

Szczęściarze epistemiczni

ZLwFiPM XII

Jerzy Pogonowski

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

Szklarska Poręba, V 2007

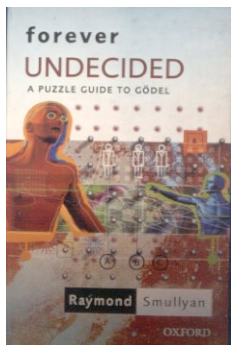
Po co tego słuchać?

Pokazujemy wybrane fragmenty tłumaczenia książki Raymonda Smullyana *Forever Undecided. A Puzzle Guide to Gödel.*, które ukaże się w 2007 roku nakładem *Książki i Wiedzy*, pod tytułem *Na Zawsze Nierozstrzygnięte. Zagadkowy Przewodnik Po Twierdzeniach Gödla.*

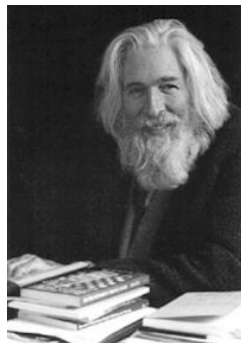
Obok zagadek o Rycerzach (mówiących zawsze prawdę) oraz Łotrach (mówiących zawsze fałsz), książka zawiera zagadki logiczne, w których w formie popularnej przedstawia się **logikę epistemiczną** oraz **logikę dowodliwości**.

Logika epistemiczna jest Państwu dobrze znana, o logice dowodliwości (**logice Gödla-Löba**) także każdy coś słyszał. Proszę zatem traktować niniejszą prezentację jako rozrywkę. Chciałbym przede wszystkim zwrócić uwagę na mistrzostwo Smullyana w popularyzowaniu wiedzy logicznej.

Forever Undecided



Forever Undecided



Raymond Smullyan

Spis treści „Forever Undecided”

Przedmowa	7
I. Możesz zostać zaskoczona!	11
1. Diabelska zagadka	13
2. Zaskoczona?	18
II. Logika kłamstwa i szczerości	23
3. Rachmistrz spisu ludności	25
4. W poszukiwaniu Ony	32
5. Międzyplanetarne zamieszanie	37
III. Rycerze, łotrzy i logika zdaniowa	45
6. Drobinka logiki zdaniowej	47
7. Rycerze, łotrzy i logika zdaniowa	57
8. Domknięcie logiczne i niesprzeczność	63

Spis treści „Forever Undecided”

IV. Bądźmy ostrożni!	73
9. Paradoksalne?	75
10. Problem się pogłębia	85
V. Kłopoty z niesprzecznością	93
11. Logicy, którzy rozumują o sobie	95
12. Kłopoty z niesprzecznością	105
13. Systemy Gödłowskie	112
14. Więcej kłopotów z niesprzecznością	117
VI. Samospełniające się przekonania i twierdzenie Löba	125
15. Samospełniające się przekonania	127
16. Diament Rajaha	138
17. Wyspa Löba	145

Spis treści „Forever Undecided”

VII. Na głębokich wodach	153
18. Myślaki typu G	155
19. Skromność, zwrotność i stabilność	165
VIII. Nie można rozstrzygnąć!	173
20. Na zawsze nierozstrzygnięte	175
21. Więcej bezrozstrzygnięć	182
IX. Światy możliwe	189
22. To niekoniecznie tak!	191
23. Światy możliwe	198
24. Od konieczności do dowodliwości	204

Spis treści „Forever Undecided”

X. Istota sprawy	209
25. Gödelowate uniwersum	211
26. Pewne godne uwagi maszyny logiczne	219
27. Samostosowalne systemy modalne	231
XI. Finale	239
28. Systemy modalne, maszyny i myślaki	241
29. Niektóre dziwaczne myślaki!	247
30. W retrospekcji	255
Słowo od tłumacza	260

- Boolos, G. 1993. *The Logic of Provability*. Cambridge University Press.
 - Smullyan, R. 1992. *Gödel's Incompleteness Theorems*. Oxford University Press.
-
- Jacek Hawranek: *Aspekty algebraiczne systemu modalnego Gödla–Löba*. Wydawnictwo Uniwersytetu Wrocławskiego, Wrocław, 1994.
 - Andrzej Indrzejczak: *Hybrydowe systemy dedukcyjne w logikach modalnych*. Wydawnictwo Uniwersytetu Łódzkiego, Łódź, 2006.
 - Jerzy Perzanowski: *Logiki modalne a filozofia*. Uniwersytet Jagielloński, Rozprawy Habilitacyjne nr 156, Kraków, 1989.
 - Kazimierz Świrydowicz: *Podstawy logiki modalnej*. Wydawnictwo Naukowe UAM, Poznań, 2004.

Plan na dziś:

- **Systemy przekonań.** Kto jest prostaczką logiczną?
- **Poziomy samoświadomości.** Kto jest szczęściarzem epistemicznym?
- **II Twierdzenie Gödla.**
Czy możesz wiedzieć, że twój system przekonań jest niesprzeczny, bez popadnięcia przy tym w sprzeczność?
- **Twierdzenie Löba i samospełniające się przekonania.**
Kiedy *wishful thinking* ma wartość?
- **I Twierdzenie Gödla i Twierdzenie Rossera (o niezupełności).**
Czy łatwy jest los *Besserwissera*?
- **Twierdzenie Tarskiego.**
Czy *dictum*: ***Doctrina multiplex, veritas una!*** jest mrzonką?
- **Dodatek:** Maszyny logiczne Smullyana.

Systemy przekonań

Notacja. Operatory epistemiczne to np.:

- B — zdanie Bp czytamy: (rozważany podmiot) *wierzy*, że p ;
- K — zdanie Kp czytamy (rozważany podmiot) *wie*, że p .

(gdzie p jest dowolnym zdaniem języka logiki epistemicznej). Zwykle zakłada się, że $Kp \equiv (p \wedge Bp)$.

Systemy epistemiczne są interesujące same przez się — w opisie systemów przekonań, w szczególności: racjonalnych świadomych przekonań. Mają one także interesującą i ważną interpretację metalogiczną:

Bp można interpretować jako *zdanie p jest dowodliwe w arytmetyce PA*.

Uwaga. Angielski termin *reasoner* oddaję przez polski neologizm *myślak*.

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 $(Bp \wedge B(p \rightarrow q)) \rightarrow Bq$;
- (3) dla dowolnego p , jeśli S wierzy w p , to wierzy w Bp ;
- (4) dla dowolnego p , S wierzy w $Bp \rightarrow BBp$.

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

Poziomy samoświadomości

Każdą osobę, która spełnia jedynie warunki (1a) i (1b) nazwiemy **prostaczkim logicznym**. Zatem, jeśli S jest prostaczkim 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 Bp ;
- **regularna**, gdy jeśli wierzy w $p \rightarrow q$, to wierzy też w $Bp \rightarrow Bq$;
- **sprzeczna**, gdy do jej systemu przekonań należy jakaś para zdań wzajemnie 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**.

Poziomy samoświadomości

Można udowodnić, że: (*) 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.

Poziomy samoświadomości

Smullyan rozważa pięć typów myślaków, o wstępujących poziomach samoświadomości:

- Typ 1: prostaczek logiczny.
- Typ 1*: prostaczek logiczny, który, jeśli uwierzył w $p \rightarrow q$, to uwierzy, że jeśli uwierzył w p , to uwierzy w q .
- Typ 2: prostaczek logiczny, który wierzy we wszystkie zdania postaci $(Bp \wedge B(p \rightarrow q)) \rightarrow Bq$.
- Typ 3: myślak typu 2, który, jeśli wierzy w p , to wierzy w Bp .
- Typ 4: szczęściarz epistemiczny, tj. normalny i regularny prostaczek logiczny, który wierzy we wszystkie zdania postaci $Bp \rightarrow BBp$, czyli wierzy, że jest normalny.

