi_1^1 - CA_0$ - teoria, w której zarówno w schemacie indukcji, jak w schemacie wyróżniania wolno nam podstawiać formuły klasy Π_1^1 - tj. formuły o postaci $\forall X\varphi$, gdzie „X" jest zmienną drugiego rzędu, a φ jest formułą arytmetyczną (żadna zmienna drugiego rzędu nie jest związana w φ).[4]

II. Teorie prawdy

Jak już mówiłem, we współczesnych dyskusjach wykracza się niejednokrotne poza definicyjne ujęcie, naszkicowane w poprzednim podrozdziale. Strategia wielu badaczy jest następująca: nie próbują definiować predykatu prawdy środkami metateorii; zamiast tego rozszerzają język teorii bazowej, wprowadzając wyrażenie „jest prawdziwe" jako nowy, pierwotny predykat. Własności tego predykatu charakteryzują następnie za pomocą odpowiedniego zestawu aksjomatów.

Moim celem będzie obecnie przedstawienie motywów, które skłaniają do tego badaczy. Głównego powodu jestem skłonny dopatrywać się w różnicy celów teoretycznych.

---

[3] Formalnie: istnieje $M \models$ PA, taki że nie istnieje $S \subseteq M$ dla którego zachodziłoby: $(M, S) \models$ (a)-(d). Fakt ten to wniosek z twierdzenia Kotlarskiego, Krajewskiego i Lachlana, na mocy którego takie interpretacje predykatu prawdy (tzw. klasy spełniania) znajdziemy wyłącznie w rekurencyjnie nasyconych modelach arytmetyki, nie każdy zaś model arytmetyki jest rekurencyjnie nasycony. Zob. Kotlarski, Krajewski, Lachlan [1981].

[4] Więcej informacji o podsystemach arytmetyki drugiego rzędu odnajdzie czytelnik w Simpson [1999].

Tarski zamierzał podać definicję prawdy dla języka przedmiotowego i używał w tym celu metajęzykowych środków. Jego zamiar wydaje się naturalny w świetle ówczesnej sytuacji problemowej: otóż pojęcie prawdy nie budziło zaufania. Nie było wiadomo, czy można posługiwać się tym pojęciem w sposób spełniający standardy naukowej precyzji.[5] Pojęcia semantyczne uważano za „nienaukowe". Kiedy w 1935 roku na kongresie w Paryżu Tarski przedstawił swoje koncepcje dotyczące prawdy, droga do sukcesu bynajmniej nie była łatwa. Jak piszą A. i S. Fefermanowie:

> Obawy Tarskiego co do negatywnego przyjęcia jego teorii prawdy przez słuchaczy na kongresie paryskim okazały się słuszne. Pewne reakcje na prace Tarskiego [...] były zdecydowanie negatywne. Kością niezgody było, czy pojęcia semantyczne można pogodzić ze ściśle empirycystycznym i antymetafizycznym stanowiskiem.[6]

Tak czy inaczej, podanie definicji pozwoliło na zrehabilitowanie pojęcia prawdy. Jest ono teraz używane w poważnym dyskursie naukowym; nie pozostało zbyt dużo miejsca na krytykę koncepcji Tarskiego w „antymetafizycznym" duchu. Obecna sytuacja jest więc zupełnie inna niż w latach trzydziestych. Wygląda na to, że wielu współczesnym teoretykom pojęcie prawdy wydaje się niewinne – w szczególności, mają oni większe zaufanie do pojęcia prawdy arytmetycznej niż do metateoretycznych środków, które wykorzystujemy do skonstruowania jego definicji (chodzi tu na przykład o logikę drugiego rzędu). Współcześni teoretycy nie próbują zatem rehabilitować pojęcia prawdy – taką rehabilitację przeprowadził już Tarski. Pragną oni raczej swobodnie używać tego pojęcia przy minimalnych kosztach własnych. Pytają na przykład: w jaki sposób moglibyśmy posługiwać się tym pojęciem bez angażowania się w ontologię, którą z jakichś powodów uważamy za wątpliwą? W bardziej szczegółowej wersji, pytanie może przybrać też formę: w jaki sposób możemy wprowadzić do naszego języka predykat prawdy arytmetycznej bez wykraczania poza język pierwszego rzędu i bez sięgania po teorię mnogości. Szukając odpowiedzi na te pytania, wybierają ujęcia aksjomatyczne. Nie próbują definiować pojęcia prawdy, lecz starają się je scharakteryzować za pomocą paru możliwie prostych zasad. Jest to jak najbardziej próba wykroczenia poza Tarskiego; pomiędzy obydwoma podejściami nie ma jednak konfliktu.

