Tytuł: **Maszyny liczące a inteligencja** [Computer Machinery and Intelligence, 1950]

Autor: Alan Mathison Turing

Za: Maszyny matematyczne i myślenie, red. E. Feigenbaum & J. Feldman, PWN, Warszawa 1972

Tłumaczenie: D. Gajkowicz

Źródło: http://www.kognitywistyka.net / mjkasperski@kognitywistyka.net

1. Gra w naśladownictwo

Proponuję rozważyć problem: „Czy maszyny mogą myśleć". Pracę nad tym zagadnieniem należy rozpocząć od zdefiniowania znaczenia terminów: 'maszyna' i 'myśleć'. Definicje mogłyby być tak zbudowane, aby odzwierciedlały taka dalece jak to jest możliwe potoczne znaczenie tych słów. Jednakże takie stanowisko jest niebezpieczne. Gdybyśmy znaczenie słów 'maszyna' i 'myśleć' mieli ustalić na drodze zbadania, w jaki sposób są one powszechnie stosowane, to trudno byłoby uzasadnić, że znaczenie pytania „Czy maszyny mogą myśleć" oraz odpowiedzi na to pytanie nie należy szukać na drodze pomiarów statystycznych, takich jak ankieta. Ale to absurd. Zamiast próby zbudowania takiej definicji, powyższy problem zastąpię innym bezpośrednio związanym z nim problemem, który wyrażę przy pomocy stosunkowo niedwuznacznych słów.

Nową postać problemu można opisać przy pomocy 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 oddzielonym 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*". Pytającemu wolno zadawać pytania A i B w ten sposób:

C: Proszę *X*, aby mi powiedział jak długie ma włosy? Teraz przypuśćmy, że *X* jest faktycznie *A*, wobec tego *A* musi odpowiedzieć. Celem *A* w grze jest dołożenie wszelkich starań, aby *C* źle go zidentyfikował. Wobec tego jego odpowiedź mogłaby być następująca:

„Moje włosy są ostrzyżone, a najdłuższe kosmyki mają około dziewięć cali długości".

Aby brzmienie głosu nie mogło pomóc pytającemu w dokonaniu identyfikacji, odpowiedzi powinny być pisane odręcznie, a jeszcze lepiej na maszynie. Idealnym środkiem porozumiewania się między pokojami jest dalekopis. Pytania i odpowiedzi mogą być też przekazywane przez pośrednika. Zadaniem trzeciego gracza w tej grze jest udzielanie pomocy pytającemu. Prawdopodobnie najlepszą dla tej osoby strategią jest udzielanie odpowiedzi zgodnych z prawdą. Może ona do swoich odpowiedzi dodawać takie rzeczy, jak: „Jestem

kobietą, nie słuchaj go", ale to nie przyniesie żadnej korzyści, ponieważ mężczyzna może robić podobne uwagi.

Teraz zapytujemy się: „Co stanie się, gdy maszyna zastąpi A w tej grze?". Czy pytający będzie decydował błędnie tak samo często jak wtedy, gdy w grze bierze udział mężczyzna i kobieta? Pytania te zastąpią nasze pytanie początkowe" „Czy maszyny mogą myśleć?".

2. Ocena nowego problemu

Nie tylko można zapytać: „Jakie jest rozwiązanie tej nowej postaci problemu", ale również: „Czy warto badać ten nowy problem?". Na to ostatnie pytanie odpowiemy od razu, aby nie wracać do niego więcej.

Zaletą nowego problemu jest ostre rozgraniczenie między fizycznymi i intelektualnymi możliwościami człowieka. Żaden inżynier ani chemik nie twierdzi, że potrafi wyprodukować materiał, który niczym by się nie różnił od skóry ludzkiej. Możliwe, że kiedyś można będzie to zrobić, ale nawet gdybyśmy rozporządzali takim wynalazkiem, to i tak nie miałoby większego sensu usiłowanie ubrania myślącej maszyny w takie sztuczne ciało w celu uczynienia jej bardziej ludzką. To nasze przekonanie znajduje odbicie w sposobie postawienia problemu, a mianowicie w postaci zakazu, który nie pozwala pytającemu widzieć, dotykać i słyszeć pozostałych uczestników gry. Niektóre inne zalety proponowanego kryterium można pokazać na przykładzie pytań i odpowiedzi. A zatem:

P: Napisz mi sonet na temat Forth Bridge.

O: Nie licz na mnie. Nigdy nie umiałem pisać wierszy.

P: Dodaj 34 957 do 70 764.

O: (Po 30-sekundowym namyśle odpowiada) 105 621.

P: Czy grasz w szachy?

O: Tak.

P: Mam K na K1 i innych figur nie mam. Ty masz tylko K na K6 i R na R1. Jest twój ruch. Jakie zrobisz posunięcie?

O: (Po 15-sekundowym namyśle) R-R8 mat.

Wydaje się, że metoda pytań i odpowiedzi nadaje się do wprowadzenia do prawie każdej dziedziny ludzkiej działalności, do której chcemy ją wprowadzić. Nie możemy winić maszyny za jej niezdolność do zwycięstwa we współzawodnictwie z człowiekiem ani winić człowieka za przegraną w wyścigu z samolotem. Dzięki warunkom naszej gry, te niemożliwości stają się nieistotne. „Świadkowie" mogą przechwalać się jeśli uważają to za wskazane, tyle ile im się podoba, gdy idzie o ich wdzięk, siłę lub bohaterstwo, ale pytający nie może zażądać praktycznych demonstracji.

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 robiłoby to bardzo złe wrażenie. Skompromitowałby się od razu swoją powolnością i

niedokładnością w arytmetyce. Czy maszyny nie mogą wykonywać czegoś, co należałoby nazwać myśleniem, ale co różni się zupełnie od myślenia człowieka? Ten zarzut jest bardzo mocny, ale nie potrzebujemy się nim przejmować, jeśli tylko mimo wszystko można będzie zbudować maszynę tak, aby grała zadowalająco w grę w imitację.

Można by argumentować, że przy rozgrywce „gry w imitację" najlepsza strategia dla maszyny mogłaby być cokolwiek inna, niż naśladownictwo zachowania się człowieka. Być może jest tak, ale myślę, że istnienie jakiegoś większego obiektu tego rodzaju jest nieprawdopodobne. W żadnym przypadku nie mam zamiaru rozpatrywać tutaj teorii gry i założę, że najlepszą strategią jest usiłowanie dawania takich odpowiedzi, które w sposób naturalny byłyby dawane przez człowieka.

3. Maszyny uczestniczące w grze

Problem, jaki postawiliśmy w punkcie 1, nie będzie w pełni określony dotąd, aż podamy znaczenie słowa „maszyna". Naturalne jest, że powinniśmy pozwolić na zastosowanie w naszych maszynach każdego rodzaju techniki inżynierskiej. Pragniemy również dopuścić możliwość skonstruowania przez inżyniera lub zespół inżynierów maszyny, która pracuje, ale której sposobu działania konstruktorzy nie mogą wystarczająco opisać, ponieważ przy tej konstrukcji zastosowali metodę w dużym stopniu eksperymentalną. W końcu, chcemy, aby do maszyn nie byli zaliczani ludzie urodzeni w zwykły sposób. Trudno jest obmyślać definicje tak, aby spełniały te trzy warunki. Można by na przykład nalegać, aby zespół inżynierów był jednej płci, ale to naprawdę nie byłoby wystarczające, ponieważ prawdopodobnie jest możliwe zbudowanie kompletnego indywiduum z pojedynczej komórki (powiedzmy) skóry człowieka. Dokonanie tego byłoby wyczynem biologicznej techniki, zasługującym na najwyższą pochwałę, ale nie bylibyśmy skłonni uważać go za przypadek „budowania myślącej maszyny". To skłania nas do zrezygnowania z wymagania, dotyczącego dopuszczalności każdego rodzaju techniki. Jesteśmy tym bardziej gotowi zrezygnować z tego warunku z uwagi na fakt, że obecne zainteresowanie „maszynami myślącymi" powstało dzięki specjalnemu rodzajowi maszyny, zazwyczaj nazywanej „elektroniczną maszyną cyfrową" lub „maszyną cyfrową". Idąc za tą myślą, w naszej grze pozwolimy brać udział tylko maszynom cyfrowym.

Ograniczenie to, na pierwszy rzut oka, wydaje się bardzo drastyczne. Spróbuję pokazać, że w rzeczywistości tak nie jest. W tym celu trzeba krótko wyjaśnić naturę i własności tych maszyn.

