WITOLD MARCISZEWSKI

Czy robot może uznać za prawdę zdanie samopotwierdzalne?

Chcąc zakłopotać robota, trzeba mu zadać pytanie: *czy uznajesz za prawdę aksjomaty arytmetyki?* Robot jest to, oczywiście, osobnik funkcjonujący na zasadzie maszyny Turinga. Można też zapytać o inne aksjomaty, ale arytmetyka to domena, w której robot jest maksymalnie kompetentny. A jeśli ktoś chciałby sprawić mu jeszcze większą trudność, niech zapyta: *czy uważasz aksjomaty arytmetyki za zdania konieczne?*

To drugie pytanie eliminuje trywialną sytuację, gdy robot dostał aksjomaty w darze jako elementy wbudowanej mu przez człowieka bazy danych. Mając je w bazie, może na pierwsze pytanie odpowiedzieć: uznaję. Ale w jego bazie znajdą się też zdania nie będące aksjomatami i nie wynikające logicznie a aksjomatów (np. o pogodzie na Grenlandii pierwszego stycznia 1900). Robot będzie je uznawał na tej samej podstawie: że ma je w bazie. Ale powinien je uznawać jakoś inaczej. Z tej inności świadomość ludzka zdaje sprawę w taki sposób, że aksjomatom przypisuje prawdziwość konieczną. Czy to samo potrafiłyby przyszłe roboty, które wedle zapowiedzi komputacjonizmu mają dorównać człowiekowi także pod względem świadomości?

Pytania o naturę przyszłych robotów nie są łatwe dla obecnych ludzi, toteż w tych rozważaniach upraszczam problem, pytając nie o stosunek robota do zdań aksjomatycznych, i zarazem koniecznych, ale o stosunek do zdań innego rodzaju. Są te drugie podobne do aksjomatów na tyle, że zyskamy przyczynek do zagadnienia postawionego na wstępie, a zarazem mają pewną osobliwość — taką, że dzięki niej można prześledzić, jak się dochodzi do przypisania im prawdziwości koniecznej.

To, że da się na gorąco uchwycić *proces* dochodzenia do pewnego stanu poznawczego jest dla tych rozważań istotne, ponieważ właściwe robotom postępowanie algorytmiczne jest zawsze jakimś procesem. Gdy się go zrekonstruuje w przypadku ludzi, umożliwi to stwierdzenie, czy mamy do czynienia z procesem algorytmicznym, czy może z procesem, w którym tylko pewne części okażą się algorytmiczne.

Jeśli w wyniku analizy niektóre elementy nie okażą się algorytmiczne, pora będzie wnikliwie rozważyć postulat komputacjonizmu (zwanego inaczej silną zasadą sztucznej inteligencji), że wynik taki bierze się wyłącznie z niedoskonałości naszych narzędzi badawczych. W rzeczywistości bowiem – głosi ów kierunek – (a) *wszystkie procesy w świecie fizycznym są algorytmiczne czyli obliczalne*, (b) ludzkie uznawanie zdań za konieczne jest procesem fizycznym, bo innych niż fizyczne w świecie nie ma (wywodzą komputacjoniści), a zatem (c) ludzkie uznawanie zdań za konieczne musi być wynikiem algorytmu. Algorytmu tego jeszcze nie znamy, a gdy przypisujemy uznawaniu konieczności charakter jakiegoś aktu intuicyjnego, to jest to jedynie przejaw niewiedzy.

W tym punkcie dyskusji *onus probandi* przechodzi na stronę komputacjonistów. To oni powinni doprowadzić do rozszyfrowania tajnego, jak dotąd, kodu neuronowego, w którym byłby zapisany ów postulowany przez nich algorytm. W ten zaś algorytm można będzie wyposażyć przyszłe roboty, tak, że już nie będą ich kłopotać pytania proponowane na wstępie. Jeśli tego nie da się uczynić, to trzeba by na innej drodze wykazać obecność algorytmu, a zarazem wyjaśnić powody jego niepoznawalności.

W celu doprowadzenia do takiej pozycji na dyskusyjnym polu gry, biorę tu na warsztat zdania, którym daję miano *samopotwierdzalnych* (nazwa ta jest naturalna, wisi niejako w powietrzu, ale nie mogę wskazać autora, który by się nią już posłużył). Najznamienitszym tej klasy przedstawicielem jest słynne zdanie gödlowskie, to jest, stwierdzające swą niedowodliwość w arytmetyce (przy założeniu niesprzeczności arytmetyki). Zdania z tej klasy podzielają z aksjomatami niedowodliwość i konieczność. Uznawanie zaś ich konieczności jest rezultatem procesu, którego opisanie da szansę porównania z procedurą algorytmiczną. Takie opisanie jest zamierzeniem niniejszego eseju.

1. Zdania z predykatem wyrażającym warunek niezbędny prawdy

1.1. Nie wydaje się, żeby w aktualnym stanie zagadnienia analiza zdań samopotwierdzalnych wyszła poza doraźne komentarze czynione na marginesie zdania gödlowskiego (szczególnie inspirujące są te, które można znaleźć o Tarskiego, Poppera i Kneale'ów). Nawet jeśli taka diagnoza stanu zagadnienia bierze się z niedoinformowania piszącego te słowa, nie będzie rzeczą daremną podjąć próbę przebadania natury zdań samopotwierdzalnych, a potem ewentualnie skonfrontować ją z wynikami systematycznej kwerendy.

