Naukoznawstwo (Etnolingwistyka V)

Jerzy Pogonowski

Zakład Logiki Stosowanej
www.logic.amu.edu.pl
pogon@amu.edu.pl

17 XI 2007

Na początek dobrego dnia

Zanim dotrzemy na szczyty refleksji metateoretycznej popatrzmy na szczyt Kościelca (7.11.2006):

Procedury poznawcze IV: refleksja metateoretyczna

Plan:

- Po co rozważania metateoretyczne?
- Teorie naukowe — definicje
- Teorie — własności metalogiczne
- Arytmetyzacja składni
- Systemy przekonań
- Twierdzenia metalogiczne

To oczywiście nie jest plan na całą resztę dzisiejszego dnia — jestem pewien, że po opuszczeniu naszego North Campus dacie sobie radę samodzielnie (zorganizować wieczór, itd.).

Refleksja metateoretyczna

Co począć ze zgromadzoną już wiedzą?
Pomijamy odpowiedź banalną: *Sprzedawać.*
Uwaga na marginesie: jest rzeczą frapującą, jaka część uzyskanej obecnie wiedzy jest wiedzą powszechnie dostępną.

Interesuje nas refleksja teoretyczna nad systemami wiedzy (systemami przekonań).
Refleksja taka ma m.in. odpowiadać na pytania o jakość zgromadzonej wiedzy (jej niesprzeczność, trafność, kompletność, itd.), dynamikę jej zmian, uwarunkowania w jej tworzeniu, itd.

Refleksja metateoretyczna: pojęcie teorii

Tak więc, zmuszeni jesteśmy (*chcę, więc muszę*) do opracowania systemu **wiedzy o wiedzy**, tj. do budowania **metateorii**.

Systematyczne naukoznawstwo jest właśnie (w zamierzeniu) tworzeniem takiej (takich) metateorii.

Dociekliwi mogą zapytać: a jaką mamy gwarancję, iż nasza wiedza o wiedzy jest trafna? Czy wiedza na poziomie metateoretycznym może być co najmniej tak samo pewna jak wiedza na poziomie przedmiotowym? Czy nie grozi nam nieskończony regres: dla objaśnienia wiedzy o wiedzy budujemy wiedzę o wiedzy o wiedzy, itd.? Odpowiedzi na niektóre tego typu pytania udziela właśnie metateoria.

Refleksja metateoretyczna: pojęcie teorii

Z wykładu logiki pamiętacie, że teorią (w sensie dedukcyjnym) nazywamy punkty stałe operatora konsekwencji. Elementy teorii nazywamy jej twierdzeniami.

Inaczej mówiąc, jeśli Cn jest operacją konsekwencji w języku L, to teorią nazywamy każdy zbiór formuł X tego języka, dla którego $Cn(X) = X$.

Mówiąc jeszcze bardziej „po ludzku", teoria to zbiór formuł domknięty na konsekwencje — X jest teorią, gdy w X są już wszelkie wnioski, które można z X wyprowadzić (przy użyciu rozważanej operacji konsekwencji).

Parafrazując to jeszcze inaczej, zbiór formuł X jest teorią, gdy nie można z formuł tego zbioru, traktowanych jako przesłanki, wyprowadzić wniosków, które nie należałyby do X.

Refleksja metateoretyczna: pojęcie teorii

Takie rozumienie pojęcia „teoria" jest — z punktu widzenia nauk empirycznych — pewną idealizacją. Jest ona konieczna, gdy chcemy poprawnie mówić o własnościach **całych** teorii.

W naukach empirycznych termin teoria ma nieco inne znaczenie niż w logice: za teorię uważa się tu pewien zespół przekonań odnoszących się do ustalonej dziedziny przedmiotowej i w miarę spójny (pojęciowo, dedukcyjnie, definicyjnie, itp.).

Będziemy używać terminu teoria w obu znaczeniach.
Wyliczymy niżej niektóre własności teorii naukowych, interesujące z metodologicznego punktu widzenia.

Logiczna rekonstrukcja koncepcji empirycznych

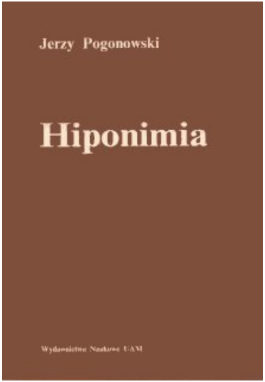
Teoretyczne koncepcje empiryczne mogą zostać poddane procedurze logicznej rekonstrukcji. Polega ona na wprowadzeniu ładu dedukcyjnego:

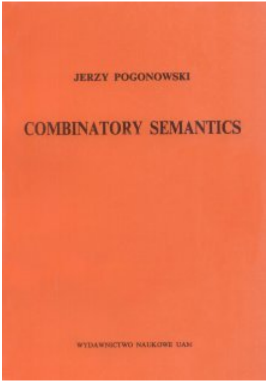
- wyszczególnieniu wszystkich pojęć pierwotnych;
- wyraźnym sformułowaniu wszystkich przyjmowanych aksjomatów;
- podaniu poprawnych definicji pozostałych (poza pierwotnymi) pojęć;
- ustaleniu, które ze stwierdzeń są twierdzeniami (posiadają dowód z aksjomatów), a które hipotezami.

Po dokonaniu logicznej rekonstrukcji rozważana koncepcja teoretyczna ma już cechy teorii sformalizowanej, do której można stosować procedury logiczne.

