

Tytuł: **Test Turinga**\*

Autor: Marek Kasperski / [mjkasperski@kognitywistyka.net](mailto:mjkasperski@kognitywistyka.net)

Recenzent: prof. dr hab. Jerzy Perzanowski (UMK), prof. dr hab. Włodzisław Duch (UMK)

Źródło: <http://www.kognitywistyka.net> / [mjkasperski@kognitywistyka.net](mailto:mjkasperski@kognitywistyka.net)

Data publikacji: 05 II 2002

Kamieniem węgielnym dla ukonstytuowania się *Artificial Intelligence* są dwa, klasyczne już teksty Alana Turinga: *Computing machinery and intelligence*<sup>1</sup> (1950) i *Can a Machine Think* (1950), z których pochodzi pojęcie testu Turinga.

**Test Turinga** czy – jak sam autor go zwał – **gra w naśladownictwo** [*imitation game*] wyrasta z zabawy towarzyskiej w stylu „retro”<sup>2</sup>. Uczestniczyły w niej cztery osoby: 1 – mężczyzna (zaznaczmy go literą *A*), 2 – kobieta (zaznaczmy ją literą *B*), 3 – goniec dowolnej płci (zaznaczmy go literą *Z*) i 4 – osoba dowolnej płci, zwana sędzią bądź zgadującym (zaznaczmy go literą *C*). Do obowiązków ostatniego z graczy należało zadawanie dowolnych pytań niewidzianym osobom 1 i 2, z którymi łączność umożliwiał goniec zanoszący pytania i dostarczający odpowiedzi. Celem tej zabawy było odgadnięcie przez (*C*) na podstawie pytań i odpowiedzi tego, kto jest kim. Tak, pokrótce, bawili się nasi dziadkowie. Turing zaś wykorzystał fakt istnienia tej zabawy do rozwiązania problemu stanowiącego odpowiedź na pytanie o to, czy maszyny mogą myśleć.

Nową postać problemu można opisać za pomocą gry, którą nazywamy „grą w naśladownictwo”. W grze tej biorą udział trzy osoby: mężczyzna (*A*), kobieta (*B*) i człowiek zadający pytania (*C*), który może być dowolnej płci. Pytający znajduje się w pokoju oddzielnym od pokoju zajmowanego przez dwu pozostałych. Jego zadaniem w grze jest rozstrzygnięcie, który z dwu pozostałych uczestników gry jest mężczyzną, a który kobietą. Zna ich on jako *X* i *Y* i przy końcu gry mówi: „*X* jest *A*, a *Y* jest *B*” lub „*X* jest *B*, a *Y* jest *A*”.<sup>3</sup>

Należy tutaj dodać, że informacja przepływająca od osoby pytającej (*C*) do osób oznaczonych jako (*A*) i (*B*) jest relacją zwrotną i odbywać się miała poprzez dalekopis (w przypadku towarzyskiej zabawy odbywało się to dzięki dodatkowej osobie gońca (*Z*)). Przy czym każda z dwu osób przepytanych musi przyjąć specyficzną dla siebie rolę: osobnik (*A*) musi ułatwiać przepytującemu (*C*) zadanie i zawsze mówić prawdę, a osobnik (*B*) wręcz

\* Tekst jest fragmentem autorskiej książki: M. J. Kasperski, *Sztuczna Inteligencja. Droga do myślących maszyn*, Helion, Gliwice 2003.

<sup>1</sup> Pierwotnie ukazał się w: "Mind", Nr (236)/1950; polskie wydanie tłum. D. Gajkowicz, w: *Maszyny matematyczne i myślenie*, red. E. A. Feigenbaum, J. Feldman, ss. 24-47, dostępny także na stronie <http://www.kognitywistyka.net>.

<sup>2</sup> Por. Hołyński 1979, ss. 16-17.

<sup>3</sup> Turing 1972, s. 24.

przeciwnie – ma to zadanie utrudniać przez możliwość kłamania. Pytania, które można zadawać, mogłyby przyjąć zaprezentowane niżej formy.

C: Jakiego koloru masz włosy?

X: Moje włosy są koloru miedzi i błyszczą tak, jakby były nawoskowane.

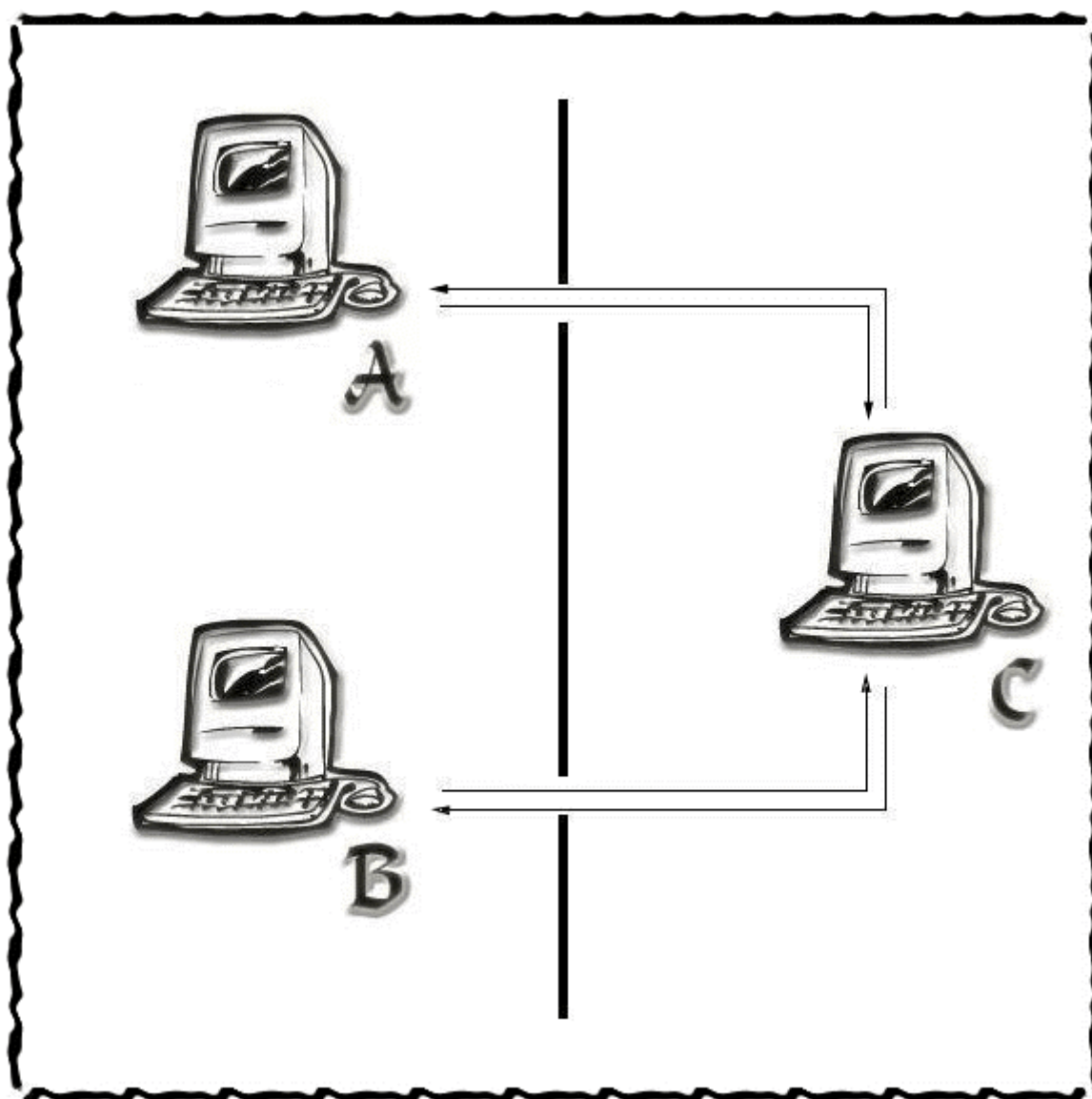
Y: A moje włosy są kruczoczarne.

C: A jakiej są długości?

X: Są tak krótkie, jak mogą tylko być, żeby nie stwierdzić, że już ich nie ma.

Y: Moje również są bardzo krótkie, choć z pewnością są dłuższe niż najkrótsze.

I tym podobne.



Rysunek 2.1. Test Turinga. Oznaczenia *A*, *B*, *C* wskazują, gdzie znajdują się w kolejności: mężczyzna (*A*), który do chwili odgadnięcia jest oznaczany jako *X*, kobieta (*B*), która do chwili odgadnięcia oznaczana jest *Y*, przepytujący (*C*), który ma odpowiednio przypisać czy  $X = A$  a  $Y = B$ , czy na odwrót  $X = B$  a  $Y = A$ . Cienkie linie wskazują na połączenia między nimi, dzięki którym są wysyłane i uzyskiwane dane. Czarna kreska widoczna po

środku jest ścianą oddzielającą pytającego od przepytanych. Dziś taki test można przeprowadzić przy dowolnym połączeniu komputerów, na przykład w sieci Internet.

Więc sytuacja wygląda tak, jak pokazano na rysunku 2.1.

Teraz – proponuje dalej Turing – zapytujemy: „Co stanie się, gdy maszyna zastąpi A w tej grze?”<sup>4</sup>

Był to rewolucyjny, jak się miało okazać, pomysł na rozwiązanie podstawowego zagadnienia filozofii Sztucznej Inteligencji! Do tej pory trudno było stawiać pytania o „myślenie maszyn”, co najmniej z dwu powodów: 1) ze względu na nieprecyzyjność słowa ‘myśleć’ i 2) ze względu na bardzo słynny fragment *Monadologii* W. G. Leibniza (§ 17)<sup>5</sup>, który podkreślał zwyczajny *automatyzm* maszyny, zaprzeczając jednocześnie możliwości myślenia *sensu stricte*.

W przypadku pierwszego aspektu – nieprecyzyjności słowa ‘myśleć’ – dominowały argumenty zaprzeczające możliwości myślenia przez maszyny. Wszystkie one, chcąc nie chcąc, wyrażały zawsze tę samą tezę:

**Maszyny nie myślą i nie mogą myśleć,  
gdyż myślenie jest domeną ludzi.**

Teza powyższa jest w prostej linii **dziejnictwem arystotelizmu**, wedle którego wyraz ‘dusza’ (w sensie „dusza człowieka posiadająca zdolność myślenia”) był terminem zhierarchizowanym, trójstopniowym i rozkładał się na<sup>6</sup>: (1) duszę wegetatywną, której cechą jest zdolność „ruchu od wewnątrz” zwana autokinezą, (2) duszę sensorywną, którą wyróżnia zdolność do zmysłowego poznania, oraz (3) duszę rozumną, którą cechuje zdolność do używania rozumu. W tym zhierarchizowanym trój-układzie zachodzą również relacje dookreślające to, z jakim rodzajem bytu mamy do czynienia: pierwszy rodzaj duszy przynależy do świata roślinnego, pierwszy i drugi rodzaj duszy przynależy do świata zwierząt, a pierwszy, drugi i trzeci rodzaj duszy przynależy do człowieka. Podział ten został dokonany przez Arystotelesa w dziele *O duszy*. Jak już zdążyliśmy się dowiedzieć z lektury pierwszego rozdziału, z podziałem tym już w XVII wieku walczyli między innymi W. G. Leibniz i J. O. de La Mettrie. R. Descartesa, co także już zostało powiedziane, powiódł on zaś do skrajnego dualizmu (tzw. **dualizmu substancji**), a przez niewyjaśniony interakcjonizm można go śmiało nazwać wyrazem mistycyzmu<sup>7</sup>, jak to zrobił współcześnie R. Penrose<sup>8</sup>.

<sup>4</sup> Turing 1972, s. 24.

<sup>5</sup> § 17. „Należy wszakże przyznać, że postrzeżenie i to, co od niego zależy, nie da się wytłumaczyć racjami mechanicznymi, tzn. przez kształty i ruchy. Przypuściwszy zaś, że istnieje maszyna, której budowa pozwala, aby myślała, czuła, miewała postrzeżenia, będzie można pomyśleć ją, z zachowaniem tych samych proporcji, tak powiększoną, by można do niej wejść jak do młyna. Założywszy to, odnaleźlibyśmy wewnątrz przy zwiedzaniu jej tylko części, które popychają siebie wzajemnie, nigdy jednak nic, co tłumaczyłoby postrzeżenie. Toteż trzeba szukać tego właśnie w substancji prostej, a nie w rzeczy złożonej czy też w machinie. I tylko to można odnaleźć w substancji prostej, tzn. postrzeżenia oraz ich zmiany. Na tym jedynie mogą polegać wszystkie czynności wewnętrzne substancji prostych.” [Leibniz 1995].

<sup>6</sup> Por. Arystoteles 1992.

<sup>7</sup> Słowo „mistycyzm” jest użyte w formie, w jakiej używa go Penrose; patrz uwagi nt. we *Wstępie*.

<sup>8</sup> Por. Penrose 1995; Penrose 2000; Penrose 1997.

Przy takim podejściu na pytanie, o to, czy maszyny mogą myśleć, należałoby odpowiedzieć:

Nie – jeśli zdefiniuje się myślenie jako działalność specjalnie i wyłącznie ludzką. Wtedy każde takie zachowanie się maszyn trzeba byłoby nazwać zachowaniem się podobnym do myślenia.

Nie – jeśli zakłada się, że w samej istocie myślenia jest coś *niezgłębionego, tajemniczego, mistycznego*.<sup>9</sup>

Jak widać zatem, implikacje takiego stanowiska w kwestii wyżej postawionego problemu są nieużyteczne — nie mówią nic poza tym, co orzekają, a nasz problem zawsze stawiają na pozycji przegranej (z podobnej metody argumentacji, obarczonej dziedzictwem arystotelizmu, korzysta argument *Chińskiego Pokoju*, J. Searle’a, o którym mowa będzie w podrozdziale 2.3.5 – „Analiza języka i mowy. Problemy z rozumieniem”; por. ss. 139-140). Na ich podstawie po pierwsze, nie jesteśmy skłonni przypisywać funkcji myślenia maszynom, ale i po drugie, jesteśmy skłonni zawsze taką funkcję przypisywać ludziom, co jak wiadomo z badań nauk o mózgu, neuropsychologii i z psychologii postaci, nie zawsze jest prawdziwe.

W ostateczności też, podsumowując ten podrozdział, z testem Turinga będziemy mieli do czynienia wówczas, gdy będą spełnione następujące warunki:

1. **Znajdzie się co najmniej dwoje ludzi i co najmniej jedna maszyna bądź program.**
2. **Maszyna bądź program muszą być takiego rodzaju, aby móc prowadzić dialog.**
3. **Żaden z listy wymienionych uczestników nie może widzieć innego uczestnika.**
4. **Każdy jeden z listy wyżej wymienionych uczestników będzie prowadził dialog z jedną, wcześniej wyznaczoną do tego osobą, zwaną od tego czasu Sędzią bądź Przepytującym. Osobą tą może być tylko człowiek.**
5. **Zadaniem Sędziego jest osądzić na podstawie przeprowadzonego dialogu z pozostałymi uczestnikami, kim jest każdy jeden z uczestników prowadzących z nim dialog. Przy czym oczekiwanymi odpowiedziami są: „Człowiek” i „Maszyna”.**
6. **Każda jedna maszyna bądź każdy jeden program, który podczas prowadzenia dialogu z Sędzią oszuka go tak, iż Sędzia stwierdzi, że ma do czynienia z człowiekiem, przejdzie pozytywnie test Turinga i będziemy mogli stwierdzić z całą stanowczością, iż jest to maszyna myśląca bądź myślący program. Oczywiście myślący w sensie Turinga.**

Doniosłość testu Turinga ze względu na możliwość rozpatrywania problemu myślenia maszyn jest zatem bezapelacyjna. Jednak warto też zapoznać się z argumentami stawianymi przeciw takiemu rozpatrywaniu interesującego nas problemu, z argumentami, z którymi musiał się zmagać sam Turing, a także – dalej – z krytyką powstałą już po ukazaniu się w druku *Computing machinery and intelligence*, czyli od przeszło pięćdziesięciu lat.

### 2.2.1. Test Turinga – zarzuty przez niego samego wyszukane i przezeń rozważone

Teraz przyszła pora, aby zaprezentować listę zarzutów, którą Turing przedstawił, zanalizował i odparł<sup>10</sup>. Możemy też poddać je ocenie z perspektywy 50 lat badań z zakresu Sztucznej Inteligencji by dodatkowo przemyśleć.

<sup>9</sup> [D] 1972, s. 16; wyróżnienie M. K.

<sup>10</sup> Turing 1972.