Pytanie, które biorę za punkt wyjścia (podczas gdy punktem dojścia jest to postawione w tytule) jest następujące: czy zdanie gödlowskie jest czymś unikalnym, czy też stanowi element szerszej klasy wypowiedzi, których sposób uzasadniania da się opisać w sposób ogólny. Jeśliby okazało się to drugie, to analiza owej ogólnej metody uzasadniania powinna wychwycić w nim elementy niewątpliwie algorytmiczne oraz inne, których algorytmiczność byłaby dopiero do wykazania. Jeśliby wszystkie operacje umysłu prowadzące do akceptacji zdania samopotwierdzalnego okazały się algorytmiczne, dostaniemy tym samym odpowiedź twierdzącą na pytanie postawione w tytule.

Nie ma powodu dopatrywania się anomalii w pewnych zdaniach samozwrotnych, które nie są samopotwierdzalne (nie są one właściwym przedmiotem analizy, ale służa do jej przygotowania). Gdy samozwrotność okaże się zjawiskiem normalnym (a nie dewiacją, którą próbuje usunąć np. teoria typów), to pozostanie wykazać, że nie tylko ta cecha rodzajowa, ale i cecha specyficzna zdań samopotwierdzalnych mieści się w pewnej normie. Wystarczy podać kilka przykładów takich zdań i zaopatrzyć każde z nich pytaniem, czy adresat potrafi stwierdzić każdego z nich prawdziwość lub fałszywość. Oto lista do rozważenia.

(1) Niniejsze zdanie jest napisane po polsku.
(2) Niniejsze zdanie nie jest napisane po polsku.
(3) Niniejsze zdanie składa się z siedmiu słów.
(4) Niniejsze zdanie składa się z trzech słów.
(5) Niniejsze zdanie zawiera błąt ortograficzny.

(6) Niniejsza wypowiedź jest zdaniem podmiotowo-orzecznikowym.
(7) Niniejsze zdanie jest wypowiedzią o pewnym zdaniu.
(8) Niniejsza wypowiedź jest przykładem zdania samozwrotnego.

Pozytywna konkluzja tego argumentu przez przykłady polegać będzie na zauważeniu, że w każdym przypadku da się rozstrzygnąć, czy wypowiedź jest prawdziwa czy fałszywa, są to więc konstrukcje językowe sensowne. A także zauważeniu, że wypowiedzi takie łatwo jest wytwarzać, a więc nie stanowią one jakiejś klasy wąskiej i ekskluzywnej.

Odrębnym zagadnieniem jest przydatność poznawcza czy komunikacyjna wypowiedzi samozwrotnych. Można odpowiedzieć, że jest to przydatność tylko sporadyczna, jak wtedy, gdy ktoś ilustruje przykładami błędy ortograficzne i obok innych przykładów, jak „s powarzaniem" czy „bapcia" wymieni też napis „błąt". Będzie też rys samozwrotności w takim jednozdaniowym podaniu: „Niniejszym zwracam się z prośbą o urlop." Oczywiście, ważkie zastosowanie znajdujemy w wywodzie Gödla. Dla obecnych jednak rozważań nie ma znaczenia, jaki stopień przydatności przypisze się zdaniom samozwrotnym. To, co się liczy, to jest uznanie ich sensowności, nie tylko w sensie syntaktycznym, ale i semantycznym; to drugie oznacza zrozumiałość rozpoznawalną po tym, że można stwierdzić, czy dane zdanie jest prawdziwe.

1.2. Prawdziwość zdań z listy 1-8 stwierdzamy na podstawie obserwacji połączonej z pewną wiedzą. O zdaniu 1 stwierdzamy prawdziwość na podstawie jego wysłuchania i znajomości języka polskiego. W przypadku zdania 6 potrzebna jest pewna znajomość gramatyki. I tak dalej. Nie da się natomiast udowodnić prawdziwości przez wnioskowanie, które wyprowadzałoby dane zdanie z jego zaprzeczenia. Taką wyprowadzalność nazwałem samopotwierdzalnością, upatrując w niej konieczność logiczną par excellence.

Zajmę się obecnie rodzajem zdań, które nie tylko są samozwrotne, lecz także samopotwierdzalne. We wszystkich rozważanych dalej przypadkach samopotwierdzalność pojawia się dzięki przyjęciu stosownej dla danego zdania zasady, która treść występującego w nim predykatu wiąże w pewien sposób (dla danego predykatu swoisty) z pojęciem prawdy. Ze względu na tego rodzaju związek każda taka zasada będzie tu określana jako *zasada aletyczna* (od gr. *aletheia* – prawda). Wyodrębnienie zasady aletycznej wśród przesłanek pozwoli, między innymi, uchwycić nerw argumentacji gödlowskiej, w której dowodliwość jest pojmowana jako warunek wystarczający, a więc pewne kryterium, prawdy arytmetycznej (w poniższym zaś przykładzie, w celach porównawczych, rozważa się zasadę aletyczną dotyczącą warunku niezbędnego zarówno prawdziwości jak i fałszywości pewnego rodzaju zdań). Oto uzasadnienie pewnej wypowiedzi, którego przebieg wskazuje na jej samopotwierdzalność, a więc i konieczność.

S1. [Zdanie] S1 jest poprawnie zbudowane.

Uzasadnienie zdania S1.