Na zasadzie przykładu, zamierzam teraz przedstawić trzy próby aksjomatycznego ujęcia teorii prawdy dla języka arytmetyki pierwszego rzędu.

---

[5] To nie znaczy, że pojęcia prawdy nie używano – logikom (za przykład może posłużyć tu Skolem) zdarzało się używać tego pojęcia (albo pokrewnego pojęcia spełniania), ale jedynie na zasadzie intuicyjnej i nieformalnej.

[6] Zob. Feferman Anita i Feferman Solomon [2009], s. 122.

1. *Teoria D(PA)*

Rozważana tu teoria jest w pewnym sensie „minimalną” teorią prawdy dla języka arytmetyki. Zacznijmy od sformułowania definicji:

Definicja 9. $D(\mathrm{PA}) = \mathrm{PA} \cup \{Tr(\ulcorner\varphi\urcorner) \equiv \varphi\colon \varphi \in L(\mathrm{PA})\}$.

Oznacza to, że D(PA) to teoria, która powstaje z arytmetyki Peano przez dodanie jako nowych aksjomatów wszystkich zdań o postaci „φ jest prawdziwe wtedy i tylko wtedy, gdy φ” jako nowych aksjomatów. Uzyskana teoria prawdy jest minimalna w tym sensie, że warunek dowodzenia wszystkich takich równoważności wydaje się podstawowym wymogiem, z którego nie chcielibyśmy zrezygnować. Teorii nie dowodzącej takich równoważności nie nazwalibyśmy po prostu teorią prawdy dla języka arytmetyki – odczytanie „*Tr*(*x*)” jako „*x* jest prawdziwe” stałoby się wówczas nienaturalne.

Zauważmy, że teorię prawdy uzyskuje się w tym podejściu niezwykle tanim kosztem. Rozszerzenie naszej teorii bazowej wydaje się zupełnie niegroźne: nie wprowadzamy żadnych nowych zasad logicznych, a nowe aksjomaty wydają się oczywiste ze względu na intuicyjną interpretację, jaką chcemy nadać nowo wprowadzonemu predykatowi.

Jednakże wady *D*(PA) jako teorii prawdy są również poważne. Omówiłem je dokładniej gdzie indziej;[7] tu sformułuję tylko podstawowy problem. Otóż żaden z warunków (a)-(d) wprowadzonych w poprzednim podrozdziale nie jest twierdzeniem *D*(PA). Wewnątrz *D*(PA) dysponujemy zatem tylko ubogim pojęciem prawdy. Nie posiadamy pojęcia kompozycjonalnego, nie jesteśmy w stanie udowodnić wielu interesujących uogólnień, które pojęcie prawdy angażują. Takie pojęcie prawdy wydaje się mało użyteczne.

2. *Teoria PA(S)⁻*

Jest to teoria, uzyskana z PA przez dodanie warunków (a)-(d) jako nowych aksjomatów, charakteryzujących pierwotny predykat „*Tr*” rozszerzonego języka. W teorii tej mamy jednak do dyspozycji tylko arytmetyczną indukcję, tzn. nie pozwalamy podstawiać w schemacie indukcji formuł z predykatem „*Tr*” (dokładniej: uzyskane w ten sposób podstawienia nie są aksjomatami $\mathrm{PA(S)}^{-}$.

[7] Zob. Cieśliński [2009].