## A. Sprzeciw teologiczny

Turing zwraca uwagę na to, że w tradycji myśli europejskiej ważną rolę odgrywa stary, przedplatoński pogląd o *nieśmiertelnej* duszy. Ponadto – zgodnie ze wspomnianą już tradycją Arystotelesa – zakłada się, że dusza ta (tzn. dusza rozumna) przysługiwać by miała tylko człowiekowi i – na podstawie połączenia tych rozważań z koncepcjami chrześcijaństwa – wierzy się, że jest dana od Boga. W rezultacie oprócz człowieka nie może posiadać jej żadne inne stworzenie na ziemi, a tym bardziej – jakkolwiek artefakt. Tak pokrótce przedstawia się ów zarzut. A co na to sam Turing?

Wydaje mi się, że cytowany wyżej argument pociąga za sobą poważne ograniczenie wszechpotęgi Boga Wszechmogącego. Przyznano, że istnieją pewne rzeczy, których On nie może zrobić, takie jak uczynienie jedności równą dwóm, ale czyż nie powinniśmy wierzyć, że może On obdarzyć duszą słonia, jeśli będzie uważał, że słoń jest tego godny? (...)

Podobny argument można sformułować w przypadku maszyn. Może on wydawać się inny, gdyż jest trudniejszy do „przełknięcia”. Ale naprawdę oznacza on jedynie nasze przekonanie o mniejszym prawdopodobieństwie uważania przez Niego tych warunków materialnych za odpowiednie do obdarzenia duszą.<sup>11</sup>

Czy nie powinniśmy zatem wierzyć, że może On obdarzyć duszą maszyny, jeśli tylko będzie uważał, że jest tego godna?!

Jednakże jest to tylko spekulacja. Teologiczne argumenty – dodaje Turing – nie wywierają na mnie głębokiego wrażenia. (...) Gdyż (...) takie argumenty często bywały niewystarczające w przeszłości: w czasach Galileusza argumentowano, że teksty: „I słońce stało jeszcze... i nie spieszyło się zejść prawie przez cały dzień” (Jozue X. 13) i „Dał ziemi podstawę, tak, że nigdy nie powinna się ona ruszyć” (Psalm CV. 5) w sposób wystarczający zbijają teorię Kopernika. Przy naszej obecnej wiedzy taki argument wydaje się bezwartościowy. Gdy ta wiedza nie była dostępna, to wywierało to zupełnie inne wrażenie.<sup>12</sup>

## B. Sprzeciw „głów w piasku”

Najkrócej mówiąc przedstawia się on następująco:

Konsekwencje myślenia maszyn byłyby zbyt okropne. Miejmy nadzieję i wiermy, że one nie mogą myśleć.<sup>13</sup>

A zatem, po pierwsze: ze względu na okrucieństwa ludzkości (jak na przykład wyeliminowanie w czasach prehistorycznej walki o byt całego gatunku *homo neandertalis*<sup>14</sup>),

<sup>11</sup> Tamże, ss. 32-33.

<sup>12</sup> Tamże, ss. 32-33.

<sup>13</sup> Tamże, s. 33.

Dobrym komentarzem do dwu powyższych argumentów – A. i B. – zdaje mi się następujący cytat z *Badania...* Hume’a: „Nie ma bardziej rozpowszechnionej, a przecież bardziej nagannej metody rozumowania, jak starać się w dyskusjach filozoficznych odeprzeć jakieś przypuszczenie przez wykazywanie jego konsekwencji niebezpiecznych dla religii i moralności. Jeżeli pogląd jakiś prowadzi do niedorzeczności, jest on niewątpliwie mylny; nie jest jednak jakiś pogląd niewątpliwie mylny dlatego, że wynikają z niego niebezpieczne konsekwencje. Takich argumentów powinno się zupełnie unikać; wszak nie przyczynią się niczem do wykrycia prawdy a służą tylko do tego, by zozydzić przeciwnika”. [Hume 2001, s. 95].

bo i z jakiej innej przyczyny, boimy się, że myślące maszyny będą sobie uzurpować bytowe pierwszeństwo i w rezultacie tego walczyć z nami. Jest to wyraz tzw. *mitu buntów robotów*, rozpowszechnianego zwłaszcza w latach 50-70. w literaturze *science fiction* (mowa o nim będzie w rozdziale 3. – „Jutro”, s. 199).

Po drugie:

Chcemy wierzyć, że Człowiek jest w jakiś subtelny sposób wyższy ponad resztę stworzenia. Najlepiej byłoby, gdyby można było wykazać, że jest on *bezwzględnie* wyższy, ponieważ wówczas nie istniałoby niebezpieczeństwo utraty jego dominującej pozycji.<sup>15</sup>

Jak widać, argument ten, ale i poprzedni – sprzeciw teologiczny, łączy nić tzw. dziedzictwa arystotelizmu, które nakreśliłem już wcześniej. Odpowiedź Turinga była następująca:

Nie myślę, że ten argument jest wystarczająco poważny, aby trzeba go było zbijać.<sup>16</sup>

Współcześnie takie poglądy sprzeciwu „głów w piasku” reprezentują między innymi: Norman Lillegard<sup>17</sup> i Bill Joy<sup>18</sup>, a w Polsce Eugeniusz Szumakowicz<sup>19</sup>.

### C. Sprzeciw matematyczny

Jest to najczęściej pojawiający się zarzut przeciwko możliwości skonstruowania maszyny myślącej. Przynajmniej się go powołując się na twierdzenia Kurta Gödla (1906-1978) (o niesprzeczności i niezupełności) z 1931 roku<sup>20</sup> czy też w rachunku lambda Alonzo Churcha (1903-1995) z 1936<sup>21</sup>, stosowne twierdzenie z teorii rekursji Stephena C. Kleene’a (1909-

<sup>14</sup> Por. chociażby: I. Tattersall, *Nie zawsze byliśmy sami*, w: „Świat Nauki”, Nr 4/2000, ss. 26-32.

<sup>15</sup> Turing 1972, s. 33.

<sup>16</sup> Tamże, s. 33.

<sup>17</sup> Wyraża się to na przykład w tezie: „W jednym z najbardziej odrażających momentów we wspomnianym odcinku serialu *Star Trek Data* (jeden z bohaterów, który jest androidem; przyp. M. K.) oświadcza, że jest on („coś”) w zażyłych stosunkach z pewną kobietą, która jest członkiem załogi. Pomysł ten jest bardziej odpychający niż zabawny”. [Lillegard 1995, s. 56].

<sup>18</sup> Wyraża się to w tezie: „Pisząc programy, projektując mikroprocesory, nigdy nie miałem poczucia, że tworzę inteligentne maszyny. *Software* i *hardware* są tak delikatne, a zdolność maszyny do „myślenia” tak nieobecna, że ewentualność taka wydawała mi się zawsze bardzo odległą przyszłością. Teraz rzecz wygląda inaczej: być może pracuję nad narzędziami, które w przyszłości zastąpią nasz gatunek. Jak się czuję? Kiepsko”. A zestawieć to można ze: „Zbudowanie pierwszego (robota, maszyny myślącej; przyp. M. K.), który będzie miał zdolność samopowielania, da początek całemu gatunkowi.” [Joy 2000, s. 93].

<sup>19</sup> Wyraża się to w tezie: (...) „najśmielsze odkrycia naukowe i wynalazki techniczne nie są w stanie dać nam niczego filozoficznie interesującego, gdy chodzi o tak zwaną sztuczną, czyli faktycznie alternatywną inteligencję (sic!, przyp. M. K.). Natomiast mogłyby spowodować katastrofę humanitarną w stylu George’a Orwella”. [Szumakowicz 2000, s. 146] Albo w jeszcze innej: „Jednakże, jeśli poważnie założyć jej (sztucznej inteligencji; przyp. M. K.) realność (a poważne filozofowanie do tego zobowiązuje!), to nie sposób nie zauważyć swego rodzaju puszki Pandory przeróżnych trudności i niebezpieczeństw: społecznych, ekonomicznych, psychologicznych i jeszcze innych”. [Szumakowicz 2000, s. 138.] Ze swej strony tylko dodam, że wyrażenie „o poważnym filozofowaniu” biorąc pod uwagę tekst wyżej wymienianego jest nadużyciem!

<sup>20</sup> K. Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und Verwandter Systeme*, w: „Monatshefte für Mathematik und Physik”, 1931, ss. 173-189.

<sup>21</sup> A. Church, *An unsolvable problem of elementary number theory*, w: „American Journal of Mathematics”, Nr (58)/1936, ss. 345-363.

Na temat „rachunku lambda” patrz między innymi: Penrose 1995, ss. 85-90; M. Nowicki, *Syntaktyczne twierdzenia limitacyjne, wyłożone sposobem Turinga, z konkluzjami Chatina*, Internet. W ostatniej pracy znajdują się również wiadomości na temat tak twierdzenia Gödla, jak i Churcha-Turinga.

1994) z 1935<sup>22</sup>, czy też analogiczne twierdzenia samego A. Turinga i J. B. Rossera z 1937<sup>23</sup>, dotyczące *Entscheidungsproblem* postawionego przez Davida Hilberta. W twierdzeniu Gödla udowadnia się,

że w każdym dostatecznie potężnym systemie logicznym można sformułować twierdzenia, których w ramach tego systemu nie można ani udowodnić, ani wykazać ich błędności, chyba że w ogóle sam system jest niekonsekwentny [niekoherentny].<sup>24</sup>

Obecnie często powołuje się nań Roger Penrose pisząc, że

Odpowiednie procesy fizyczne w mózgu powodują powstanie świadomości, ale tych procesów nie można nawet symulować obliczeniowo.<sup>25</sup>

Czyli, konkludując:

nie może istnieć algorytm dowodzenia twierdzeń w arytmetyce – przy założeniu, że jest ona systemem niesprzecznym. Innymi słowy, nie można podać procedury tak ogólnej, by mieć pewność, że przy danym układzie aksjomatów da się podać (w ramach tej procedury, przyp. M. K.) dowód każdego prawdziwego twierdzenia arytmetyki.<sup>26</sup>

No i, jak by to nie było, Turing akceptuje swój dorobek zaprezentowany w tekście *On computable numbers, with an application to the Entscheidungsproblem* tak samo, jak z wynikiem swojej pracy *Über formal unentscheidbare Sätze der Principia Mathematica und Verwandter Systeme* nie polemizowałby Kurt Gödel. Obaj sądzili, że są takie prawdy, do których nie będzie mogła dojść maszyna Turinga. Lecz ani pierwszy, ani drugi

Nie sądził jednak, żeby były one bardziej osiągalne dla człowieka niż dla maszyny. Głęboko był przekonany, że ludzie i maszyny „jadą na jednym wozie”.<sup>27</sup>

Jak pisze sam Turing:

Krótką ripostą na ten argument jest to, że **choć ustalono, że istnieją granice możliwości każdej poszczególnej maszyny, to jednak jedynie bez dowodu stwierdzono, że żadne takie ograniczenia nie stosują się do ludzkiego intelektu.** Nie jestem jednak zdania, że ten pogląd można zbyć tak łatwo. Za każdym razem, gdy jednej z tych maszyn zadaje się odpowiednio krytyczne pytanie i daje ona określoną odpowiedź, to wiemy, że ta odpowiedź musi być błędna i daje nam to pewne poczucie wyższości. Czyżby to uczucie było złudne? Jest ono bez wątpienia zupełnie nieklamane, ale myślę, że nie należy zbyt wielkiej wagi do niego przywiązywać. My sami zbyt często dajemy błędne odpowiedzi na pytania, aby można było usprawiedliwić nasze zadowolenie z takiego dowodu omyłności części maszyn. (...) Tak więc, krótko mówiąc, mogliby być ludzie zdolniejsi od każdej danej maszyny, ale z kolei mogłyby być inne zdolniejsze maszyny itd..<sup>28</sup>

<sup>22</sup> S. C. Kleene, *General recursive functions of natural numbers*, w: "American Journal of Mathematics", Nr (57)/1935, ss. 153-157, 219-244.

<sup>23</sup> A. M. Turing, *On computable numbers, with an application to the Entscheidungsproblem*, w: "Proceedings of the London Mathematics Society", Nr (42)/1936, ss. 230-265; Nr (43)/1937, s. 544.

<sup>24</sup> Turing 1972, s. 33.

<sup>25</sup> Penrose 2000, s. 31.

<sup>26</sup> Marciszewski 1998, s. 21.

<sup>27</sup> Tamże, s. 21.

<sup>28</sup> Turing 1972, s. 33; wyróżnienie M. K.

Czyli jedno jest pewne – tym, co się nam nie uda (ale przecież nie mamy takich zamiarów), jest zbudowanie środkami finistycznymi maszyny wszechwiedzącej, o której pisał między innymi Lem.

Współcześnie na powyższy argument powołuje się przede wszystkim jeden z zagorzałych krytyków silnej wersji AI (*strong AI*), Roger Penrose<sup>29</sup>.

#### D. Argument świadomości

Jest to drugi z najczęściej pojawiających się zarzutów, który współcześnie odżył zapewne pod wpływem badań nad fenomenem świadomości prowadzonych w obrębie nauk o mózgu między innymi przez: Francisę Cricka i Christofa Kocha<sup>30</sup>, Semira Zeki<sup>31</sup>, Nikosa K. Logotheisa<sup>32</sup>, Susan Greenfield<sup>33</sup> i wielu innych badaczy. Ale nie tylko przez nich. Również przyczyniły się do owego zainteresowania pisma współczesnych filozofów, takich jak: Daniel Dennett<sup>34</sup>, John R. Searle<sup>35</sup>, Paul Churchland i Patricia Smith Churchland<sup>36</sup>, czy też Hilary Putnam<sup>37</sup>. Argument ów

Jest bardzo dobrze wyrażony w mowie profesora Jeffersona wygłoszonej w 1949 r., z której cytuję: „Dotąd nie będziemy mogli zgodzić się z poglądem, że maszyna jest równa mózgowi, dopóki maszyna nie potrafi napisać sonetu lub skomponować koncertu, dzięki odczuwanym myślom i emocjom, a nie dzięki szansie natrafienia na odpowiednie symbole, to znaczy – potrafi nie tylko napisać je, ale także wiedzieć, że je napisała. Żaden mechanizm nie może odczuwać (a nie jedynie sztucznie sygnalizować łatwy fortel) przyjemności ze swojego sukcesu, zmartwienia, gdy jej lampy topią się, nie może podniecać się pochlebstwem, cierpieć z powodu swoich

<sup>29</sup> Por.: Penrose 1995; Penrose 2000; Penrose 1997. Patrz wcześniejszy cytat z Penrose 2000.

Należy tutaj nadmienić, iż tezy Penrose’a są nieco bogatsze, lecz ze względu na to, iż ich autor wydał do tej pory trzy książki, gdzie je wnikliwie omawia zainteresowanych odsyłam do nich. Nie mogę jednak powstrzymać się przed pewnym komentarzem ostatecznym, wyrwanym z Trefil 1998, w którym zamieszczony jest taki komentarz: „Przypuśćmy jednak, że hipoteza Penrose’a okazuje się w pełni trafna. Przypuśćmy, że (1) mózg rzeczywiście nie jest cyfrowym komputerem i (2) dlatego nim nie jest, że działa według praw nowej nauki (przypominam pogląd C Penrose’a; przyp. M. K.), której miejsce znajduje się na przecięciu fizyki klasycznej, mechaniki kwantowej i teorii unifikacji pola. (...) co by się stało, gdyby teorię unifikacji pola już napisano i moglibyśmy śmiało wejść w lukę między fizyką kwantową a klasyczną. Wtedy, jeżeli Penrose ma rację, moglibyśmy zrozumieć działanie mózgu na poziomie cząsteczek i komórek. (...) Skoro już zrozumiemy mózg w kategoriach nowej nauki Penrose’a, jest, jak sądzę, bardzo prawdopodobne, że ktoś wpadnie na pomysł zbudowania nowej maszyny – jakiegoś metakomputera, działającego zgodnie z prawami owej nauki. Tak jak komputer cyfrowy działa według praw mechaniki kwantowej, metakomputer będzie działał według praw metanauki.” [Trefil 1998, ss. 144-145].

<sup>30</sup> Por.: Crick, Koch 1992; F. Crick, *Zdumiewająca hipoteza, czyli nauka w poszukiwaniu duszy*, tłum. B. Chacińska-Abrahamowicz, M. Abrahamowicz, Prószyński i S-ka, Warszawa 1997.

<sup>31</sup> Por. Zeki 1992.

<sup>32</sup> Por. Logotheis 2000.

<sup>33</sup> Por. *O neuronalnym podłożu świadomości. Rozmowa Stephena Jonesa z Susan Greenfield*, Internet.

<sup>34</sup> Por. zwłaszcza: D. C. Dennett, *Consciousness Explained*, Boston 1991, a następnie w: Penguin Books, London 1993; D. C. Dennett, *Content and Consciousness*, Routledge and Kegan Paul, London 1969; D.C. Dennett, *Animal Consciousness: What Matters and Why*, w: "Social Research", Nr (62)/1995, ss. 691-710; Dennett 1997. Zob. też: P. Czarniecki, *Koncepcja umysłu w filozofii D. C. Dennetta*, Internet.