(1) S1 nie jest poprawnie zbudowane. — *założenie dowodu nie wprost.*
(2) Jeśli S1 nie jest poprawnie zbudowane, to wypowiedź „S1 jest poprawnie zbudowane" nie jest prawdziwa. — *z Tarskiego definicji prawdy.*

(3) Wypowiedź „S1 jest poprawnie zbudowane” nie jest prawdziwa. — *1,2; reguła odrywania.*
(4) Wypowiedź „S1 jest poprawnie zbudowane” jest fałszywa. — *3; zasada dwuwartościowości.*
(5) Wypowiedź „S1 jest poprawnie zbudowane” $=$ S1. — *zdanie obserwacyjne.*
(6) S1 jest fałszywe. — *4,5; reguła zastępowania.*
(7) S1 jest ciągiem wyrazów, o którym da się wykazać, że jest on zdaniem fałszywym. — *zdanie obserwacyjne dotyczące wierszy 1-6.*
(8) Jeśli o jakimś ciągu wyrazów da się wykazać, że jest on zdaniem fałszywym, to jest on poprawnie zbudowany. — *zasada aletyczna: poprawność budowy jako warunek niezbędny prawdziwości lub fałszywości.*
(9) S1 jest poprawnie zbudowane. — *7,8; reguła odrywania z uszczegółowieniem.*

Wniosek 9 został uzyskany z założenia 1 będącego negacją zdania 9 i z dołączonych w toku wnioskowania twierdzeń 2, 4, 5, 7, 8. Żeby oddać zwięźle schemat logiczny tego wnioskowania, oznaczmy Tezę dowodzoną 9 przez t, a Koniunkcję wymienionych pięciu twierdzeń przez k. Otrzymujemy wtedy formułę:

$$((\neq t \ \wedge \ k) \Rightarrow t) \Rightarrow (k \Rightarrow t)$$

Jak widać, wniosek t wolno uznać za prawdziwy pod warunkiem uznania twierdzeń K, co w powyższym wywodzie zostało uczynione, kolejno, w wymienionych wierszach 2, 4, 5, 7, 8.

Analogiczne rozumowanie można przeprowadzić w odniesieniu do predykatu „jest zdaniem oznajmującym” i do każdego innego, który by wyrażał warunek niezbędny prawdziwości, jak i fałszywości, zdania określonego języka.

Uzyskujemy w ten sposób dane, żeby podjąć pytanie tytułowe o zakres możliwości robota. Wiadomo, że łatwo mu przychodzi posługiwanie się logiką pierwszego rzędu, a więc wykona on kroki wnioskowania polegające na przejściu od wierszy 1 i 2 do 3, następnie od 4 i 5 do 6, wreszcie od 7 i 8 do 9. Problemem jest to, czy istnieją i dostępne są dlań algorytmy, których wykonanie doprowadziłoby do uznania twierdzeń z wierszy 2, 4, 5, 7, 8.

Odpowiedź komputacjonizmu na obecnym etapie, gdy stan badań nad mózgiem nie wystarcza do rozpoznania w nim takich algorytmów, przybierze postać alternatywy: albo algorytmy takie istnieją albo wniosek 9 nie jest należycie uzasadniony.

Pierwszy człon tej alternatywy może okazać się impulsem dla badań neurobiologicznych, ukazuje bowiem dobrze określony cel poszukiwań. Jeśli urządzeń elektronicznych nie da się tak zaprogramować, żeby konstruowały wnioskowania w rodzaju 1-9, to może wchodziłyby w grę algorytmy wykonalne dla ludzkich mózgów. Tak dostalibyśmy wskazówkę, czego w nich szukać. Wiemy bowiem, że warunkiem koniecznym algorytmiczności procedury jest dyskretność, to znaczy, złożenie z dających się oddzielić symboli, a ponadto sterowanie przekształcaniami symboli przez reguły prowadzące do rozwiązania danego problemu; na przykład, do skonstruowania Tarskiego definicji prawdy (potrzebnej w wierszu 2). Jeśli takie procedury algorytmiczne uda się odkryć w mózgach wykonujących rozumowania Gödla, Tarskiego czy Turinga, świadczyć to będzie, że są one robotami (nawet gdy mają konsystencję gęstej owsianki – jak wyraził się kiedyś Turing).

2. Zdania z predykatem wyrażającym warunek dostateczny prawdy

2.1. Zdanie S1 orzeka o sobie pewną własność, mianowicie poprawność syntaktyczną, a więc orzeka w sposób pozytywny. Obecnie rozważymy zdania odmawiające sobie pewnej własności, a więc orzekające o sobie negatywnie. Należy do nich zdanie gödlowskie jako odmawiające sobie dowodliwości.

Żeby skoncentrować się wpierw na właściwościach całej tej klasy, a po takim przygotowaniu zająć się pytaniem, co wyróżnia wsród jej elementów zdanie gödlowskie, rozważmy następującą wypowiedź.

S2. [Zdanie] S2 nie jest objawione [przez istotę nieomylną].

Uzasadnienie zdania S2.

(1) S2 jest objawione. — *założenie dowodu nie wprost.*
(2) Jeśli S2 jest objawione, to S2 jest prawdziwe. — *zasada aletyczna dla pojęcia objawienia.*
(3) S2 jest prawdziwe. — *1,2; reguła odrywania.*
(4) Jeśli S2 jest prawdziwe, to prawdą jest „S2 nie jest objawione". — *zdanie obserwacyjne.*
(5) Prawdą jest „S2 nie jest objawione". — *3,4; reguła odrywania.*
(6) Jeśli prawdą jest „S2 nie jest objawione", to S2 nie jest objawione. — *z Tarskiego definicji prawdy.*
(7) S2 nie jest objawione. — *5,6; reguła odrywania.*

Negacja zdania S2 także nie jest objawiona, bo skoro – jak okazaliśmy – S2 jest prawdą, to jego negacja, czyli zdanie „S2 jest objawione", musi być fałszywa, a zdanie fałszywe nie może być objawione. Jak widać, doktryna objawiona musi w sobie zawierać tezę o własnej niezupełności (wbrew skrajnym fundamentalistom, którzy by głosili, że objawienie zawiera wszystkie prawdy).