Dla tak określonej teorii, zachodzi:

Fakt 10. PA(S)$^-$ jest konserwatywnym rozszerzeniem PA,

czyli: każde zdanie języka arytmetyki (bez predykatu prawdy) mające dowód w PA(S)$^-$, daje się udowodnić środkami samej tylko arytmetyki Peano – nie potrzebujemy w jego dowodzie używać aksjomatów, charakteryzujących pojęcie prawdy.[8] Własność ta wydaje się atrakcyjna: mimo że w obrębie PA(S)$^-$ posługujemy się kompozycjonalnym pojęciem prawdy, to jednak okazuje się ono wciąż w pewnym sensie niewinne. Odpowiada to intuicji, zgodnie z którą pojęcie prawdy arytmetycznej służy nam zasadniczo do mówienia o języku arytmetyki, nie zaś do uprawiania arytmetyki – nie do dowodzenia nowych twierdzeń.

Choć jednak intuicja może się wydawać przekonująca, są tu pewne problemy. Wielu teoretyków uważa, że od teorii prawdy powinniśmy mimo wszystko oczekiwać określonej mocy teoriodowodowej – taka teoria powinna tłumaczyć, dlaczego jesteśmy gotowi zaakceptować szereg bardzo podstawowych uogólnień, angażujących pojęcie prawdy (spodziewane wytłumaczenie brzmiałoby zaś: jesteśmy gotowi zaakceptować wspomniane uogólnienia, gdyż wynikają one z niekontrowersyjnych aksjomatów naszej teorii prawdy). Rzecz jasna, zwrot „bardzo podstawowe uogólnienia" nie jest precyzyjny i nie jest w związku z tym do końca wyraźne, czego od teorii prawdy powinniśmy oczekiwać w tym zakresie. Rozważmy jednak przykład. Jednym z podstawowych uogólnień, do sformułowania których wykorzystujemy predykat prawdy, jest zdanie stwierdzające prawdziwość logiki. W tej właśnie kwestii PA(S)$^-$ okazuje się bardzo słaba, a jej słabość dzielą wszystkie konserwatywne teorie prawdy dla języka arytmetyki, rozszerzające PA(S)$^-$. Stwierdza to poniższy fakt:

Fakt 11. PA(S)$^-$ + „Logika pierwszego rzędu (dla języka z dodawaniem i mnożeniem) jest prawdziwa" nie jest konserwatywnym rozszerzeniem PA.[9]

Istnieją zatem powody, by odrzucić PA(S)$^-$ jako zadowalającą teorię prawdy. Dodatkowo należy zauważyć, że PA(S)$^-$ jest w pewnym sensie nienaturalną teorią – nie widać wyraźnych motywacji, które zabraniałyby nam odrzucać indukcję dla formuł rozszerzonego języka. Wręcz przeciwnie, rozszerzenie teorii o dodatkowe aksjomaty indukcji wydaje się bardzo naturalnym posunięciem. Można by wręcz twierdzić, że takie rozszerzenie nie ma nic

[8] Zob. Kotlarski, Krajewski, Lachlan [1981]; zob. także Engström [2002].
[9] Zob. Cieśliński [w druku], zob. także Cieśliński [2009], s. 164.

wspólnego z naszymi szczególnymi koncepcjami dotyczącymi prawdy – bylibyśmy gotowi zaakceptować indukcję dla dowolnego nowego predykatu, opisującego liczby naturalne. Prowadzi to nas do trzeciej – najsilniejszej z rozważanych tu teorii.