Uwaga. Terminy: prostaczek logiczny oraz szczęściarz epistemiczny nie występują w Forever Undecided; wprowadzamy je na użytek tej prezentacji.

Poziomy samoświadomości

Z podanych definicji wynika, że:

- Każdy prostaczek logiczny jest myślakiem typu 1^* .
- Każdy myślak typu 1^* jest regularnym prostaczkiem logicznym (i *vice versa*).
- Każdy myślak typu 2 wie, że jest typu 1^* .
- Myślaki typu 3 to dokładnie normalne myślaki typu 2.
- Dla $1 \leq n < 4$: każdy myślak typu n jest też myślakiem typu $n + 1$.
- $1 < n \leq 4$: każdy myślak typu n wierzy, że jest myślakiem typu $n - 1$.

Uwaga. Ponieważ każdy szczęściarz epistemiczny **wie**, że jest szczęściarzem epistemicznym, więc stanowi on zwieńczenie hierarchii samoświadomych myślaków. Inaczej mówiąc, gdybyśmy chcieli zdefiniować myślaka typu 5 jako takiego, który jest typu 4 i wierzy, iż jest typu 4, to otrzymalibyśmy jedynie myślaka typu 4.

II Twierdzenie Gödla

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

Twierdzenie 1.

Przypuśćmy, że normalny prostacek logiczny S wierzy w zdanie postaci $p \equiv \neg Bp$. 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 $Bp \rightarrow B \perp$.
- (c) Jeśli S jest szczęściarzem epistemicznym i wierzy, że nie może być sprzeczny, to stanie się sprzeczny.

II Twierdzenie Gödla

Dowód Twierdzenia 1.

(a) Przypuśćmy, że S wierzy w p .

Będąc normalnym, uwierzy w Bp .

Nadto, ponieważ wierzy w p oraz wierzy w $p \equiv \neg Bp$,
więc musi uwierzyć w $\neg Bp$
(bo jest prostaczką logiczną).

A więc uwierzy jednocześnie w Bp oraz w $\neg Bp$,
a stąd stanie się sprzeczny.

II Twierdzenie Gödla

(b) Przypuśćmy, że S jest szczęściarzem epistemicznym. Ponieważ jest wtedy prostaczkim logicznym i wierzy w $p \equiv \neg Bp$, więc musi także wierzyć w $p \rightarrow \neg Bp$.

Nadto, S jest regularny, a stąd uwierzy w $Bp \rightarrow B\neg Bp$. Wierzy też w $Bp \rightarrow BBp$ (ponieważ wie, że jest normalny).

Zatem S uwierzy w $Bp \rightarrow (BBp \wedge B\neg Bp)$, które jest logiczną konsekwencją ostatnich dwóch zdań.

Wierzy również w $(BBp \wedge B\neg Bp) \rightarrow B \perp$ (na mocy (*), ponieważ dla dowolnego zdania X , S wierzy w $(BX \wedge B\neg X) \rightarrow B \perp$, a więc wierzy w jego szczególny przypadek, gdzie X jest zdaniem Bp).

Gdy S już uwierzy jednocześnie w $Bp \rightarrow (BBp \wedge B\neg Bp)$ oraz w $(BBp \wedge B\neg Bp) \rightarrow B \perp$, będzie musiał uwierzyć w $Bp \rightarrow B \perp$ (ponieważ jest prostaczkim logicznym).

II Twierdzenie Gödla

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

Założmy teraz, że S wierzy w $\neg B \perp$ (wierzy, że nie może być sprzeczny).

Ponieważ wierzy też w $\neg B \perp \rightarrow \neg Bp$ (jak właśnie widzieliśmy), więc uwierzy w $\neg Bp$.

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

II Twierdzenie Gödla

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*.

Przykład teologiczny

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 $g \equiv \neg \mathbf{B}g$, 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.

Modalna interpretacja dowodliwości

Przy modalnej interpretacji **dowodliwości** nie mamy jednak takiej możliwości ucieczki, jak w powyższym przykładzie.

Wiadomo, że formuła $god(\bar{n})$, stwierdzająca swoją własną niedowodliwość w PA, jest prawdziwa, lecz dowodu w PA nie posiada.

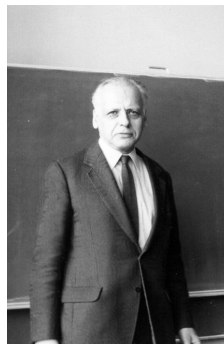
Można pokazać, że twierdzeniem stosownego systemu modalnego (w którym reprezentujemy dowodliwość w PA) jest:

$$god(\bar{n}) \equiv \neg \mathbf{B} god(\bar{n}).$$

Wishful thinking

Pokażemy teraz, co
wystarcza, aby każda z
Uroczych Pań została —
powiedzmy — **Miss World
2007**.

Będzie to przykład
samospełniającego się
przekonania.



Martin Hugo Löb

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 2007, to zostaniesz Miss World 2007.*
- wierzysz też mnie (JP), który mówi:
(**) *Jeśli kiedykolwiek uwierzysz, że ja zawsze mówię prawdę, to zostaniesz Miss World 2007.*

Twierdzenie 2.

Przy powyższych założeniach *zostaniesz Miss World 2007*. 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 2007.

Dowód składa się z dwóch części.

1. W pierwszej pokazujemy, że nasze założenia implikują $B\alpha$. Jest to dowód założeniowy, dostępny dla każdej szczęściary epistemicznej.

Mamy udowodnić formułę:

$$(\star) \quad ((B\alpha \rightarrow \alpha) \wedge (k \equiv (Bk \rightarrow \alpha))) \rightarrow B\alpha.$$

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 $k \equiv (Bk \rightarrow \alpha)$.

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

Samospełniające się przekonania

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

Zostaniesz Miss World 2007!!!

Cieszysz się???

Uwaga. Powyższy dowód był przykładem dowodu wprost. Aby pokazać, że zostaniesz Miss World 2007 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 2007. Tekst odczytu dostępny na stronie:

www.logic.amu.edu.pl/seminarium.html

I Twierdzenie Gödla

Myślak jest nazywany **stabilnym**, jeśli dla każdego zdania p , jeśli wierzy on w Bp , to wierzy też w p .

Powiemy, że system przekonań myślaka jest **niezupełny**, jeśli istnieje co najmniej jedno zdanie p takie, że myślak nigdy nie uwierzy w p ani też nigdy nie uwierzy w $\neg p$ (pozostanie na zawsze niezdecydowany, czy p jest prawdziwe, czy fałszywe).

Systemy przekonań, które nie są niezupełne, nazywamy **zupełnymi**. Osoby, które władają takimi systemami przekonań, są dość uciążliwe w kontaktach społecznych — każda taka osoba jest **Besserwisserem**, kimś kto na każdy pogląd ma wyrobione zdanie, pozbawiony jest wątpliwości.

Gdy zajmujemy się **systemami twierdzeń** raczej niż **zespołami przekonań**, to **systemami typu 1** nazwiemy te, które spełniają warunki 1a i 1b podane wyżej.

I Twierdzenie Gödla

Normalny prostaczek logiczny przybywa na Wyspę Rycerzy i Łotrów i wierzy w reguły wyspy. (To, czy reguły wyspy rzeczywiście obowiązują, czy nie, jest bez znaczenia.)

Spotyka tubylca, który mówi:

„Nigdy nie uwierzysz, że jestem rycerzem.”

Udowodnimy, że zachodzi wtedy:

Twierdzenie 3.