Logiczna rekonstrukcja koncepcji empirycznych

O logicznych rekonstrukcjach wybranych koncepcji lingwistycznych poczytać możesz np. w:





Własności teorii

Powiemy, że teoria T jest NIESPRZECZNA, gdy nie zawiera ona jednocześnie: jakiegoś zdania oraz jego zaprzeczenia.
W przeciwnym przypadku powiemy, że T jest SPRZECZNA.

Niesprzeczność jest podstawowym wymogiem metodologicznym.
Jeśli jakaś teoria T nie jest niesprzeczna, tj. zawiera parę zdań postaci A oraz $\neg A$, to wyprowadzić w niej można dowolne zdanie B, na mocy Prawa Dunsa Scotusa: $A \rightarrow (\neg A \rightarrow B)$. Inaczej mówiąc, teoria sprzeczna zawiera wszystkie zdania (rozważanego języka), a więc jest bezużyteczna — nie niesie żadnej informacji.

Własności teorii

Powiemy, że teoria T jest ω-NIESPRZECZNA, gdy, jeśli (dla dowolnej formuły A o jednej zmiennej wolnej) jej twierdzeniami są: $A(0)$, $A(1)$, $A(2)$, ..., to jej twierdzeniem **nie jest** $\exists x \ \neg A(x)$.
Jeśli teoria nie jest ω-niesprzeczna, to mówimy, że jest ω-SPRZECZNA.

Własność ω-niesprzeczności jest mocniejsza niż własność niesprzeczności: jeśli teoria jest ω-niesprzeczna, to jest też niesprzeczna, ale niekoniecznie na odwrót.
Mówienie o własności ω-niesprzeczności ma sens w przypadku teorii, w których można używać (nazw dla) liczb naturalnych.
Własność ω-niesprzeczności związana jest z ω-regułą. W jednym z omówionych dalej metatwierdzeń zobaczymy, jaką rolę pełni własność ω-niesprzeczności.

Własności teorii

Teoria T jest ZUPEŁNA, gdy z każdej pary wzajem sprzecznych zdań jedno jest twierdzeniem tej teorii.
Ta własność wydaje się być ideałem metodologicznym. Gdy teoria jest zupełna (i niesprzeczna), to każdy z problemów sformułowanych w jej języku (czyli każde pytanie rozstrzygnięcia) znajduje w niej odpowiedź (choć może nie być możliwe jej efektywne (rekurencyjne) uzyskanie).

Zupełne są np.: elementarna teoria nierówności, elementarna arytmetyka liczb naturalnych (bez mnożenia), elementarna dwuwymiarowa geometria Euklidesa.

Jednak tylko niewiele bogatych, interesujących teorii jest zupełnych, (jak dowiemy się pod koniec tego wykładu).

Własności teorii

Teoria T jest AKSJOMATYZOWALNA, gdy daje się przedstawić w postaci zbioru wniosków otrzymanych za pomocą obliczalnych reguł wnioskowania z obliczalnego zbioru aksjomatów.

Zamiast (intuicyjnego!) pojęcia obliczalności w precyzyjnej definicji teorii aksjomatyzowalnych użyć należy któregoś z matematycznych odpowiedników tego pojęcia (a więc np. pojęcia zbioru rekurencyjnego).

„Ludzką mową", aksjomatyzowalność danej teorii oznacza, że można w sposób efektywny wyliczyć wszystkie jej założenia czynione dogmatycznie (bez dowodu) oraz iż używane w niej reguły wnioskowania również mają charakter efektywny.

Własności teorii

Teoria T jest ROZSTRZYGALNA, gdy zbiór jej twierdzeń jest obliczalny (precyzyjnie mówiąc: rekurencyjny). W przeciwnym przypadku T nazywamy NIEROZSTRZYGALNĄ.
Jeśli T jest rozstrzygalna, to istnieje efektywna metoda pozwalająca, dla każdego problemu (pytania rozstrzygnięcia) sformułowanego w jej języku, rozstrzygnąć, czy problem ten ma w T odpowiedź.

Uwaga: Rozstrzygalność i zupełność to dwie różne własności. Pewien związek między nimi ustala następujące twierdzenie:

Jeśli teoria T jest niesprzeczna, zupełna i aksjomatyzowalna, to jest rozstrzygalna.

Własności teorii

Teoria T jest KATEGORYCZNA, gdy wszystkie jej interpretacje (modele) są izomorficzne.
To bardzo mocna własność; niewiele teorii ją posiada. Kategoryczność to możliwość jednoznacznego wyznaczenia modelu zamierzonego teorii.

Teoria T jest KATEGORYCZNA W MOCY κ (gdzie κ jest nieskończoną liczbą kardynalną), gdy wszystkie jej interpretacje (modele) mocy κ są izomorficzne.
Dla przykładu: teoria gęstego liniowego porządku bez elementu najmniejszego oraz największego jest kategoryczna w mocy $\aleph_0$. Inaczej mówiąc: istnieje jeden (z dokładnością do izomorfizmu) taki porządek.
Pomyśl o liczbach wymiernych i relacji $\leqslant$.

Własności teorii

Własności teorii zależą od języka, w którym jest formułowana oraz od własności rozważanej operacji konsekwencji (a więc od branego pod uwagę systemu logiki). Standardem logicznym, także w zastosowaniach jest (Klasyczna) Logika Pierwszego Rzędu, w skrócie FOL (od *First Order Logic*; ewentualny skrót polski źle mi się kojarzy).

Z wykładu logiki pamiętacie, że FOL ma dwie ważne własności metalogiczne:

- FOL jest TRAFNA (*sound*), tj. wszystkie tezy FOL są jej tautologiami;
- FOL jest PEŁNA (*complete*), tj. wszystkie tautologie FOL są jej tezami.