Podobnie jak w przypadku S1, wnioskowanie pokazuje, że zdanie S2 wynika z założenia będącego jego zaprzeczeniem przyjętego łącznie z koniunkcją twierdzeń, która tym razem składa się ze zdań 2, 4, 6. Schemat wnioskowania tylko tym się różni od poprzedniego, że inne jest rozłożenie negacji, mianowicie:

$$((t \wedge k) \Rightarrow\neq t) \Rightarrow (k \Rightarrow\neq t)$$

Jeden i drugi schemat czyni widocznym, że przyjęcie zdania t zależy od przyjęcia dołączonych w toku wnioskowania założeń.

W powyższym wywodzie zasadą aletyczną jest zdanie 2 podające pewne kryterium prawdy o charakterze warunku wystarczającego. W rozważanym przypadku jest ta zasada jakby cytatem z doktryny wyrażającej wiarę w istnienie prawd objawionych. Przypomnijmy (por. ustęp 1.2), że to, która z zasad aletycznych zostanie użyta w roli przesłanki zależy od predykatu użytego w zdaniu uzasadnianym. Skoro predykatem w S2 jest zwrot „jest objawione", to w rozumowaniu trzeba skorzystać z zasady przyjmującej objawienie za warunek wystarczający prawdziwości zdania.

2.2. Gdy rozważamy inne niż S2 zdania samopotwierdzalne odmawiające sobie pewnego przymiotu (tzn. różniące się od S2 zanegowanym predykatem), za każdym razem przeprowadzamy takie samo rozumowanie, zmieniając tylko w wierszu 2 zasadę aletyczną, odpowiednio do treści danego predykatu.

Godny uwagi jest predykat „jest tautologią [logiczną]”, jako że wystarczalność tego przymiotu do zapewnienia zdaniu prawdziwości (stwierdzona w zasadzie aletycznej) jest poza dyskusją, będąc niezależna od stanowiska filozoficznego. Oto jak, idąc tym tropem, wykazuje się samopotwierdzalność zdania:

ST. Zdanie ST nie jest tautologią.

Jeśli ST jest tautologią, to jest zdaniem prawdziwym (zasada aletyczna), a więc jest prawdą, że zachodzi ST, to znaczy, ST nie jest tautologią. A gdy z założenia, że ST jest tautologią wynika jego zaprzeczenie, to założenie to należy odrzucić jako fałszywe czyli przyjąć owo zaprzeczenie.

Negacja zdania ST także nie jest tautologią, bo skoro SP – jak właśnie wykazaliśmy – jest prawdą, to jego negacja jest fałszywa, a żadna tautologia nie jest fałszywa.

A oto kilka przykładów zasad aletycznych z klasycznych systemów epistemologii. Związane z każdą z nich zdanie oznaczone jest etykietą zawierającą literę „S” i inicjał filozofa reprezentującego w sposób klasyczny daną zasadę, podczas gdy etykiety złożone z „S” i symbolu cyfrowego (jak stosowane wyżej) są zarezerwowane dla zdań wyposażonych w pełne (a nie tylko szkicowe) wykazanie samopotwierdzalności.

Według Platona sądy dotyczące odwiecznych idei zawdzięczane anamnezie, czyli przypomnieniu wiedzy ze stadium preegzystencji duszy, są niezawodnie prawdziwe. Jest to platońska zasada aletyczna. Tak więc, zdanie odmawiające sobie cechy pochodzenia z anamnezy będzie z konieczności prawdziwe. Ma ono postać:

SP. Zdanie SP nie pochodzi z anamnezy.

Analogicznie jak w poprzednich wywodach, zaprzeczając w wyjściowym założeniu zdaniu SP, tym samym przyznajemy zdaniu SP pochodzenie z anamnezy, a ponieważ to pochodzenie jest wystarczającym gwarantem prawdziwości, zdanie SP musi być prawdziwe – wbrew założeniu, które wobec tego musi zostać odrzucone.

Negacja zdania SP także nie pochodzi z anamnezy, bo skoro – jak okazaliśmy – SP jest prawdą, to jego negacja jest fałszywa, a zdanie fałszywe nie może być gwarantowane anamnezą.

Zmieniając nieco sąsiedztwo w czasie i przechodząc do Arystotelesa, spotkamy predykat *jest intelektualnie oczywiste*, odnoszony np. do pierwszych zasad myślenia. Odpowiada mu zasada aletyczna:

SA. Zdanie SA nie jest [intelektualnie] oczywiste.

reszta zaś rozumowania przebiega jak wyżej: jeśli SA jest oczywiste, to jest prawdziwe (zasada aletyczna), a skoro głosi prawdziwie głosi, że nie jest oczywiste, to nie jest oczywiste.

Także zaprzeczenie SA czyli zdanie „SA jest oczywiste” nie jest oczywista, bo skoro SA jest jest prawdą, to jego negacja jest fałszywa, a więc nie może być oczywista.

Tak można kolejno prześledzić zasady aletyczne i odpowiednie rozumowania nawiązujące do innych klasyków. Oto zdania z predykatami wyrażającymi wystarczające (wedle danego systemu) warunki prawdziwości.