<sup>35</sup> Por.: Searle 1995; Searle 1999a; Searle 1999b.

<sup>36</sup> Por. zwłaszcza: P. M. i P. S. Churchland, *Matter and Consciousness*, Bradford Books, MIT Press, Cambridge 1984, Mass; Churchland 1995.

<sup>37</sup> Por. Putnam 1998.



błędów, być oczarowanym przez seks, być złym lub przybitym, gdy nie może dostać tego, co chce”.<sup>38</sup>

I znów, jak się wydaje, zarzut związany z nieposiadaniem przez maszynę świadomości zmierza w kierunku wypowiedzenia stanowczej niemożności myślenia przez maszyny. Zresztą tak samo będzie przy okazji kolejnego – *E* – argumentu. Lecz tak naprawdę, nie ustalono jeszcze, że maszyna faktycznie nie może myśleć (wystarczy przeanalizować cytaty Turinga ze ss. 49-50; patrz również cyt. poniższy i cyt. W. Marciszewskiego, s. 49) i w rezultacie faktycznie nie może być świadoma. Tym bardziej, że dostęp do świadomości dany jest nam zawsze subiektywnie.

Według krańcowej (solipsystycznej; przyp. M. K.) postaci tego poglądu jedynym sposobem upewnienia się, że maszyna myśli, jest *być* maszyną i odczuwać, że się myśli. Można by opisać te uczucia świata, ale naturalnie nikt nie byłby usprawiedliwiony, gdyby wziął tego rodzaju wiadomość pod uwagę. Podobnie – według tego poglądu jedynym sposobem przekonania się, że jakiś *człowiek* myśli, jest być tym człowiekiem.<sup>39</sup>

Jako rozwiązanie naszego problemu Turing zaproponował zmodyfikowaną wersję gry w imitację (bez gracza *B*), stosowaną również w praktyce, o nazwie *viva voce*.

Ma ona na celu przekonanie się, czy ktoś rzeczywiście rozumie coś, czy też „nauczył się tego na pamięć jak papuga” (zapamiętajmy to zdanie dla dalszych rozważań).<sup>40</sup>

A kojarzyć się nam może z tak dobrze znanymi sytuacjami jak egzaminy w czasie studenckiej sesji czy matury, w których ciąg pytań ma za zadanie zweryfikować w odpowiedni sposób zdanie naszego egzaminatora o stanie naszej wiedzy, a niekiedy i inteligencji, oraz – rzecz jasna – świadomości (przypisanej nam wcześniej).

Możemy przyjąć trochę inny niż Turinga sposób obrony przed owym argumentem odwołując się do fragmentów *Dociekań filozoficznych* Ludwiga Wittgensteina (1889-1951):

Cóż się dzieje, gdy osoba (...) czyta na przykład gazetę? – Jej wzrok przesuwają się – jak mówimy – po drukowanych słowach i wypowiada je głośno albo tylko do siebie; przy czym niektóre słowa ujmuje w ich drukowanym kształcie jako jedną całość, w innych ogarnia wzrokiem tylko pierwsze sylaby, a czasem odczytuje je sylaba po sylabie lub nawet litera po literze. Powiedzielibyśmy też, że osoba ta przeczytała zdanie, jeżeli czytając zdanie nie mówiła wprawdzie ani głośno, ani do siebie, ale była potem w stanie powtórzyć to zdanie dosłownie lub w przybliżeniu. – Może ona zwracać uwagę na to, co czyta, a może też – jak by można rzec – funkcjonować jedynie jako maszyna do czytania, tzn. czytać głośno i poprawnie, ale bez zwracania uwagi na to, co czyta.<sup>41</sup>

Czyta bez świadomości, jakbyśmy powiedzieli, ale

Trzeba przyznać, że jeśli chodzi o wymawianie jakiegoś *pojedynczego* drukowanego wyrazu, to w świadomości ucznia „udającego”, że czyta – czyli *funkcjonującego*

<sup>38</sup> Turing 1972, s. 35.

<sup>39</sup> Tamże, s. 35.

<sup>40</sup> Tamże, s. 35.

<sup>41</sup> Wittgenstein 2000, § 156, s. 92.

*tylko jako maszyna do czytania* – może zachodzić to samo co w świadomości „czytającego” ów wyraz wprawnego czytelnika.<sup>42</sup>

Po zastanowieniu się nad fenomenem tego, czy ktoś rzeczywiście czyta, moglibyśmy prawdziwie orzec: „On czyta”, ale sprawiana iluzja czytania – poprzez funkcjonalne (bez użycia świadomości) wyrażanie poszczególnych słów – zdaniem Wittgensteina, może być trudna do wykrycia.

A to, co stoi na przeszkodzie, że nie potrafimy tego jeszcze stwierdzić, może wynikać z

naszej niedostatecznej znajomości zjawisk zachodzących w mózgu i w układzie nerwowym. Gdybyśmy je znali lepiej, to dostrzegliśmy, jakie połączenia wytworzone zostały w wyniku ćwiczenia; a wtedy, zajrzawszy do mózgu, moglibyśmy orzec: „Teraz *przeczytał* ów wyraz, teraz powstało połączenie, na którym polega czytanie”.<sup>43</sup>

Zatem badania nad świadomością i procesami świadomymi powinny zmierzać (i zmierzają, czego dowodem są badania współczesnych naukowców)<sup>44</sup> w kierunku wyznaczenia szlaków i zjawisk w mózgu występujących bądź współwystępujących przy procesach świadomej uwagi, jak na przykład świadomego czytania. Z wiedzą tą, z korzyścią dla naszych rozważań, winny pójść prace nad możliwością zweryfikowania tego, czy odpowiednie zjawisko w maszynie cyfrowej może odpowiadać zjawisku zachodzącemu w mózgu człowieka. Jeśli tak, wówczas nie stoi nic na przeszkodzie, aby stwierdzić, iż:

**zachowanie x maszyny jest zachowaniem świadomym, gdyż identyczne zjawiska zachodzą w mózgu człowieka co we wnętrzu maszyny przy jego zachowaniu x.**

Przywołajmy jeszcze na moment fragment z *papugą*. Jak pamiętamy, Turing, wcielając wcześniej w życie grę *viva voce*, twierdzi, iż taka właśnie gra może sprawdzić, niczym egzamin, czy maszyna jest świadoma. Lecz rozpatrzmy taką oto dosłowną sytuację, która wydarzyć się miała naprawdę.

<sup>42</sup> Tamże, s. 93.

<sup>43</sup> Tamże, s. 95.

Choć są też i tacy, którzy 1) zaliczają *neuroscience* do nauk niepewnych, takich jak psychologia („W środowisku lekarskim dość rozpowszechniona jest opinia, że neurologia (obok może psychiatrii) jest najbardziej enigmatyczną i niepewną gałęzią medycyny (...) Czy jest to jej permanentna charakterystyka? Moja diagnoza w tej kwestii brzmi: zasadniczo tak!” [Szumakowicz 2000, ss. 135-136]) i 2) twierdzą, że już wszystko z zakresu *neuroscience* wiemy („Otóż wiele praw funkcjonowania mózgu już znamy (co najmniej od czasów Pawłowa,...” (sic!, przyp. M. K.)), a ponadto 3) uważają, że wiedza ta nie ma żadnej wartości poznawczej („...dalsze będziemy poznawać, być może, w sposób nieograniczony w czasie (...), ale trudno mi pojąć, dlaczego miałoby to przynieść coś filozoficznie istotnego.” [cyt. Tamże, s. 147]). Oczywiście z wiadomego względu takimi „rewelacyjnymi” poglądami nie będę się ani zajmować, ani przejmować w tej pracy. Zaś samemu zainteresowanemu, E. Szumakiewiczowi, na jego „*trudności z pojmowaniem*” tak zasadniczej kwestii, jaką jest wpływ wiedzy z zakresu nauk o mózgu na filozofię (choćby epistemologię czy nauki kognitywne), należałoby przypomnieć słowa Bertranda Russella wypowiedziane do jednego z filozofów, który nie rozumiał pewnej kwestii głoszonej przez Russella: „*Młody człowieku, za twoje braki w pojmowaniu, to ja już nie odpowiadam!*”

<sup>44</sup> Do listy badaczy z zakresu neuronauk zajmujących się świadomością w tym miejscu należałoby dodać: Michaela S. Gazzanigę (por. Gazzaniga 1997; Gazzaniga 1998), Garalda M. Edelmanna (por. Edelman 1998) oraz Antonio R. Damasio (por. Damasio 1992; Damasio 1999; Damasio 2000; Damasio 2000b. Zob. również referat wygłoszony na „Symposium Kognitywnym: Subiektywność a świadomość”, Obrzycko 2000, zorganizowany przez UAM: D. Wiener, *Koncepcja świadomości Antonio Damasio przedstawiona w książce The Feeling of what happens: body and emotion in making of consciousness*, Internet).

Locke nigdy zapewne niepodejrzewany o łatwowierność, nie zawahał się uwierzyć w historię pewnej papugi, o której opowiada w swoich *Pamiętnikach* kawaler Temple. Papuga ta rzekomo odpowiadała do rzeczy na pytania i nauczyła się prowadzić, tak jak my, swego rodzaju rozmowę, mającą pewną ciągłość.<sup>45</sup>

Fragment ten, po pierwsze, wywołuje w nas pobłażliwy uśmiech na myśl o naiwności tak poważnego filozofa, jakim był z pewnością J. Locke, lecz z drugiej strony trwoży ze względu na tematykę w niniejszej pracy poruszaną. Papuga zachowała się bowiem tak, jak chcielibyśmy, aby zachowała się maszyna myśląca, która zdaje test Turinga (a jej przecież rozumności czy inteligencji nie przypiszemy). Jest to w rezultacie taki sam zarzut jak rozpatrywany nieco dalej zarzut S. Lema z „kosmicznym gramofonem” (por. ss. 64-65), czyli biernie odtwarzanym zapisem. Lecz – podobnie jak tamten – niemożliwy do zaistnienia.

Współcześnie argumentem związanym ze świadomością posługują się: John Searle, Hilary Putnam i Alwyn Scott<sup>46</sup>.

### E. Argumenty wypływające z różnych niemożności

Jest to w zasadzie argument bardzo zbliżony do poprzedniego. Powiada on:

Zgadzam się z tobą, że możesz zrobić maszyny wykonujące to wszystko, o czym wspominałeś, ale nigdy nie będziesz w stanie zrobić maszyny, która by zrobiła X.<sup>47</sup>

Przy czym za X można podstawić takie czynności jak:

Być uprzejmym, pomysłowym, pięknym, przyjacielskim, mieć inicjatywę, mieć zmysł humoru, odróżniać dobro od zła, robić błędy, zakochiwać się, lubić truskawki ze śmietaną, stać się obiektem czyjejś miłości, uczyć się z doświadczenia, używać właściwych słów, być przedmiotem swojej własnej myśli, potrafić zachowywać się w tak rozmaity sposób jak człowiek, robić coś naprawdę nowego.<sup>48</sup>

Obecnie najczęściej poruszonym przykładem w ramach X jest kwestia czucia bólu i możliwości (czy raczej w tej debacie niemożności) jego odczuwania przez maszyny (por. s. 74)<sup>49</sup>.

Bardzo często też – ze względu na pokrewieństwo dwu ostatnich argumentów – podaje się je w jednym ciągu (poniższy przykład).

Maszyna nie posiada świadomości. Zatem potrafi grać w szachy z Kasparowem, ale nie posiada świadomości, że: 1) gra teraz w szachy, 2) Kasparow jest mistrzem świata, 3) wygrała i, ostatecznie, 4) cieszyć się z wygranej.<sup>50</sup>

\* William Temple (1628-1699) był wybitnym dyplomatą znienawidzonym przez teologów anglikańskich za wolnomyślnie przekonania. Biskup Burnet oskarżał go o materializm, ateizm i hołdowanie nauce Konfucjusza. Temple napisał, prócz *Pamiętników*, szereg dzieł o treści politycznej i historycznej oraz szkiców literackich i filozoficznych. Zebranych później w dwóch tomach pt. *Miscellanea*. (Przyp. za cyt.).

<sup>45</sup> La Mettrie 1984, s. 33.

<sup>46</sup> Wyraża się to w tezie: „ Sądzę, że naprawdę inteligentna maszyna musi być zdolna do wzajemnego oddziaływania z kulturą, a do tego jest potrzebne posiadanie świadomości. A świadomość nie pojawi się w sposób cudowny w komputerze w dniu, w którym jakieś maszyny przejdą pozytywnie test Turinga, jak to sugerują niektórzy funkcjonalści. Stałoby się tak tylko wtedy, gdyby inżynierowie budujący komputery byli w stanie zaprogramować w komputerze świadomość”. [Scott 1999, ss. 186-187].

<sup>47</sup> Turing 1972, s. 36.

<sup>48</sup> Tamże, s. 36.

<sup>49</sup> Por. chociażby: Lillegard 1995.

Skąd biorą się podobne poglądy na ograniczone w swym zachowaniu maszyny, którym odbiera się prawo do śmiechu? Najczęściej są one wynikiem indukcji wynikającej z obserwacji potocznych – przez całe życie mieliśmy do czynienia z licznymi maszynami i na ich podstawie wyrobiliśmy sobie obraz czegoś takiego jak *MASZYNA*, zapominając o poszczególnych jej modelach – uczyniliśmy z nich w rezultacie model sztandarowy, definicję.

Maszyny są brzydkie, każda z nich jest przeznaczona do bardzo ograniczonego celu, są one beżycieczne w przypadku cokolwiek innego celu, różnaitość zachowania się każdej z nich jest bardzo mała itd..<sup>51</sup>

Ale zapominamy – po pierwsze – o tym, że nigdy jeszcze nie mieliśmy *de facto* do czynienia z maszyną myślącą, o której przecie jest mowa, więc w rezultacie indukcja nasza jest cokolwiek nie na miejscu! A po drugie – o tym, że na podstawie codziennych doświadczeń ktoś również może wysnuć taką tezę w stosunku do ludzi. W rezultacie **parafrazując zdanie:**

**„Zgadzam się z tobą, że możesz zrobić maszyny wykonujące to wszystko, o czym wspomniałeś, ale nigdy nie będziesz w stanie zrobić maszyny, która by zrobiła X”**

na:

**„Zgadzam się z tym, że znasz osobnika (bądź jesteś osobnikiem), który potrafi zrobić to wszystko, o czym wspomniałeś, ale nie potrafi on (bądź nie potrafisz) zrobić X”.**

Naszym X mogą być rozmaite czynności: zamaszysty zamach w powietrzu z półobrotem wokół własnej osi, który potrafił wykonać mistrz wschodnich sztuk walki, Bruce Lee; stepowanie w rytm padającego deszczu, co potrafił zrobić Fred Aster w musicalu *Deszczowa piosenka*; zaśpiewanie w taki sposób urodzinowej piosenki *Happy birthday* (polskie *Sto lat*), że o mała wersja ta nie stała się hymnem narodowym, skądinąd wielkiego narodu, bo tego właśnie dokonała Marylin Monroe itd. A co na to autor gry w naśladownictwo?

Nie wydaje mi się, aby prace i zwyczaje rodzaju ludzkiego stanowiły odpowiedni materiał, do którego można by stosować naukową indukcję. Bardzo dużą część czasoprzestrzeni trzeba by zbadać, aby móc otrzymać wiarygodne wyniki. Inaczej możemy (tak jak większość angielskich dzieci) rozstrzygnąć, że każdy mówi po angielsku i że (w rezultacie, przyp. M. K.) głupio jest uczyć się francuskiego.<sup>52</sup>

Rozpatrzmy jednak dla uzupełnienia rozważań przypadek specjalny. Powiedzmy, że ktoś pod nasze X podstawiał „popęlniać błędów” i otrzymał w rezultacie: „Ta maszyna może wszystko, o czym wspomniałeś, ale jednego nie może, popęlniać błędów!”. Dlaczego ten przypadek jest taki ważny? Ano dlatego, że ma ciekawe implikacje w przypadku gry w imitację.

Żądanie: „maszyny nie mogą popęlniać błędów” wydaje się dziwne. Ktoś może zapytać: „Czy są one z tego powodu cokolwiek gorsze?”. Ale przyjmijmy bardziej życzliwe stanowisko i spróbujmy przekonać się, co to żądanie naprawdę oznacza. Myślę, że ten głos krytyczny można wyjaśnić w terminach gry w naśladownictwo. Wymaga się, aby pytający mógł odróżnić maszynę od człowieka po prostu dając im do rozwiązania pewną ilość problemów arytmetycznych. Maszyna zostałaby

<sup>50</sup> M. Kasperski, *Sztuczny człowiek czy sztuczny Bóg?*, Internet, s. 3.

<sup>51</sup> Turing 1972, s. 36.