Jeśli myślak jest jednocześnie niesprzeczny i stabilny, to jego system przekonań jest niezupełny. Dokładniej mówiąc, znajdziemy zdanie p takie, że zachodzą następujące dwa warunki:

- (a) Jeśli myślak jest niesprzeczny, to nigdy nie uwierzy w p .
- (b) Jeśli myślak jest jednocześnie niesprzeczny i stabilny, to nigdy nie uwierzy w $\neg p$.

I Twierdzenie Gödla

Zdanie p o które chodzi jest po prostu zdaniem k — zdaniem stwierdzającym, że tubylec jest rycerzem.

Tubylec wygłosił $\neg Bk$, a więc myślak uwierzy w $k \equiv \neg Bk$.

(a) Przypuśćmy, że myślak wierzy w k . Wtedy, będąc normalnym, uwierzy w Bk . Uwierzy też w $\neg Bk$ (ponieważ wierzy w k oraz wierzy w $k \equiv \neg Bk$ i jest prostaczką logiczną), a stąd stanie się sprzeczny. Zatem, jeśli jest niesprzeczny, to nigdy nie uwierzy w k .

(b) Przypuśćmy, że myślak jest prostaczką logiczną i wierzy w $k \equiv \neg Bk$, wtedy wierzy też w $\neg k \equiv Bk$. Przypuśćmy teraz, że kiedykolwiek uwierzy on w $\neg k$. Wtedy uwierzy w Bk . Jeśli jest stabilny, to uwierzy w k i stąd stanie się sprzeczny (ponieważ wierzy w $\neg k$). Zatem, jeśli jest jednocześnie stabilny i niesprzeczny, to nigdy nie uwierzy w $\neg k$. Podsumowując, jeśli jest on jednocześnie stabilny i niesprzeczny, to nigdy nie uwierzy że tubylec jest rycerzem i nigdy nie uwierzy, że tubylec jest łotrem.

I Twierdzenie Gödla

To samo rozumowanie, którego użyto w rozwiązaniu powyższego problemu, gdy zastosować je do systemów matematycznych raczej niż do myślaków, ustanawia następującą postać **Pierwszego Twierdzenia Gödla o Niezupełności**:

Twierdzenie 4. Dowolny niesprzeczny, normalny, stabilny system Gödłowski musi być niezupełny. Dokładniej, jeśli S jest normalnym systemem typu 1, a p jest zdaniem takim, że $p \equiv \neg \mathbf{B}p$ jest dowodliwe w S , to jeśli S jest niesprzeczny, to p nie jest dowodliwe w S , a jeśli S jest dodatkowo stabilny, to $\neg p$ również nie jest dowodliwe w S .

Zdanie p nazywamy **nierozstrzygalnym** w systemie S , jeśli ani p ani jego negacja $\neg p$ nie jest dowodliwe w S . Zatem Pierwsze Twierdzenie Gödla o Niezupełności mówi nam, że dla dowolnego niesprzecznego, normalnego, stabilnego systemu Gödłowskiego S , musi zawsze istnieć co najmniej jedno zdanie p , które, choć **wyrażalne** w języku S , nie jest **rozstrzygalne** w S — nie można w S udowodnić ani tego zdania, ani jego negacji.

I Twierdzenie Gödla

Dla dowolnej własności P liczb, zdanie stwierdzające, że istnieje co najmniej jedna liczba n mająca własność P zapisujemy: $\exists nP(n)$.

Przypuśćmy, że mamy system matematyczny i własność P taką, że zdanie $\exists nP(n)$ jest dowodliwe w systemie, a jednak dla każdego poszczególnego n zdanie $\neg P(n)$ jest dowodliwe — to jest, wszystkie z nieskończenie wielu zdań $\neg P(0), \neg P(1), \neg P(2), \dots, \neg P(n), \dots$ są dowodliwe.

Oznacza to, z jednej strony, że w systemie można udowodnić zdanie ogólne stwierdzające, że jakaś liczba ma własność P , a jednak o każdej *poszczególnej* liczbie n można udowodnić, że liczba ta owej własności nie posiada!

Systemy takie nazywane są *ω -sprzecznymi*.

I Twierdzenie Gödla

Pojęcie ω -sprzeczności zostało kiedyś zabawnie scharakteryzowane przez matematyka Paula Halmosa, który zdefiniował ω -sprzeczną matkę jako taką, która mówi swojemu dziecku: „Jest coś, co możesz robić, ale nie możesz robić tego, nie możesz robić tamtego, nie możesz robić owego, ...” Dziecko pyta: „Ale, mamusiu, czy jest *cokolwiek* co mógłbym robić?” Matka odpowiada: „O tak, ale nie jest to to, ani tamto, ani owo, ...”

System jest nazywany ω -niesprzecznym, jeśli nie jest on ω -sprzeczny. Tak więc dla systemu ω -niesprzecznego, jeśli $\exists n P(n)$ jest dowodliwe, to istnieje co najmniej jedna liczba n taka, że zdanie $\neg P(n)$ *nie jest* dowodliwe.

Sprzeczny system typu 1 jest również ω -sprzeczny, ponieważ w sprzecznym systemie typu 1 wszystkie zdania są dowodliwe.

I Twierdzenie Gödla

We wszystkich dotąd rozważanych problemach, *kolejność* w której myślak wierzył w różnorakie zdania nie odgrywała roli. W pozostałych problemach w tej części, kolejność ta odgrywa rolę pierwszorzędą.

Myślak przybywa na Wyspę Rycerzy i Łotrów pewnego dnia, który nazwiemy dniem numer 0. Następny dzień jest dniem numer 1, następny dniem numer 2, i tak dalej.

Dla każdej liczby naturalnej n mamy więc dzień numer n (n -ty dzień) i zakładamy, że myślak jest nieśmiertelny i ma przed sobą nieskończenie wiele dni.

I Twierdzenie Gödla

Dla każdej liczby naturalnej n i dowolnego zdania p niech $B_n p$ będzie zdaniem stwierdzającym, że myślak uwierzył w p w jakimś momencie n -tego dnia.

Zdanie Bp jest, jak zwykle, zdaniem stwierdzającym, że myślak uwierzy w p tego lub innego dnia, lub, co na jedno wychodzi, zdaniem $\exists n B_n p$ (istnieje n takie, że myślak uwierzy w p n -tego dnia).

Nazwiemy myśłaka *ω -sprzecznym*, jeśli istnieje co najmniej jedno zdanie p takie, że myślak (kiedyś) wierzy w Bp , a jednak dla każdego n wierzy on (kiedyś) w $\neg B_n p$.

Myśłaka nazywamy *ω -niesprzecznym*, jeśli nie jest on ω -spreczny.

I Twierdzenie Gödla

Rozważmy teraz myślaka, który spełnia następujące trzy warunki.

- **Warunek C_1 .** Jest on typu prostaczkim logicznym.
- **Warunek C_2 .** Dla dowolnej liczby naturalnej n i dowolnego zdania p : (a) jeśli myślak wierzy w p n -tego dnia, to (prędzej czy później) uwierzy w $B_n p$; (b) jeśli nie wierzy on w p n -tego dnia, to (prędzej czy później) uwierzy w $\neg B_n p$. (Oddajemy w ten sposób, że myślak śledzi to, w jakie zdania wierzył, a w jakie nie wierzył we wszystkich dniach poprzednich.)
- **Warunek C_3 .** Dla dowolnych n oraz p myślak wierzy w zdanie $B_n p \rightarrow B p$ (które, oczywiście, jest zdaniem prawdziwym).

Następujący problem jest bardzo zbliżony do oryginalnego sformułowania Gödla jego Pierwszego Twierdzenia o Niezupełności.

I Twierdzenie Gödla

Myślak spełniający powyższe trzy warunki przybywa na Wyspę Rycerzy i Łotrów i wierzy w reguły wyspy. Spotyka tubylca, który mówi mu:

„Nigdy nie uwierysz, że jestem rycerzem.”