Można również powiedzieć, że ta identyfikacja maszyn z maszynami cyfrowymi, analogicznie do naszego kryterium „myślenia" będzie niewystarczająca tylko wtedy, gdy (w przeciwieństwie do mojego przekonania) okaże się, że maszyny cyfrowe nie potrafią wypaść dobrze w grze.

Mamy już w eksploatacji pewną liczbę maszyn cyfrowych i można zapytać: „dlaczego by nie przeprowadzić eksperymentu natychmiast? Łatwo byłoby spełnić warunki gry. Można by eksperymentować z pewną ilością pytających i opracować statystykę, pokazującą częstość występowania prawidłowych identyfikacji". Krótka odpowiedź jest następująca: nie pytamy czy wszystkie maszyny cyfrowe wypadłyby dobrze w grze, ani czy obecnie dostępne maszyny cyfrowe wypadłyby dobrze, ale czy są do pomyślenia maszyny cyfrowe, które z

powodzeniem brałyby udział w grze. Jest to jednak tylko krótka odpowiedź. Później ujrzymy ten problem w innym świetle.

4. Maszyna cyfrowa

Ideę działania maszyn cyfrowych można wyjaśnić, mówiąc, że te maszyny są przeznaczone do przeprowadzania dowolnych operacji, które mogłaby wykonać ludzka maszyna cyfrowa. Wychodzimy z założenia, że ludzka maszyna cyfrowa postępuje według stałych reguł i nie może odbiec od nich w żadnym szczególe. Możemy założyć, że te reguły zawarte są w książce, którą wymienia się za każdym razem, gdy ma być wykonana nowa praca. Dysponuje ona również nieograniczonym zapasem papieru, na którym wykonujemy swoje obliczenia. Może również swoje mnożenia i dodawania wykonywać na arytmometrze, ale ot nie ma znaczenia.

Jeśli powyższe wyjaśnienie potraktujemy jako definicję, to zajdzie niebezpieczeństwo argumentacji w kółko. Niebezpieczeństwa tego unikniemy, nakreślając sposoby, przy pomocy których uzyskuje się żądany efekt. Zazwyczaj można uważać, że maszyna cyfrowa składa się z trzech części: pamięci, jednostki wykonawczej, sterowania.

Pamięć jest magazynem informacji i odpowiada papierowi ludzkiej maszyny cyfrowej, przy czym jest to bądź papier, na którym wykonuje ona swoje obliczenia, bądź papier, na którym wydrukowano jej książkę reguł. W tym stopniu, w jakim ludzka maszyna cyfrowa wykonuje obliczenia w swojej głowie, część pamięci maszyny cyfrowej będzie odpowiadała pamięci ludzkiej.

Jednostka wykonawcza przeprowadza różnorodne pojedyncze operacje, zawarte w obliczeniu. Te pojedyncze operacje będą różne dla różnych maszyn. Zwykle można wykonywać dosyć długie operacje, takie jak: „Pomnóż 3 540 675 445 przez 7 076 345 687”, ale w niektórych maszynach dopuszczalne są tylko bardzo krótkie operacje, takie jak: „Zapisz 0”.

Wspomnieliśmy już, że dostarczona maszynie cyfrowej „książka reguł” zostaje zastąpiona w maszynie przez część jej pamięci. Nazywa się wtedy „tablicą instrukcji”. Zadaniem sterowania jest dopilnowanie, aby te instrukcje były wykonywane prawidłowo i we właściwej kolejności. Sterowanie jest tak zbudowane, aby dobrze wywiązywało się z tego zadania.

Informacja w pamięci zazwyczaj jest podzielona na średniej wielkości paczki. W jednej maszynie, na przykład, paczka może składać się z dziesięciu cyfr dziesiętnych. Liczby przydzielane są w pewien systematyczny sposób częściom pamięci, w której znajdują się różne paczki informacji. Typowa instrukcja mogłaby mówić: „Dodaj liczbę, znajdującą się pod adresem 6809 do liczby, znajdującej się pod adresem 4302 i wynik prześlij z powrotem pod ten ostatni adres w pamięci”.

Nie trzeba dodawać, że w maszynie taka instrukcja nie jest wyrażona np. w języku angielskim. Z większym prawdopodobieństwem mogłaby ona być zakodowana w następującej postaci: 6 809 430 217. Tutaj liczba 17 wyznacza (z pośród wielu możliwych) operację, która ma być wykonana na tych dwóch liczbach. W tym wypadku operacja jest taka, jak opisano wyżej, a mianowicie” „Dodaj liczbę...”. Zauważmy, że instrukcja zajmuje miejsce 10 cyfr i w ten bardzo wygodny sposób tworzy jedną paczkę informacji. Normalnie sterowanie będzie pobierało instrukcje do wykonania po kolei według adresów, pod którymi

są one zapamiętane, ale od czasu do czasu można spotkać instrukcję taką, jak: „Teraz wykonaj instrukcję zapamiętaną pod adresem 5606 i kontynuuj pracę od tego miejsca”, albo: „Jeśli komórka 4505 zawiera 0, to następnie wykonaj instrukcję, znajdującą się w komórce 6707, w przeciwnym razie kontynuuj pracę po kolei”. Te ostatnie rodzaje instrukcji są bardzo ważne, ponieważ umożliwiają wielokrotne powtarzanie sekwencji operacji, aż do chwili spełnienia pewnego warunku, przy czym nie muszą przy tym być wykonywane ciągle nowe instrukcje, lecz stałe te same od początku. Weźmy pod uwagę rodzinną analogię. Przypuśćmy, że matka chce, aby Tomek wstępował do szewca codziennie rano po drodze do szkoły, aby dowiedzieć się, czy jej buty są zreperowane, więc może ona prosić go o to codziennie rano na nowo. Albo może ona raz wywiesić w hallu zawiadomienie dla wszystkich, które Tomek zobaczy, gdy będzie wychodził do szkoły i które powie mu, aby wstąpił po buty i aby zniszczył zawiadomienie, gdy wróci z butami.

Czytelnik musi przyjąć to jako fakt, że można budować maszyny cyfrowe, że rzeczywiście zostały one zbudowane według opisanych przez nas zasad i, że faktycznie mogą one bardzo dokładnie naśladować działania ludzkiej maszyny cyfrowej.

Książka reguł, którą, jak pisaliśmy, stosuje ludzka maszyna cyfrowa jest, oczywiście, wygodną fikcją. W rzeczywistości prawdziwie ludzkie maszyny cyfrowe pamiętają, co muszą zrobić. Jeśliby ktoś chciał, aby maszyna przy wykonywaniu pewnej złożonej operacji naśladowała zachowanie ludzkiej maszyny cyfrowej, to należy zapytać człowieka, jak on to robi i następnie odpowiedź przetłumaczyć na język maszynowy i przedstawić w postaci tablicy instrukcji. Budowanie tablicy instrukcji nazywamy zwykle „programowaniem”. „Zaprogramowanie maszyny tak, aby przeprowadziła operację *A*” oznacza wprowadzenie odpowiedniej tablicy instrukcji do maszyny w ten sposób, że ona wykona *A*.

Interesującym wariantem idei maszyny cyfrowej jest „maszyna cyfrowa z elementem przypadkowym”. Takie maszyny posiadają instrukcje, wymagające rzucenia kości lub jakiegoś równoważnego procesu elektronicznego; jedna z tego rodzaju instrukcji mogłaby być na przykład następująca: „Rzuć kości i otrzymaną liczbę prześlij pod adres 1000”. Czasami pisze się, że taka maszyna ma wolną wolę (chociaż ja sam tego nie powiedziałbym). Normalnie z obserwacji maszyny nie można określić, czy zawiera ona element przypadkowy, ponieważ podobny efekt mogą dawać takie urządzenia jak uzależnienie wyboru cyfr od ułamków dziesiętnych π.

Większość obecnych maszyn cyfrowych posiada tylko skończoną pamięć. Idea maszyny cyfrowej z nieograniczoną pamięcią nie przedstawia żadnych teoretycznych trudności. Oczywiście, w określonym czasie można używać tylko skończoną część pamięci. Podobnie można zbudować tylko pamięć o skończonej wielkości, ale możemy sobie wyobrazić, że w miarę potrzeby można dodawać jej więcej. Takie maszyny są specjalnie interesujące z teoretycznego punktu widzenia i będziemy je nazywać maszynami cyfrowymi o nieskończonej wielkiej pamięci.