3. *Teoria PA(S)*

Definiujemy teorię PA(S) jako rozszerzenie $\mathrm{PA(S)}^{-}$ o nowe aksjomaty indukcji – dołączamy mianowicie jako aksjomaty wszystkie podstawienia schematu indukcji w rozszerzonym języku, z predykatem prawdy. Otrzymujemy w ten sposób silną teorię, która (w przeciwieństwie do $\mathrm{PA(S)}^{-}$) nie jest konserwatywnym rozszerzeniem arytmetyki Peano. W szczególności, zachodzi:

Fakt 12. $\mathrm{PA(S)} \vdash \mathrm{Con}_{\mathrm{PA}}$

gdzie „$\mathrm{Con}_{\mathrm{PA}}$” to zdanie języka arytmetyki, które przy naturalnej interpretacji odczytujemy „Nie istnieje *d* będące dowodem sprzeczności (np. „$0 \neq 0$”) z aksjomatów PA”. Z Faktu 12 wynika niekonserwatywność PA(S) względem PA, gdyż sama PA, na mocy drugiego twierdzenia Gödla, nie dowodzi własnej niesprzeczności.

Idea dowodu Faktu 12: przez indukcję dla formuł z predykatem prawdy pokazujemy, że:

$$\mathrm{PA(S)} \vdash \forall\psi\, [Pr_{\mathrm{PA}}(\psi) \Rightarrow Tr(\psi)],$$

gdzie „$Pr_{\mathrm{PA}}(\psi)$” to formuła języka arytmetyki pierwszego rzędu, która przy naturalnym odczytaniu głosi „ψ jest twierdzeniem PA”. Okazuje się zatem, że PA(S) dowodzi prawdziwości wszystkich twierdzeń PA. Z tego wewnątrz PA(S) wyprowadzamy wniosek o niesprzeczności PA. Gdyby mianowicie PA była sprzeczna, to dowodziłaby, że $0 \neq 0$. Pokazujemy w PA(S), że PA tego nie dowodzi. Mamy:

$$\forall\psi\, [Pr_{\mathrm{PA}}(\psi) \Rightarrow Tr(\psi)],$$

z czego wnioskujemy:

$$Pr_{\mathrm{PA}}(\ulcorner 0 \neq 0 \urcorner) \Rightarrow Tr(\ulcorner 0 \neq 0 \urcorner).$$

Faktem jest jednak, że:

$$\neg\, 0 \neq 0,$$

zatem:

$$\neg\, Tr(\ulcorner 0 \neq 0 \urcorner).$$

Jako końcowy wniosek uzyskujemy:

$$\neg\, Pr_{\mathrm{PA}}(\ulcorner 0 \neq 0 \urcorner).$$

Właśnie to zdanie oznaczyliśmy jako Con_{PA} (czyli PA jest niesprzeczna).

□

Warto zauważyć, że podany przykład nie wyczerpuje arytmetycznej mocy teorii PA(S) – jest to teoria arytmetycznie silniejsza niż PA z dodaną informacją o niesprzeczności PA (czyli: PA(S) dowodzi więcej zdań arytmetycznych niż teoria PA + $\mathrm{Con}_{\mathrm{PA}}$). Środkami PA(S) można np. również udowodnić niesprzeczność PA + $\mathrm{Con}_{\mathrm{PA}}$; ponownie, na mocy drugiego twierdzenia Gödla, PA + $\mathrm{Con}_{\mathrm{PA}}$ nie dowodzi zaś swojej własnej niesprzeczności. Okazuje się, że pod względem siły arytmetycznej PA(S) odpowiada systemowi arytmetyki drugiego rzędu ACA – uzyskujemy go z poznanego już przez nas ACA_0 przez dodanie pełnej indukcji dla formuł drugiego rzędu. A zatem także i w tym przypadku okazuje się, że istnieje sposób uzyskania arytmetycznie mocnej teorii poprzez (na pozór) słabe i bardzo intuicyjne wzmocnienie wybranej aksjomatyki dla teorii prawdy.

Podsumowanie

W literaturze proponowano rozmaite aksjomatyzacje, charakteryzujące sens predykatu prawdy.[10] Takie aksjomatyzacje pozwalały rozważać zagadnienia związane z teorią prawdy bez angażowania silnych – i w wielu kontekstach niepotrzebnych - środków (takich jak logika drugiego rzędu albo teoria zbiorów). Nie ma jednak konfliktu pomiędzy podejściem aksjomatycznym a podejściem klasycznym – oba podejścia raczej uzupełniają się nawzajem niż sobie przeczą.