Udowodnimy, że zachodzi wtedy:

Twierdzenie 5.

- (a) Jeśli myślak jest (prosto) niesprzeczny, to nigdy nie uwierzy, że tubylec jest rycerzem.
- (b) Jeśli myślak jest ω -niesprzeczny, to nigdy nie uwierzy, że tubylec jest łotrem.

Zatem jeśli myślak jest ω -niesprzeczny (a stąd także prosto niesprzeczny), to pozostanie na zawsze niezdecydowany co do tego, czy tubylec jest rycerzem, czy też łotrem.

I Twierdzenie Gödla

Najłatwiejszym sposobem rozwiązania obecnego problemu będzie pokazanie, że dowolny myślak spełniający warunki 1, 2 oraz 3 musi być normalny, a jeśli jest ω -sprzeczny, to musi być też stabilny.

(a) Pokazujemy, że jest on normalny.

Przypuśćmy, że wierzy on w p .

Wtedy dla pewnego n , wierzy on n -tego dnia w p .

Wtedy, na mocy punktu (a) z warunku 2, uwierzy w $B_n p$.

Wierzy także w $B_n p \rightarrow B p$ (na mocy warunku 3), a więc będąc typu 1 (warunek 1) uwierzy w $B p$.

Zatem jest normalny.

I Twierdzenie Gödla

(b) Przypuśćmy teraz, że jest on ω -niesprzeczny.

Pokażemy, że jest stabilny.

Przypuśćmy, że wierzy on w Bp .

Jeśli nigdy nie uwierzy w p , to dla każdej liczby n , nie wierzy on w p n -tego dnia, a stąd na mocy punktu (b) z warunku 2, dla każdego n wierzy on w $\neg B_n p$.

Ale ponieważ wierzy on w Bp , więc stanie się wtedy ω -sprzeczny.

Zatem, jeśli jest on ω -niesprzeczny i wierzy w Bp , to musi wierzyć w p tego lub innego dnia.

Dowodzi to, że jeśli jest on ω -niesprzeczny, to musi być stabilny (zakładając, że spełnia on warunki C_1 , C_2 , C_3 — lub nawet tylko (b) z warunku C_2).

Zatem, na mocy Twierdzenia 4, pozostanie on na zawsze niezdecydowany.

Twierdzenie Rossera

Dla dowolnych zdań p oraz q , powiemy, że myślak *uwierzył w p wcześniej niż (zanim) uwierzył w q* , jeśli jest taki dzień, w którym wierzy on w p , a jeszcze nie uwierzył w q . Jeśli myślak *nigdy* nie uwierzy w q , ale uwierzył w p (tego lub innego dnia), to uznajemy, iż *prawdziwe* jest, że uwierzył w p wcześniej, niż uwierzył w q . (Innymi słowy, nie musi on wcale kiedykolwiek uwierzyć w q , aby uwierzyć w p wcześniej niż uwierzyć w q .) Niech $Bp < Bq$ będzie zdaniem stwierdzającym, że myślak uwierzył w p wcześniej niż uwierzył w q . Jeśli $Bp < Bq$ jest prawdziwe, to oczywiście $Bq < Bp$ jest fałszywe.

Zdefiniujemy myślaka *Rosserowskiego* jako prostaczkę logicznego, dla którego zachodzi następujący warunek:

Warunek R. Dla dowolnych zdań p oraz q , jeśli myślak uwierzył w p pewnego dnia, w którym jeszcze nie uwierzył w q , to (wcześniej czy później) uwierzy on w $Bp < Bq$ oraz w $\neg(Bq < Bp)$.

Twierdzenie Rossera

Myślak Rosserowski przybywa na Wyspę Rycerzy i Łotrów i wierzy w reguły wyspy. Spotyka tubylca, który mówi mu:

„Nigdy nie uwierzysz wcześniej, że jestem rycerzem, niż uwierzysz, że jestem łotrem.”

(Oddając to symbolicznie, tubylec wygłasza zdanie $\neg(Bk < B\neg k)$.)
Udowodnimy:

Twierdzenie 6.

Jeśli myślak jest *po prostu* niesprzeczny, to musi na zawsze pozostać niezdecydowany, czy tubylec jest rycerzem, czy łotrem.

Twierdzenie Rossera

Ponieważ tubylec stwierdził $\neg(\mathbf{B}k < \mathbf{B}\neg k)$, więc myślak uwierzy w $k \equiv \neg(\mathbf{B}k < \mathbf{B}\neg k)$. Przypuśćmy, że myślak jest (prosto) niesprzeczny. Mamy pokazać, że nigdy nie uwierzy w k i nigdy nie uwierzy w $\neg k$.

(a) Przypuśćmy, że kiedyś uwierzył w k . Ponieważ jest niesprzeczny, więc nigdy nie uwierzy w $\neg k$, a stąd uwierzy w k wcześniej niż uwierzy w $\neg k$. Stąd, uwierzy w $\mathbf{B}k < \mathbf{B}\neg k$ (na mocy warunku R). Ale wierzy też w $k \equiv \neg(\mathbf{B}k < \mathbf{B}\neg k)$, a więc uwierzy w $\neg k$, a wierząc już w k stanie się sprzeczny! Tak więc, jeśli jest niesprzeczny, to nigdy nie uwierzy w k .

(b) Przypuśćmy, że kiedyś uwierzył w $\neg k$. Będąc niesprzeczny, nigdy nie uwierzy w k , a stąd uwierzy w $\neg k$ wcześniej niż uwierzy w k , a stąd na mocy warunku R uwierzy w $\neg(\mathbf{B}k < \mathbf{B}\neg k)$. Ale wierzy on w $k \equiv \neg(\mathbf{B}k < \mathbf{B}\neg k)$, a więc uwierzy wtedy w k i stanie się sprzeczny. A zatem, jeśli jest niesprzeczny, to nie może także uwierzyć w $\neg k$.

Twierdzenie Rossera

Dowodliwe zdania systemów matematycznych są dowodliwe na różnych etapach.

Moglibyśmy myśleć o systemie matematycznym jako o komputerze zaprogramowanym tak, aby dowodzić różnorodnych zdań *kolejno*.

Powiemy, że p jest **dowodliwe wcześniej (zanim) niż q** (w danym systemie matematycznym), jeśli p zostało udowodnione na pewnym etapie, na którym q jeszcze nie zostało udowodnione (q może być lub też nie być udowodnione na jakimś późniejszym etapie).

Twierdzenie Rossera

Dla dowolnych zdań p oraz q wyrażalnych w systemie, zdanie $Bp < Bq$ (p jest dowodliwe wcześniej niż q) również jest wyrażalne w systemach typu tych rozpatrywanych przez Gödla, a Rosser pokazał, że jeśli p jest dowodliwe wcześniej niż q , to zdania $Bp < Bq$ oraz $\neg(Bq < Bp)$ są oba dowodliwe w systemie.

Rosser znalazł także zdanie p takie, że $p \equiv \neg(Bp < B\neg p)$ jest dowodliwe w systemie. (Takie zdanie p odpowiada tubylcowi z pierwszego rozważanego w tej części problemu, który mówi: „Nigdy nie uwierzysz wcześniej, że jestem rycerzem, niż uwierzysz, że jestem łotrem.”)

Wtedy, na mocy rozumowania z rozwiązania wspomnianego problemu, jeśli p jest dowodliwe, to system jest sprzeczny, a jeśli $\neg p$ jest dowodliwe, to system także jest sprzeczny.

A zatem, jeśli system jest niesprzeczny, to zdanie p jest nierozstrzygalne w systemie.

Twierdzenie Rossera

Zdanie Gödla może zostać sformułowane jako:

„Nie jestem dowodliwe na żadnym etapie.”