Idea maszyny cyfrowej jest stara. Charles Babbage, profesor matematyki w Cambridge w latach 1828-1839, zaprojektował taką maszynę, zwaną Maszyną Analityczną, ale nigdy jej nie ukończył. Chociaż Babbage dysponował wszystkimi podstawowymi ideami, to jednak jego maszyna nie była w owym czasie dużą atrakcją. Jej szybkość byłaby zdecydowanie większa niż szybkość ludzkiej maszyny cyfrowej, ale około 100 razy mniejsza niż szybkość maszyny Manchester, która jest jedną z wolniejszych nowoczesnych maszyn. Pamięć miała być czysto mechaniczna, zbudowana z kółek i kart.

Fakt, że Maszyna Analityczna Babbage'a miała być całkowicie mechaniczna, pomoże nam pozbyć się pewnego przesądu. Dużą wagę często przywiązuje się faktu, że nowoczesne maszyny cyfrowe są elektryczne i że system nerwowy jest także elektryczny. Ponieważ maszyna Babbage'a nie była elektryczna i ponieważ wszystkie maszyny cyfrowe są w pewnym sensie równoważne, wobec tego widzimy, że stosowanie elektryczności nie może mieć teoretycznego znaczenia. Naturalnie, elektryczność zwykle wkracza tam gdzie wchodzi w grę szybka sygnalizacja. Nie jest więc dziwne, że znajdujemy ją w obu tych dziedzinach. W systemie nerwowym zjawiska chemiczne są co najmniej tak samo ważne jak elektryczne. W pewnych maszynach cyfrowych system pamięci jest głównie akustyczny. Widać wobec tego, że stosowanie elektryczności jest bardzo powierzchownym podobieństwem. Gdybyśmy chcieli znaleźć takie podobieństwa, to powinniśmy szukać raczej analogii matematycznych.

5. Uniwersalność maszyn cyfrowych

Maszyny cyfrowe, o których mówiliśmy w poprzednim paragrafie, można zaliczyć do „maszyn o stanach dyskretnych". Są to maszyny, które przechodzą gwałtownymi skokami z jednego zupełnie określonego stanu do innego. Stany te na tyle różnią się od siebie, że można pominąć możliwość pomylenia ich między sobą. Ściśle mówiąc nie ma takich maszyn. Naprawdę, wszystko zmienia się w sposób ciągły. Ale istnieje wiele rodzajów maszyn, o których z powodzeniem można *myśleć* jako o maszynach o stanach dyskretnych. Na przykład, przy rozpatrywaniu wyłączników światła, wygodną fikcją jest stwierdzenie, że każdy wyłącznik musi być definitywnie otwarty albo zamknięty. Wprawdzie muszą istnieć położenia pośrednie, ale dla większości celów możemy to pominąć. Jako przykład maszyny o stanach dyskretnych mogłoby służyć koło, które jest zatrzymywane zapadką raz na sekundę po obrocie o 120°, ale może być również zatrzymane przez dźwignię, którą można sterować z zewnątrz; ponadto w jednej z pozycji koła świeci lampa. Tę maszynę można opisać abstrakcyjnie w sposób następujący. Wewnętrzny stan maszyny (określony położeniem koła) może być q_1, q_2 lub q_3. sygnałami wejściowymi są: i_0 lub i_1 (położenie dźwigni). W każdej chwili stan wewnętrzny jest określony przez stan poprzedni i sygnał wejściowy zgodny z zestawieniem:

	q_1	q_2	q_3
i_0	q_2	q_3	q_1
i_0	q_1	q_2	q_3

Tablica określa sygnały wyjściowe – jedynie widoczne na zewnątrz wskazanie wewnętrznego stanu (światło):

stan	q_1	q_2	q_3
wyjście	o_0	o_0	o_0

Ten przykład jest typowy dla maszyn o dyskretnych stanach. Można je opisywać takimi tablicami pod warunkiem, że mają one tylko skończoną liczbę możliwych stanów.

Wydawałoby się, że mając dany stan początkowy maszyny i sygnały wejściowe zawsze można przewidzieć wszystkie przyszłe stany. Jest to reminiscencja poglądu Laplace'a, który sądził, że na podstawie znajomości kompletnego stanu wszechświata w jednej chwili,

opisanego położeniami i szybkościami wszystkich jego cząstek powinno być możliwe przewidzenie wszystkich przyszłych stanów. Jednakże rozważane przez nas przewidywanie jest znacznie bliższe praktyki niż przewidywanie Laplace'a. System „wszechświata jako całości" ma taką właściwość, że całkiem małe błędy występujące w warunkach początkowych mogą wywierać decydujący wpływ w czasie późniejszym. Przesunięcie w pewnej chwili pojedynczego elektronu o bilionową część centymetra mogłoby spowodować tak wielką różnicę, jak różnica występująca między człowiekiem zabitym przez lawinę rok później lub wychodzącym z wypadku cało. Podstawową własnością systemów mechanicznych, zwanych „maszynami o stanach dyskretnych" jest niewystępowanie tego zjawiska. Nawet, gdy rozpatrujemy prawdziwe fizyczne maszyny zamiast maszyn wyidealizowanych, to odpowiednio dokładna znajomość stanu w danej chwili, daje odpowiednio dokładną znajomość każdej liczby stanów następnych.

Jak wspominaliśmy, maszyny cyfrowe należą do klasy maszyn o dyskretnych stanach. Jednak, liczba możliwych dla takiej maszyny stanów jest ogromna. Na przykład: maszyna obecnie pracująca w Manchesterze posiada około $2^{165\,000}$, to znaczy około $10^{50\,000}$ możliwych stanów. Porównajmy tę liczbę z trzema stanami opisanego wyżej przykładowo koła zapadkowego. Nietrudno zorientować się dlaczego liczba stanów maszyny cyfrowej musi być tak ogromna. Maszyna cyfrowa posiada pamięć, która stanowi odpowiednik papieru, używanego przez ludzką maszynę cyfrową. Musi być możliwe zapisanie w pamięci każdej kombinacji symboli, które można byłoby zapisać na papierze. Dla uproszczenia załóżmy, że jako symboli używamy tylko cyfr od 0 do 9. ręcznie pisane warianty pomijamy. Przypuśćmy, że maszyna cyfrowa może zapamiętać 100 kartek papieru, z których każda zawiera 50 linii, a na każdej linii mieści się 30 cyfr. Wtedy ilość stanów wynosi $10^{100 \cdot 50 \cdot 30}$, to jest $10^{150\,000}$. jest to liczba stanów trzech Manchesterskich maszyn razem wziętych. „Pojemnością pamięci" maszyny zazwyczaj nazywamy logarytm o podstawie 2 liczby stanów. Tak więc, maszyna z Manchesteru posiada pojemność pamięci około 165 000, a nasza przykładowa maszyna z kołem zapadkowym około 1,6. Jeśli dwie maszyny zestawia się razem, to pojemność zestawu tych dwóch maszyn jest sumą pojemności maszyn składowych. To prowadzi do możliwości formułowania takich twierdzeń, jak: „Maszyna Manchesterska zawiera 64 ścieżki magnetyczne, z których każda ma pojemność 2560, osiem lamp elektronowych o pojemności 1280 każda. Razem z innymi pamięciami o łącznej pojemności 300 ogólna pojemność pamięci maszyny w Manchester wynosi 174 380".

Działanie maszyny o stanach dyskretnych można przewidzieć na podstawie odpowiadającej jej tablicy. Nie ma powodu, dla którego nie można by tego rachunku przeprowadzić na maszynie cyfrowej. Maszyna cyfrowa mogłaby naśladować zachowanie się każdej maszyny o dyskretnych stanach pod warunkiem, że odbywałoby się to wystarczająco szybko. W takim razie, grę w imitację można byłoby rozgrywać ze wspomnianą maszyną (jako *B*) i z naśladującą ją maszyną cyfrową (jako *A*), a pytający nie potrafiłby rozróżnić ich. Naturalnie, maszyna cyfrowa musiałaby mieć zarówno odpowiednią pojemność pamięci jak i musiałaby pracować dostatecznie szybko. Ponadto, trzeba byłoby ją programować na nowo dla każdej nowej maszyny, którą miałaby naśladować.