Własności teorii

Jeśli w rozważanej teorii empirycznej wymagane jest stosowanie innego niż FOL systemu logiki (np.: logiki epistemicznej, modalnej, deontycznej, wielowartościowej, itd.), to zwykle żąda się, aby rozważana logika owe dwa powyższe warunki spełniała.

Pewien kłopot jest ze stosowaniem logiki drugiego rzędu — jak wiadomo, ten system logiczny nie jest pełny.

Pod koniec tego wykładu rozważymy problem trapiący niektórych filozofów: Która logika jest przydatna jako logika nauki?

Własności teorii

W ostatniej części wykładu podamy pewne ustalenia dotyczące własności teorii, o fundamentalnym znaczeniu dla epistemologii. Będą one sformułowane bez pedanterii formalnej, w znacznym uproszczeniu. Wykorzystamy dwie drogi przedstawienia tych ustaleń:

- metodę arytmetyzacji składni;
- interpretację modalną (pojęcia dowodliwości).

Obie metody zostaną tu podane w wersji popularnej.

Własności teorii

Precyzyjne przedstawienie tej problematyki wykracza poza możliwości (czasowe) tego wykładu. Zainteresowanym polecam np. lekturę bardziej zaawansowanych pozycji:

- Hofstaedter, D.: *Gödel, Escher, Bach: An Eternal Golden Braid.*
- Hunter, G.: *Metalogika.*
- Krajewski, S.: *Twierdzenie Gödla i jego interpretacje filozoficzne.*
- Murawski, R. *Funkcje rekurencyjne i elementy metamatematyki.*

Własności teorii

Polecam także znakomite książki z zagadkami logicznymi, w których popularyzuje się wiedzę na temat metalogiki i jej zastosowań:

- Smullyan, R.: *Jaki jest tytuł tej książki?*
- Smullyan, R.: *Dama czy tygrys?*
- Smullyan, R.: *Szatan, Cantor i nieskończoność.*
- Smullyan, R.: *Przedrzeźniać przedrzeźniacza.*
- Smullyan, R.: *Na zawsze nierozstrzygnięte.*

Arytmetyzacja składni

Zakładam, że pamiętacie (z kursu logiki) co to jest język I rzędu. W szczególności, korzystać będziemy z następujących pojęć logicznych:

- aksjomat, reguła wnioskowania, dowód, twierdzenie, teoria;
- interpretacja, spełnianie (formuły w interpretacji przez wartościowanie);
- podstawienie termu za zmienną w formule;
- metajęzyk.

Na wykładzie poświęconym definicjom podano aksjomatykę Peana dla arytmetyki. Dziś wykorzystywać będziemy (*implicite*) inną aksjomatykę dla PA, w której symbolami pierwotnymi są, oprócz zera oraz następnika, także symbole funkcyjne: dodawania i mnożenia.

Arytmetyzacja składni

Aksjomaty specyficzne systemu Arytmetyki Peana:

- A_1: $\forall x \forall y\, (x \neq y \rightarrow S(x) \neq S(y))$
- A_2: $\forall x\, (0 \neq S(x))$
- A_3: $\forall x\, (x \neq 0 \rightarrow \exists y\, (x = S(y)))$
- A_4: $\forall x\, (x + 0 = x)$
- A_5: $\forall x \forall y\, (x + S(y) = S(x + y))$
- A_6: $\forall x\, (x \cdot 0 = 0)$
- A_7: $\forall x \forall y\, (x \cdot S(y) = (x \cdot y) + x)$
- A_8: $(A(0) \wedge \forall x\, (A(x) \rightarrow A(S(x)))) \rightarrow \forall x\, A(x)$

(dla dowolnej formuły A, o jednej zmiennej wolnej, języka Arytmetyki Peana).
A_8 nie jest jednym aksjomatem, lecz schematem (przeliczalnie wielu) aksjomatów. P_8 nazywamy **zasadą indukcji**.

Arytmetyzacja składni

W języku arytmetyki PA możemy „mówić" nie tylko o liczbach naturalnych (interpretacja zamierzona), lecz także o samej arytmetyce PA. Możliwość tę uzyskujemy poprzez kodowanie symboli, wyrażeń oraz ciągów wyrażeń języka PA liczbami naturalnymi.

Kodować można na wiele sposobów. Dla przykładu: niech zmienna indywiduowa x_i ma numer $2 \cdot i$, a pozostałe symbole jakieś numery nieparzyste:

- $\neg$ — 3,
- $\forall$ — 5,
- $+$ — 7,
- S — 9,
- 0 — 11,
- (— 13,
-) — 15,
- $=$ — 17, itp.

Arytmetyzacja składni

Wtedy ciągi symboli języka PA można (jednoznacznie!) kodować przez iloczyny kolejnych liczb pierwszych podniesionych do stosownych potęg. Np. kodem wyrażenia:

$$\forall x_2 \ \neg(x_2 = S(0))$$

będzie:

$$2^5 \cdot 3^4 \cdot 5^3 \cdot 7^{13} \cdot 11^4 \cdot 13^{17} \cdot 17^9 \cdot 19^{13} \cdot 21^{11} \cdot 23^{15} \cdot 29^{15}$$

Cóż, nie jest to mała liczba... Istotna jest jednak jedynie możliwość jednoznacznego oraz obliczalnego kodowania ciągów symboli.

Arytmetyzacja składni

Liczbę, która koduje wyrażenie języka PA (term, formułę) nazywamy numerem Gödlowskim tego wyrażenia (w skrócie: numerem). Liczebniki to numery termów postaci: 0, $S(0)$, $S(S(0))$, $S(S(S(0)))$,... (a więc numery nazw kolejnych liczb naturalnych).