Bardziej wyszukane zdanie Rossera może zostać sformułowane jako:

„Nie mogę być dowiedzione na żadnym etapie, chyba że moja negacja została już wcześniej udowodniona.”

Zdanie Gödla, chociaż prostsze, wymaga założenia ω -niesprzeczności dla przeprowadzenia rozumowania. Zdanie Rossera, chociaż bardziej skomplikowane, dostarcza szukanego rezultatu przy słabszym założeniu prostej niesprzeczności.

Twierdzenie Tarskiego

Przypuśćmy, że mamy myślaką — nazwijmy go Paul — który jest zawsze *ściśły* w swoich przekonaniach (nigdy nie wierzy w zdania fałszywe). Nie musi on być prostaczką logiczną, ani normalną, nie jest też konieczne, aby rzeczywiście odwiedzał Wyspę Rycerzy i Łotrów. Wszystko, co musimy o nim wiedzieć to to, że jest *ściśły*.

Pewnego dnia tubylec mówi o nim:

„Paul nigdy nie uwierzy, że jestem rycerzem.”

Wtedy logicznie wynika stąd:

Twierdzenie 7.

System przekonań Paula jest niezupełny.

Twierdzenie Tarskiego

Jeśli Paul kiedykolwiek uwierzy, że tubylec jest rycerzem, to sfalsyfikuje to tym samym to, co powiedział tubylec, czyniąc tubylca łotrem, a tym samym czyniąc Paula nieścisłym z powodu jego wiary, że tubylec jest rycerzem.

Ale powiedziano nam, że Paul jest ścisły, a więc nigdy nie uwierzy on, że tubylec jest rycerzem.

Stąd, to co powiedział tubylec jest prawdziwe, a więc tubylec rzeczywiście jest rycerzem.

Wtedy, ponieważ Paul jest ścisły, nigdy nie będzie żywił fałszywego przekonania, że tubylec jest łotrem.

A zatem Paul nigdy nie dowie się, czy tubylec jest rycerzem, czy łotrem.

Twierdzenie Tarskiego

Komentarz. Treść matematyczna powyższej zagadki jest następująca. W systemach rozważanych przez Gödla mamy nie tylko pewne zdania nazywane zdaniami *dowodliwymi*, lecz również obszerniejszą klasę zdań nazywanych zdaniami *prawdziwymi* systemu.

W klasie zdań prawdziwych systemu obowiązują reguły tabliczek prawdziwościowych dla spójników logicznych.

Nadto, dla każdego zdania p systemu, zdanie Bp jest *prawdziwym* zdaniem systemu wtedy i tylko wtedy, gdy p jest zdaniem *dowodliwym* systemu.

Gödel znalazł godne uwagi zdanie g takie, że zdanie $g \equiv \neg Bg$ było zdaniem *prawdziwym* systemu (było ono nawet faktycznie dowodliwe w systemie, ale ten mocniejszy fakt nie jest potrzebny dla obecnego rozumowania).

Twierdzenie Tarskiego

Gdyby g było fałszywe, to Bg byłoby prawdziwe, a stąd g byłoby dowodliwe, a stąd prawdziwe, i mielibyśmy sprzeczność.

Zatem g jest prawdziwe, a stąd $\neg Bg$ jest prawdziwe, czyli g nie jest dowodliwe w systemie.

Tak więc, g jest prawdziwe, ale niedowodliwe w systemie.

Ponieważ g jest prawdziwe, więc $\neg g$ jest fałszywe, a stąd także niedowodliwe w systemie (ponieważ wszystkie dowodliwe zdania są prawdziwe).

A zatem g jest nierozstrzygalne w systemie.

Dawniejsza opozycja filozoficzna wobec logiki modalnej była osadzona w przybliżeniu w trzech różnych (i nieporównywalnych) przekonaniach. Po pierwsze, są tacy, którzy są przekonani, że wszystko, co jest prawdziwe jest koniecznie prawdziwe, a stąd nie ma żadnej różnicy między prawdą a prawdą konieczną. Po drugie, są tacy, którzy wierzą, że nic nie jest koniecznie prawdziwe, a stąd dla dowolnego zdania p , zdanie Np (p jest koniecznie prawdziwe) jest po prostu fałszywe! A po trzecie, są i tacy, którzy twierdzą, że słowa „koniecznie prawdziwe” nie niosą jakiegokolwiek sensu. Tak więc, każde z tych nastawień filozoficznych odrzuca logikę modalną ze swoich własnych powodów. W istocie, pewien bardzo znany filozof wstawił się sugestią, że nowoczesna logika modalna została poczęta w grzechu. Na co Boolos bardzo stosownie odpowiedział: „Jeśli nowoczesna logika modalna została poczęta w grzechu, to została wybawiona przez Gödłowskość”. [W oryginale: *If modern modal logic was conceived in sin, then it has been redeemed through Gödliness.*]

Koniec

Prezentacja nie rości sobie pretensji do kompletności:

- ani jako przedstawienie wszystkich treści *Forever Undecided*,
- ani jako wprowadzenie do logiki dowodliwości.

Staraliśmy się jedynie pokazać próbkę możliwości popularyzacji wiedzy o logice modalnej i jej zastosowaniach.

Zachęcamy do lektury książki!

Jeśli starczy czasu, to pokazujemy jeszcze [Dodatek](#).

Maszyny logiczne Smullyana

Smullyan skonstruował cały szereg maszyn logicznych, które drukują zdania „mówiące” coś o nich samych.

Maszyny: Craiga, Fergussona i McCullocha, przedstawione w *Jaki jest tytuł tej książki?* oraz *Dama czy tygrys?* są już znane polskiemu czytelnikowi. Tu przedstawimy pewną maszynę Smullyana, opisaną w *Forever Undecided*.

Dla pełnego zrozumienia jej działania potrzebna jest znajomość wybranych logik modalnych: logiki epistemicznej oraz logiki dowodliwości (logiki Gödla-Löba).

Zakładamy u audytorium znajomość tego materiału.

Maszyny logiczne Smullyana

Malcolm Fergusson, gdy usłyszał o twierdzeniach Gödla i Löba, z miejsca zabrał się za konstrukcję maszyny, którą z zachwytem pokazał swoim przyjaciołom.

Ku ich zadowoleniu udowodnił, że maszyna jest niesprzeczną i stabilną maszyną typu G, a szczególne upodobanie znalazł w demonstracji, że maszyna, chociaż niesprzeczną, nigdy nie może dowieść własnej niesprzeczności!

Maszyna ilustruje w niezwykle prosty i pouczający sposób podstawowe idee zawarte w Pierwszym oraz Drugim Twierdzeniu Gödla jak również w Twierdzeniu Löba.

Niżej podajemy opis działania maszyny Fergussona oraz pewne ważne fakty jej dotyczące.

Opis pochodzi z rozdziału 26 *Forever Undecided*. W rozdziale tym znajdujemy też opis dwóch innych maszyn, który tu pominiemy.

Maszyna drukuje różnorakie zdania zbudowane z siedemnastu symboli. Pierwsze siedem z tych symboli to następujące:

$$\begin{array}{ccccccc}
 P & \perp & \rightarrow & (&) & d & , \\
 1 & 2 & 3 & 4 & 5 & 6 & 7
 \end{array}$$

Pod każdym z tych symboli podpisano jego *numer Gödlowski*.

Pozostałe dziesięć symboli to znane cyfry 1, 2, 3, 4, 5, 6, 7, 8, 9, 0. Tym cyfrom przyporządkujemy numery Gödlowskie w następujący sposób. Numerem Gödlowskim cyfry 1 jest 89 (8 po której następuje jedna 9); numerem Gödlowskim cyfry 2 jest 899 (8 po której następują dwie 9); i tak dalej, aż do cyfry 0, której numerem Gödlowskim jest 89999999999 (8 po której następuje dziesięć 9).