Tę specjalną własność maszyn cyfrowych, polegającą na możliwości naśladowania każdej maszyny o dyskretnych stanach, opisujemy mówiąc, że maszyny cyfrowe są maszynami *uniwersalnymi*. Ważną konsekwencją, wypływającą z faktu istnienia maszyn, posiadających taką własność jest – pomijając kwestię szybkości – brak potrzeby projektowania różnych procesów obliczeniowych. Wszystkie te procesy można przeprowadzić na jednej maszynie

cyfrowej, odpowiednio w każdym przypadku zaprogramowanej. Zobaczymy, że w konsekwencji tego faktu wszystkie maszyny cyfrowe są w pewnym sensie równoważne.

Możemy teraz ponownie rozważyć myśl wysuniętą przy końcu punktu 3. Sugerowaliśmy tytułem próby, że pytanie „Czy maszyny mogą myśleć?" należy zastąpić pytaniem: „Czy są możliwe maszyny cyfrowe, które wypadłyby dobrze w grze w imitację?". Możemy zrobić to w sposób, zdawałoby się, ogólniejszy i zapytać: „Czy istnieją maszyny o dyskretnych stanach, które wypadłyby dobrze w tej grze?". Ale, z punktu widzenia uniwersalności, widzimy, że każde z tych pytań jest równoważne następującemu: „Ustalmy naszą uwagę na jednej specyficznej maszynie cyfrowej *C*. Czy jest prawdą, że modyfikując tę maszynę tak, aby miała odpowiednią pamięć, odpowiednio powiększając jej szybkość działania i dostarczając jej odpowiedni program, można spowodować, aby maszyna *C* grała zadowalająco rolę *A* w grze w naśladownictwo, przy czym rolę *B* grałby człowiek?".

6. Przeciwne poglądy na temat zasadniczego pytania

Aby wyjaśnić nasze stanowisko możemy teraz zastanowić się nad podłożem naszego problemu i wtedy będziemy gotowi do kontynuowania rozważań na temat naszego pytania: „Czy maszyny mogą myśleć?" oraz na temat jego wariantu, cytowanego przy końcu ostatniego punktu. Nie możemy całkowicie porzucić początkowej postaci problemu, ponieważ na temat poprawności dokonanego zastąpienia będą istniały poglądy niezgodne z naszymi i musimy przynajmniej wziąć pod uwagę to, co w związku z tym powiedziano.

Zrozumienie istoty rzeczy stanie się łatwiejsze dla Czytelnika, jeśli najpierw wyjaśnię moje własne przekonanie w tek kwestii. Rozważmy najpierw bardziej ścisłą postać pytania. Wierzę, że za około pięćdziesiąt lat stanie się możliwe programowanie maszyn cyfrowych o pojemności pamięci rzędu 10^9 tak, aby grały w grę w naśladownictwo tak dobrze, że przeciętny pytający po pięciu minutach zadawania pytań nie będzie miał więcej niż 70 procent szansy dokonania prawidłowej identyfikacji. Jestem przekonany, że pierwotne pytanie: „Czy maszyny mogą myśleć?" oznacza zbyt mało, aby zasługiwało na dyskusję. Niemniej, wierzę, że pod koniec tego stulecia używanie słów i ogólna opinia ludzi wykształconych zmieni się tak bardzo, że można będzie mówić o maszynach myślących, nie spodziewając się sprzeciwu. Jestem przekonany ponadto, że zatajenie tych przekonań nie służyłoby żadnemu pożytecznemu celowi. Zupełnie błędny jest rozpowszechniony pogląd, że naukowcy poruszają się nieubłaganie od dobrze ustalonego faktu do dobrze ustalonego faktu, nigdy nie korzystając w swojej pracy z żadnych postępowych przypuszczeń. Jeśli tylko wiadomo, które fakty są udowodnione, a które są jedynie przypuszczeniami, to żadna szkoda nie może z tego postępowania wyniknąć. Przypuszczenia mają znaczenie, gdyż sugerują użyteczne linie badań.

Obecnie przejdę do rozważenia poglądów odmiennych od moich własnych.

a. Sprzeciw teologiczny

Myślenie jest funkcją nieśmiertelnej duszy człowieka. Bóg dał nieśmiertelną duszę każdemu mężczyźnie i każdej kobiecie, ale nie dał jej żądnemu innemu stworzeniu ani maszynom. Wobec tego żadne zwierzę, ani żadna maszyna nie może myśleć.

Nie mogę zgodzić się z żadnym powyższym twierdzeniem, ale spróbuję odpowiedzieć na nie w terminach teologicznych. Sądziłbym, że argument byłby bardziej przekonujący, gdyby zwierzęta zakwalifikowano do jednej grupy razem z ludźmi, ponieważ, moim zdaniem, większa różnica istnieje między typowym stworzeniem żywym a tworem nieożywionym, niż między człowiekiem a innymi zwierzętami. Dowolny charakter ortodoksyjnego poglądu stanie się jaśniejszy, jeśli rozważymy jak mógłby on przedstawić się członkowi jakiejś innej społeczności religijnej. Dlaczego chrześcijanie odrzucili muzułmański pogląd, że kobiety nie mają dusz? Ale odłóżmy tę kwestię na bok i powróćmy do głównego argumentu. 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? Moglibyśmy oczekiwać, że użyłby On swojej siły w połączeniu z mutacją, która dałaby słoniowi odpowiednio ulepszony mózg do służenia potrzebom tej duszy. 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. Wspomniane warunki zostaną przedyskutowane w pozostałej części tego artykułu. Usiłując zbudować takie maszyny nie powinniśmy bez szacunku uzurpować sobie Jego mocy tworzenia dusz; nasza zasługa nie jest większa niż przy płodzeniu dzieci: w każdym przypadku jesteśmy raczej narzędziami Jego woli, dostarczającymi siedzib dusz, które On tworzy.

Jednakże jest to tylko spekulacja. Teologiczne argumenty nie wywierają na mnie głębokiego wrażenia, jakkolwiek można je pomocniczo stosować. Stwierdzono, że 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.

b. Sprzeciw „głów w piasku"

„Konsekwencje myślenia maszyn byłyby zbyt okropne. Miejmy nadzieję i wierzmy, że one nie mogą myśleć". Ten argument rzadko jest wyrażany tak otwarcie. Ale działa on na większość z nas, którzy w ogóle o tym myślimy. Chcemy wierzyć, że Człowiek jest w jakiś subtelny sposób wyższy ponad resztę stworzenia. Najlepiej byłoby, gdyby można było wykazać, że jest on *bezwarunkowo* wyższy, ponieważ wówczas nie istniałoby niebezpieczeństwo utraty jego dominującej pozycji. Popularność argumentu teologicznego jest wyraźnie związana z tym uczuciem. Uczucie to może być bardzo silne w ludziach intelektu, ponieważ oni cenią potęgę myślenia wyżej niż inni i są bardziej skłonni do oparcia swojej wiary w wyższość człowieka na tej potędze.

Nie myślę, że ten argument jest wystarczająco poważny, aby trzeba go było zbijać. Pociecha byłaby bardziej odpowiednia: być może powinno się jej szukać w wędrówce dusz.

c. Sprzeciw matematyczny

Opierając się na pewnych wynikach logiki matematycznej można wykazać, że istnieją granice możliwości maszyn o stanach dyskretnych. Najlepiej znanym z tych wyników jest

twierdzenie Gödela (1931), który pokazuje, 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. Istnieją inne, pod pewnymi względami podobne wyniki, które zawdzięczamy Churchowi (1936), Kleene'owi (1935), Rosserowi i Turingowi (1937). Ten ostatni wynik jest najwygodniejszy do rozpatrywania, ponieważ odnosi się bezpośrednio do maszyn, podczas gdy inne można stosować tylko w stosunkowo pośrednim argumencie: np., gdybyśmy chcieli zastosować twierdzenie Gödela to musielibyśmy poza tym podać jakieś sposoby opisu systemów logicznych w terminach maszyn i maszyn w terminach systemów logicznych. Wspomniany wynik odnosi się do pewnego rodzaju maszyny, która jest zasadniczo maszyną cyfrową o nieograniczonej pamięci. Stwierdza on, że istnieją pewne rzeczy, których taka maszyna nie może zrobić. Jeśli ze względu na swoją konstrukcję maszyna jest przeznaczona do odpowiadania na pytania jak w grze w naśladownictwo, to będą istniały takie pytania, na które udzieli ona błędnej odpowiedzi, bądź nie da w ogóle odpowiedzi bez względu na dany jej na odpowiedź czas. Naturalnie może być dużo takich pytań, przy czym na pytania, na które jedna maszyna nie potrafi odpowiedzieć, inna maszyna może odpowiedzieć w sposób zadawalający. Oczywiście na razie zakładamy, że pytania są tego rodzaju, że wymagają odpowiedzi „tak" lub „nie" zamiast takich pytań, jak: „Co myślisz o Picassie?". Wiemy, że maszyny nie potrafią odpowiadać na tego rodzaju pytania: „Weź pod uwagę maszynę określoną jak następuje... Czy ta maszyna odpowie kiedykolwiek „tak" na jakieś pytanie?". Kropki należy zastąpić standardowym opisem jakieś maszyny, który mógłby być czymś w rodzaju opisu zastosowanego w punkcie 5. gdy między opisaną maszyną, a maszyną, której zadajemy pytania występuje jakaś względnie prosta relacja, to można pokazać, że odpowiedź jest albo błędna, albo nie nadchodzi wcale. Jest to wynik matematyczny: argumentuje się, że to dowodzi niezdolności maszyn, dla których ludzki intelekt nie jest odpowiednim przedmiotem badań.