Umowa notacyjna. W dalszym ciągu, niech:

- x, y, z oznaczają zmienne indywidualne;
- m, n, r oznaczają liczby naturalne;
- $\overline{m}$, $\overline{n}$, $\overline{r}$ oznaczają liczebniki.

Kodowanie ciągów wyrażeń (np.: dowodów) odbywa się tak samo, jak kodowanie wyrażeń (iloczyny kolejnych liczb pierwszych podniesionych do potęg będącym numerami elementów ciągu). Także to kodowanie jest jednoznaczne i obliczalne.

Arytmetyzacja składni

Można pokazać, że formuły $dow(x, y)$ oraz $podst(x, y) = z$ dają się precyzyjnie określić w PA tak, aby:

- $dow(\overline{m}, \overline{n})$ wyrażała fakt, że ciąg formuł o numerze m jest dowodem formuły o numerze n;
- w konsekwencji, $\exists x \;\; dow(x, \overline{n})$ stwierdzała, że formuła o numerze n jest twierdzeniem PA;
- $podst(\overline{m}, \overline{n}) = \overline{r}$ stwierdzała, że r jest numerem formuły otrzymanej z formuły (o jednej zmiennej wolnej) o numerze n przez podstawienie w miejsce tej zmiennej liczebnika $\overline{m}$.

Nieważne, jak skomplikowane są formuły $dow(x, y)$ oraz $podst(x, y) = z$; ważne, że wyrażane przez nie pojęcia są obliczalne.

Arytmetyzacja składni

Niech T będzie dowolną niesprzeczną teorią zawierającą arytmetykę PA (a więc umożliwiającą wyrażenie formuł $dow(x, y)$ oraz $podst(x, y) = z$). Korzystając z wyżej (skrótowo!) omówionego kodowania można pokazać:

- **Twierdzenie Gödla-Rossera.** T **nie jest** zupełna (istnieje zdanie jej języka takie, że ani ono, ani jego negacja nie jest twierdzeniem T);
- **Twierdzenia Tarskiego.** Prawdziwość formuł teorii T **nie jest** definiowalna w T.

Uwaga: definicja prawdy (autorstwa Alfreda Tarskiego), którą poznaliście na kursie logiki, była sformułowana w metajęzyku — języku istotnie silniejszym od języka przedmiotowego.

Arytmetyzacja składni

Dowód Twierdzenia Gödla-Rossera.

Niech $god(y)$ będzie skrótem dla formuły $\neg\exists x \;\; dow(x, podst(y, y))$ oraz niech n będzie numerem formuły $god(y)$.

Wtedy: formuła $god(\overline{n})$ stwierdza, że formuła o numerze, który jest denotacją termu $podst(\overline{n}, \overline{n})$ nie ma dowodu.

Ale term $podst(\overline{n}, \overline{n})$ oznacza właśnie numer formuły $god(\overline{n})$.

Widać więc, że formuła $god(\overline{n})$ stwierdza o sobie samej, że nie jest twierdzeniem.

Arytmetyzacja składni

Zatem: formuła $god(\overline{n})$ jest prawdziwa wtedy i tylko wtedy, gdy term $podst(\overline{n}, \overline{n})$ oznacza numer formuły nie mającej dowodu, tzn. wtedy i tylko wtedy, gdy $god(\overline{n})$ nie ma dowodu.

Gdyby więc $god(\overline{n})$ miała dowód, to byłaby nieprawdziwa, a więc w teorii T dałoby się udowodnić formułę nieprawdziwą, co jest sprzeczne z założeniem, że T jest niesprzeczna.

Stąd: formuła $god(\overline{n})$ jest prawdziwa, lecz nie posiada dowodu.

Z prawdziwości $god(\overline{n})$ wynika, ż $\neg god(\overline{n})$ jest fałszywa.

Stąd i z niesprzeczności T otrzymujemy, że również $\neg god(\overline{n})$ nie ma dowodu.

Arytmetyzacja składni

Dowód Twierdzenia Tarskiego.

Przypuśćmy, że język teorii T zawiera formułę $pr(x)$ wyrażającą własność prawdziwości formuł tej teorii, tj. taką, że:

$pr(\overline{m})$ jest prawdziwa wtedy i tylko wtedy, gdy m jest numerem formuły prawdziwej języka teorii T.

Pokażemy, że przypuszczenie to prowadzi do sprzeczności, a więc musi być odrzucone (w konsekwencji, dostajemy tezę Twierdzenia Tarskiego).

Niech $alf(x)$ będzie formułą: $\neg pr(podst(x, x))$ oraz niech r będzie numerem formuły $alf(x)$.

Arytmetyzacja składni

Wtedy term $podst(\overline{r}, \overline{r})$ oznacza numer formuły $alf(\overline{r})$.
Ale formuła $alf(\overline{r})$ jest tożsama z formułą $\neg pr(podst(\overline{r}, \overline{r}))$.

Formuła $alf(\overline{r})$ stwierdza, że formuła o numerze oznaczanym przez term $podst(\overline{r}, \overline{r})$ jest fałszywa.
Zatem $alf(\overline{r})$ stwierdza o sobie samej, że jest fałszywa.

Zatem formuła $alf(\overline{r})$ jest prawdziwa wtedy i tylko wtedy, gdy formuła o numerze oznaczanym przez term $podst(\overline{r}, \overline{r})$ jest nieprawdziwa.

Ale liczba oznaczana przez term $podst(\overline{r}, \overline{r})$ jest właśnie numerem formuły $alf(\overline{r})$.
Stąd: $alf(\overline{r})$ prawdziwa wtedy i tylko wtedy, gdy $alf(\overline{r})$ fałszywa.
SPRZECZNOŚĆ.