SD. Zdanie SD nie jest wypowiedzią *jeśli myślę, to jestem*. — Descartes.

SL. Zdanie SL nie jest prawdą rozumu. — Leibniz.
SK. Zdanie SK nie jest zdaniem analitycznym. — Kant.

W podobny sposób jak dla S2, ST, SP i SA pokazujemy, że negacja każdego z powyższych zdań nie jest, odpowiednio, zdaniem kartezjańskim, prawdą rozumu, zdaniem analitycznym.

Tą drogą dochodzimy do wniosku, że żaden z rozważanych zbiorów zdań, jak zdania objawione, tautologie, zdania oczywiste itd. nie ma cechy pełności, tak się ją rozumie w logice w odniesieniu do systemów dedukcyjnych. Nie jest więc tak, że dla każdego wziętego pod uwagę zdania języka, powiedzmy, logiki pierwszego rzędu jest ono tautologią lub zaprzeczeniem tautologii; istnieją formuły w tym względzie neutralne, jak choćby pojedyncza zmienna „p”. Nie jest też tak (jeśli uwierzymy w platońską anamnezę), że każde zdanie języka polskiego wyraża albo prawdę pochodzącą z anamnezy albo zaprzeczenie takiej prawdy; i tutaj mamy zdania neutralne, na które Platon miał nawet specjalne określenie: gr.) *doxa*, to jest, mniemanie. To samo dotyczy pozostałych predykatów powiązanych z zasadami aletycznymi.

Wśród takich predykatów jest jeden, który nie był dotąd brany pod uwagę, jak też nie była brana pod uwagę przyporządkowana mu zasada aletyczna. A ma on związek szczególny z zapisanym w tytule głównym problemem tych rozważań: czy robot potrafi uzasadnić prawdę konieczną? Jest to orzekany o pewnych zdaniach predykat: *dowodliwe w arytmetyce.* Ów związek szczególny na tym polega, że z czym, jak z czym, ale z arytmetyką, robot powinien sobie radzić dobrze. Na tym więc gruncie najbardziej dlań, by tak rzec, swojskim trzeba w pierwszej kolejności zbadać jego kompetencje.

3. Czy robot potrafi uzasadnić zdanie gödlowskie?

3.1. Samozwrotne zdania samopotwierdzalne są zdaniami koniecznymi *par excellence*, to znaczy, są tego gatunku wybitnymi reprezentantami. Są nimi z tego tytułu, że zdanie samopotwierdzalne wynika logicznie z własnego zaprzeczenia, trudno więc o wyższy rodzaj konieczności.

Z drugiej jednak strony, proces wywodzenia takiego zdania z jego negacji zawiera wśród przesłanek pewną zasadę aletyczną. Ta zaś nie musi należeć do zdań koniecznych. Może nawet wyrażać pogląd, który nie należy do bezspornych, jak ten z przykładu S2, gdzie zasada aletyczna za kryterium prawdy podaje objawienie. Także zasady aletyczne proponowane (ustęp 2.2) przez klasyków filozofii są siłą rzeczy kontrowersyjne, bo dla każdego klasyka istnieje inny klasyk, który się z nim nie zgadza.

Z tego powodu wysoka konieczność zdań samopotwierdzalnych jest zarazem względna; to znaczy, jest wysoka przy danym założeniu, którym jest odpowiednia zasada aletyczna. A zatem, żeby tę względność zredukować do zera lub jego bliskiej okolicy, trzeba mieć zasadę aletyczną, która sama się cechuje odpowiednio wysokim stopniem konieczności.

Idealnym w tym względzie kandydatem jest zdanie stwierdzające niesprzeczność arytmetyki. Jest to kandydatura, która zrazu może się wydać podejrzana, skoro nie ma i być nie może dowodu niesprzeczności arytmetyki (o ile nie sięgniemy po środki dowodowe uchodzące za ryzykowne), ale jej trafność da się wykazać przy pewnym pragmatycznym kryterium konieczności pochodzącym od Quine’a (i podzielanym przez piszącego te słowa).

Quine jest znanym przeciwnikiem dychotomii analityczne-syntetyczne, w której termin „analityczne” reprezentuje pojęcie konieczności. Nie dlatego, żeby tym terminom odmawiał

wszelkiego sensu, ale dlatego, że zamiast owej ostrej dychotomii dostrzega stopniowalność cechy konieczności w zbiorze tych zdań, które się składają na całość wiedzy naukowej.

Myśl Quine'a można oddać architektoniczną metaforą konstrukcji, z której pewne elementy dadzą się usunąć bez rujnowania budowli, podczas gdy usunięcie innych grozi zawaleniem się jakiejś części i pociąga potrzebę rekonstrukcji. Są też takie elementy, że ich usunięcie powodowałoby ruinę całego gmachu. Mamy tu na uwadze gmach wiedzy, którego elementami są teorie i ich twierdzenia.

Stopień niezbędności twierdzenia dla zachowania integralności konstrukcji to stopień jego konieczności. Jest to pragmatyczna interpretacja pojęcia wiedzy koniecznej, obecnego w epistemologii i w filozofii nauki co najmniej od Platona. W myśl tej interpretacji, na szczycie hierarchii konieczności znajdują się prawa logiki.