Tak więc, każdy z siedemnastu symboli uzyskuje numer Gödlowski.

Dla danego wyrażenia złożonego, odnajdujemy jego numer Gödłowski przez zastąpienie każdego symbolu jego numerem Gödłowskim — dla przykładu, numerem Gödłowskim wyrażenia $(P \perp \rightarrow \perp)$ jest 412325. Inny przykład: numerem Gödłowskim $P35$ jest 18999899999.

Dla dowolnego wyrażenia E , przez \bar{E} rozumiemy numer Gödłowski E (zapisany jako ciąg cyfr 1, 2, ..., 0).

Nie każda liczba jest numerem Gödłowskim jakiegoś wyrażenia (na przykład, 88 nie jest numerem Gödłowskim żadnego wyrażenia).

Jeśli n jest numerem Gödłowskim jakiegoś wyrażenia, to będziemy czasem odwoływać się do tego wyrażenia jako do n -tego wyrażenia. (Dla przykładu, Pd jest szesnastym wyrażeniem, \perp jest drugim wyrażeniem.)

Maszyna jest *samoodnosząca się* (do siebie) w tym sensie, że wyrażenia drukowane przez maszynę stwierdzają, co maszyna może, a czego nie może wydrukować. Wyrażenie nazywamy *drukowalnym*, jeśli maszyna może je wydrukować.

Symbol „ P ” oznacza „drukowalne” i dla dowolnego wyrażenia E zbudowanego z podanych siedemnastu symboli, jeśli chcemy zapisać zdanie stwierdzające, że E jest drukowalne, to piszemy nie PE , lecz $P\bar{E}$ (tj., P po którym następuje numer Gödłowski E).

Dla przykładu, zdaniem stwierdzającym, że $(P \perp \rightarrow \perp)$ jest drukowalne jest $P(P \perp \rightarrow \perp)$ — tj. $P412325$.

Dla dowolnych wyrażeń X oraz Y , Fergusson zdefiniował *diagonalizację X względem Y* jako wyrażenie $(X(\bar{X}, \bar{Y}) \rightarrow Y)$.

Symbol „ d ” jest skrótem dla „diagonalizacja” — i dla dowolnych wyrażeń X oraz Y , wyrażenie $Pd(\bar{X}, \bar{Y})$ jest zdaniem stwierdzającym, że diagonalizacja X względem Y jest drukowalna.

Zdefiniujemy teraz, co to znaczy, że wyrażenie jest *zdaniami (maszynowym)* i co to znaczy, że zdanie jest *prawdziwe*.

- (1) \perp jest zdaniem i \perp jest fałszywe.
- (2) Dla dowolnego wyrażenia X , wyrażenie $P\bar{X}$ jest zdaniem i jest ono prawdziwe wtedy i tylko wtedy, gdy wyrażenie X jest drukowalne.
- (3) Dla dowolnych wyrażeń X oraz Y , wyrażenie $Pd(\bar{X}, \bar{Y})$ jest zdaniem i jest ono prawdziwe wtedy i tylko wtedy, gdy wyrażenie $(X(\bar{X}, \bar{Y}) \rightarrow Y)$ — które jest diagonalizacją X względem Y — jest drukowalne.
- (4) Dla dowolnych zdań X oraz Y , wyrażenie $(X \rightarrow Y)$ jest zdaniem i jest ono prawdziwe wtedy i tylko wtedy, gdy albo X nie jest prawdziwe, albo Y jest prawdziwe.

Rozumie się, że żadne wyrażenie nie jest zdaniem (maszynowym), jeśli nie zostało otrzymane zgodnie z powyższymi regułami. Spójniki logiczne $\neg, \wedge, \vee, \equiv$ są definiowane z \rightarrow oraz \perp w znany sposób.

Podamy teraz reguły ustalające, co maszyna może wydrukować. Maszyna jest zaprogramowana do kolejnego drukowania nieskończonej listy zdań. Pewne zdania, nazywane *aksjomatami* mogą zostać wydrukowane na każdym etapie tego procesu. Wśród aksjomatów są wszystkie tautologie. (tak więc, dla dowolnej tautologii X , maszyna może wydrukować X kiedy tylko chce, niezależnie od tego, co dotąd wydrukowała lub czego nie wydrukowała w poprzednich etapach.)

Dalej, maszyna jest zaprogramowana tak, że dla dowolnych zdań X oraz Y , jeśli na pewnym etapie maszyna wydrukowała już X oraz $X \rightarrow Y$, to może wydrukować Y . Tak więc, maszyna jest **typu 1** (w tym sensie, że zbiór zdań drukowalnych jest typu 1).

Ponieważ jest prawdą, że jeśli X oraz $X \rightarrow Y$ są oba drukowalne, to Y też jest drukowalne, to zdanie $(\overline{P\overline{X}} \wedge \overline{P(X \rightarrow Y)}) \rightarrow \overline{P\overline{Y}}$ jest prawdziwe; lub, co na jedno wychodzi, zdanie $\overline{P(X \rightarrow Y)} \rightarrow (\overline{P\overline{X}} \rightarrow \overline{P\overline{Y}})$ jest prawdziwe. Maszyna „wie” zatem o prawdziwości wszystkich zdań postaci $\overline{P(X \rightarrow Y)} \rightarrow (\overline{P\overline{X}} \rightarrow \overline{P\overline{Y}})$ i przyjmuje je jako aksjomaty. Tak więc, maszyna jest **typu 2**.

Następnie, jeśli maszyna kiedykolwiek wydrukuje zdanie X , to „wie” ona, że wydrukowała X i prędzej czy później wydrukuje prawdziwe zdanie $P\bar{X}$. (Zdanie $P\bar{X}$ jest prawdziwe, ponieważ X zostało wydrukowane.) A więc maszyna jest normalna, a stąd jest **typu 3**.

Ponieważ maszyna jest normalna, więc dla dowolnego zdania X , zdanie $P\bar{X} \rightarrow P\overline{P\bar{X}}$ jest prawdziwe. Czyli maszyna jest początkowo „świadoma” prawdziwości wszystkich takich zdań oraz przyjmuje je jako aksjomaty. Zatem maszyna jest **typu 4**.

Jest jeszcze jedna rzecz, którą maszyna potrafi robić, a jest to rzecz dość istotna. Dla dowolnych wyrażeń X oraz Y , zdanie $Pd(\bar{X}, \bar{Y})$ jest prawdziwe wtedy i tylko wtedy, gdy $(X(\bar{X}, \bar{Y}) \rightarrow Y)$ jest drukowalne, co z kolei zachodzi wtedy i tylko wtedy, gdy zdanie $P(X(\bar{X}, \bar{Y}) \rightarrow Y)$ jest prawdziwe. Zatem następujące zdanie jest prawdziwe: $Pd(\bar{X}, \bar{Y}) \equiv P(X(\bar{X}, \bar{Y}) \rightarrow Y)$. Maszyna „wie” o prawdziwości wszystkich takich zdań i przyjmuje je jako aksjomaty. Te aksjomaty nazywane są **aksjomatami przekątniowymi**.

Aksjomaty i reguły maszyny

Aksjomaty:

- *Grupa 1.* Wszystkie tautologie.
- *Grupa 2.* Wszystkie zdania postaci $P(\overline{X \rightarrow Y}) \rightarrow (P\overline{X} \rightarrow P\overline{Y})$.
- *Grupa 3.* Wszystkie zdania postaci $P\overline{X} \rightarrow \overline{P\overline{X}}$.
- *Grupa 4 (aksjomaty przekątniowe).* Wszystkie zdania postaci $Pd(\overline{X}, \overline{Y}) \equiv P(\overline{X(\overline{X}, \overline{Y}) \rightarrow Y})$, gdzie X oraz Y są dowolnymi wyrażeniami (niekoniecznie zdaniami).