Arytmetyzacja składni

Uwagi.

- Twierdzenie udowodnione w 1930 roku przez Gödla zakładało ω-niesprzeczność teorii T. Założenie to osłabił (do niesprzeczności T) Rosser.
- Oba powyższe dowody wykorzystywały rozumowanie przekątniowe.
- Możliwe jest uprawianie metalogiki (metamatematyki) bez arytmetyzacji składni. Pokazał to niedawno Pan Profesor Andrzej Grzegorczyk, rozwijając oryginalne pomysły Alfreda Tarskiego dot. teorii konkatenacji.
- Ciekawostka prowincjonalna: Uniwersytet Poznański nie był zainteresowany zatrudnieniem Alfreda Tarskiego, jednego z największych logików wszystkich czasów.

Systemy przekonań

WIEDZA jest systemem przekonań. Systemy takie opisuje się m.in. przy użyciu pojęć logiki epistemicznej.
Operatory epistemiczne to np.:

- $\boldsymbol{B}$ — zdanie $\boldsymbol{B}p$ czytamy: *(rozważany podmiot) wierzy, że p*;
- $\boldsymbol{K}$ — zdanie $\boldsymbol{K}p$ czytamy *(rozważany podmiot) wie, że p*.

(gdzie p jest dowolnym zdaniem języka logiki epistemicznej). Zwykle zakłada się, że $\boldsymbol{K}p \equiv (p \wedge \boldsymbol{B}p)$.

Systemy epistemiczne są interesujące same przez się — w opisie systemów przekonań, w szczególności: racjonalnych świadomych przekonań.
Wspominamy tu o nich, ponieważ mają one także interesującą i ważną interpretację metalogiczną:
$\boldsymbol{B}p$ można interpretować jako *zdanie p jest dowodliwe w arytmetyce PA*.

Systemy przekonań

Wiesz już, że wiedza to system uzasadnionych, prawdziwych przekonań. Nadto wiesz, że ponieważ żyjesz na wolnej planecie, to możesz mieć przekonania, jakie sobie wybierzesz: nie ma zakazu, abyś wierzyła w co tylko zechcesz. Jako że już niedługo będziesz magistrem, a więc będziesz należeć do Elity Intelektualnej Rzeczpospolitej Polskiej, Unii Europejskiej i nie wiadomo, czego jeszcze, może czas na zadanie sobie kilku pytań, np.:

- Czy moje przekonania są trafne?
- Czy moje przekonania są niesprzeczne?
- Czy moje przekonania są zupełne?
- Co mogę sądzić o moich przekonaniach?
- Co mogę wiedzieć o moich przekonaniach?
- Czy moje przekonania są racjonalne?
- Czy jestem w pełni świadoma swoich przekonań?

Systemy przekonań

Przypuśćmy, że jesteś racjonalną, samoświadomą Istotą. Jak to przypuszczenie przełożyć na język logiki epistemicznej? Oto propozycja. Nazwiemy szczęściarzem epistemicznym każdą osobę S, której system przekonań spełnia warunki:

- (1a) S wierzy we wszystkie tautologie klasycznego rachunku zdań;
- (1b) system przekonań S jest domknięty na regułę *modus ponens*: jeśli S wierzy w p oraz wierzy w $p \rightarrow q$, to wierzy także w q;
- (2) dla dowolnych p oraz q, S wierzy w $(\boldsymbol{B}p \wedge \boldsymbol{B}(p \rightarrow q)) \rightarrow \boldsymbol{B}q$;
- (3) dla dowolnego p, jeśli S wierzy w p, to wierzy w $\boldsymbol{B}p$;
- (4) dla dowolnego p, S wierzy w $\boldsymbol{B}p \rightarrow \boldsymbol{B}\boldsymbol{B}p$.

Uwaga: rozważamy tylko osoby, które albo zawsze mówią prawdę, albo zawsze mówią fałsz.

Systemy przekonań

Każdą osobę, która spełnia jedynie warunki (1a) i (1b) nazwiemy prostaczkiem logicznym. Zatem, jeśli S jest prostaczkiem logicznym, to jego/jej system przekonań zawiera klasyczną logikę zdaniową, ale S może być tego nieświadom(a).

Powiemy, że osoba S jest:

- normalna, gdy jeśli wierzy w p, to wierzy też w $\boldsymbol{B}p$;
- regularna, gdy jeśli wierzy w $p \rightarrow q$, to wierzy też w $\boldsymbol{B}p \rightarrow \boldsymbol{B}q$;
- sprzeczna, gdy do jej systemu przekonań należy jakaś para zdań wzajem sprzecznych, lub — co na jedno wychodzi — *fałsz logiczny*, który oznaczamy przez $\perp$.

Uwaga: może bardziej właściwe byłoby mówienie o własnościach systemów przekonań, a nie osób.

Systemy przekonań

Można udowodnić, że: ($\bigstar$) dowolny szczęściarz epistemiczny S wie, że jeśli uwierzy w jakieś zdanie p oraz w jego negację $\neg p$, to stanie się sprzeczny.

O szczęściarzach epistemicznych można udowodnić wiele innych ciekawych rzeczy. Nie wszystkie z nich będą nam dalej potrzebne. Dodajmy może jedynie, że:

- każdy szczęściarz epistemiczny jest normalny, a nawet wie, że jest normalny;
- każdy szczęściarz epistemiczny jest regularny i o tym także wie;
- wreszcie, każdy szczęściarz epistemiczny jest przekonany o tym, że jest szczęściarzem epistemicznym; a zatem to jego przekonanie jest trafne i, w konsekwencji, każdy szczęściarz epistemiczny wie, że jest szczęściarzem epistemicznym.