<sup>52</sup> Tamże, s. 37

zdemaskowana z powodu swojej szalonej celności (por. argument nazwany *problemem naturalnej celowości*; strona 67 niniejszej pracy). Odpowiedź na to jest prosta. Maszyna (zaprogramowana do grania w grę) nie usiłowałaby udzielać *prawidłowych* odpowiedzi na problemy arytmetyczne, natomiast rozmyślnie wprowadzałaby błędy w sposób obliczony na zmylenie pytającego.<sup>53</sup>

I drugi, równie specjalny przypadek, dotyczący kwestii „bycia przedmiotem swojej własnej myśli”, czyli – krótko mówiąc – przypadek dotyczący aspektu *samoświadomości*. Tym razem odpowiadamy krótko, acz stanowczo:

W kwestii, że maszyna nie może być przedmiotem swojej własnej myśli można, naturalnie, odpowiedzieć tylko wtedy, gdy można będzie wykazać, że maszyna *trochę* myśli na temat jakiegoś przedmiotu.<sup>54</sup>

Co sprowadza się do przytaczanego wcześniej, choć nie wprost, stwierdzenia: „Poczekajmy, zobaczymy”.

Dobrze jest teraz zwrócić uwagę, iż cały argument „E. Argumenty wypływające z różnych niemożności”, ma w rezultacie bardzo wiele wspólnego z argumentem, który się pojawi później pod nazwą: *problem sędziego*. W rezultacie indukcja, którą przeprowadził wcześniej sędzia (C), a w wypadku powyższego argumentu oponent, i jej wyniki przynależą do (C) jako część jego skłonności, stanu osobistych poglądów.

Argument ten współcześnie wyznaje między innymi cytowany wcześniej A. Scott<sup>55</sup>.

## F. Zarzut lady Lovelace

Z sylwetką lady Ady Lovelace spotkaliśmy się już w pierwszym rozdziale przy okazji omawiania maszyny analitycznej Babbage’a. Lecz oprócz znanych jej prac nad programami do wyliczania poszczególnych liczb  $B_k$  szeregu Bernoulliego<sup>56</sup> ważna jest dla nas jej praca z 1842 roku<sup>57</sup>, w której to pojawiła się następująca teza:

Maszyna analityczna nie rości sobie pretensji do *oryginalności* rozwiązań. Może ona wykonać *wszystko to, co wiemy w jaki sposób zlecić jej do wykonania*” (jej kursywa).<sup>58</sup>

<sup>53</sup> Tamże, s. 37 (porównaj również drugi cytat z A. Turinga, s. 68 tej pracy).

<sup>54</sup> Turing 1972, s. 38.

<sup>55</sup> Wyraża się to w tezie: „Cechą układu inteligentnego nie jest jedynie przekonanie kogoś w czasie krótkiej konwersacji, że jest się układem inteligentnym. (...) Prawdziwym zadaniem jest zrobić coś inteligentnego, na przykład: przetłumaczyć piosenkę, podać wymyślone twierdzenie matematyczne, podać przekonujące argumenty lub jeszcze coś innego. (...) [Scott 1999, s. 186] Por. również z fragmentem wypowiedzi Jeffersona, z Turing 1972.

<sup>56</sup> Por. Marciszewski 1998, s. 63. Liczby szeregu Bernoulliego tworzone są następująco:

$1 + \frac{1}{2^{2k}} + \frac{1}{3^{2k}} + \frac{1}{4^{2k}} + \dots = \prod_{i=1}^{2k} \frac{2^{2k-1}}{(2k)!} B_k$ , gdzie  $k$  jest kolejną dowolną liczbą naturalną, a  $B_k$  liczbą Bernoulliego z niej

utworzoną, np.  $k = 1, B_k = \frac{1}{6}$ ;  $k = 2, B_k = \frac{1}{30}$ ;  $k = 3, B_k = \frac{1}{42}$ ; itd. Por. I. N. Bronsztejn, K. A. Siemiendajew,

*Matematyka*, PWN, Warszawa 1986, t. 2/3, ss. 382-383.

<sup>57</sup> Por. s. 33, przyp. 46.

<sup>58</sup> A. Lovelace; za: Turing 1972, s. 38.

Jest to bardzo znany i powszechnie stosowany argument mówiący o tym, iż w rezultacie maszyna, nawet maszyna myśląca, nie może/nie będzie mogła niczego więcej, niż program (więc i wcześniej programiści) będzie przewidywał. Będzie czynić to, co zechce ich operator i tylko w ramach tego, co sama będzie miała zakodowane w programie – poza tym nic więcej. Także na przykład w czasie wolnym od pracy nie będą wymyślały zabaw ani niczego innego, w stosunku do tego, co zostało już wymyślone (jak czyni to człowiek hołdując powiedzeniu: „Wyobraźnia jest matką wszystkich rzeczy”).

Turing broniąc się przed tym zarzutem cytuje swojego kolegę Hartree’go<sup>59</sup>, który zauważa, że owszem, maszyna Babbage’a nie może zrobić nic nowego, ale:

To nie znaczy, że w ogóle nie jest możliwe zbudowanie elektronicznego urządzenia, które będzie „myślało dla siebie” lub w którym, w terminach biologicznych, można by zainstalować odruch warunkowy, który stanowiłby podstawę „uczenia się”. Problem czy to jest w zasadzie możliwe czy też nie, jest zarówno stymulujący, jak i interesujący.<sup>60</sup>

Tekst ten był pisany w 1949 roku, a tekst Turinga w 1950. Jeszcze wówczas nie było maszyn, które potrafiłyby sprostać zadaniu stworzenia czegoś *nowego* (w sensie – nie zaprogramowanego wcześniej). Dziś jednak czasy nieco się zmieniły. Maszyny dowodzą twierdzeń, nad którymi trudzili się wcześniej pierwszorzędni matematycy (tak było z hipotezą H. Robbinsa (1936), której nikt nie mógł dowieść aż do roku 1996, kiedy tego dokonał komputerowy system EQP – dowód był krótki, bo liczył 17 kroków<sup>61</sup> (patrz: rozdział 2.3.2 – „Dowodzenie twierdzeń logiki i matematyki”, ss. 92-112). Maszyny również generują nowe rozwiązania z innych dziedzin, jak na przykład w ramach medycyny, przy wykorzystaniu tzw. systemów eksperckich (o których mowa będzie trochę później), a nawet stwarzają nowe patenty, często z nimi samymi związane. A przykładem niech będzie taka oto historia:

Amerykańscy informatycy z Brandies University w Waltham przeprowadzili komputerową symulację ewolucji robotów. Komputer miał do dyspozycji jedynie trzy rodzaje wirtualnych elementów: proste pałeczki jako elementy szkieletu, przeguby do ich łączenia i sztuczne neurony do przesyłania sygnałów. Skomplikowany program miał z tych prostych części zaprojektować sprawnie poruszającego się robota. Już po 300 próbach komputer zaczął modelować urządzenia zdolne do ruchu, zaś najbardziej zaawansowany projekt mógłby nawet pełzać na trzech nogach!

Badacze zrobili więc kolejny krok. Podłączyli komputer do urządzenia stosowanego przez inżynierów do szybkiego wytwarzania prototypów projektowanych urządzeń i podzespołów. Przy drobnej pomocy autorów eksperymentu sprzężona z komputerem maszyna zbudowała rzeczywisty model małego robota, który mógł się poruszać tak jak jego wirtualny poprzednik.<sup>62</sup>

<sup>59</sup> D. R. Hartree, *Calculating Instruments and Machines. Urbana III*, University of Illinois Press, 1949.

<sup>60</sup> D. R. Hartree; za: Turing 1972, s. 38.

<sup>61</sup> Por.: W. McCune, *EQP, Internet*; W. McCune, *Robbins Algebras Are Boolean*, Internet.

Swoją drogą to dziwne, że jesteśmy skłonni uważać pewnych ludzi za geniuszy, a żadnej maszyny za genialną: Marian Rejewski, jeden z trzech inżynierów polskich biorących czynny udział w złamaniu kodu niemieckiej maszyny Enigma, został poproszony przez jeden z amerykańskich instytutów o złamanie pewnego trudnego kodu. Kodu, który powstał najprawdopodobniej około 1905 roku, a którym posługiwali się w kontaktach gen. J. Piłsudski i partia PPS. Kod ów Amerykanie próbowali przez zgoła 70 lat złamać i żadnemu z ich kryptografów rzecz ta się nie udała. M. Rejewski przyjąwszy zadanie złamał go w dwa tygodnie! Najlepszym określeniem Rejewskiego będzie przyznanie mu geniuszu i, jak to określił jeden z polskich znawców tematu, geniuszu pierwszej kategorii. Jednakże jak w tym świetle określić fakt udowodnienia hipotezy Robbinsa przez program EQP? Pytanie to pozostawiam na razie bez odpowiedzi.

<sup>62</sup> *Robot, syn robota* 2000, s. 16.

W to samo zdaje się wierzył Turing. Sam zwracał uwagę (za Hartree'm), że zrobienie takiej maszyny mogącej niejako „wyjść” poza to, co ma w programie, będzie możliwe, gdy tylko zostaną skonstruowane odpowiednio mocne maszyny, wymieniając przy tym dużą pojemność pamięci i szybkość. Liczył, że maszyny z pamięcią  $10^9$  bitów dadzą już sobie z tym radę – szacuje się na przykład, że pamięć mózgu człowieka wynosi  $10^{20}$  bitów<sup>63</sup>, a współczesnego superkomputera ASCI White<sup>64</sup> (od ang. *Advanced Strategic Computing Initiative*) firmy IBM, ponad 160 TB, czyli  $160 \cdot 10^{12}$ .

Innym rodzajem obrony przed zarzutem lady Lovelace, byłoby sprawdzenie tego, czy człowiek może wyjść poza siebie samego. I jeśli chodzi o robienie rzeczy oryginalnych, to i owszem, nawet często jesteśmy świadkami takich rzeczy (przykłady można przytaczać chociażby za *Księgą rekordów Guinnessa*), ale jeśli potraktujemy program maszyny jak oprogramowanie naszego umysłu, to wówczas stwierdzić będziemy musieli, że poza nie wyjść nie możemy także<sup>65</sup>. Mało tego, jeśli traktować poważnie dorobek nauk o mózgu, a nie mamy powodu by tak nie czynić, to musimy przyjąć za Michaeliem Gazzanigą na przykład, że możliwości nasze są skończone, a wyznaczają je DNA i – dalej za podwójną helisą – możliwości rozwoju mózgu ludzkiego, które niczym *implicite* z naszego DNA wynikają<sup>66</sup>. Zatem zarzut ten na dwojaki sposób można odrzucić i wszystko zależy od tego, co potraktujemy jako program, poza który wyjść nie możemy. Bo:

1. **jeżeli program to z góry narzucone dyrektywy postępowania**, nazwijmy go programem w wąskim znaczeniu, gdyż konkretne rozkazy odpowiadają za użycie ich w konkretnych sytuacjach, **to maszyna może wyjść poza program**.

Zaś:

2. **jeżeli program to tak ogólne oprogramowanie, że pozwala nam generować nowe dyrektywy postępowania** (dzięki kombinatoryce) w ramach ogólnie ujętych praw, np. praw biologicznych (DNA, mózg), **to poza nie nawet my ludzie nie możemy wyjść**, np. nie wyjdziemy poza percepcję człowieka stając się (czy też raczej czując się, albo czując świat) nietoperzem<sup>67</sup>.

Taki program nazwijmy programem w szerokim znaczeniu<sup>68</sup>.

<sup>63</sup> Por. Hołyński 1979, s. 140.

<sup>64</sup> Por. *Nowy superkomputer firmy IBM 2000*, s. 12.

<sup>65</sup> Nawet, gdy weźmie się za przykład ludzi genialnych, to należałoby stwierdzić, iż zjawisko geniuszu, chociaż rzadkie, to zachodzi w ramach tego samego programu, co równie rzadkie przypadki debilizmu. Tyle, że są to przypadki graniczne amplitudy w ramach sinusoidy na osi rozwoju intelektu.

<sup>66</sup> Wyrażone jest to na przykład w tezie: „Jeśli, jak staram się udowodnić w tej książce, większość naszych zdolności psychicznych jest wynikiem selekcji, dojrzały mózg, gdzie mieszczą się połączenia nerwowe umożliwiające zachodzenie złożonych, ludzkich procesów psychicznych, musi rozwijać się w genetycznie zaprogramowany, ściśle określony sposób.” [Gazzaniga 1997, s. 40]. Takiego samego zdania jest również większość współczesnych badaczy z kręgu neuronauk, na przykład G.M. Edelman ze swoją koncepcją „neuralnego darwinizmu”. Por. Gazzaniga 1997, s. 64 i Edelman 1998.

<sup>67</sup> Por. T. Nagel, *Jak to jest być nietoperzem*, w: T. Nagel, *Pytania ostateczne*, tłum. A. Romaniuk, Spacja, Warszawa 1997, ss. 203-219.

<sup>68</sup> Ktoś mógłby mi zarzucić przy tej okazji, że teza powyżej sformułowana jest nieprawdziwa. Przy tym mógłby powołać się na rozwój cywilizacyjny – wychodzenie poza program kultury człowieka. Powoływanie się na wpływ kultury nie powoduje jednak falsyfikacji powyższej tezy. Widziałbym taki wpływ, tak samo jak zmieniające się pasmo DNA, jako domenę metody kombinatorycznej. Wówczas też wpływ kultury, czy tzw. wyjście poza kulturę, odbywa się dalej w obrębie kultury (na zasadzie rozszerzania się podzbiorów zbioru).

Jeszcze inna forma obrony przed zarzutem lady Ady Lovelace wyłoni się po przeczytaniu poniższego fragmentu:

„Maszyna wykonująca jedynie to, co zawiera stworzony przez nas program” – dla niektórych jest to może „intuicyjnie oczywiste”, ale czy „niewątpliwie prawdziwe”? Jeśli program dla maszyny będzie pisać nie człowiek, lecz inna maszyna? Jeśli i tę maszynę programowała z kolei inna maszyna zaprogramowana również przez maszynę, dla której oprogramowanie sporządziła jakaś  $n$ -ta maszyna? I cóż z tego, że w końcu tego łańcucha zawsze odnajdziemy inicjatywę człowieka? Jego rola będzie tak znikoma, jak wpływ wiadomości nabytych w szkółce niedzielnej na pracę pięćdziesięcioletka o ukształtowanej osobowości.

Dwadzieścia lat temu (a rzecz była pisana w 1978 roku; przyp. M. K.) R. M. Friedberger\* potraktował zagadnienie pisania programów jako szczególny przypadek rozwiązywania ogólnych problemów przez maszynę zdolną do uczenia się metodą prób i błędów. Dla prostych przykładów obliczeniowych komputer wypisywał dowolny program i sprawdzał przydatność jego rozkazów. Wybór najbardziej skutecznych rozkazów dawał w końcu gotowe programy.<sup>69</sup>

Jest to doskonały przykład maszyn wychodzących poza swój program (oczywiście, jak ustaliliśmy, program w wąskim tego słowa znaczeniu). Jeśli chodzi o wykorzystywanie komputerów do podejmowania decyzji, którym to zagadnieniem zajmę się nieco później, na przykład w podrozdziale: 2.3.3 – „Rozpoznawanie obrazów” (por. ss. 113-128). Teraz zaznaczę tylko, że metoda wykorzystywania komputerów jako specyficznie pojętych doradców czy też sędziów mających ocenić sytuację, stosowana jest w programach NASA<sup>70</sup>. Na promach kosmicznych, w przestrzeni pozaziemskiej, istnieją tak nieprzyjazne warunki (w tym i dla sprzętu komputerowego), że zamiast jednego *MEGA-komputara*, którego przy dzisiejszym stanie technicznym nie sposób tam utrzymać, montuje się pięć, porównywalnych mocą obliczeniową do zwykłego Pentium II firmy Intel. W sytuacjach kryzysowych komputery te „głosują” nad podjętą decyzją.

Jeśli chodzi zaś o pierwszy fragment cytatu, mówiący o maszynach piszących programy dla maszyn i... tak  $n$ -razy, to dzisiaj stosuje się tak zwane **algorytmy genetyczne**, które w pewien sposób wypełniają tak zarysowaną ideę.

Programowanie genetyczne ma na celu stworzenie programu na wzór doboru naturalnego<sup>71</sup>. Postępuje się w tym przypadku według poniższej ogólnej procedury.

1. Tworzymy „chromosomy” (poszczególne cechy) systemu dla programu genetycznego, z których ma się składać.
2. Na podstawie (1) realizujemy opisywane przez nie systemy.
3. Badamy wyniki z (2). Badanie to ma stwierdzić, które z systemów najlepiej się nadają do spełniania ostatecznych zadań. To postępowanie decyzyjne może także przejąć maszyna

\* R. M. Friedberger, *A Learning Machine*, w: "IBM Journal of research and Development", Nr 2/1958. (Przyp. za cyt.).

<sup>69</sup> Hołyński 1979, ss. 185-186.

<sup>70</sup> Por. Dębek 2001.