Reguły operowania:

- (1) Aksjomaty mogą zostać wydrukowane na każdym etapie.
- (2) Dla dowolnych już wydrukowanych zdań X oraz $(X \rightarrow Y)$, maszyna może wydrukować Y .
- (3) Dla dowolnego wydrukowanego już zdania X , maszyna może wydrukować $P\overline{X}$.

Rozumie się, że jedynym sposobem wydrukowania przez maszynę jakiegoś zdania X na pewnym etapie jest zastosowanie się do powyższych reguł. Zatem, X jest drukowalne na danym etapie *tylko* wtedy, gdy zachodzi jeden z następujących trzech warunków: (1) X jest aksjomatem; (2) istnieje zdanie Y takie, że Y oraz $(Y \rightarrow X)$ zostały już wydrukowane na etapie wcześniejszym; (3) istnieje zdanie Y takie, że X jest zdaniem $P\bar{Y}$ oraz Y zostało już wydrukowane na etapie wcześniejszym.

Uwagi. Dla każdego zdania X , niech BX będzie zdaniem $P\bar{X}$. Symbol „ B ” *nie* należy do języka maszyny; używamy go do *mówienia o* maszynie. Używamy „ B ” jako odpowiadającej operacji, która przyporządkowuje każdemu zdaniu X zdanie $P\bar{X}$.

Gdy mówimy, że maszyna jest typu 4, rozumiemy przez to, że jest ona typu 4 ze względu na tę operację B . W istocie, bez aksjomatów przekątniowych, system aksjomatyczny tej maszyny jest *systemem modalnym K_4* .

Zobaczymy wkrótce, że dodanie aksjomatów przekątniowych daje nam pełną moc *systemu modalnego G* .

Dowodliwość. Zdefiniowaliśmy dla każdego zdania maszyny co to znaczy, że zdanie to jest *prawdziwe*, a więc każde zdanie maszyny wyraża określone zdanie, które może być prawdziwe lub może być fałszywe.

Uwaga. Dotąd *proposition* oddawaliśmy zawsze jako *zdanie*. Teraz mamy:

- *zдания (maszyny)* (w oryginale *sentences*) — zdania języka przedmiotowego,
- oraz zdania *metajęzyka* (w oryginale *propositions*), tj. języka, w którym *mówimy o* maszynie, jej zdaniach (maszynowych), itp.

W przypadkach, gdy mogłoby to prowadzić do nieporozumień, w dalszym ciągu będziemy dodawać określenie *maszynowe*, gdy mowa będzie o zdaniach drukowanych przez maszynę.

Powiemy, że maszyna *dowodzi* danego zdania, gdy drukuje ona zdanie maszynowe, które wyraża to dane zdanie. Dla przykładu, zdanie maszynowe $\neg P2$ wyraża zdanie stwierdzające, że maszyna jest niesprzeczna (ponieważ 2 jest numerem Gödłowskim \perp), a więc jeśli maszyna wydrukowała $\neg P2$, to udowodniła swoją własną niesprzeczność. Gdyby maszyna wydrukowała $P2$, to udowodniłaby swoją własną *sprzeczność*.

Powiemy, że maszyna jest *ściśła*, jeśli wszystkie zdania dowodliwe przez maszynę są prawdziwe.

Powiemy, że maszyna jest *niesprzeczna*, jeśli nie może ona dowieść \perp , oraz że jest *stabilna*, jeśli dla każdego zdania (maszynowego) X , jeśli $P\bar{X}$ jest drukowalne, to drukowalne jest też X .

Zwrotność. Przechodzimy teraz do dowodu, że maszyna jest Gödłowska, a faktycznie, zwrotna.

Ważne własności maszyny Fergussona

- (1) Znajdziemy zdanie G takie, że zdanie $G \equiv \neg P\bar{G}$ — tj. zdanie $G \equiv (P\bar{G} \rightarrow \perp)$ — jest drukowalne.
- (2) Pokażemy, że dla dowolnego zdania Y istnieje zdanie X takie, że zdanie $X \equiv (P\bar{X} \rightarrow Y)$ jest drukowalne.

Uwaga. Problem 1 jest szczególnym przypadkiem problemu 2, a więc najpierw rozwiążemy problem 2.

Przypomnijmy, że:

- warunek wspomniany w problemie (2) nazywamy **zwrotnością**;
- **systemem typu G** nazywamy system modalny typu 4, w którym dowodliwe są wszystkie zdania postaci $B(Bp \rightarrow q) \rightarrow Bp$.

Niech Y będzie dowolnym zdaniem. Dla dowolnego wyrażenia Z , zdanie $Pd(\overline{Z}, \overline{Y}) \equiv P(Z(\overline{Z}, \overline{Y}) \rightarrow Y)$ jest drukowalne (ponieważ jest jednym z aksjomatów przekątniowych).

Weźmiemy za Z wyrażenie Pd i otrzymujemy wtedy, że $Pd(\overline{Pd}, \overline{Y}) \equiv P(Pd(\overline{Pd}, \overline{Y}) \rightarrow Y)$ jest drukowalne.

Ponieważ maszyna jest typu 1, więc wynika stąd, że następujące zdanie jest drukowalne:

$$(Pd(\overline{Pd}, \overline{Y}) \rightarrow Y) \equiv \overline{(P(Pd(\overline{Pd}, \overline{Y}) \rightarrow Y) \rightarrow Y)}$$

Tak więc, zdanie $X \equiv (P\overline{X} \rightarrow Y)$ jest drukowalne, gdzie X jest zdaniem $(Pd(\overline{Pd}, \overline{Y}) \rightarrow Y)$.

Problem 1 jest szczególnym przypadkiem problemu 2, gdy za Y weźmiemy \perp . Tak więc, zdaniem Gödla G dla tej maszyny jest $Pd(\overline{Pd}, \overline{\perp}) \rightarrow \perp$ — tj., zdanie $(Pd(16, 2) \rightarrow \perp)$.

Co stwierdza zdanie $Pd(16, 2)$?

Mówi ono, że diagonalizacja szesnastego wyrażenia względem drugiego wyrażenia jest drukowalna. Wyrażeniem szesnastym jest Pd , a wyrażeniem drugim jest \perp , a więc $Pd(16, 2)$ mówi, że diagonalizacja Pd względem \perp jest drukowalna, ale ta diagonalizacja to zdanie $(Pd(16, 2) \rightarrow \perp)$ — tj. właśnie samo zdanie G !

A więc $Pd(16, 2)$ mówi, że G jest drukowalne, a stąd $(Pd(16, 2) \rightarrow \perp)$ — które jest zdaniem G — mówi, że G *nie jest* drukowalne (lub, co na jedno wychodzi, że drukowalność G implikuje fałsz logiczny). Tak więc, G mówi, że G nie jest drukowalne; G jest *prawdziwe* wtedy i tylko wtedy, gdy G nie jest drukowalne.

Zatem G stwierdza swoją własną niedrukowalność. Oto, w miniaturce, pomysłowa idea Gödla otrzymywania samoodniesienia.

Zdanie $G \equiv \neg P\bar{G}$ — tj. zdanie $G \equiv (P\bar{G} \rightarrow \perp)$ — jest nie tylko prawdziwe, ale także drukowalne (jest ono jednym z aksjomatów przekątniowych).

Ponieważ maszyna jest normalna i jest typu 1, wynika stąd na mocy Pierwszego Twierdzenia Gödla o Niezupełności, że jeśli maszyna jest niesprzeczna, to G nie jest drukowalne, a jeśli maszyna jest dodatkowo stabilna, to również $\neg G$ nie jest drukowalne.