Krótką ripostą na ten argument jest to, że chociaż 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 odpowiednie 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 niekłamane, 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 omylności części maszyn. Ponadto naszą wyższość z takiego powodu możemy odczuwać jedynie w związku z jedną maszyną, nad którą uzyskaliśmy nasz drobny triumf. Nie występowałaby kwestia jednoczesnego triumfowania nad *wszystkimi* maszynami. Tak więc, krótko mówiąc, mogliby być ludzie zdolniejsi od każdej danej maszyny, ale i z kolei mogłyby być inne zdolniejsze maszyny itd.

Myślę, że ci, którzy obstają przy argumencie matematycznym na ogół przyjęliby grę w naśladownictwo za podstawę dyskusji. Tych, którzy wierzą w dwa poprzednie argumenty prawdopodobnie nie interesowałyby żadne kryteria.

d. Argument świadomości

Argument ten 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 swego sukcesu, zmartwienia, gdy jej lampy topią się, nie może podniecać się pochlebstwem, cierpieć z powodu swoich błędów, być oczarowanym przez sex, być złym lub przybitym, gdy nie może dosta tego, co chce".

Wydaje się, że ten argument jest zaprzeczeniem słuszności naszego testu. Według krańcowej postaci tego poglądu jedynym sposobem upewnienia się, że maszyna myśli jest *być* maszyną i odczuwać, że się myśli. Można by wtedy opisać te uczucia światu, 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 właśnie człowiekiem. Faktycznie jest to punkt widzenia solipsysty. Może być to najlogiczniejszy pogląd do utrzymania, ale utrudnia komunikację idei. *A* jest przekonany, że „*A* myśli, ale *B* nie myśli", podczas gdy *B* wierzy, że „*B* myśli, ale *A* nie myśli". Zamiast ciągłego spierania się co do tej kwestii, zazwyczaj przyjmuje się grzeczną konwencję, że każdy myśli.

Jestem pewny, że profesor Jefferson nie chciałby przyjąć krańcowego i solipsystycznego punktu widzenia. Prawdopodobnie zechciałby on zaaprobować jako test grę w imitację. Grę (bez gracza *B*) często stosuje się w praktyce pod nazwą *viva voce*. Ma ona na celu przekonanie się, czy ktoś rzeczywiście rozumie coś, czy też „nauczył się tego na pamięć jak papuga". Posłuchajmy fragmentu takiego *viva voce*:

Pytający: Czy w pierwszej linii twojego sonetu, która brzmi: „Czy mam porównać cię do letniego dnia" sformułowanie „wiosenny dzień" nie byłoby tak samo dobre lub lepsze?

Świadek: Nie byłoby ono do rymu.

Pytający: A co myślisz o „dniu zimowym"? To byłoby do rymu.

Świadek: Tak, ale nikt nie chce być porównanym do dnia zimowego.

Pytający: Co byś powiedział na to, gdyby pan Piekwiek przypomniał ci o Bożym Narodzeniu?

Świadek: Nic szczególnego.

Pytający: A jednak Boże Narodzenie jest dniem zimowym i nie myślę, że pan Piekwiek miałby na myśli porównanie poetyckie.

Świadek: Nie sądzę, że mówisz serio. Przez dzień zimowy rozumie się raczej typowy dzień zimowy, niż specjalny dzień, jak Boże Narodzenie.

I tak dalej. Co powiedziałby profesor Jefferson, gdyby maszyna pisząca sonety potrafiła odpowiadać w ten sposób *in viva voce*? Nie wiem, czy uważałby on, że maszyna „jedynie sztucznie sygnalizuje" te odpowiedzi, ale nie sądzę, że opisywałby ją jako „łatwy fortel" gdyby jej odpowiedzi były tak wystarczające i odpowiednie, jak w powyższym ustępie. Myślę, że to jędrne powiedzenie dotyczyło takich urządzeń, jak wprowadzenie do maszyny

zapisu czyjegoś głosu czytającego sonet, z odpowiednim przełącznikiem, który można włączać od czasu do czasu.

Tak więc w skrócie, myślę, że większość tych, którzy popierają argument świadomości, dałaby się raczej skłonić do zaniechania tego argumentu, niż być zmuszona do przejścia na pozycję solipsystyczną. Oni prawdopodobnie zachcieliby przyjąć nasz test.
Nie chciałbym, aby odnosiło się wrażenie, że jestem przekonany o tym, że świadomość nie jest wcale tajemnicza. Na przykład istnieje coś w rodzaju paradoksu, związanego z każdą próbą jej lokalizacji. Ale nie myślę, że te tajemnice koniecznie trzeba rozwiązać zanim będziemy mogli odpowiedzieć na pytanie, którym zajmujemy się w tym artykule.

e. Argumenty wypływające z różnych niemożności

Te argumenty mają następującą postać: „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*”. W związku z tym sugeruje się liczne cechy *X*. Podam niektóre z nich: Być uprzejmym, pomysłowym, pięknym, przyjacielskim, mieć inicjatywę, mieć zmysł humoru, odróżnić 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.

Zazwyczaj niczym nie popiera się tych twierdzeń. Wierzę, że najczęściej znajduje się je na zasadzie indukcji naukowej. Człowiek widział tysiące maszyn w swoim życiu. Z tego, co zobaczył wyciąga pewną ilość ogólnych wniosków. Maszyny są brzydkie, każda z nich jest przeznaczona do bardzo ograniczonego celu, są one bezużyteczne w przypadku cokolwiek innego celu, rozmaitość zachowania się każdej z nich jest bardzo mała itd. Naturalnie wnioskuje on, że są to niezbędne własności maszyn w ogóle. Wiele z tych ograniczeń jest związanych z bardzo małą pojemnością pamięci większości maszyn. (Przypuszczam, że idea pojemności pamięci rozciąga się w pewien sposób i na maszyny inne niż maszyny o stanach dyskretnych.) Dokładna definicja nie ma znaczenia, ponieważ w obecnej dyskusji nie jest wymagana żadna matematyczna dokładność. Kilka lat temu, gdy bardzo niewiele słyszało się o maszynach cyfrowych, można było wywołać duże niedowierzanie, mówiąc o ich własnościach bez opisywania ich budowy. Działo się tak przypuszczalnie dzięki podobnemu zastosowaniu zasady naukowej indukcji. Te zastosowania owej zasady są, naturalnie, przeważnie podświadome. Gdy oparzone dziecko boi się ognia i okazuje swój lęk unikając go, to powiedziałbym, że zastosowało ono naukową indukcję. (Mógłbym, naturalnie, opisać także na wiele innych sposobów, jego zachowanie się.) Nie wydaje się, aby prace i zwyczaje rodzaju ludzkiego stanowiły odpowiedni materiał, do którego można by stosować naukową indykcję. 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 głupie jest uczyć się francuskiego.

O wielu spośród wspomnianych niemożności można wypowiedzieć specjalne uwagi. Niemożność lubienia truskawek ze śmietaną może wydawać się czytelnikowi błaha. Być może można byłoby zrobić maszynę tak, aby lubiła tę wyborną potrawę, ale każda tego rodzaju próba byłaby idiotyczna. W związku z tą niemożnością ważne jest to, że wnosi ona swój wkład do niektórych innych niemożności, np. do trudności zachodzenia tego samego rodzaju życzliwości między człowiekiem a maszyną, jak między ludźmi.