<sup>71</sup> Bardzo dużo o tej metodzie modelowania sztucznej inteligencji, Czytelnik znajdzie na stronie domowej dr hab. Haliny Kwaśnickiej.



4. Na podstawie (3) eliminujemy systemy najsłabsze i pozostawiamy najsprawniejsze, które w kolejnej fazie będą pełnić znowu rolę poszczególnych „chromosomów”. To postępowanie również może przejąć maszyna!

Procedura, która tu została przedstawiona nazywa się **programowaniem genetycznym**, a wykorzystywana jest między innymi przy jednym z najbardziej ambitnych współczesnych projektów w ramach Sztucznej Inteligencji, przy projekcie belgijskiej firmy Starlab, CAM-Brain Machine, mającym na celu zbudowanie sztucznego mózgu o pojemności biliona neuronów (będzie o tym mowa w podrozdziale 2.4.1 – „Od sztucznych sieci neuronowych do sztucznego mózgu. Maszyna typu B w świetle projektu CAM-Brain”, ss. 165-169).

### G. Argument wypływający z ciągłości systemu nerwowego

System nerwowy na pewno nie jest maszyną o stanach dyskretnych. Mały błąd w informacji o wielkości nerwowego impulsu wchodzącego do neuronu może spowodować różnicę wielkości impulsu wyjściowego. Można argumentować, że ponieważ tak jest nie można oczekiwać, aby można było naśladować zachowanie się systemu nerwowego przy pomocy systemu o stanach dyskretnych<sup>72</sup>.

Argument ów, ze względu na naszą grę w naśladownictwo, jest jak najbardziej chybiony. **Nie interesuje nas tu porównanie mózgu i maszyny pod względem strukturalnym**, do którego się zarzut ów odwołuje, **a tylko podobieństwo funkcjonalne**. Przecie chodzi o to, aby maszyna spełniając funkcje językowe, generowała nam „coś na kształt” umysłu. A można by było to jeszcze ograniczyć – nie musi nam wcale, idąc za koncepcją Turinga, generować umysłu ze wszystkimi jego atrybutami, wystarczy, że będzie przemawiać w bardzo specyficzny sposób: **ma generować sensownie zdanie  $x$  w odpowiedzi na poprzedzające je – ze strony dowolnego współmówiącego ( $C$ ) – zdanie  $y$** <sup>73</sup> (o kwestii rozróżnienia strukturalnych i funkcjonalnych podobieństw i różnic pomiędzy mózgiem a maszynami będę jeszcze mówić przy omawianiu metafory komputerowej).

Współcześnie można jeszcze inaczej odeprzeć ten zarzut. Istnieją już maszyny analogowe, które spełniają warunek stanów dyskretnych<sup>74</sup>. Mało tego. Buduje się już tzw. **maszyny neuronalne**, których struktura reprezentuje strukturę znaną nam z mózgu ludzkiego. Przykładem może być CAM8 (CAM, to skrót od ang. **C**ellular **A**utomata **M**achine, zaś cyfra 8 określa liczbę modułów pamięci komórkowej wchodzących w jej skład), maszyna zbudowana na przełomie lat 80. i 90. przez inżynierów z Massachusetts Institute of Technology (MIT) za cenę 40 tysięcy dolarów (taka była cena prototypu)<sup>75</sup>. Jest to pierwszy wieloprocessorowy komputer, który został skonstruowany do przetwarzania informacji w sposób zgodny z ideą komórkowego automatu. Będziemy szerzej mówić o nim i jego zastosowaniu przy analizie projektu CAM-Brain.

Współcześnie wyrazicielami argumentu wypływającego z ciągłości systemu nerwowego są zwłaszcza badacze z obrębu neuronauk, na przykład Susan Greenfield<sup>76</sup> (choć nie wszyscy, na

<sup>72</sup> Turing 1972, s. 40.

<sup>73</sup> Por. M. Kasperski, *Sztuczny człowiek czy sztuczny Bóg?*, Internet, s. 2.

<sup>74</sup> Por. też: W. Duch, *Wstęp do kognitywistyki*, Internet.

<sup>75</sup> Por. Buller 1998, ss. 48-49, 56.

<sup>76</sup> Wyraża się to w bardzo krótkim, acz dobitnym zdaniu: (...) „trudno pojąć, czemu niektórzy ludzie uważają mózg za komputer, czy chociażby próbują snuć między nimi analogie.” [Greenfield 1999, s. 140]. Jednak już w swojej następnej książce (Greenfield 1998) wycofuje się z poprzednio głoszonej „trudności w pojmowaniu” metafory komputerowej, chociażby w zdaniu: „Porównanie mózgu z komputerem wykazuje pewne ich

pewno nie jest nim M.S. Gazzaniga; patrz podrozdziały: 2.3.5 – „Analiza mowy i języka” (ss. 151-153) i 2.4 – „Sztuczny mózg – sztuczne sieci neuronowe (*Neural Network*). Ku metaforze komputerowej” (ss. 154-164)), ale też i inni, m.in. A. Scott<sup>77</sup>.

## H. Argument wpływający z nieformalności zachowania się

Współcześnie ten zarzut przyjął nazwę *frame problem* (problem ramy)<sup>78</sup>. Dotyczy on wszystkich takich czynności, z którymi spotykamy się na co dzień. Jakich? Przeanalizujmy go na przykładzie jednego z najbardziej popularnych, choć, dodajmy, niekoniecznie najsztudniejszych, żartów algorytmicznych lat 60-70. ubiegłego stulecia – na tzw. algorytmie dnia „wytężonej pracy”<sup>79</sup>.

1. Start – przychodzisz do pracy. Przejdź do 2.
2. Czy szef patrzy na ciebie? Tak – przejdź do 3; Nie – przejdź do 4.
3. Udawaj, że pracujesz. Po czym przejdź do 5.
4. Zrób sobie herbaty. Po czym przejdź do 5.
5. Czy czas pracy minął? Tak – przejdź do 6; Nie – przejdź do 2.
6. Stop – idź do domu.

W schemacie tym są jednak znane nam z codziennego życia poważne luki, na przykład: co będzie, gdy szef będzie miał dobry humor i zamiast chcieć, żebyś pracował, będzie się chciał napić z tobą herbaty? Albo: co będzie, gdy szef będzie miał zły humor i wówczas, jakbyś wydajnie nie pracował i tak nie będzie go to satysfakcjonować? Może być też tak, że chociaż szef ma zły humor, to żeby go sobie poprawić będzie chciał się z tobą napić (wówczas też niekoniecznie herbaty). Może się też zdarzyć, że herbata się dawno skończyła – i co wówczas?

Lata, z których żart ów pochodzi, przyniosły pewne rozwiązanie. Było nim zapisywanie programów nie w sposób algorytmiczny, ale **heurystyczny** (tzn. przy wykorzystaniu zestawu pewnych bardzo ogólnych zasad działania, które były oparte na takim zestawianiu faktów już znanych i nowych, że można było znaleźć między nimi występujące zależności). Podobną linię argumentacji – przy wsparciu heurystyki – proponował Turing twierdząc, że:

jeśli podlegamy prawom zachowania się, to jesteśmy jakąś maszyną. (...) – I dalej, co dla nas jest bardzo istotne, a może i najistotniejsze ze względu na powyższy zarzut. – (...) nie możemy tak łatwo dać się przekonać o nieistnieniu kompletnych

---

podobieństwa”. Albo: „Najbardziej owocne są porównania całości procesów toczących się w tych dwóch rodzajach struktur: biologicznej i mechanicznej” [Greenfield 1998, rozdział *Mózg elektrony*, ss. 22-23]. Będę jeszcze o tym mówić przy okazji podrozdziału 2.4 – „Sztuczny mózg – sztuczne sieci neuronowe (*Neural Network*). Ku metaforze komputerowej”, ss. 154-164, zwłaszcza rys. 24, s. 154.

<sup>77</sup> Wyraża się to w tezie: „Trzeba koniecznie podkreślić, że mózg ludzki działa na poziomie, którego subtelność i złożoność leży daleko, daleko poza czymkolwiek, co może sobie obecnie wyobrazić inżynieria komputerowa” [Scott 1999, s. 187]. Należy oczywiście podkreślić, i tak już podkreślone, słowo „obecnie”. I dodać, iż pomimo utrzymywania niektórych zarzutów względem testu Turinga (patrz zarzuty: *D* i *E*), jak i brakiem zgody z funkcjonalnym podejściem, Scott zgadza się z tezą, że sztuczna inteligencja jest możliwa do urzeczywistnienia: „Niemniej jednak nie widzę żadnej możliwości udowodnienia, że nie jest możliwe zbudowanie maszyny wyposażonej w inteligencję. Czyż można wyznaczyć jakiegokolwiek granice wspaniałości lub genialności przyszłych inżynierów?”. [Scott 1999, s. 187].

<sup>78</sup> Zob. między innymi: J. McCarthy, *Some Philosophical Problems from the Standpoint of Artificial Intelligence*, <http://www-formal.stanford.edu/jmc/mcchay69/mcchay69.html>.

<sup>79</sup> Za: Hołyński 1979, ss. 134-135.

praw zachowania się, mających postać kompletnych reguł postępowania (heurezy; przyp. M. K.).

... a...

jedyną znaną drogą, która może nas doprowadzić do znalezienia takich praw jest obserwacja naukowa i z pewnością nie znamy takich przypadków, w których moglibyśmy powiedzieć „Szukaliśmy dosyć. Nie ma takich praw”.<sup>80</sup>

Tak więc na podstawie naszych osobistych doświadczeń i tego, co zostało już powiedziane, możemy stwierdzić, że **do sytuacji przewidzianych** (starych, z którymi się już zetknęliśmy osobiście bądź poprzez obserwację), czyli – możemy rzec – **objętych programem, my, jako ludzie, stosujemy się (najczęściej) tak, jak dzieje się to w przypadku algorytmu, tzn. wykorzystując swoją „bazę danych” zachowań i samo to „oprogramowanie”** (czyli swoje doświadczenie)<sup>81</sup>. **W sytuacjach nieprzewidzianych** (tzn. nowych) **poszukujemy podobieństw i zależności ze starymi stosując wypróbowane w tych ostatnich techniki reakcji, bądź ucząc się nabywamy nowych**. Przykładem na to, jak dalece może być posunięta mechaniczna strona naszego działania, jest tzw. „bezduszość urzędników”, zaczynająca się od bibliotekarek czy farmaceutów niesprzedających bez recepty lekarstwa, mimo że wyraźnie jest potrzebne, a na wydawcach książek skończywszy. Zaś istotą nieznalesienia zależności między sytuacjami może być  *płacz*  bądź  *śmiech*  jako wyraz bezradności<sup>82</sup>.

Dzisiaj kwestię zachowania się w nowych sytuacjach podejmują programy uczące się i samouczące się (będzie o nich mowa w rozdziale 2.3.4 – „Uczenie maszyn (*machine learning*)”, ss. 129-130) oraz tzw. algorytmy stochastyczne i probabilistyczne, które rozwiązują zagadnienia zbyt trudne dla algorytmów deterministycznych<sup>83</sup>.

## I. Argument wpływający z pozazmysłowej percepcji

Jest to ostatni z listy zarzutów w stosunku do gry w imitację, które zostały omówione przez Turinga w *Computing machinery and intelligence* i najmniej poważny (chyba wraz z pierwszym). Odwołuje się on do rzekomego oddziaływania na człowieka całego spektrum zachowań pozazmysłowych, takich jak: telepatia, jasnowiedzenie, wiedza uprzednia czy lewitacja.

Jako kontrargument, do takiej na przykład telepatii, można zadać pytanie, które swego czasu zadał Stanisław Lem: *Czy myśli w umyśle są wyrażone w jednym z języków naturalnych, czy występują w jakieś innej formie kodowania niż język naturalny?* Jeżeli występują w formie języka naturalnego to, w jaki sposób są przekładalne myśli Anglika na myśli Japończyka, których sposób myślenia jest determinowany przez ich własny język etniczny? A o tym wiemy dzięki językoznawstwu kognitywnemu. Jeżeli zaś są kodowane w jakiś inny sposób to, w jaki? Jak wówczas są wyrażane przez takiego domniemanego telepatę (kwestią trudności

<sup>80</sup> Turing 1972, s. 41.

<sup>81</sup> Według wyników badań z zakresu nauk o mózgu za takie maszynowe działanie człowieka odpowiedzialny jest mózdzek. Głównie za motorykę człowieka, ale też za pewnego rodzaju mechaniczność mowy.

<sup>82</sup> *Śmiech i płacz* traktowane są jako wyraz bezradności w wielu koncepcjach antropologicznych (na przykład Helmut Plessner traktuje je jako „sytuacje graniczne”; H. Plessner, *Pytanie o condito humana*, tłum. M. Łukasiewicz, Z. Krasnodębski, A. Załuska, PIW, Warszawa 1998, rozdz. *Uśmiech*).

<sup>83</sup> Por. W. Duch, *Wstęp do kognitywistyki*, Internet.

językowych zajmę się przy okazji podrozdziału 2.3.5 – „Analiza mowy i języka”, ss. 131-153)? Jednak parapsychologia (nauka zajmująca się zjawiskami pozazmysłowymi) albo nie stawia takich pytań, albo odwołuje się do „praw przyrody, które jeszcze nie zostały poznane”. I być może w drugim przypadku jeszcze odpowiedź taka – na zasadzie odwoływania się do nieznanego – byłaby do przyjęcia gdyby nie fakt, iż te domniemane prawa przeczą współczesnym, sprawdzonym prawom fizyki.

Inną formą obrony przed zarzutem dotyczącym pozazmysłowej percepcji jest ta, którą zaprezentował sam Turing:

Bardziej charakterystyczny argument oparty na pozazmysłowej percepcji mógłby być następujący: „Zagrajmy w grę w naśladownictwo, biorąc za świadków człowieka, który jest dobrym odbiornikiem telepatycznym i maszynę cyfrową. Pytający może zadawać takie pytania jak: „Jakiego koloru jest karta, którą trzymam w prawej ręce?”. Człowiek, dzięki telepatii lub jasnowidzeniu daje prawidłową odpowiedź 130 razy na 400 kart. Maszyna może tylko zgadywać przypadkowo i może zyskać 104 prawidłowe odpowiedzi, tak, że pytający dokona prawidłowej identyfikacji”. Tutaj otwiera się interesująca możliwość. Załóżmy, że w maszynie cyfrowej znajduje się generator liczb przypadkowych. Wtedy naturalną rzeczą byłoby korzystanie z niego przy dawaniu odpowiedzi. Ale wówczas na ten generator liczb przypadkowych oddziaływałyby lewitacyjne moce pytającego. Może dzięki tej lewitacji maszyna zgadywałaby prawidłowo częściej niż można byłoby oczekiwać z rachunku prawdopodobieństwa, tak że pytający nadal nie potrafiłby dokonać prawidłowej identyfikacji.<sup>84</sup>

Tych, którzy nadal upieraliby się przy argumencie dotyczącym pozazmysłowej percepcji i pytali, jak to jest możliwe, że pole telepaty oddziałuje na maszynę, śpieszę poinformować, że jest to zawsze możliwe biorąc pod uwagę „prawa przyrody, których jeszcze nie znamy”, a na które to zwolennicy pozazmysłowej percepcji tak często się powołują!

W ten sposób doszliśmy do końca analizy testu Turinga i listy zarzutów pojawiających się przeciwko jego trafności, a znanych już samemu Turingowi i przez niego analizowanych. Pora zatem najwyższa przejść do zarzutów, które pojawiły się nieco później, by zaraz po tym poznać konsekwencje wyżej zaprezentowanego funkcjonalnego stanowiska Turinga, jakie miały dla dalszych zadań *Artificial Intelligence*, które podejmowane były w ciągu minionych 50 lat.

## 2.2.2. Test Turinga – zarzuty przez krytyków postawione i obrona przez Autora tej pracy czyniona

Jednymi z pierwszych, którzy zaczęli zastanawiać się konstruktywnie nad pomysłem Turinga, byli: klasyk *Artificial Intelligence*, Marvin Minsky (1966) oraz Stanisław Lem (1966). Minsky zwrócił uwagę, że sędzia (C) może ulec zbytnej sugestii tego, co sam uważa za inteligencję<sup>85</sup>. Tego samego zdania jest również współczesny polski logik Witold Marciszewski<sup>86</sup>. Píše on:

<sup>84</sup> Turing 1972, s. 42.

<sup>85</sup> Por. M. Minsky, *Artificial Intelligence*, w: "Scientific American", Nr 10/1966; polski przekład w zbiorze: *Dziś i jutro maszyn cyfrowych*, PWN, Warszawa 1969, s. 290. Zob. również: Marciszewski 1998, s. 77.

<sup>86</sup> Mało kto wie, że w latach 1987-1990 przyczynił się on między innymi do prac badawczych A. Trybulca nad programem do dowodzenia twierdzeń z zakresu logiki, znanym jako Mizar! Do dziś program ten cieszy się na całym świecie bardzo dobrą opinią.