Systemy przekonań

Za chwilę dowiesz się czegoś naprawdę frapującego o swoim systemie przekonań. Udowodnimy mianowicie:

Twierdzenie 1.
Przypuśćmy, że normalny prostaczek logiczny S wierzy w zdanie postaci $p \equiv \neg \boldsymbol{B} p$. Wtedy:

- (a) Jeśli S kiedykolwiek uwierzy w p, to stanie się sprzeczny.
- (b) Jeśli S jest szczęściarzem epistemicznym, to wie, iż jeśli kiedykolwiek uwierzy w p, to stanie się sprzeczny — tj. uwierzy w $\boldsymbol{B} p \rightarrow \boldsymbol{B} \perp$.
- (c) Jeśli S jest szczęściarzem epistemicznym i wierzy, że nie może być sprzeczny, to stanie się sprzeczny.

Systemy przekonań

Dowód Twierdzenia 1.

(a) Przypuśćmy, że S wierzy w p.

Będąc normalnym, uwierzy w $\boldsymbol{B}p$.

Nadto, ponieważ wierzy w p oraz wierzy w $p \equiv \neg \boldsymbol{B}p$,
więc musi uwierzyć w $\neg \boldsymbol{B}p$
(bo jest prostaczkiem logicznym).

A więc uwierzy jednocześnie w $\boldsymbol{B}p$ oraz w $\neg \boldsymbol{B}p$,
a stąd stanie się sprzeczny.

Systemy przekonań

(b) Przypuśćmy, że S jest szczęściarzem epistemicznym. Ponieważ jest wtedy prostaczkiem logicznym i wierzy w $p \equiv \neg \boldsymbol{B}p$, więc musi także wierzyć w $p \rightarrow \neg \boldsymbol{B}p$.

Nadto, S jest regularny, a stąd uwierzy w $\boldsymbol{B}p \rightarrow \boldsymbol{B}\neg \boldsymbol{B}p$. Wierzy też w $\boldsymbol{B}p \rightarrow \boldsymbol{B}\boldsymbol{B}p$ (ponieważ wie, że jest normalny).

Zatem S uwierzy w $\boldsymbol{B}p \rightarrow (\boldsymbol{B}\boldsymbol{B}p \wedge \boldsymbol{B}\neg \boldsymbol{B}p)$, które jest logiczną konsekwencją ostatnich dwóch zdań.

Wierzy również w $(\boldsymbol{B}\boldsymbol{B}p \wedge \boldsymbol{B}\neg \boldsymbol{B}p) \rightarrow \boldsymbol{B}\perp$ (na mocy ($\bigstar$), ponieważ dla dowolnego zdania X, S wierzy w $(\boldsymbol{B}X \wedge \boldsymbol{B}\neg X) \rightarrow \boldsymbol{B}\perp$, a więc wierzy w jego szczególny przypadek, gdzie X jest zdaniem $\boldsymbol{B}p$).

Gdy S już uwierzy jednocześnie w $\boldsymbol{B}p \rightarrow (\boldsymbol{B}\boldsymbol{B}p \wedge \boldsymbol{B}\neg \boldsymbol{B}p)$ oraz w $(\boldsymbol{B}\boldsymbol{B}p \wedge \boldsymbol{B}\neg \boldsymbol{B}p) \rightarrow \boldsymbol{B}\perp$, będzie musiał uwierzyć w $\boldsymbol{B}p \rightarrow \boldsymbol{B}\perp$ (ponieważ jest prostaczkiem logicznym).

Systemy przekonań

(c) Ponieważ S wierzy w $\boldsymbol{B}p \rightarrow \boldsymbol{B}\perp$ (jak właśnie udowodniliśmy), więc wierzy także w $\neg\boldsymbol{B}\perp \rightarrow \neg\boldsymbol{B}p$.

Załóżmy teraz, że S wierzy w $\neg\boldsymbol{B}\perp$ (wierzy, że nie może być sprzeczny).

Ponieważ wierzy też w $\neg\boldsymbol{B}\perp \rightarrow \neg\boldsymbol{B}p$ (jak właśnie widzieliśmy), więc uwierzy w $\neg\boldsymbol{B}p$.

A ponieważ wierzy również w $p \equiv \neg\boldsymbol{B}p$, więc uwierzy w p, a stąd stanie się sprzeczny, na mocy (a).

Systemy przekonań

Udowodniliśmy przed chwilą nie byle co, bo modalną (epistemiczną) wersję **II Twierdzenia Gödla** (o niedowodliwości niesprzeczności arytmetyki w samej arytmetyce).
Oczywiście był to dowód w postaci wielce uproszczonej — precyzyjny dowód wymagałby, powiedzmy, jednosemestralnego wykładu wstępnego.

W tej prezentacji korzystaliśmy z rozdziału 12 tłumaczenia książki Raymonda Smullyana *Na zawsze nierozstrzygnięte* (które ukazało się w 2007 roku nakładem *Książki i Wiedzy*).
Książka ta, podobnie jak inne z wymienionych uprzednio prac Smullyana, jest przykładem mistrzostwa w popularyzacji wiedzy (logicznej). Zachęcam do lektury.

Systemy przekonań

Przykład.
Przypuśćmy, że jesteś studentką teologii i że Twój Ulubiony Profesor teologii mówi do Ciebie:
Bóg istnieje wtedy i tylko wtedy, gdy nigdy nie uwierzysz, że Bóg istnieje.