Żądanie: „maszyny nie mogą popełniać 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 zdemaskowana z powodu swojej szalonej celności. Odpowiedź na to jest prosta. Maszyna (zaprogramowana do grania w grę) nie usiłowałaby udzielić *prawidłowych* odpowiedzi na problemy arytmetyczne, natomiast rozmyślnie wprowadzałaby błędy w sposób obliczony na zmylenie pytającego. Mechaniczny defekt prawdopodobnie by się ujawnił sam poprzez nieodpowiednią decyzję, dotyczącą rodzaju błędu, jaki można popełnić w arytmetyce. Nawet ta interpretacja krytyki nie jest dostatecznie życzliwa. Ale ze względu na miejsce nie możemy sobie pozwolić na wniknięcie w to głębiej. Wydaje mi się, że ta krytyka polega na pomieszaniu dwóch rodzajów błędów. Możemy nazwać je „błędami działania” i „błędami wnioskowania”. Błędy działania są spowodowane pewnymi mechanicznymi lub elektrycznymi usterkami, które powodują, że maszyna zachowuje się inaczej, niż w sposób wynikający z jej konstrukcji. W dyskusjach filozoficznych pragnie się uniknąć możliwości występowania takich błędów; z tego względu rozważa się „abstrakcyjne maszyny”. Te abstrakcyjne maszyny są to raczej fikcje matematyczne, niż obiekty fizyczne. Z definicji są one niezdolne do popełniania błędów działania. W tym znaczeniu możemy zgodnie z prawdą powiedzieć, że „maszyny nigdy nie mogą popełniać błędów”. Błędy wnioskowania mogą powstać tylko wtedy, gdy do wyjściowych sygnałów maszyny przywiązane jest pewne znaczenie. Maszyna mogłaby na przykład wypisywać na maszynie równania matematyczne lub angielskie zdania. Gdy na maszynie zostanie napisane fałszywe zdanie, to mówimy, że maszyna popełniła błąd wnioskowania. Oczywiście, nie ma żadnego powodu, aby twierdzić, że maszyna nie może popełnić tego rodzaju błędu. Wystarczy by maszyna wypisywała tylko wielokrotnie „0 – 1”. Biorąc mniej perwersyjny przykład mogłaby ona posiadać jakąś metodę wyciągania wniosków na drodze naukowej indukcji. Musimy oczekiwać, że taka metoda będzie prowadzić sporadycznie do błędnych wyników.

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. Niemniej „przedmiot działań maszyny” wydaje się coś oznaczać przynajmniej dla ludzi, którzy mają z tym do czynienia. Gdyby na przykład maszyna próbowała znaleźć rozwiązanie równania: $x^2-40x-11=0$, to można byłoby pokusić się określić równanie, jako część przedmiotu, jakim zajmuje się w owej chwili maszyna. W tym znaczeniu maszyna niewątpliwie może być sowim własnym przedmiotem. Można ją wykorzystać do pomocy w sporządzaniu jej własnych programów lub w celu przewidzenia efektu zmian jej własnej struktury. Obserwując wyniki swojego własnego zachowania się, może ona modyfikować swoje własne programy, tak aby efektywniej osiągnąć pewne cele. Są to raczej możliwości bliskiej przyszłości niż utopijne marzenia.

Krytyka, że maszyna nie może mieć dużej rozmaitości zachowania się jest dokładnie tym samym, co powiedzenie, że nie może mieć ona dużej pojemności pamięci. Aż do naprawdę ostatnich czasów rzadko spotykało się pamięć o pojemności 1000 cyfr.

Głosy krytyczne, które tutaj rozważamy są często zamaskowanymi postaciami argumentu świadomości. Zazwyczaj, jeśli ktoś utrzymuje, że maszyna może zrobić jedną z tych rzeczy i opisuje rodzaj metody, którą maszyna mogłaby zastosować, to nie wywiera to na słuchaczu wielkiego wrażenia. Myśli się na ogół, że metoda (jak by nie była, ponieważ musi być

mechaniczna) jest w istocie raczej prostą. Porównaj zawartość nawiasów z twierdzeniu Jeffersona cytowanym na stronie 11.

f. Zarzut lady Lovelace

Nasze najszczegółowsze informacje o maszynie analitycznej Babbage'a pochodzą z rozprawy lady Lovelace (1842). Stwierdza się w niej: „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). Hartee (1949) cytuje tę wypowiedź i dodaje: „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. Problem ten wyniknął z ostatnich odkryć. Ale nie wydaje się, aby maszyny obecnie budowane lub projektowane miały tę własność".

Całkowicie zgadzam się co do tego z Harteem. Zauważmy, że nie twierdzi on, że omawiane maszyny nie posiadały tej własności, ale raczej, że dowody dostępne lady Lovelace nie zachęcały jej do wierzenia, że one ją miały. Jest zupełnie możliwe, że omawiane maszyny posiadałaby w pewnym sensie tę własność. Dlatego załóżmy, że pewne maszyny o stanach dyskretnych posiadają tę własność. Maszyna analityczna była uniwersalną maszyną cyfrową i wobec tego, gdyby posiadała wystarczająco dużą pojemność pamięci i szybkość, to mogłaby przy odpowiednim zaprogramowaniu naśladować omawianą maszynę. Prawdopodobnie ten argument nie przyszedł na myśl hrabinie ani Babbage'owi. W każdym razie nie mieli obowiązku żądać od maszyny wszystkiego, co można było zażądać. Całe to zagadnienie zastanie rozpatrzone ponownie pod nagłówkiem maszyn uczących się. Wariant zarzutu lady Lovelace stwierdza, że maszyna „nie może nigdy zrobić nic naprawdę nowego". Na razie można go odparować powiedzeniem: „Nie ma nic nowego pod słońcem". Któż może mieć pewność, że wykonana przez niego „oryginalna praca" nie jest tylko rozwojem nasienia, zasadzonego w nim przez nauczanie lub rezultatem stosowania dobrze znanych ogólnych reguł. Zręczniejszy wariant tego zarzutu mówi, że maszyna nie może nigdy „zaskoczyć nas". To twierdzenie jest bardziej otwartym wyzwaniem i można przeciwstawić się mu bezpośrednio. Maszyny często mnie zaskakują. Dzieje się tak przeważnie dlatego, że nie dokonałem niezbędnych obliczeń, aby móc określić co można od nich oczekiwać, albo raczej ponieważ chociaż dokonałem obliczeń, to jednak wykonałem je w pośpieszny, niedbały sposób, podejmując ryzyko takiego podejścia. Być może mówię sobie: „Przypuszczam, że napięcie tutaj powinno być takie samo jak tam, w każdym razie załóżmy, że tak jest". Oczywiście, często nie mam racji i rezultatem tego jest moje zaskoczenie, ponieważ od czasu wykonania eksperymentu zapomniałem o tych założeniach. Te założenia narażają mnie na wymówki na temat mojego nieprawidłowego sposobu postępowania, ale nie rzucają wątpliwości na moją wiarygodność, gdy mówię o doznawanym przez siebie zaskoczeniu.

Nie spodziewam się, że ta replika ucieszy mego krytyka. Powie on prawdopodobnie, że takie zaskoczenia są właściwe pewnemu twórczemu działaniu mojego umysłu i nie przynoszą zaszczytu maszynie. To prowadzi nas z powrotem do argumentu świadomości i odciąga daleko od idei zaskoczenia. Jest to linia argumentacji, która musimy uważać za zamkniętą, ale być może, warto zauważyć, że zrozumienie czegoś takiego jak zaskoczenie wymaga tyle samo „twórczej czynności umysłowej" bez względu na to, czy zaskakujące wydarzenie pochodzi od człowieka, książki, maszyny lub czegoś jeszcze innego.

Pogląd, że maszyny nie mogą spowodować zaskoczenia powstał z powodu fałszywego rozumowania, na które są narażeni zwłaszcza filozofowie i matematycy. Jest to założenie, że skoro tylko jakiś fakt zostanie przedstawiony umysłowi, to równocześnie z nim wprowadzone zostają do umysłu wszystkie jego konsekwencje. W wielu wypadkach jest to bardzo użyteczne założenie, ale zbyt łatwo zapomina się, że jest ono fałszywe. Naturalną konsekwencją takiego postępowania jest dodatkowe założenie, że samo wypracowanie konsekwencji z danych i ogólnych reguł nie jest wcale zasługą.

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 dużą różnice 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.