Był w naszym kraju dostojnik znany z wyjątkowo nieinteligentnych wypowiedzi. Pogląd o tym defekcie podzielał piszący te słowa (czyli W. Marciszewski; przyp. M. K.), a miał on bardzo inteligentnego kolegę, wyśmienitego logika, który darzył umysł owego dostojnika wielkim respektem. Skąd ta ostra różnica ocen? Widać wrażliwość na stylistykę nie była składnikiem wysokiej inteligencji mojego kolegi, a właśnie nieszczęsna stylistyka była tym, czym wyróżniał się ów dostojnik. Tenże kolega, we wczesnych latach 80., uznał za mało rozcigniętego swego rozmówcę, który wtedy przewidywał upadek Związku Radzieckiego.<sup>87</sup>

Komentarz do powyższego jest dość jasny:

Jedni z nas uważali – pisze dalej autor powyższego fragmentu – że Związek Radziecki pokona cały świat, drudzy przewidywali jego rychły upadek. Któryś z tych pierwszych, gdyby zarazem żywił przekonanie o wyższości człowieka nad komputerem, przypisałby człowiekowi jako inteligentniejszemu, odpowiedź gloryfikującą Związek Radziecki. Ale gdyby, przypadkiem, wiedza komputera była w nim implementowana przez jakiegoś marksistę, to właśnie komputer dałby tę gloryfikującą odpowiedź, o co by go nie podejrzewał (jako głupszego) marksista w roli Obserwatora (naszego C, przyp. M. K.). Gdyby natomiast Obserwator – antymarksista rozumował przy założeniu o wyższej inteligencji człowieka, to właśnie komputer, jako uważany za głupszego, byłby podejrzany o wiarę w niebanalność komunistycznego imperium.<sup>88</sup>

Podobny zarzut postawiłem też w jednej ze swoich prac nazywając go **problemem sędziego**<sup>89</sup>. Tak jak Witold Marciszewski w powyższych fragmentach, a wcześniej jeszcze Marvin Minsky, idąc ich śladem, zwróciłem uwagę na fakt dysponowania przez sędziogo wyrobionymi przez lata skłonnościami do przypisywania komuś i nieprzypisywania komuś innemu faktu posiadania inteligencji. A zróżnicowanie skłonności może być różne, co zresztą we fragmencie *Sztucznej inteligencji* Marciszewskiego jest widoczne, a co ukaże się nam jeszcze wyraźniej po analizie poniższego przykładu.

W listopadzie 1991 roku w Bostonie w Muzeum Komputerów odbyła się próba *ograniczonego testu Turinga* (to jest takiego, w którym można zadawać pytania z jednej, określonej przez twórcę SI, dziedziny). Szerzej idea ograniczonego testu Turinga znana jest jako *konkurs Loebnera*, który to został powołany właśnie dzięki staraniom dr Hugh'a Loebnera za zgodą The Cambridge Center for Behavioral Studies (więcej o konkursie Loebnera patrz s. 209).

Wynik był zaskoczeniem. Wygrał program „PC-Therapist”. Jego autor – Joseph Weintraub – jako temat rozmów zadeklarował „pogaduszki” (*whimsical conversation*). Pogaduszki te zmyliły pięcioro sędziów (z dziesięciu; przyp. M. K.)! Myśleli, że rozmawiali z człowiekiem. Nie obyło się bez sytuacji humorystycznych. – I teraz proszę uważać – „Znawca sztuk Szekspira” został przez troje sędziów uznany za program komputerowy. Okazało się, że była to kobieta z krwi i kości.<sup>90</sup>

Można by zapytać, jak to się dzieje, że jeden sędzia stwierdza, iż (A) jest człowiekiem, a drugi, iż ów (A) jest komputerem? Dla nas odpowiedź jest oczywista i wypływa w rezultacie z komentarza samego autora powyższego fragmentu:

<sup>87</sup> Marciszewski 1998, s. 77-78.

<sup>88</sup> Tamże, s. 78.

<sup>89</sup> Zob. M. Kasperski, *Sztuczny człowiek czy Sztuczny Bóg?*, Internet, ss. 4-5.

<sup>90</sup> Buller 1998, s. 12.

Na usprawiedliwienie sędziów trzeba przyznać, że (niewiasta owa; przyp. M. K.) potrafiła cytować z marszu tak obszerne fragmenty dzieł mistrza ze Stratfordu, iż ich pamiętanie skojarzono raczej z zapisem na dysku.<sup>91</sup>

Zatem, jak widać, wystarczą proste *skłonności* bądź *nastawienie*, aby w rezultacie poddać w wątpliwość przypisanie przez (C) inteligencji danemu (A). Rzecz jest warta wymienienia, gdyż:

Jeżeli tak się dzieje, a dzieje się tak, co powyższy cytat pokazuje dość dobitnie to, który właściwie ma rację – czy *sędzia numer 1*, czy też *sędzia numer 2*? Co stanowi zatem o tym, czy *X* jest *maszyną myślącą*? Z powyższego można wnioskować, że prawo większości – odwołanie się w pytaniu o prawdę do koncepcji zgody powszechnej. I już nawet mniejsza o kompetencję [samego sędziego], bo powyższy przykład tego nie zakłada, – i teraz uwaga – ale wówczas ilu przepytujących (C) miałoby uczestniczyć w takim teście? Dziesięciu jak w powyższym przypadku? Dwudziestu, stu? A jeśli jeden to, kto ma nim być? Jakie warunki musi spełniać, na czym się znać?<sup>92</sup>

To właśnie w jednej z wcześniejszych prac nazwałem problemem sędziego.

Innym, co należałoby podkreślić, bardzo oryginalnym zarzutem, jaki został postawiony w stosunku do gry w naśladownictwo, jest koncepcja *wszystko wiedzącej maszyny*, lecz *wiedzącej* w specyficzny sposób, co zostało wyrażone przez Stanisława Lema w *Summa technologiae*<sup>93</sup>:

Zauważmy ze swej strony, że grę można skomplikować. – Pisze Lem o grze w naśladownictwo. – Mianowicie są do pomyślenia dwa rodzaje maszyn. Pierwsza jest „zwykłą” maszyną cyfrową, która jest złożona jak mózg ludzki (w sposób funkcjonalny oczywiście, czyli tak jak przewiduje to pomysł Turinga; przyp. M. K.); można z nią grać w szachy, rozmawiać o książkach, o świecie, na wszelkie w ogóle tematy. Gdybyśmy ją otworzyli, ujrzelibyśmy ogromną ilość obwodów sprzężonych tak, jak są sprzężone obwody neuronów w mózgu, poza tym – jej bloki pamięci itd., itp.

Druga maszyna jest zupełnie inna. – I teraz, co istotne dla naszych rozważań. – Jest to do planety (albo do kosmosu) powiększony Gramofon. Posiada ona bardzo dużo, na przykład sto trylionów nagranych odpowiedzi na wszelkie możliwe pytania<sup>94</sup>. Tak więc, gdy pytamy, maszyna wcale niczego „nie rozumie”, a tylko forma pytania, tj. kolejność drgań naszego głosu, uruchamia przekaźnik, który puszcza w obroty płytę czy taśmę z nagraniem odpowiedzi. Mniejsza o stronę techniczną. Rozumie się, że maszyna taka jest nieekonomiczna, że jej nikt nie zbuduje, bo i to właściwie niemożliwe, i głównie, nie wiadomo po co, by to robić. Ale nas interesuje strona teoretyczna. Bo **jeśli o tym, czy maszyna ma świadomość, decyduje zachowanie, a nie budowa wewnętrzna, czyż nie dojdziemy pochopnie do wniosku, że „kosmiczny gramofon” ją posiada – i tym samym wypowiemy nonsens? (A raczej nieprawdę).**<sup>95</sup>

<sup>91</sup> Tamże, s. 12.

<sup>92</sup> M. Kasperski, *Sztuczny człowiek...*, Internet, s. 5.

<sup>93</sup> Przypominam, że *Summa technologiae* Lema była pisana w latach 1951-1966.

<sup>94</sup> Oczywiście, rzecz taka jest niemożliwa, gdyż powstaje zasadnicze pytanie: jak coś nieskończenie wielkiego – wszystkie możliwe odpowiedzi – zamknąć w skończoności?

<sup>95</sup> Lem 1996a, s. 168; wyróżnienie M. K. Ten sam problem niedawno przypomniał S. Lem w: Lem 1999, s. 196.

Analizując przytoczony fragment dzieła Lema należy zwrócić uwagę, iż: pierwszy rodzaj maszyny bardzo przypomina strukturalnie tę z fragmentu paragrafu 17 *Monadologii* G. W. Leibniza (patrz: s. 45, przyp. 58), tylko, oczywiście, w bardziej współczesnej wersji – miast *popychających się elementów* – trybów i dźwigni – *obwody sprzężone* i *bloki pamięci*. Trzeba podkreślić także to, że o świadomości miałyby decydować budowa wewnętrzna (ściślej – mózg), co nie wydaje mi się prawdą, a tym bardziej paradygmatem, gdyż, jak sądzę, to nie sama budowa jest znacząca (tzn., co, z jakich „części” się składa) dla świadomości, a raczej poziom komplikacji organu<sup>96</sup>. Skrajnie wyraża to teza W. Ross Ashby’ego:

Wszystkie układy przekazujące informację w ilości powyżej  $10^8$  bitów na sekundę powinny być świadome.<sup>97</sup>

Jeżeli zaś okazałoby się inaczej, to wówczas należałoby stwierdzić, iż tylko idea sztucznego mózgu – i to na sposób strukturalny – ma szansę na urzeczywistnienie maszyny myślącej (patrz: podrozdział 2.4 – „Sztuczny mózg – sztuczne sieci neuronowe (*Neural Network*). Ku metaforze komputerowej”, ss. 154-176). Jednak zajmijmy się drugim pomysłem – „Wszechwiedzącym” Gramofonem, bo warto zwrócić uwagę na jego funkcjonalny aspekt.

Nie dziwi fakt niewiary Lema w realne zaistnienie swojej własnej fantazji, nawet w okrojonej wersji. Pisany przeszło 40 lat temu tekst, z przyczyn oczywistych, nie brał pod uwagę dokonań późniejszych, jak na przykład tzw. **obliczanie rozproszone** czy też **wirtualny komputer**<sup>98</sup>.

Obliczanie rozproszone wykorzystywane jest do bardzo złożonych zadań, jak na przykład do deszyfracji:

Rząd amerykański korzysta z systemu szyfrowania DES (**D**igital **E**ncryption **S**tandard, czyli cyfrowy system kodowania). W tym systemie informacja jest szyfrowana za pomocą jednego z  $2^{56}$  kluczy kodujących, które są ogólnie dostępne.

Aby odczytać zaszyfrowaną wiadomość bez znajomości klucza, należy sprawdzić wszystkich  $2^{56}$  możliwości. Dla jednego komputera jest to zadanie praktycznie niewykonalne. Kilku kryptografów wpadło na pomysł, żeby złamać DES, wykorzystując Internet. Zaprosili do zabawy setki tysięcy Internautów i po kilku miesiącach pracy setek tysięcy komputerów kod został złamany.<sup>99</sup>

Tego rodzaju *wirtualny komputer*, który wykorzystuje połączenie wielu prywatnych komputerów, pozwala na częściowe zrealizowanie, pod kierownictwem człowieka oczywiście, pomysłu Lema. Niemożliwość zrealizowania Wszechwiedzącego Gramofonu w sposób absolutny związana jest z niemożliwością zamknięcia nieskończoności pytań i odpowiedzi w skończonej ilości zdań (jak już wiemy, wypływa to po części z twierdzenia Gödla). Zatem wyobraźmy sobie, że grupa informatyków zdobyła się na taki krok, a maszyna zwana *kosmicznym gramofonem* pozwala na swobodną wymianę zdań tak, jak zakłada to test Turinga. To wcale nie musi *być sto trylionów nagranych odpowiedzi na wszelkie możliwe pytania*, jak pisze Lem, aby można było przeprowadzić przeciętną rozmowę, z przeciętnym

<sup>96</sup> Blżej mi do *emergentyzmu*, wg którego istota komplikujących się funkcji umysłowych miałyby przyczynę w komplikacji strukturalnej mózgu. Ale... przechodzenie na kolejne piętra komplikacji nie odbywałoby się *stopniowo*, a *skokowo*. Por. na ten temat chociażby: Trefil 1998, ss. 162-163.

<sup>97</sup> W. R. Ashby, *Cybernetics*, w: "Recent Progress in Psychiatry", Nr 3/1958; za: Hołyński 1979, s. 152.

<sup>98</sup> Por. Wallich 2001.

<sup>99</sup> Hołyński 2000, s. 19.

człowiekiem. Czy znaczyłyby to, że układ ów myśli? Czy raczej „myśli”? Prawdę mówiąc, problem jest bardzo interesujący i zupełnie otwarty.

Jeśli będziemy analizować maszynę, która będzie miała nagrane odpowiedzi i tylko je będzie odtwarzała, to wówczas nawet gramofon czy zwykły magnetofon może nas wprowadzić w błąd w kwestii tego, czy myśli. Zapisy na płycie czy kasecie byłyby jego zdarzeniami mentalnymi (*sic!*). I faktycznie – mówiąc, że taka maszyna myśli, wypowiadamy nonsens (można znów, choć tylko częściowo, powołać się na przykład z papugą; por. ss. 51, 53).

Ale, jak mi się wydaje, wystarczyłoby, żeby taka maszyna zamiast gotowych odpowiedzi posiadała schematy językowe, na przykład schematy fonemów (bo dzięki nim się porozumiewamy) czy difonemów (złożeń fonemów) i za ich pomocą, na podstawie owych:

1. **schematów językowych – fonemów/difonemów;**
2. **schematów gramatycznych – syntaktyki;**
3. **i bazy, na przykład podstawowych słów** (takich podstawowych słów, których używa przeciętny człowiek jest około 10 tysięcy<sup>100</sup>);
4. **oraz przy udziale oprogramowania generującego kolejne zdania** — niczym młyn mielący ziarna słów;

można by otrzymać w rezultacie program odpowiadający na pytania i – co teraz ważne ze względu na powyższy fragment Lema – używający odpowiedzi, których nie posiadał w swojej pamięci (bazy wiedzy)! To właśnie określenie **GENEROWAĆ** jest istotą myślenia i inteligencji, które świetnie pasuje do ram pytań i odpowiedzi wyznaczonych przez Turinga. A, co ciekawe, na podstawie takich koncepcji i charaktersystycznych ich hipotez jak: A) *gramatyka generatywna* N. Chomsky’ego, B) *język mentalezyjski* J. Fodora<sup>101</sup>, C) *neurojęzyk* P. M. i P. S. Churchlandów, a nawet D) tzw. *neuronalny darwinizm*<sup>102</sup> G.M. Edelmana (czy zbliżona do niego wersja *selektywnego modelu mózgu* autorstwa M.S. Gazzanigi) można sądzić, iż w taki sposób można by skonstruować maszynę myślącą, a zasadę generowania ze schematów nowych zdań, zastosować do wszystkich maszyn imitujących myślenie – a dalej, mowę – człowieka (por. podrozdział 2.3.5 – „Analiza języka i mowy”, ss. 131-153)<sup>103</sup>. Jednak zarzutów to jeszcze nie koniec.

Witold Marciszewski w swej książce analizuje jeszcze inny zarzut wynikający, jak się okaże, z tego, że część ludzi hołduje przekonaniu: „Dociekać jest sprawą ludzką!”.

Pytania w teście Turinga zadaje Obserwator (nasz (C); przyp. M. K.), podczas gdy komputer jest jednym z dwóch respondentów. Teraz niech sytuacja wygląda następująco. Obserwator dostaje od rozmówców pytania, na które odpowiada wedle swej wiedzy. Po udzieleniu przezeń odpowiedzi każdy z jego rozmówców reaguje w

<sup>100</sup> Por. Hołyński 1979, s. 120.

<sup>101</sup> Język mentalezyjski (*mentalese language*), którego istnienie głosi Jerry Fodor, miałby wedle niego być językiem, którym się posługujemy myśląc (w przeciwieństwie do – mówiąc). Przy tym byłby językiem o konkretnej budowie:

1. posiadałby skończoną bazę prostych terminów semantycznych;
2. nieskończenie wiele semantycznych złożonych terminów na podstawie (1) i syntaktycznych reguł;
3. nieskończenie wiele zdań generowanych na podstawie (1) i (2) i reguł syntaktycznych.

Por. również ss. 208-209.

<sup>102</sup> Por.: Gazzaniga 1997, s. 64; Edelman 1998.