Dysponując taką miarą konieczności, zastosujmy ją do poglądu, że arytmetyka liczb naturalnych jest teorią niesprzeczną. Niezależnie od tego, jakie i jak silne argumenty da się na rzecz tego podać, istotne jest to, że gdyby arytmetyka okazała się sprzeczna, cały gmach ludzkiej wiedzy runąłby w gruzy. Wierze przeto w niesprzeczność arytmetyki atrybut konieczności – w świetle kryterium pragmatycznego – przysługuje w stopniu porównywalnym ze statusem praw logiki.

A zatem, jeśli chcemy zadać robotowi do uzasadnienia prawdę *par excellence* konieczną w postaci zdania samopotwierdzalnego, ale takiego, żeby wywód samopotwierdzalności wspierał się na koniecznie prawdziwej zasadzie aletycznej, to na taką zasadę nadaje się teza o niesprzeczności arytmetyki. Wypowiedzią zaś, której samopotwierdzalność, a więc i konieczność, w ten sposób się wywodzi jest sławne zdanie gödlowskie:

G. *Zdanie G jest niedowodliwe* [w arytmetyce].

W wywodzie uzasadniającym samopotwierdzalność zdania G, sąd o niesprzeczności arytmetyki będzie formułowany jako zasada aletyczna w postaci: *Jeśli zdanie jest dowodliwe* [*w arytmetyce*], *to jest prawdziwe.* Gdyby bowiem było dowodliwe, a zarazem fałszywe, czyniłoby to z arytmetyki zbiór zdań wewnętrznie sprzeczny.

3.2. Wywód, analogiczny co do struktury z wywodem zdania S2 (ustęp 2.1), podany jest niżej, samo zaś zdanie (dla zachowania ciągłości numeracji wywodów) opatrzone jest etykietą S3.

S3. [Zdanie] S3 nie jest dowodliwe [w arytmetyce].

Uzasadnienie zdania S3.

(1) S3 jest dowodliwe. — założenie dowodu nie wprost.
(2) Jeśli S3 jest dowodliwe, to S3 jest prawdziwe. — *zasada aletyczna dla pojęcia dowodu w arytmetyce.*
(3) S3 jest prawdziwe. — *1,2; reguła odrywania.*
(4) Jeśli S3 jest prawdziwe, to prawdą jest „S3 nie jest dowodliwe". — *zdanie obserwacyjne.*
(5) Prawdą jest „S3 nie jest dowodliwe". — *3,4; reguła odrywania.*
(6) Jeśli prawdą jest „S3 nie jest dowodliwe", to S3 nie jest dowodliwe. — *z Tarskiego definicji prawdy.*

(7) S3 nie jest dowodliwe. — *5,6; reguła odrywania.*

Powołanie się w wierszu 4 na obserwację (podobnie jak w wywodzie S2) należy tak rozumieć, że kierujemy wzrok na wiersz czwarty obecnego ustępu 3.2 i wtedy zauważamy, że zdanie figurujące pod etykietą S3 jest identyczne ze zdaniem, o którym mówi S3, skoro więc prawdą jest jedno z nich, to i drugie.

Powyższy wywód wywód można by uczynić krótszym, chodzi jednak o doprowadzenie go do takiej postaci, żeby zdanie w ostatnim wierszu było negacją zdania w pierwszym rozpoznawalną jako negacja po samym kształcie napisu. Wtedy jest wyraźnie widoczne, że rozumowanie podpada pod schemat (ten sam, co zastosowany do S2) wskazujący na samopotwierdzalność naszego zdania, mianowicie: $((t \wedge k) \Rightarrow\neq t) \Rightarrow (k \Rightarrow\neq t)$.

Negacja zdania S3 także nie jest dowodliwa, bo skoro – jak okazaliśmy – S2 jest prawdą, to jego negacja jest fałszywa, a zdanie fałszywe nie może być dowodliwe w arytmetyce, jeśli arytmetyka ma być teorią niesprzeczną.

Nie da się zaprzeczyć, że wywód (1)-(7) spełnia kryteria dowodu formalnego. Powstaje on przez trzykrotne zastosowanie reguły odrywania, a więc operacji czysto syntaktycznej, a struktura całości określona jest regułą opartą na wyżej cytowanym prawie logiki zdań. Reguła ta pozwala uznać za tezę danego systemu zdanie wynikające z własnego zaprzeczenia (i ewentualnie z innych zdań, o ile uznaje się te inne); mamy więc tu znowu postępowanie czysto syntaktyczne. A jeśli ta procedura jest dowodem, to czy fakt jej zastosowania nie przeczy treści zdania S3? Sprzeczność jednak nie zagraża, bo S3 mówi o swej niedowiedlności w arytmetyce. Powyższy zaś dowód przebiega nie w języku arytmetyki lecz w innym, z poziomu „meta", należącym do piętra wyższego niż język przedmiotowy arytmetyki.

To, że wynik ten otwiera drogę do wykazania, iż istnieje przynajmniej jedno zdanie niedowodliwe w samej arytmetyce, jest zasługą genialnego chwytu arytmetyzacji składni zastosowanego przez Gödla. Przejście przez ten drugi, wysoce techniczny, etap dowodu niepełności arytmetyki jest także tym, co czeka naszego robota, jeśli ma się on wykazać – w myśl doktryny komputacjonizmu – zdolnością rozwiązywania problemów dorównującą ludzkiej. Tu jednak koncentrujemy się na będącej do wykonania części wstępnej, mianowicie rozumowaniu (1)-(7).