Prawdą jest, że maszyna o stanach dyskretnych musi być inna od maszyny o stanach ciągłych. Ale jeśli będziemy stosować się do reguł gry w naśladownictwo, to pytający nie będzie w stanie skorzystać z tej różnicy. Tę sytuację można wyjaśnić, jeśli rozpatrzymy jaką inną prostszą maszynę o stanach ciągłych. Bardzo odpowiednią maszyną jest analizator różniczkowy. (Analizator różniczkowy, używany do pewnego rodzaju obliczeń nie jest rodzajem maszyny o stanach dyskretnych). Niektóre z tych maszyn drukują swoje odpowiedzi, a więc nadają się do wzięcia udziału w grze. Nie jest możliwe, aby maszyna cyfrowa przewidziała dokładnie jakie odpowiedzi dawałby analizator różniczkowy, ale z pewnością potrafiłaby dawać prawidłowo odpowiedzi. Na przykład na żądanie podania wartości π (która faktycznie wynosi około 3,1416) postąpiłaby rozsądnie dokonując wyboru na chybił trafił spomiędzy 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. w tych warunkach pytającemu byłoby trudno odróżnić analizator różniczkowy od maszyny cyfrowej.

h. Argument wypływający z nieformalności zachowania się

Niemożnością jest napisanie takiego zbioru reguł, według których człowiek mógłby postępować w każdych możliwych do pomyślenia okolicznościach. Można by na przykład mieć regułę, która by mówiła, że trzeba się zatrzymać, gdy się zobaczy czerwone światło regulujące ruch uliczny, i iść, jeśli zobaczy się zielone, ale co będzie, gdy na skutek jakiegoś uszkodzenia będą palić się oba światła? Może można by zdecydować, że najbezpieczniej jest zatrzymać się. Lecz na skutek tej decyzji może później z łatwością powstać nowa trudność. Okazuje się, że podanie reguł postępowania obejmujących każdą ewentualność jest niemożliwe choćby nawet były to reguły, dotyczące świateł, regulujących ruch uliczny.

Na podstawie powyższych rozważań dowodzi się, że my nie możemy być maszynami. Spróbuję odtworzyć ten dowód, ale obawiam się, że trudno mi będzie usprawiedliwić go. Wydaje się, że ten dowód brzmi następująco: „Gdyby każdy człowiek posiadał określony zbiór reguł postępowania, przy pomocy których regulowałby swoje życie, wówczas nie byłby wcale lepszy od maszyny. Ale ponieważ takich reguł nie ma – ludzie nie mogą być maszynami". W tym rozumowaniu rzuca się w oczy niewyłączny środek. Nie sądzę, że ten argument kiedykolwiek został postawiony zupełnie tak samo jak tutaj, ale jestem przekonany, że tym niemniej się go stosuje. Jednakże kwestię tę zaciemnia pewne pomieszanie pojęć mogące wystąpić między „regułami postępowania", a „prawami zachowania się". Przez

„reguły postępowania” rozumiem takie spostrzeżenia, jak: „Zatrzymaj się, gdy zobaczysz czerwone światła”, na które to spostrzeżenia można reagować i z których można zdawać sobie sprawę. Przez „prawa zachowania się” rozumiem prawa natury stosujące się do ciała ludzkiego, takie jak, „jeśli uszczypniesz go, to on piśnie”. Jeśli w cytowanym argumencie zastąpić „prawa postępowania, przy pomocy których reguluje on swoje życie” przez „prawa zachowania się, które regulują jego życie” to niewyłączny środek stanie się możliwy do przezwyciężenia. Dzieje się tak, ponieważ wierzymy, że nie tylko jest zgodne z prawdą twierdzenie, że jeśli podlegamy prawom zachowania się, to jesteśmy jakąś maszyną (niekoniecznie maszyną o stanach dyskretnych), ale że i odwrotnie, jeśli jesteśmy taką maszyną, to podlegamy takim prawom. Jednakże nie możemy tak łatwo dać się przekonać o nieistnieniu kompletnych praw zachowania się, mających postać kompletnych reguł postępowania. Jedyną znaną nam 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”.

Możemy pokazać bardziej przekonująco, że nie można usprawiedliwić żadnego takiego twierdzenia. Przypuśćmy, że moglibyśmy mieć pewność znalezienia takich praw w razie gdyby istniały. Wtedy mając maszynę o stanach dyskretnych, na pewno moglibyśmy dowiedzieć się o niej wystarczająco dużo na drodze obserwacji, tak aby móc przewidzieć jej przyszłe zachowanie się i w rozsądnym czasie, powiedzmy równym tysiącu lat. Ale nie wydaje się, aby sprawa przedstawiała się w ten sposób. Ułożyłem na maszynie cyfrowej z Manchesteru mały program, wykorzystujący tylko 1000 miejsc w pamięci. Przy pomocy tego programu maszyna, której dostarczono jedną szesnastocyfrową liczbę, podaje w przeciągu dwóch sekund inną liczbę. Twierdzę, że nikt nie potrafi dowiedzieć się z tych odpowiedzi wystarczająco dużo o programie, tak aby potrafić przewidzieć wszystkie odpowiedzi na niewypróbowane wartości.

i. Argument wypływający z pozazmysłowej percepcji

Zakładam, że czytelnik jest obeznany z pojęciem pozazmysłowej percepcji i ze znaczeniem czterech jej elementów, a mianowicie: telepatią, jasnowidzeniem, wiedzą uprzednią i lewitacją. Te niepokojące zjawiska zdają się przeczyć wszystkim naszym zwyczajnym pojęciom naukowym. Jednakże chcielibyśmy je zdyskredytować! Na nieszczęście świadectwo statystyczne, przynajmniej dla telepatii, jest nieodparte. Bardzo trudno jest przekształcić swoje sądy tak, aby pasowały do nich te nowe fakty. Skoro raz zostały one przyjęte, to nie wydaje się dużym krokiem naprzód wiara w duchy i strachy. Wyobrażenie, że nasze ciała poruszają się po prostu według znanych praw fizyki, byłoby jednym z pierwszych wyobrażeń, które trzeba byłoby odrzucić.

Ten argument jest, moim zdaniem, silny. Można odpowiedzieć na niego, że wiele teorii naukowych można zrealizować w praktyce, pomimo sprzeczności z pozazmysłową percepcją; że naprawdę dobrze można dawać sobie radę, jeśli się o niej zapomni. Jest to dosyć słaba pociecha i można obawiać się, że myślenie jest właśnie tego rodzaju zjawiskiem, dla którego pozazmysłowa percepcja może mieć specjalne znaczenie.

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 uzyskać 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. Z drugiej strony, mógłby on zgadnąć prawidłowo, w ogóle bez pytania, na drodze jasnowidzenia. Z pozazmysłową percepcją wszystko może się zdarzyć.

Jeśli uznamy istnienie telepatii, to trzeba będzie zaostrzyć nasz test. Sytuacja mogłaby uchodzić za analogiczną do tej, która miałby miejsce, gdyby pytający mówił do siebie, a jeden z konkurentów podsłuchiwałby a uchem przyłożonym do ściany. Umieszczenie konkurentów w „pokoju zabezpieczonym od telepatii" spełniłoby wszystkie wymagania.

7. Maszyny uczące się

Czytelnik z pewnością odgadł, że nie mogę poprzeć sowich poglądów bardzo przekonującymi pozytywnymi argumentami. Gdybym miał takie argumenty, to nie zadawałbym sobie tyle trudu, aby wykazać fałszywość rozumowania w poglądach przeciwnych. Obecnie przedstawię takie dowody, jakie posiadam.

Powróćmy na chwilę do zarzutu lady Lovelace, zgodnie z którym maszyna może robić tylko to, co powiemy, że ma zrobić. Można by powiedzieć, że człowiek może „wstrzyknąć" ideę do maszyny, na co ona zareaguje w pewnym stopniu i następnie uspokoi się tak, jak uderzona młoteczkiem struna fortepianu. Innym porównaniem byłby stos atomowy o mniej niż krytycznej wielkości: wstrzyknięta idea odpowiadałaby neutronowi, wchodzącemu do stosu z zewnątrz. Każdy taki neuron spowoduje pewne zakłócenie, które w końcu zaniknie. Jeśli jednakże powiększyć w wystarczający sposób wielkość stosu, to istnieje duże prawdopodobieństwo, że zakłócenie wywołane przez taki nadchodzący neutron będzie powiększało się dalej, aż do zniszczenia całego stosu. Czy istnieje podobne zjawisko dla umysłów i czy istnieje ono dla maszyn? Wydaje się, że takie zjawisko występuje w przypadku umysłu ludzkiego. Wydaje się, że większość umysłów ludzkich jest „podkrytyczna", to znaczy, że w tej analogii odpowiada stosowi wielkości podkrytycznej. Idea przedstawiona takiemu umysłowi przeciętnie powoduje powstanie w odpowiedzi mniej niż jednej idei. Mniejsza część umysłów ludzkich jest nadkrytyczna. Idea przedstawiona takiemu umysłowi może wywołać całą „teorię" złożoną z drugorzędnych, trzeciorzędnych i bardziej odległych idei. Wydaje się, że umysły zwierząt są zdecydowanie podkrytyczne. Obstając przy tej analogii zapytajmy „Czy można zrobić maszynę nadkrytyczną?".