<sup>103</sup> Częściowe omówienie powyższych teorii znajduje się w odpowiednich miejscach poniższej pracy (patrz: *Indeks*), lecz z przyczyn ograniczonych tematyką nie będę się bardziej wgłębiał w ich struktury. Zainteresowanych jednocześnie odsyłam do odpowiedniej literatury (patrz: *Bibliografia*).



ten sposób, że albo uznaje tę odpowiedź za wystarczającą, albo stawia dalsze pytania. Na przykład prosi o lepsze uzasadnienie czy wyjaśnienie któregoś z punktów czy też pyta, jak teoria zakładana w odpowiedzi radzi sobie z takim to a takim zagadnieniem. Po tego rodzaju reakcjach ludzki Obserwator powinien by rozpoznać, który z reagujących jest jego bliźnim, a który maszyną.<sup>104</sup>

Spostrzeżenie to, moim zdaniem, jest mieczem obosiecznym. Z faktu tego, że niektórzy ludzie są dociekliwi nie wynika wprost, że maszyna nie podoła zadaniu bycia dociekliwym (wręcz przeciwnie – fakt ów można wykorzystać do tego, by maszyna była jeszcze bardziej wiarygodna w swoim zadaniu imitacji ludzkiego myślenia poprzez zaimplementowanie jej dozy dociekliwości). Poza tym może zdarzyć się też tak, że fakt dociekliwości ze strony *X* może zostać odebrany przez (*C*) jako chęć wprowadzenia go w błąd, wyrażony w formie: „A co on taki ciekawski?”. Wówczas też kontrargument Marciszewskiego może zadziałać w stronę drugą.

Innego rodzaju zarzutem jest odwołanie się do pewnych „zdolności”, które licząca maszyna posiada z racji swego istnienia.

Jest pewien rodzaj pytań, na podstawie których odpowiedzi zawsze możemy wywnioskować poprawnie, czy po drugiej stronie znajduje się maszyna. Jaki to typ pytań? Zastanówmy się. Wyróżnikiem dla maszyn obliczających jest moc obliczeniowa, której nie ma nikt z ludzi<sup>105</sup>. Wystarczy teraz powołać się na znajomość tego faktu i zadawać pytania w stylu: „Ile to jest 79% z 284,013 albo 34 251<sup>7</sup>?”, a komputer obliczy to w mig, a prawie każdy człowiek (wyłączając genialnych rachmistrzów) miałby z tym problem. Ten kruczek nazwałbym **problemem naturalnej celowości**, gdyż właśnie ze względu na nią (na ową celowość) maszyny liczące dobrze liczą, a ludzie mają w zamian dobre pomysły, jak na przykład zbudowanie maszyny realizującej to, z czym sami się borykają (jak w tym przypadku maszyny liczącej).

Ktoś może twierdzić, iż taki test (test Turinga; przyp. M. K.) jest niesprawiedliwy w stosunku do komputera. Gdyby odwrócić role i nakazać człowiekowi, aby udawał komputer, oraz zaprogramować komputer tak, aby udzielał prawdziwych odpowiedzi, to przesłuchująca maszyna bez trudu mogłaby ustalić, kto jest kto. Wystarczyłoby żądanie, by przesłuchiwanymi wykonali dostatecznie skomplikowany rachunek arytmetyczny. Dobry komputer powinien natychmiast udzielić poprawnej odpowiedzi, natomiast człowiek mógłby mieć z tym poważne problemy.<sup>106</sup>

Sam Turing również doskonale zdawał sobie z tego faktu sprawę, lecz, co ciekawe, nie widział w takich zdaniach niczego niebezpiecznego:

Grę można, być może skrytykować z tego powodu, że maszynie dano w niej o wiele mniejsze szanse niż człowiekowi. Gdyby człowiek starał się udawać maszynę, to oczywiście robiłby bardzo złe wrażenie. Skompromitowałby się od razu swoją powolnością i niedokładnością w arytmetyce. (...) Ten zarzut jest bardzo mocny, ale nie potrzebujemy się nim przejmować, jeśli tylko mimo wszystko można będzie zbudować maszynę tak, aby grała zadawalająco w grę w imitację.<sup>107</sup>

<sup>104</sup> Marciszewski 1998, s. 79.

<sup>105</sup> Oczywiście nie dotyczy to znanych przypadków genialnych rachmistrzów. Na przykład Niemca Johanna Martina Zachariasa Dase'a (1824-1861), czy działającego w latach pięćdziesiątych profesora matematyki na Uniwersytecie Edynburskim, Aleksandra Aitkena. W tej sprawie por. R. Penrose 1995, s. 21. Również w niniejszej pracy cyt. z Penrose'a, ss. 68-69.

<sup>106</sup> Penrose 1995, s. 21.

<sup>107</sup> Turing 1972, s. 25.

Lecz dalej, jak analizował w „*G. Argumencie wyphywającym z ciągłości systemu nerwowego*” (por. ss. 59-60) rzecz o identycznych skutkach co w wyżej wymienionym problemie, to rozwiązaniem miało być:

Na przykład na żądanie podania wartości  $\pi$  (która faktycznie wynosi około 3,1416) postąpiłaby rozsądnie dokonując wyboru na chybił trafił spośród wartości 3,12; 3,13; 3,14; 3,15; 3,16 z prawdopodobieństwami wynoszącymi (powiedzmy) 0,05; 0,15; 0,55; 0,19; 0,06.<sup>108</sup>

Tylko przecie nie po to będziemy konstruować maszynę myślącą, żeby się myliła (nawet jeśli te pomyłki będą kontrolowane jak w powyższym przypadku). Także swój test Turing obronił, ale moim zdaniem za cenę pozbawienia, w dalszym etapie, maszyny rozumności (w tym sensie trafności odpowiedzi  $X$  w stosunku do odpowiedniości (prawdziwości) tej odpowiedzi na pytanie  $C$ ).

Inną metodę obrony testu Turinga może stanowić fakt przypadków rachunkowych geniuszy (choćby znanego z filmu *Rain Men*, z Dustinem Hoffmanem w roli głównej), którzy posiadają zdolności porównywalne z szybkością obliczania zdań arytmetycznych przez komputery. O tym też pisze R. Penrose:

Zdarzają się rachunkowi geniusze, zdolni do wykonywania zdumiewających obliczeń w pamięci, i to bez błędów i bez widocznego wysiłku. Na przykład żyjący w Niemczech w latach 1824-1861 Johann Martin Zacharias Dase, syn niepiśmiennego chłopca, potrafił pomnożyć w pamięci dwie ośmiocyfrowe liczby w niecałą minutę, a dwie dwudziestocyfrowe w sześć minut! Takie osiągnięcia można łatwo przypisać komputerowi. Osiągnięcia rachunkowe Aleksandra Aitkena, który był profesorem matematyki na Uniwersytecie Edynburskim w latach pięćdziesiątych, są równie imponujące. Znane są również inne podobne przypadki.<sup>109</sup>

Wprawdzie chodzi nam o przypadki sztamkowe, a nie wyjątkowe, ale przecie przepytujący ( $C$ ) nie musi wcale wiedzieć, a nawet więcej – nie wie – z kim ma do czynienia po drugiej stronie. To może stanowić linię obrony przed zarzutem, lecz nie przy współudziale z wcześniej wspomnianymi (przy okazji problemu sędziego) ludzkimi skłonnościami do uśredniania możliwości umysłowych innych osób (w tym maszyn). Skojarzenie przez sędziów wydarzeń w Bostonie z 1991 roku faktu pamiętania przez człowieka *obszernych fragmentów dzieł mistrza ze Stanfordu z zapisem na dysku* może tylko dowodzić, iż taki sam byłby finał rozegrania powyższego przypadku funkcjonowania rachunkowych geniuszy – tu przypadki poprawnych wyników obliczeń (czy szybkości obliczeń) dokonanych przez ludzi przypisywanoby zapisom na dysku komputera.

Ostatnim z listy zarzutów kierowanych w stronę gry w naśladownictwo niech będzie argument związany z ludzką umiejętnością zmieniania płaszczyzny rozważań.

<sup>108</sup> Tamże, s. 40.

Przykład taki też analizowałem w pracy *Sztuczny człowiek...* Cytuję, co następuje: „Maszyna nie posiada *wolnej woli*. Zatem jak realizuje procedurę kłamstwa? Za pomocą  $A$ ) wyuczenia,  $B$ ) ewentualnego doboru losowego. Przykład  $A$ : Maszyna nauczyła się, że jeżeli zadane jest pytanie „Czy jesteś sztuczną inteligencją?”, ma odpowiedzieć „Nie”. Przykład  $B$ : Maszyna spytana „Czy podoba ci się obraz Salvadora Dali *Trwałość pamięci*?”, może odpowiedzieć „Tak” bądź „Nie”, w zależności od tego, która z opcji została wylosowana.” [Tamże, s. 3]. Przykład  $B$  jest świadectwem takiego rozumowania, jak w przedstawionym wyżej fragmencie z: Turing 1972.

<sup>109</sup> Penrose 1995, s. 21.

Taka strategia daje dodatkowe wyjście z sytuacji: nie trzeba mówić prawdy, nie trzeba kłamać, wystarczy zmienić płaszczyznę rozważań. Przykład? Wyobraźmy sobie następujące sytuacje:

Mała córka o imieniu Matylda (*M*) prosi ojca (*O*) o zgodę na wyjście na noc do koleżanki. Pomimo nalegań i licznych argumentów ze strony (*M*), (*O*) się nie zgadza. (*M*) w tym momencie mówi: *Wiesz co, tatusiu, dawno z mamusią nie byliście sami, może byś jej zrobił niespodziankę? Ja bym ci pomogła w przygotowaniu kolacji.* – I, chciałoby się dodać, przy okazji opuszczenia towarzystwa na noc. – Co robi (*M*)? Zmienia płaszczyznę w myśl swojej strategii.

Mała córka o imieniu Matylda (*M*) przychodzi ze szkoły do domu, w którym jest ojciec (*O*). W szkole nie poszło jej najlepiej, bo dostała ocenę niedostateczną z klasówki z matematyki, nad którą zeszłego dnia siedziała wraz z (*O*), by się odpowiednio przygotować. Ojciec się pyta: *No i jak ci poszło?* Córka odpowiada: *Wiesz co zauważyłam? Że dawno nie byliśmy nigdzie razem. Taka ładna pogoda, może wybralibyśmy się nad jezioro?* – Albo: *Teraz nie mam czasu. Później ci powiem.* A, jak wiadomo, owo „później” przy takich okazjach nigdy nie następuje.<sup>110</sup>

Najprostszą metodą obrony gry w imitację, która jednak słyca całość problemu, jest stwierdzenie, że *maszyny myślące będą myślały*, a skoro tak, zatem nie widać powodu, dlaczego nie miałyby zmieniać płaszczyzny rozważań w myśl swojej strategii.

Przykład drugi ilustruje jeszcze jedną rzecz. Zdanie: *No i jak poszło?* jest w zasadzie niedokończone – nie mówi nam poszło z czym?/z kim?, a to, że córka zrozumiała to zdanie zawdzięcza tylko temu, iż odwołała się do tła rozważań. Miała pewne nastawienie, o co (*O*) może się spytać. Maszyny nie posiadają również i tego.<sup>111</sup>

Ten dodatkowy aspekt jeszcze bardziej utrudnia nam obronę testu Turinga. Jest to w rezultacie zarzut odnoszący się przy okazji do rozumienia przez maszynę semantyki, a nie tylko syntaktyki (będzie o tym mowa w podrozdziale 2.3.5 – „Analiza języka i mowy”, ss. 131-153). Zarzut – o podobnej konsekwencji jak przedstawiony tutaj – nazwany *Chińskim Pokojem* autorstwa J. Searle’a, będzie omówiony dalej. Obronić jednak się przed nim można na dwa sposoby: po pierwsze tak, jak uczyniłem to w innej zgoła pracy:

Niektórzy wierzą w to, że można skonstruować odpowiednio bogatą syntaktykę, w której dowiedne są zdania, tj. prawdziwe pod względem semantycznym – żargonowo: takie, gdzie syntaktyka realizowałaby semantykę. Takiego zdania był chociażby sam Kurt Gödel.<sup>112</sup>

A jako sposób drugi proponuję przeanalizować poniższy tekst:

Nie wdając się w bliższe szczegóły wyobraźmy sobie komputer kierujący pracą robota, który przedstawia dwie piramidki i cztery prostopadłościennymi bloki. Dzieje się to na małym obszarze, gdzie znajduje się jeszcze spore pudełko. (...) Człowiek wydaje pierwsze polecenie:

Człowiek: Wybierz wysoką zieloną piramidkę.

Maszyna: OK.

Komentarz: *Maszyna odpowiada „OK.” po wykonaniu rozkazu.*

Człowiek: Odszukaj blok i wóź go do pudełka.

<sup>110</sup> M. Kasperski, *Sztuczny człowiek...*, Internet, s. 4.

<sup>111</sup> Tamże, s. 4.

<sup>112</sup> Kasperski 2001b, s. 72.

Maszyna: NIE ROZUMIEM O JAKI BLOK CHODZI.  
 Komentarz: *Na planie są cztery bloki i maszyna jest na razie zdezorientowana.*  
 Człowiek: Czerwony.  
 Maszyna: OK.  
 Człowiek: Co jest w pudełku?  
 Maszyna: ZIELONA PIRAMIDA I CZERWONY BLOK.  
 Człowiek: Gdzie jest blok?  
 Maszyna: W PUDEŁKU.  
 Komentarz: *Poprzednio program nie wiedział, o jaki blok chodzi, teraz zakłada, że mieliśmy na myśli wymieniony ostatnio blok czerwony*<sup>113</sup>.

Zachowanie maszyny w tym konkretnym przypadku ilustruje, że maszyna również może posiadać nastawienie i „domyślać” się na podstawie kontekstu. A maszyna, która tak zręcznie „wywnioskowała”, o jaki blok chodziło człowiekowi (na co wskazuje ostatnia część dialogu) jest programem z 1976 roku autorstwa Terry’ego Winograda. Więc od tamtego czasu, sprawności w interpretowaniu i przetwarzaniu struktur językowych wszystkich maszyn, miast utracić, jeszcze sporo zyskały (o czym przekonamy się przy okazji podrozdziału 2.3.5 – „Analiza języka i mowy”, ss. 131-153).

\*\*\*

Podsumowując dwa ostatnie podrozdziały trudno nie oprzeć się pokusie, iż większość, jeśli nie wszystkie, z argumentów przeciw testowi Turinga i tym samym możliwości powstania maszyny myślącej (w sensie Turinga) jest całkowicie chybiona. Tym bardziej, iż nowe rozwiązania technologiczne ciągle weryfikują głosy informatycznych malkontentów w stylu: *Ale maszyny wciąż nie mogą...*

### 2.2.3. Test Turinga – konsekwencje

Pierwszą zauważalną konsekwencją gry w imitację jest funkcjonalny charakter zachowania się człowieka, który tu pełni kluczową rolę – sprowadzenie człowieka do jego umysłowości, dodajmy, umysłowości wyrażanej za pomocą języka – poprzez pytania i odpowiedzi. Koncepcja ta została nazwana **funkcjonalizmem**. Wyraża się ona, tak jak arytmetyka, przez to, że w metodzie tej wyznacza się uniwersum, na którym przeprowadza się z góry ustalone operacje. To ostatnie powoduje, że metodę ową zwie się również **operacjonizmem**<sup>114</sup>. W końcu to, że operacje te są w rezultacie obliczeniami, sygnuje dla niej nazwę **komputacjonizm** (od ang. *compute*, co znaczy „obliczać”). Przy czym zakłada się tu spełnienie dwu szczególnych warunków<sup>115</sup>: pierwszy polega na tym, że odpowiednio bogata struktura obliczeniowa w pełni wystarczy, aby można było stwierdzić, że przysługuje jej umysł (tzw. teza o obliczeniowej wystarczalności (ang. *computational sufficiency*)) i drugi – procesy obliczeniowe w pełni pozwalają nam na badanie i poznawanie procesów poznawczych człowieka (tzw. teza o obliczeniowej wyjaśnialności, ang. *computational explanation*).

Wpływ na takie podejście metodologiczne Turinga do teorii umysłu odegrały dwa zasadnicze fakty:

<sup>113</sup> Hołyński 1979, ss. 106-110.

<sup>114</sup> Nie mylić z operacjonizmem ani w fizyce (P. W. Bridgman (1882-1961)), ani w psychologii, który został zapoczątkowany w 1930 roku przez amerykańskiego psychologa, E. G. Boringa.

<sup>115</sup> Por. Chuderski, *Wykorzystanie metod sztucznej inteligencji do badań nad umysłem*, Internet.

1. kiedy chodzi o *inteligentne maszyny*, nie zwraca się uwagi na inne aspekty człowieka niż inteligencja, jak na przykład podobieństwo ciał;
2. w tamtych czasach (lata 50) w psychologii dominujący był **behawioryzm**, w prostej linii wywodzący się z **psychologii funkcjonalnej**, a jeszcze w tamtych czasach<sup>116</sup> sprawiający wrażenie, że da szansę na naukowe badanie psychicznych aspektów człowieka.