Nasz robot, żeby dojść do konkluzji, musi po drodze dołączyć do założenia wyjściowego trzy przesłanki — te z wierszy 2, 4 i 6. Jeśli te przesłanki są prawdziwe, a przy tym komputacjoniści mają rację, to dojście do każdej z nich jest wynikiem procesu sterowanego algorytmem. Nie szkodzi, mówią oni, że nie powstały jeszcze uczynione przez człowieka roboty, które posiadłyby tę sztukę. Widać, że posiadł ją mózg ludzki, a więc on jest owym robotem, tyle, że wyprodukowanym przez Ewolucję a nie, powiedzmy, firmę Sun Microsystems.

To prawda, że jeśli w tym przypadku mózg realizuje jakiś algorytm, to się nie różni od robota. Skąd jednak mamy wiedzieć, że ów proces jest algorytmiczny? Rzecznicy komputacjonizmu nie mają na to, jak dotąd innego uzasadnienia, jak przyjęte a priori dwa założenia natury filozoficznej: (A) wszystkie procesy umysłowe są fizyczne; (B) wszystkie procesy fizyczne są algorytmiczne.

Poprawność sylogizmu powstałego z przesłanek A i B jest ponad wszelką wątpliwość, ale zasadność przesłanek wymaga zbadania. Ograniczę się do jednego z problemów związanych z przesłanką B.

3.3. Weźmy na warsztat, jako dostatecznie reprezentatywne dla problemu, zdanie (2) — zasadę aletyczną dotyczącą dowodu: *Jeśli zdanie jest dowodliwe w arytmetyce, to jest ono prawdziwe.*

Odpowiednio zaprogramowany robot ma pojęcie dowodu w tym sensie, że wykonuje procesy dowodzenia. Czy może mieć on także pojęcie prawdy? Jest to kwestia kluczowa dla oceny zasadności sformułowanej wyżej (ustęp 3.3) tezy B.

Proces dowodzenia tym się różni od aktu ujmowania prawdy, że jest sekwencją operacji na symbolach, podczas gdy ujęcie prawdy polega na uznaniu trafności reakcji układu na pewien stan jego otoczenia. Mamy więc w takim ujęciu dwa splecione ze sobą akty: reakcję i dotyczącą jej refleksję. Reakcja bywa bezpośrednia, jak w przypadku spostrzeżeń zmysłowych, lub pośrednia, jak w przypadku teorii będącej jakby przedłużeniem danych obserwacyjnych w kierunku tego, co już nie jest w świecie obserwowalne. W każdym przypadku refleksja nad reakcją dotyczy tego, czy jest ona trafnym rozwiązaniem problemu, przed jakim stał dany układ.

Tak więc, od tego rodzaju refleksji nieodłączne jest uprzednie wejście w to, co bywa nazywane sytuacją problemową. Ten drugi termin jest potrzebny, żeby odróżnić problemy artykułowane słownie, a przynajmniej będące przedmiotem świadomości, od takich, które rozwiązuje pierwotniak czy bakteria; jest przeto pojęcie sytuacji problemowej na tyle szerokie, żeby objąć nim coś, co się przydarza wszelkim istotom czującym, a nie tylko myślicielom umiejącym artykułowac swoje problemy. Podstawowe problemy istot czujących, to jak uniknąć pożarcia, jak samemu zaspokoić głód, jak zapewnić sobie niezbędne do życia warunki temperatury, oświetlenia itp.

Trafne rozwiązanie tego rodzaju problemu stanowi elementarne kryterium prawdy. A trafne rozwiązanie, czyli znalezienie prawdy, jest podstawowym probierzem inteligencji. Skoro mierzymy inteligencję częstością i jakością trafnych rozwiązań, to nie może odznaczać się inteligencją ten, kto nie żywi żadnych problemów.

> Przeoczenie tej zależności jest nieuleczalną słabością testu Turinga. Trudno tu oprzeć się potrzebie dokonania pewnych ocen. To, że Turing ułożył taki test w swym jedynym w życiu artykule napisanym dla periodyku filozoficznego (*Mind*, 1950), które to pismo jak i jego czytelników miał w nie największej może estymie (jakże ich umysły mogły się równać z odkrywcą nieobliczalności i zwycięzcą pojedynku z Enigmą), to da się psychologicznie zrozumieć (artykuł ten w paru miejscach wydaje się być „pisany na kolanie”, np. traktowanie równie serio argumentów przeciw sztucznej inteligencji pochodzących od Gödla i argumentów czerpanych z parapsychologii jest jawnym błędem w konstrukcji tekstu). To jednak, że nikt potem nie kusił się o testowanie komputera od strony jego zdolności stawiania pytań, a zarazem tylu autorów głosi jego równość, a nawet wyższość w stosunku do umysłu ludzkiego, to jest fakt mocno zagadkowy, który trzeba pewnie wyjaśniać w kategoriach socjologicznych. Jedno z podstawowych praw w tej dziedzinie odkrył już Andersen w bajce o nowych szatach cesarskich (do tej bajki nawiązuje Roger Penrose tytułem *Nowy umysł cesarza*, ale i on nie posunął się do wskazania na „nagość” testu Turinga).

Do narodzenia się pojęcia prawdy w czyimś umyśle nie wystarczy, że umysł ten trafnie rozwiązuje swe problemy, czy będą to owe elementarnie życiowe, czy te z wyższych stadiów rozwoju, gdy jako bodziec działa, mówiąc językiem Peirce’a, podrażnienie niepewnością (*irritation of doubt*). Ale i to wysublimowane podrażnienie wymaga podłoża biologicznego,

trudno by przydarzyło się ono płytce krzemu. Od praktyki rozwiązywania problemów do pojęcia prawdy droga prowadzi przez doświadczenie własnych błędów, a także cudzych zmyśleń lub kłamstw (słusznie Popper podkreśla, jak do rozwoju pojęcia prawdy przyczynili się łgarze). Dopiero w takiej sytuacji pojawia się potrzeba odróżnienia osobnymi określeniami błędów i kłamstw od rozwiązań trafnych i opowieści wiarogodnych. W wyniku takiej refleksji powstają terminy: *prawda* i *fałsz*.