Jeśli wierzysz profesorowi, to wierzysz w zdanie $g \equiv \neg \boldsymbol{B} g$, gdzie g jest zdaniem stwierdzającym, że Bóg istnieje.
Wtedy, zgodnie z Twierdzeniem 1, nie możesz wierzyć w swoją własną niesprzeczność bez popadnięcia w sprzeczność.

Oczywiście, możesz wierzyć we własną niesprzeczność, bez popadnięcia przy tym w sprzeczność — wystarczy, że przestaniesz ufać Twojemu Ulubionemu Profesorowi.
Coś za coś.

Systemy przekonań

Przy modalnej interpretacji dowodliwości nie mamy jednak takiej możliwości ucieczki, jak w powyższym przykładzie.
Widzieliśmy, że formuła $god(\overline{n})$, stwierdzająca swoją własną niedowodliwość w PA, jest prawdziwa, lecz dowodu w PA nie posiada.
Można pokazać, że twierdzeniem stosownego systemu modalnego (w którym reprezentujemy dowodliwość w PA) jest:

$$god(\overline{n}) \equiv \neg \boldsymbol{B} god(\overline{n}).$$

Na koniec tych rozważań o logice epistemicznej pokażę jeszcze, co wystarcza, aby każda z Was została — powiedzmy — **Miss World 2008**.
Będzie to przykład samospełniającego się przekonania.

Samospełniające się przekonania

Przypuśćmy, że:

- jesteś szczęściarą epistemiczną;
- osoby, które rozważamy albo zawsze mówią fałsz, albo zawsze mówią prawdę (i Ty wiesz, że tak jest);
- wierzysz swojemu chłopakowi, który prawdziwie (!) mówi:
  ($*$) *Jeśli uwierzysz, że zostaniesz Miss World 2008, to zostaniesz Miss World 2008.*
- wierzysz też mnie (JP), który mówi:
  ($**$) *Jeśli kiedykolwiek uwierzysz, że ja zawsze mówię prawdę, to zostaniesz Miss World 2008.*

Wtedy **zostaniesz Miss World 2008.** Cieszysz się?

Samospełniające się przekonania

Dla skrótu, przyjmijmy oznaczenia:

- k zastępuje zdanie stwierdzające, iż ja (JP) zawsze mówię prawdę;
- α zastępuje zdanie stwierdzające, że zostaniesz Miss World 2008.

Dowód składa się z dwóch części.
1. W pierwszej pokazujemy, że nasze założenia implikują $\boldsymbol{B}\alpha$. Jest to dowód założeniowy, dostępny dla każdej szczęściary epistemicznej.
Mamy udowodnić formułę:

$$(\bigstar) \qquad ((\boldsymbol{B}\alpha \rightarrow \alpha) \wedge (k \equiv (\boldsymbol{B}k \rightarrow \alpha))) \rightarrow \boldsymbol{B}\alpha.$$

Uwaga: zdanie k stwierdza, iż JP zawsze mówi prawdę; a więc prawdą jest, że JP wypowiada $(**)$ dokładnie wtedy, gdy prawdziwe jest $k \equiv (**)$, czyli dokładnie wtedy, gdy prawdziwe jest $k \equiv (\boldsymbol{B}k \rightarrow \alpha)$.

Samospełniające się przekonania

1.	$(\boldsymbol{B}\alpha \rightarrow \alpha) \wedge (k \equiv (\boldsymbol{B}k \rightarrow \alpha))$	założenie
2.	$\boldsymbol{B}\alpha \rightarrow \alpha$	OK: 1
3.	$k \equiv (\boldsymbol{B}k \rightarrow \alpha)$	OK: 1
4.	$k \rightarrow (\boldsymbol{B}k \rightarrow \alpha)$	OR: 3
5.	$(\boldsymbol{B}k \rightarrow \alpha) \rightarrow k$	OR: 3
6.1.	k	założenie dodatkowe
6.2.	$\boldsymbol{B}k \rightarrow \alpha$	MP: 4, 6.1.
6.3.	$\boldsymbol{B}k$	6.1. i warunek **(3)**
6.4.	α	MP: 6.2., 6.3.
7.	$k \rightarrow \alpha$	6.1.→6.4.
8.	$\boldsymbol{B}(k \rightarrow \alpha)$	7 i warunek **(3)**
9.	$\boldsymbol{B}k \rightarrow \boldsymbol{B}\alpha$	8 i warunki **(1a)** i **(2)**
10.	$\boldsymbol{B}k \rightarrow \alpha$	2, 9 i warunki **(1b)**, **(1a)** (prawo sylog. hipotet.)
11.	k	MP: 5, 10
12.	$\boldsymbol{B}k$	11 i warunek **(3)**
13.	α	MP: 10, 12
14.	$\boldsymbol{B}\alpha$	13 i warunek **(3)**.

Samospełniające się przekonania

2. Ponieważ proroctwo $(*)$ Twojego chłopaka (tj. zdanie $\boldsymbol{B}\alpha \rightarrow \alpha$) jest z założenia prawdziwe, a powyższy dowód formuły $(\bigstar)$ pokazuje, iż nasze założenia implikują $\boldsymbol{B}\alpha$, więc na mocy reguły odrywania otrzymujemy α, czyli tezę.

Zostaniesz Miss World 2008!!!
Cieszysz się???

Uwaga: powyższy dowód był przykładem dowodu wprost. Aby pokazać, że zostaniesz Miss World 2008 nie musieliśmy odwoływać się do **absurdu**. Cieszysz się?