Pierwszy aspekt jest dosyć oczywisty i jasny: możliwość widzenia maszyny, czyli, że tak się wyrażę, naocznego doświadczenia różnic pomiędzy nią a człowiekiem od razu dyskwalifikuje grę w naśladownictwo (przynajmniej przy dzisiejszych rozwiązaniach z zakresu robotyki kognitywnej i możliwości nadania całkowicie antropomorficznych kształtów maszynie!<sup>117</sup>). Można by oczywiście stwierdzić, iż grę winno się poprowadzić w taki sposób, aby rozmiar imitacji był jak najpewniejszy (na przykład angażował imitację ludzkiego głosu bądź ludzkiego ciała), lecz, jak to wyżej zostało powiedziane, w maszynach mających za zadanie naśladować mentalny aspekt człowieka, taka budowa jest zbyt cenna dopóty, dopóki nie wykaże się, że ma istotny wpływ na funkcję myślenia. Dziś oczywiście już w pewien sposób wiemy, że aby maszyna była zdolna naśladować człowieka winna posiadać chociaż część swojej cielesności, na przykład *sensorium*<sup>118</sup> – czyli całość pięciu zmysłów (będzie o tym mowa w podrozdziałach: 2.3.3 – „Rozpoznawanie obrazów” (ss. 113-128) i 2.4.2 – „Cog, Aibo... Problem intencjonalności i problem ciała” (ss. 170-183). Lecz, powtórzę to stanowczo, ze względu na intencje Turinga, dla jego testu brak cielesności maszyny nie stanowi przeszkody dla rozwoju maszyn myślących, wręcz przeciwnie – cielesność jest postrzegana jako zbyt cenna, tak jak dla wszystkich programów typu IDS (*Intelligent Dialogue Systems*) mających prowadzić inteligentnie dialog.

Drugi aspekt wypływa w zasadzie z pierwszego, a na jego zastosowanie miały z pewnością wpływ czasy, w których tworzył Turing. Chodzi o to, że jeżeli wygląd zewnętrzny dyskwalifikuje maszynę przy próbie udawania przez nią człowieka, to jaką obrać inną drogę, by imitacja ta była w miarę doskonała? Lekarstwem, jak się okazało, była behawiorystyczna, a jeszcze wcześniej funkcjonalna, koncepcja człowieka.

Do czasów psychologii funkcjonalnej psychologowie zwracali uwagę i badali treści umysłowe. Zmienili to radykalnie: Amerykanin J. R. Angell, związany z Katedrą Psychologii w Chicago wraz z uczelnianym kolegą, filozofem J. Deweyem (1859-1952) oraz

<sup>116</sup> Stwierdzenie „jeszcze w tamtych czasach” odwołuje się do późniejszej druzgoczącej krytyki behawioryzmu dokonanej przez Noama Chomsky’ego (Chomsky, *Review of B. F. Skinner's „Verbal Behaviour”*, w: „Language”, 1959).

<sup>117</sup> Oczywiście związane jest to tylko z dzisiejszym stanem technicznym, a objawia się na przykład w mechanicznym sposobie zachowania się robotów, wolnym od gracji ruchów ludzi, czy też po prostu w mało antropomorficznym wyglądzie (choć już i nad tym trwają prace, przykładem są badania nad antropomorficznymi dłońmi prowadzone przez Jessicę Lauren Banks z MIT). Jednak są spore szanse, że perspektywy tej gałęzi inżynierii szybko przyniosą jakieś zadawalające w tej materii rezultaty. Osobiście liczę, że za jakieś 50 lat stanie się możliwym pomylenie człowieka z maszyną na podstawie zewnętrznego wyglądu ich ciał.

<sup>118</sup> Jak pisał Stanisław Lem: „Słowa tego nie znajdzie się ani w słowniku obcych wyrazów, ani w encyklopedii, nawet w *Wielkim Warszawskim Słowniku* opatrzone jest wykrzyknikiem, oznaczającym, że lepiej go nie używać. *Mnie jest jednak potrzebne*. Sensorium to całość wszystkich zmysłów oraz wszystkich dróg (zazwyczaj nerwowych), jakimi informacje, powiadamiające nas o „istnieniu czegokolwiek”, mkną do ośrodkowego układu nerwowego.” [Lem 1999, s. 22; wyróżnienie, M. K.]

szwajcarskim psychologiem, É. Claparède (1873-1940)<sup>119</sup>. Zaproponowali oni aby badać funkcje umysłu. Dominującymi, i w tamtych czasach nowatorskimi, tezami były<sup>120</sup>:

1. psychologia winna badać: w psychologii widzenia – formy widzenia, a nie barwy; w zagadnieniu wyobraźni – wyobraźnię właśnie, miast wyobrażenia; w zagadnieniach emocji – odczuwanie, a nie jak dotąd uczucia (czyli: nadrzędną rolę pełni badanie funkcjonowania umysłu, a nie wytworów tego funkcjonowania);
2. psychologia winna w badaniu zastosować metodę historyzmu, czyli badania umysłu w czasie, badania dynamicznego, a nie, jak wcześniej, zawężyć się do tu i teraz jego wytworów.

Warunki te zostały przejęte również i później przez behawiorystów – Johna B. Watsona (1878-1958) i Burrhusa F. Skinnera (1904-1990) i radykalizowane do tez następujących<sup>121</sup>:

1. funkcje umysłu przejawiają się w zachowaniu, a dokładniej – modelu bodziec → reakcja, zatem winno badać się zachowanie (lecz – uwaga – zachowanie jest w zasadzie tylko mechanizmem jak w przypadku odruchu Pawłowa);
2. zachowanie, rozwijając się w czasie, rozwija się w grupach, w których dany osobnik uczestniczy(ł) i grupa ta na niego wpływa, na przykład poprzez podniety, zatem metodę historyzmu można rozszerzyć do badania grup społecznych.

Dla ułatwienia identyfikacji powyższych tez można je sprowadzić do następujących:

1. **funkcjonalizm: stan mentalny  $x$  jest funkcją stanu mózgu bądź/i układu nerwowego  $x$** , na przykład „odczuwać ból” znaczy „być w stanie spowodowanym przez bodziec na nocyceptory”<sup>122</sup>;
2. **behawioryzm: stan mentalny  $x$  przejawia się w zachowaniu i jest funkcją organizmu** (ewentualnie jego dyspozycją do bycia w stanie  $x$ ).

Jedyną zauważalną różnicą pomiędzy funkcjonalistyczną koncepcją umysłu, a jego odpowiednikiem w behawioryzmie jest rozszerzenie „sprawcy” umysłu z mózgu (w przypadku pierwszego) do organizmu (w przypadku drugiego)<sup>123</sup>. Jednak jest to, jak zaraz wykażę, rozszerzenie, które to w ogóle nie jest znaczące dla naszej gry w imitację. Dlaczego? Gdyż:

<sup>119</sup> Funkcjonalizm, prócz w psychologii, rozwijał się również w innych dziedzinach: w filozofii i metodologii – jako kierunek przyznający szczególną rolę pojęciu funkcji i poszukujący zależności funkcjonalnych w naukach, które decydują o statusie nauk, w antropologii z głównymi przedstawicielami: Bronisławem Malinowskim (1884-1942) i R. A. Radcliffe-Brownem, którzy wyjaśniali fakty antropologiczne przez ustalanie funkcji, jakie pełnią one w danym systemie kulturowym.

<sup>120</sup> Por. Tatarkiewicz 1990, t. 3/3, ss. 281-298.

<sup>121</sup> Por. Blackman 2000, ss. 147-174.

<sup>122</sup> Nocyceptory, czyli receptory bólowe.

<sup>123</sup> Por. następujący fragment na to wskazujący: „Watson inaczej, niż było to w zwyczaju, pojmował reakcje, a zwłaszcza rolę systemu nerwowego: odmówił mu, a także i samemu mózgowi tej wyjątkowej roli, jaką mu dotąd przypisywała psychologia. Dotąd było tak: psychologia, nawet najbardziej fizjologiczna, liczyła się w organizmie tylko z systemem nerwowym, a inne jego składniki miała za pozbawione znaczenia dla zjawisk psychicznych. Watson zaś sądził zupełnie inaczej: to system nerwowy ma dla nich znaczenie stosunkowo najmniejsze. Właściwie nie robi on nic innego jak tylko to, że przenosi podniety; i bez niego wszystko działałoby się tak samo, a tylko nieco później, bo organizm i tak by przekazał podniety gruczołom i mięśniom. O reakcji jednostki stanowi nie tylko mózg i nerwy, ale również organy wewnętrzne, gruczoły, mięśnie i w ogóle całe ciało.” [Tatarkiewicz 1990, s. 294].

1. Rozszerzona sfera psychiczna w przypadku behawioryzmu, która sięga do zachowań bądź dyspozycji do zachowań, ma się nijak do myślącej maszyny w sensie Turinga. Obserwowanie jej zachowania ogranicza się przecie do zachowań językowych. Jej **zachowanie inne niż w języku**<sup>124</sup> nie jest związane z maszyną Turinga!
2. Nawet **zachowanie w języku** maszyny myślącej jest ograniczone. Jak wspomniałem już wyżej, sfera cielesności nie wpływa w żaden deprymujący sposób na podstawową tezę Turinga. Także nie interesuje nas tu na przykład wpływ ciała na emocjonalne podejście w rozmowie, bo wyrażane jest poprzez ton i intonację głosu, a te cechy przynależą do cielesności. To, co nas interesuje, to wyrażanie „myśli” maszyny i, jak wiemy, odbywać się ono będzie poprzez terminal. Jak więc analizować wpływ cielesności, objawiający się na przykład zdenerwowaniem po zadaniu intymnego pytania? Jedyne w języku, co stwierdzimy w toku naszych obserwacji.

Zatem, biorąc pod uwagę dwa ostatnie punkty znoszące granice pomiędzy funkcjonalistycznym a behawiorystycznym podejściem do umysłu, należałoby stwierdzić:

**funkcjonalizm = behawioryzm  
z punktu widzenia podejścia zaprezentowanego w teście Turinga.**

Warto jeszcze zwrócić uwagę na najczęściej pojawiający się zarzut wobec funkcjonalizmu. Funkcjonalne potraktowanie relacji mózg-umysł każe nam przypuszczać, iż maszyna spełniająca funkcjonalnie zadanie na przykład prowadzenia dialogu – odpowiadania na pytania jak to ma miejsce w teście Turinga – w rezultacie myśli. Ta konsekwencja, niestety, pokutuje, co można wyrazić następująco:

Funkcjonalizm charakteryzuje stany psychiczne za pomocą ich ról przyczynowych. Można to funkcjonalistyczne pojęcie zilustrować dzięki tabeli maszyny sprzedającej *Coca-Colę* (założeniem jest tutaj to, że *Coca-Cola* kosztuje 10 centów).

S1	S2 (przy założeniu, że zostało wrzucone 5 centów)
5 centów włożono nic nie wyrzucaj przejdź do S2	wyrzuć <i>Coca-Colę</i> przejdź do S1
10 centów włożono wyrzuć <i>Coca-Colę</i> zostań w S1	wyrzuć <i>Coca-Colę</i> i 5 centów przejdź do S1

Możemy powiedzieć (i nie wszyscy uznaliby to za żart), że S1 = pragnienie 10 centów, a S2 = pragnienie 5 centów. Znaczy to, że pragnienie jest stanem zdefiniowanym funkcjonalnie poprzez powyższą tabelkę. W ten sposób scharakteryzowane typy stanów mentalnych nie mogą być zredukowane na sposób behawiorystyczny. I tak, to S1 jest przyczyną S2, a nie jakieś „dane wejściowe”. A także to S1 razem z S2 są przyczyną danych wyjściowych. Dla behawiorystów „stany psychiczne” nic takiego nie

<sup>124</sup> Pojęcie „zachowania w języku” ukułem na potrzeby wysłowienia jednego sposobu wyrażania się (zachowania) maszyny Turinga. Sensu mojej obronie takiego stanowiska w behawioryzmie, który mówił o całościowym zachowaniu, nadaje książka Skinnera *Verbal Behavior* (New York: Appleton-Century-Crofts, 1957), czyli *Zachowanie werbalne*, w której stworzył analizę konceptualną, zajmującą się zachowaniem werbalnym i świadomością. Por. Blackman 2000, ss. 152-154.

powodują (jeżeli z ich punktu widzenia mówienie o „stanach psychicznych” jest w ogóle możliwe).<sup>125</sup>

Oczywistą rzeczą jest, że z naszego punktu widzenia twierdzenie, że ta konkretna maszyna posiada pragnienia czy przekonania, jest li tylko czystą fantasmagorią bądź żartem (przynajmniej wtedy, gdy bierzemy pod uwagę automat do sprzedawania *Coca Coli*). W przyszłości jednak może okazać się, iż podobnego rodzaju żarty trzeba będzie potraktować całkowicie poważnie. I choć nie jest to argumentem przeciw wyraźnie widocznej nonsensowności tak sformułowanego funkcjonalizmu, prezentowanego przecież przez tak zwaną silną wersję Sztucznej Inteligencji (*strong AI*), to nie można jeszcze dziś wykluczyć, że maszyny, które będą myśleć, będą także posiadać przekonania, a nawet w swoim czasie pragnienia! Jedyna rzecz, która zatem może okazać się nonsensowna, biorąc powyższe pod uwagę, może wypływać z braku precyzji naszego języka, a mówiąc ściślej, z braku precyzji języka *psychologicznego*. Przecież weźmy pod uwagę fakt, że mówienie o myślących maszynach przed zdefiniowaniem przez Turinga tego, co to jest *myślenie* (oczywiście myślenie w sensie Turinga), mogło wydawać się, i do dziś niektórym tak się wydaje<sup>126</sup>, nonsensowne. Zaś odpowiednie zdefiniowanie wyklucza tę nonsensowność, gdy tylko wykaże się, że jest takie *X*, które tę definicję wypełnia — realizuje to, co z definicji realizować miało. Zatem: **stwierdzenie, że coś wydaje się trudne do pojęcia** (czy też w dzisiejszym aparacie pojęciowym — nonsensowne) **nie wyklucza tego, że jest to prawdą!**<sup>127</sup> Na końcu warto też zauważyć jedną rzecz: test Turinga nie tylko pozwala orzec, czy dana maszyna myśli, ale pozwala także na to, że jej formy myślenia mogą być wyżej ocenione przez ewentualnych sędziów niż myślenie samego człowieka (tak będzie wówczas, gdy sędzia (*C*) pomyli się na korzyść maszyny, ze szkodą dla człowieka). A to może być ważny krok przy próbach skategoryzowania inteligencji wyższego stopnia niżli inteligencja człowieka.

\*\*\*

Podsumowując: wszystko to, co zostało powiedziane o maszynie przechodzącej grę w imitację, zostało tym samym powiedziane o maszynie myślącej w sensie Turinga<sup>128</sup>. Zatem:

**O maszynie myślącej w sensie Turinga będziemy mówić, gdy:**

1. będzie porozumiewać się z dowolnym *X* będącym człowiekiem. Przy czym porozumiewanie będzie odbywać się w języku pisanim;
2. porozumienie z dowolnym *X* nie będzie wychodzić poza język pisany tak, że nawet nie będzie przejawiać się w mowie, jeżeli będzie to miało zdecydowanie negatywny wpływ na wynik przeprowadzanego eksperymentu;
3. porozumienie zakłada, że strony będą się rozumieć, tzn. 1) zakłada, że język ten musi być rozumiany przez obydwie zainteresowane strony oraz 2) rozumienie przez maszynę będzie oznaczać: sensowne generowanie zdania/bądź zdań *x*, na poprzedzające je zdanie/bądź zdania *y*, kierowane ze strony dowolnego *X*, będącego człowiekiem.

<sup>125</sup> Lillegard 1995, ss. 44-45.

<sup>126</sup> Por. Szumakowicz 2000.

<sup>127</sup> Historia nauki zna przypadki, gdy coś wydawało się nonsensowne: teoria względności Alberta Einsteina i paradoks bliźniaków; superpozycja w mechanice kwantowej, w słynnym eksperymencie kota Erwina Schrödingera itd.

<sup>128</sup> Oczywiście odróżniam tu pojęcia „maszyny Turinga” od „maszyny myślącej w sensie Turinga”. Por. to, co już napisałem o maszynie Turinga (ss. 36-37) z powyższym.





Wyżej rozpisane trzy punkty w zasadzie nie stwierdzają niczego innego niż sam test Turinga, określają jednak bardzo ważną rzecz, a mianowicie to, że

**żadna inna funkcja przypisywana człowiekowi nie musi być spełniana przez maszynę myślącą w sensie Turinga, jeżeli wychodzi poza powyższe!**

Jeżeli test będzie zdany pomyślnie, śmiało powinniśmy przypisać maszynie zdolność myślenia – oczywiście myślenia w sensie wyżej nakreślonym. Jeżeli jednak ktoś nie zaakceptuje owych trzech punktów, to oczywistym jest, że nie będzie skłonny przypisywać funkcji myślenia maszynie (nawet wtedy, gdy test owo zda pomyślnie).

Druga poł. 2001