Typową teorią bazową, z którą pracowano, była arytmetyka Peano. W rozważaniach dotyczących prawdy, pełniła rolę teorii składni – daje nam ona wystarczające środki, abyśmy mogli opisać podstawowe pojęcia syntaktyczne. Naturalne pytanie wiąże się z tym, jak dużo zależy tu od wyboru teorii bazowej – które rezultaty uzyskane w ramach prac nad aksjomatycznymi teorii prawdy są uniwersalne, a które mają charakter lokalny (związany z

[10] Aksjomatyzacje, które tu rozważałem miały służyć tylko jako przykłady. Proponowanych w literaturze aksjomatyzacji jest znacznie więcej; niektóre z nich (zgodnie z intencjami autorów) miały formalizować raczej podejście Kripkego (zob. Kripke [1975]) niż idee Tarskiego. O tego typu aksjomatykach, zob. Sheard [1994].

wyborem tej a nie innej teorii bazowej). W tym ostatnim pytaniu kryje się moim zdaniem perspektywa dalszych, owocnych badań. Z jednej strony, należałoby rozważyć teorie silne, takie jak teoria mnogości. Z drugiej – warto by opisać, co przynoszą próby budowania teorii prawdy nad teoriami arytmetycznymi znacznie słabszymi niż PA. Na szczególną uwagę zasługiwałby zwłaszcza przypadek $I\Delta_0$ (arytmetyka z niepełną indukcją, w schemacie indukcji wolno nam podstawiać wyłącznie formuły z kwantyfikatorami ograniczonymi), a nawet jeszcze słabsza arytmetyka Robinsona. Wciąż nie jest jednak do końca jasne, w jaki sposób powinno się formułować aksjomatykę teorii prawdy przy tak słabych teoriach bazowych.

Bibliografia

[1] Cieśliński, Cezary [2009] *Deflacyjna koncepcja prawdy. Wybrane zagadnienia logiczne*, Semper, Warszawa.

[2] Cieśliński, Cezary [2010] „Truth, conservativeness, and provability", *Mind* 119 (474), 409-422.

[3] Engström, Fredrik [2002] *Satisfaction classes in nonstandard models of first-order arithmetic*, Chalmers University of Technology and Göteborg University, Göteborg.

[4] Feferman Anita i Feferman Solomon [2009] *Alfred Tarski. Życie i logika*, Wydawnictwa Akademickie i Profesjonalne, Warszawa, przeł. Joanna Golińska-Pilarek i Marian Srebrny.

[5] Halbach, Volker [1999] "Conservative Theories of Classical Truth", *Studia Logi*ca 62, 353–70.

[6] Halbach, Volker [2001] „Disquotational truth and analyticity", *Journal of Symbolic Logic* 66, 1959-1973.

[7] Horwich Paul [1990] *Truth*, Basil Blackwell, Oxford.

[8] Ketland Jeffrey [1999] „Deflationism and Tarski's paradise", *Mind* 108, 64-94.

[9] Kotlarski Henryk, Krajewski Stanisław i Lachlan Alistair [1981] „Construction of satisfaction classes for nonstandard models", *Canadian Mathematical Bulletin* 24, 283-293.

[10] Kripke, Saul [1975] "Outline of a Theory of Truth", *Journal of Philosophy* 72, 690–716.

[11] Shapiro Stewart [1998] „Proof and truth: through thick and thin”, *Journal of Philosophy* 95, 493-521.

[12] Sheard Michael [1994] „A guide to truth predicates in the modern era”, *Journal of Symbolic Logic* 59, 1032-54.

[13] Sheard, Michael [2001] “Weak and strong theories of truth”, *Studia Logica* 68, 89–101.

[14] Simpson, Stephen [1999] *Subsystems of Second Order Arithmetic*, Springer Verlag, Berlin.

[15] Tarski, Alfred [1933] „Pojęcie prawdy w językach nauk dedukcyjnych”, *Prace Towarzystwa Naukowego Warszawskiego, wydział III*, nr. 34, Warszawa.