3.4. Taka rekonstrukcja genezy pojęcia prawdy, sięgająca do biologicznego podłoża powstawania problemów, ustala linię graniczną między dwoma rodzajami czynnika fizycznego (*hardware'*u) — biologicznym i np. elektronicznym. Powstaje pytanie, czy ta różnica ma wpływ na wchodzący w grę czynnik logiczny. Ten drugi obejmuje dziedzinę programów (*software'*u), ale nie musi się do niej ograniczać. Programy, jako algorytmy, są reprezentowane przez liczby obliczalne, da się jednak pomyśleć sterowanie czynnikiem fizycznym przez czynnik logiczny nie będący algorytmicznym, czyli taki, który byłby reprezentowany przez liczby nieobliczalne. Nie wiedząc, czy taki istnieje, możemy go rozważać hipotetycznie; a na użytek tych rozważań zarezerwować termin *czynnik logiczny* jako nadrzędny wobec tych dwóch, z których jeden jest związany z liczbami obliczalnymi (maszyna Turinga i wszelka maszyna cyfrowa), drugi zaś z nieobliczalnymi.

Przy takim sformułowaniu problemu, komputacjonizm ma gotową odpowiedź: czynnik logiczny drugiego rodzaju nie istnieje, wobec tego czynnik pierwszego rodzaju wystarczy, żeby sterować wszelkim czynnikiem fizycznym, w tym biologicznym. Jest to odpowiedź przytoczona na samym początku tych rozważań; trzeba więc teraz, zmierzając do ich konkluzji, pokazać, na ile one posunęły dyskusję. Liczy się nie tylko ostateczna odpowiedź, ale i każde przybliżenie do odpowiedzi przez dalej idące sprecyzowanie zagadnienia.

Sprecyzowanie, które zawdzięczamy analizie zdań samopotwierdzalnych, w szczególności zdania gödlowskiego, lokalizuje dokładniej onus probandi, czyli gdzie spoczywa ciężar dowodzenia, po stronie komputacjonizmu. Żeby wykazać, że także robot może dojść do zdania gödlowskiego, komputacjonizm powinien wziąć za przedmiot badań proces dochodzenia do pojęcia prawdy, ten mianowicie, który zaczyna się na szczeblu rozwiązywania przez organizm elementarnych problemów. Tak dokładnie mając wskazany kierunek natarcia, badacz miałby za zadanie zidentyfikować ów algorytm neuronowy sterujący ewolucją pojęcia prawdy aż po wywód gödlowski.

Jeśli algorytm taki istnieje, nic nie będzie stać na przeszkodzie, żeby go implementować w dowolnym układzie działającym na modłę maszyny Turinga, pod warunkiem zapewnienia temu układowi odpowiedniej interakcji z otoczeniem niosącym dlań zagrożenia jak i szanse. Nie można jednak wykluczyć z góry takiego wyniku, że jakieś składające się na ten proces czynności poznawcze mają charakter analogowy, co może (choć nie musi) otwierać drogę liczbom nieobliczalnym. Wstępna refleksja dostrzega charakter analogowy choćby percepcji wzrokowej czy słuchowej; trzeba wyjaśnić, czy da się ona zastąpić należycie dokładną symulacją cyfrową. Jeśli się to uda, będzie to punkt dla komputacjonizmu.

Są jednak następne pytania. Na przykład, w procesach podejmowania decyzji pojawiają się oszacowania wielkości korzyści i wielkości prawdopodobieństwa. Według komputacjonizmu, takie oszacowania są procesami fizycznymi. Jak je scharakteryzować co do algorytmiczności? Matematyczna funkcja prawdopodobieństwa może przybierać ex definitione

wartości nieobliczalne, ale mózg nie musi korzystać z tej całej obfitości. Być może, te jego stany fizyczne które polegają na instynktownym szacowaniu czegoś takiego jak prawdopodobieństwo (instynktownym, to jest, nie jawiącym się jako określona liczba na poziomie świadomości) przybierają tylko wartości wymierne, odpowiadające jakimś wymiernym wartościom prawdopodobieństwa z przedziału od 0 do 1. Cokolwiek by się twierdziło w tej materii, na tak, czy na nie, trzeba tego dowieść. Tego właśnie — przyznania, że sprawa jest otwarta — wymaga także od komputacjonistów analiza stanu zagadnienia w obecnym stadium badań.

Werdykt przyznający słuszność komputacjonistom zapadłby wtedy, gdyby został odkryty algorytm prowadzący do ukształtowania się w naszych mózgach stanu fizycznego będącego zapisem pojęcia prawdy. Wtedy robotowi nie będzie trudno wykazać także prawdziwość zdań koniecznych z gatunku samopotwierdzalnych. Autor zaś tego eseju nie będzie mógł wystąpić z pozwem o zniesławienie, jeżeli zostanie przez kogoś nazwany maszyną, tak samo bowiem wolno będzie nazwać Gödla. Tymczasem jednak wydaje się rozsądne wstrzymywać się od tego rodzaju określeń wobec jakichkolwiek osobników rasy ludzkiej.