Samospełniające się przekonania

Ciekawostka prowincjonalna. 16 maja 2005 roku odbyły się demokratyczne wybory Dyrektora Instytutu Językoznawstwa UAM.
Dwa tygodnie wcześniej, na Seminarium Zakładu Logiki Stosowanej UAM, odczyt ***Kto będzie Dyrektorem Instytutu Językoznawstwa UAM?*** wygłosiła Pani Dr **Alice Ann Hunter** (*Department of Independent Logic*, **King David University**, Negev Desert).
Korzystając z twierdzeń logiki epistemicznej (z Twierdzenia Löba), Dr Hunter trafnie przewidziała wynik wyborów. Jak się domyślasz, dowód był podobny do podanego wyżej dowodu, że zostaniesz Miss World 2008.
Tekst odczytu dostępny na stronie:

www.logic.amu.edu.pl

W 2008 roku odbędą się wybory nowego Dyrektora IJ UAM. Postaramy się wcześniej zaprosić dr Hunter. Może znów udowodni coś ciekawego?

Twierdzenia metalogiczne

Metatwierdzenia to twierdzenia o teoriach, systemach przekonań, itp. Metatwierdzenia logiczne mówią więc coś o systemie logiki klasycznej, o innych systemach logicznych, o poszczególnych teoriach, itd.
Na kursie logiki poznaliście przykłady metatwierdzeń:

- Twierdzenie o pełności FOL: zbiór tez FOL jest identyczny ze zbiorem praw FOL.
- Twierdzenie o istnieniu modelu: teoria T jest niesprzeczna wtedy i tylko wtedy, gdy ma model.
- Twierdzenie o zwartości: zbiór formuł X ma model wtedy i tylko wtedy, gdy każdy skończony podzbiór X ma model.
- Twierdzenie Churcha: Klasyczny Rachunek Predykatów nie jest rozstrzygalny.
- Twierdzenie Löwenheima – Skolema: jeśli teoria T ma model, to ma model przeliczalny.

Twierdzenia metalogiczne

Niektóre metatwierdzenia logiczne mają charakter limitacyjny: wskazują na pewne ograniczenia obiektywnej (!) natury, którym podlega metoda dedukcyjna.

Omówione dziś: Twierdzenie Tarskiego, I oraz II Twierdzenie Gödla, Twierdzenie Churcha — to przykłady twierdzeń limitacyjnych. Wskazują one, że **pełna** realizacja tzw. Programu Hilberta nie jest możliwa. W programie tym postulowano, m.in.:

- dokonanie aksjomatyzacji wiedzy matematycznej;
- wykazanie, za pomocą metod efektywnych niesprzeczności matematyki;
- wykazanie, iż wiedza matematyczna jest zupełna.

Twierdzenia metalogiczne

Nadzieje na realizację tego programu wypływały m.in. z wiary, iż każdy problem matematyczny można rozstrzygnąć (metodami efektywnymi). Mamy jednak:

- Twierdzenie Churcha: nierozstrzygalność KRP;
- Twierdzenie Tarskiego: niedefiniowalność pojęcia prawdy (formuł teorii w niej samej);
- Twierdzenie Löwenheima – Skolema: FOL nie odróżnia mocy nieskończonych;
- I twierdzenie Gödla: niezupełność PA;
- II twierdzenie Gödla: niedowodliwość niesprzeczności PA w niej samej.;
- Twierdzenie o Niewspółmożliwości: kategoryczny opis modelu zamierzonego nie jest współmożliwy z pełnością systemu logiki.

Czy zatem powinniśmy posługiwać się jakąś inną, mocniejszą niż FOL, logiką?

Twierdzenia metalogiczne

Moc wyrażania logiki można zwiększyć na różne sposoby, np.:

- uznać za dopuszczalne reguły wnioskowania o nieskończonej liczbie przesłanek (jak np. w ω-regule);
- wprowadzić nowe stałe logiczne (np. nowe kwantyfikatory);
- dopuścić możliwość kwantyfikacji predykatów (logiki wyższych rzędów).

Jednak większa moc wyrażania jest ujemnie skorelowana z ufnością w aparat inferencyjny: np. w logice II rzędu można opisać modele zamierzone w sposób kategoryczny, ale nie zachodzi w tej logice twierdzenie o pełności.

Twierdzenia metalogiczne

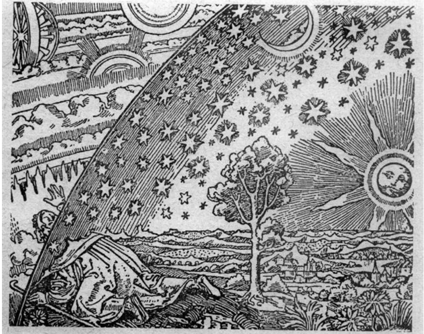
Przy stosownym rozumieniu pojęcia logiki udowodnić można Twierdzenia Lindströma:

- Każda logika, w której zachodzi Twierdzenie o Pełności oraz Twierdzenie Löwenheima – Skolema, jest równoważna FOL.
- Każda logika, w której zachodzi Twierdzenie o Zwartości oraz Twierdzenie Löwenheima – Skolema, jest równoważna FOL.

Uważa się, że powyższe twierdzenia stanowią mocny argument za tzw. Tezą Pierwszego Rzędu, głoszącą, iż to właśnie FOL jest WŁAŚCIWĄ logiką, której powinniśmy używać.
To normatywne zalecenie **nie dotyczy** Matematyków.

Koniec!

W następnym, ostatnim już odcinku naszego Flirtu Intelektualnego pogwarzymy o Granicach Poznania:



Biblijny obraz budowy niebios przedstawiony zgodnie z opowiadaniem średniowiecznego zakonnika, który rzekomo znalazł „kraniec Ziemi"