A więc, *jeśli* maszyna jest jednocześnie niesprzeczna i stabilna, to zdanie G jest nierozstrzygalne w systemie zdań, które maszyna może wydrukować.

Maszyna jest faktycznie typu 4, a ponieważ jest Gödłowska — zdanie $G \equiv \neg P\bar{G}$ jest drukowalne — więc z Drugiego Twierdzenia Gödla o Niedowodliwości niesprzeczności wynika, że jeśli maszyna jest niesprzeczna, to nie może ona dowieść swojej własnej niesprzeczności — tj. nie może wydrukować zdania $\neg P2$.

Nadto, jeśli maszyna jest niesprzeczna, to zdanie $\neg P2$ jest *prawdziwe*, a stąd jest innym przykładem zdania prawdziwego, którego maszyna nie może wydrukować.

Co więcej, maszyna jest zwrotna (problem 2), a będąc typu 4, musi być Löbowska (na mocy Twierdzenia Löba), a więc dla dowolnego zdania X , jeśli $P\bar{X} \rightarrow X$ jest drukowalne, to drukowalne jest X . Ponieważ każdy zwrotny Löbowski system typu 4 jest typu G, więc wynika stąd, że maszyna jest typu G.

Czy Maszyna Fergussona jest niesprzeczna?

Poprawność, ścisłość i niesprzeczność Maszyny Fergussona.

Pokazaliśmy, że *jeśli* maszyna Fergussona jest niesprzeczna, to nie może udowodnić swojej własnej niesprzeczności.

Ale skąd wiemy, czy maszyna jest, czy nie jest niesprzeczna?

Udowodnimy teraz, że maszyna jest nie tylko niesprzeczna, ale że jest też całkowicie ścisła — tj., że każde zdanie wydrukowane przez maszynę jest prawdziwe.

Pokazaliśmy już, że wszystkie *aksjomaty* maszyny są prawdziwe, ale prześledźmy uważnie to rozumowanie.

Aksjomaty Grupy 1 są wszystkie tautologiami, a stąd są z pewnością prawdziwe.

Jeśli chodzi o aksjomaty Grupy 2, to powiedzieć, że $\overline{P(X \rightarrow Y)} \rightarrow (\overline{PX} \rightarrow \overline{Y})$ jest prawdziwe to tyle, co powiedzieć, że jeśli oba $\overline{P(X \rightarrow Y)}$ oraz \overline{PX} są prawdziwe, to takie jest też \overline{PY} , czyli to samo, co powiedzieć, że jeśli $(X \rightarrow Y)$ oraz X są oba drukowalne, to takie jest też Y .

A tak oczywiście jest, na mocy Operacji 2.

Tak więc, aksjomaty Grupy 2 są wszystkie prawdziwe.

Jeśli chodzi o aksjomaty Grupy 3, powiedzieć, że $P\bar{X} \rightarrow \overline{P\bar{X}}$ jest prawdziwe, to tyle, co powiedzieć, że jeśli $P\bar{X}$ jest prawdziwe, to takie jest też $\overline{P\bar{X}}$.

To z kolei jest tym samym, co powiedzenie, że jeśli X jest drukowalne, to takie jest też $\overline{P\bar{X}}$ — a tak jest rzeczywiście, na mocy Operacji 3.

Jeśli chodzi o aksjomaty przekątniowe, to $Pd(\bar{X}, \bar{Y})$ jest prawdziwe wtedy i tylko wtedy, gdy $(X(\bar{X}, \bar{Y}) \rightarrow Y)$ jest drukowalne, a tak jest wtedy i tylko wtedy, gdy $\overline{P(X(\bar{X}, \bar{Y}) \rightarrow Y)}$ jest prawdziwe.

Zatem $Pd(\bar{X}, \bar{Y}) \equiv \overline{P(X(\bar{X}, \bar{Y}) \rightarrow Y)}$ jest prawdziwe.

Wiemy teraz, że wszystkie aksjomaty maszyny są prawdziwe, ale musimy pokazać, że wszystkie zdania *drukowalne* są prawdziwe.

Przypomnijmy, że maszyna drukuje zdania na pewnych *etapach*.
Chcemy teraz ustanowić następujący lemat, twierdzenie i wniosek:

- **Lemat.** Jeśli X jest zdaniem wydrukowanym na pewnym etapie i wszystkie zdania wydrukowane do tego etapu są prawdziwe, to X jest prawdziwe.
- **Twierdzenie.** Każde zdanie wydrukowane przez maszynę jest prawdziwe.
- **Wniosek.** Maszyna jest jednocześnie niesprzeczna i stabilna.

Dowody.

Najpierw udowodnimy lemat. Załóżmy, że wszystkie dotąd wydrukowane zdania są prawdziwe; mamy pokazać, że X jest prawdziwe.

Przypadek 1. X jest aksjomatem. Wtedy X jest prawdziwe (jak już udowodniliśmy).

Przypadek 2. Istnieje zdanie Y takie, że Y oraz $(Y \rightarrow X)$ zostały już wydrukowane. Wtedy z przyjętego założenia Y oraz $(Y \rightarrow X)$ są oba prawdziwe, a więc X jest prawdziwe.

Przypadek 3. X jest postaci $P\bar{Y}$, gdzie Y jest zdaniem, które już zostało wydrukowane. Ponieważ Y zostało wydrukowane, więc $P\bar{Y}$ jest prawdziwe — tj. X jest prawdziwe.

To kończy dowód lematu.

Dowód Twierdzenia.

Maszyna jest zaprogramowana tak, aby wydrukować wszystkie drukowalne zdania w jakimś określonym ciągu $X_1, X_2, \dots, X_n, \dots$. Przez X_n rozumiemy zdanie wydrukowane na etapie n .

Pierwsze zdanie wydrukowane przez maszynę (zdanie X_1) musi być aksjomatem (ponieważ dotąd maszyna nie wydrukowała żadnych zdań), a stąd X_1 musi być prawdziwe.

Jeśli powyższa lista zawierałaby jakiegokolwiek zdanie fałszywe, to musiałyby istnieć *najmniejsza* liczba n taka, że X_n jest fałszywe — to jest, musiałyby istnieć *pierwsze* zdanie fałszywe wydrukowane przez maszynę. Wiemy, że n nie jest równe 1 (ponieważ X_1 jest prawdziwe), a zatem n jest większe od 1. Znaczy to, że maszyna drukuje zdanie fałszywe na etapie n , ale na wszystkich wcześniejszych etapach drukowała wyłącznie zdania prawdziwe. Przeczy to jednak lematowi.

Zatem maszyna nigdy nie może wydrukować jakichkolwiek zdań fałszywych.

Dowód Wniosku.

Ponieważ maszyna jest ścista (na mocy Twierdzenia), więc \perp nigdy nie może zostać wydrukowane, ponieważ \perp jest fałszywe. Zatem maszyna jest niesprzeczna.

Następnie, przypuśćmy, że $P\bar{X}$ jest drukowalne. Wtedy $P\bar{X}$ jest prawdziwe (na mocy Twierdzenia), co oznacza, że X jest drukowalne. Zatem maszyna jest stabilna.

Widzimy teraz, że maszyna Fergussona *jest* niesprzeczna, ale nigdy nie potrafi dowieść swojej niesprzeczności. Tak więc i ty i ja (równie dobrze jak Fergusson) wiemy, że maszyna jest niesprzeczna, ale biedna maszyna wiedzy tej nie ma!

Koniec dodatku

O dalszych wynikach związanych z „maszynową” interpretacją twierdzeń metalogicznych traktuje rozdział 28 *Forever Undecided*.

W szczególności, podane są związki między maszynami logicznymi a samostosowanymi systemami modalnymi.