Analogia do „łupiny od cebuli" jest również pomocna. Rozważając funkcje umysłu lub mózgu, znajdujemy pewne operacje, które możemy wyjaśnić w czysto mechanicznych terminach. One, mówimy, nie odpowiadają prawdziwemu umysłowi, ale stanowią coś w rodzaju łupiny, którą musimy zdjąć, aby znaleźć prawdziwy umysł. Ale później w tym co pozostało natrafiamy na dalszą łupinkę do zdarcia i tak dalej. Czy postępując w ten sposób dojdziemy kiedykolwiek do „prawdziwego" umysłu, czy też w końcu dojdziemy do łupinki, w której nic nie ma? W tym drugim wypadku cały umysł byłby mechaniczny. (Jednakże nie byłaby to maszyna o stanach dyskretnych. Przedyskutowaliśmy to uprzednio).

Ostatnie dwa punkty nie roszczą sobie pretensji do przedstawienia przekonujących argumentów. Należałoby je raczej opisać jako „deklamacje mające na celu wzbudzenie wiary".

Pogląd wyrażony na początku punktu f. można poprzeć w jedyny, naprawdę wystarczający sposób, który polega na doczekaniu końca stulecia i wykonaniu wtedy opisanego eksperymentu. Ale cóż możemy powiedzieć w międzyczasie? Jakie należałoby obecnie przedsięwziąć kroki, gdyby eksperyment miał zostać uwieńczony sukcesem?

Jak wyjaśniłem, problem tkwi głównie w programowaniu. Postęp w dziedzinie wiedzy inżynierskiej będzie miał również miejsce, ale wydaje się nieprawdopodobne, aby nie stanął on na wysokości tych wymagań. Pojemność pamięci mózgu szacuje się na: od 10^{10} do 10^{15} cyfr binarnych. Ja osobiście skłaniam się do niższych wartości i sądzę, że tylko bardzo mała część pojemności pamięci mózgu służy do myślenia wyższego rodzaju. Większa jej część prawdopodobnie służy do zapamiętywania wrażeń wzrokowych. Byłbym zdziwiony, gdyby do dostatecznie poprawnej rozgrywki gry w naśladownictwo potrzeba było więcej, niż 10^9 cyfr binarnych, przynajmniej w grze z niewidomym człowiekiem. (*Zauważ*: Pojemność 11 wydania *Encyklopedia Britannica* wynosi $2 \cdot 10^9$). Pojemność pamięci rzędu 10^7 byłaby absolutnie możliwa do zrealizowania, nawet przy obecnych technikach realizacji. Prawdopodobnie w ogóle nie jest potrzebne powiększenie szybkości działania maszyn. Te części nowoczesnych maszyn, które można uważać za analogi komórek nerwowych pracują od nich około tysiąc razy szybciej. Powinno to zapewnić „margines bezpieczeństwa", który mógłby pokryć występujące z wielu powodów straty szybkości. Wobec tego nasz problem polega na wymyśleniu sposobu programowania tych maszyn tak, aby grały w grę. Przy mojej obecnej szybkości pracy (pisząc około tysiąc cyfr programu dziennie) około sześćdziesięciu pracowników pracując pilnie przez pięćdziesiąt lat, mogłoby wykonać tę pracę gdyby nic nie poszło do kosza na śmieci. Wydaje się, że pożądana byłaby jakaś bardziej szybka metoda.

Podczas prób naśladowania dojrzałego umysłu ludzkiego jesteśmy zmuszeni dużo myśleć o procesie, który doprowadził go do stanu, w którym się aktualnie znajduje. Możemy zauważyć trzy elementy:

1. początkowy stan umysłu, powiedzmy urodzenie,
2. edukacja, której był poddawany umysł,
3. inne doświadczenia nie określane mianem edukacji, którym był poddawany umysł.

Zamiast programu symulującego dorosły umysł, dlaczego raczej nie spróbować zbudować program symulujący umysł dziecka? Gdyby następnie poddać go odpowiedniemu procesowi edukacji, to można by otrzymać umysł dojrzały. Przypuszczalnie, umysł dziecka jest czymś w rodzaju notesu, jaki kupuje się w sklepie z artykułami piśmienniczymi. Raczej niewielki mechanizm i dużo pustych kartek. (Mechanizm i sztuka pisania są, z naszego punktu widzenia, niemal synonimami). Mamy nadzieję, że umysł dziecka posiada tak niewielki mechanizm, że coś w tym rodzaju można łatwo zaprogramować. W pierwszym przybliżeniu możemy przyjąć, że ilość pracy włożona w edukację maszyny jest prawie taka sama jak w przypadku dziecka ludzkiego.

Tak więc podzieliliśmy nasz problem na dwie części: program dziecka i proces edukacji. Te dwie części są bardzo ściśle powiązane ze sobą. Nie możemy oczekiwać, że już w pierwszej próbie opracujemy dobry automat dziecka. Trzeba będzie przeprowadzić eksperyment z nauczaniem jednej takiej maszyny i zobaczyć, jak dobrze ona się uczy. Następnie można

wypróbować inną maszynę i przekonać się, czy jest lepsza czy gorsza. Występuje oczywisty związek między tym procesem a ewolucją stosownie do następujących tożsamości:

struktura automatu dziecka = materiał dziedziczny,
zmiany automatu dziecka = mutacje,
selekcja naturalna = opinia eksperymentatora.

Można jednakże mieć nadzieję, że ten proces będzie szybciej działał, niż ewolucja. Ewolucja drogą doboru naturalnego jest powolną metodą nawarstwiania się zalet. Eksperymentator powinien potrafić ją przyspieszyć na drodze ćwiczenia inteligencji. Równie ważny jest fakt, że oddziaływanie eksperymentatora nie ogranicza się do przypadkowych mutacji. Jeśli on potrafi wyśledzić przyczynę jakieś słabości, to prawdopodobnie będzie mógł obmyśleć rodzaj mutacji, która ją poprawi.

Nie będzie można zastosować dokładnie tego samego procesu nauczania do maszyny, co do normalnego dziecka. Nie będą przewidziane na przykład nogi, tak że nie będzie można zlecić maszynie, aby wyszła i napełniła wiadro na węgiel. Być może, mogłaby ona nie posiadać oczu. Ale choćby można było najlepiej przezwyciężyć te braki zręczną wiedzą inżynierską, to nie można by było wysłać tego tworu do szkoły tak, aby inne dzieci zbytnio się z niego nie śmiały. Trzeba mu dać jakieś lekcje. Nie powinniśmy zbytnio interesować się nogami, oczami itd. Przykład panny Heleny Keller pokazuje, że edukacja może mieć miejsce pod warunkiem, że w jakiś sposób jest możliwa obukierunkowa komunikacja między nauczycielem i uczniem.

Normalnie z procesem nauczania kojarzymy kary i nagrody. Jakieś proste automaty dzieci można by zbudować lub zaprogramować na tego rodzaju zasadzie. Maszyna musiałaby być tak zbudowana, aby było mało prawdopodobne powtórzenie się wypadków, które zaszły na krótko przed pojawieniem się sygnału kary, podczas gdy sygnał nagrody powiększałby prawdopodobieństwo powtórzenia wypadków, które do niego doprowadziły. Te definicje nie zakładają z góry żadnych uczuć ze strony maszyny. Wykonałem pewne eksperymenty z jednym takim automatem-dzieckiem i udało mi się nauczyć go kilku rzeczy, ale metoda nauczania byłą zbyt mało ortodoksyjna, aby można uważać, że ten eksperyment został naprawdę uwieńczony powodzeniem.

Stosowanie kar i nagród może w najlepszym razie stanowić część procesu nauczania. Z grubsza mówiąc, jeśli nauczyciel nie ma innych sposobów komunikowania się z uczniem, to ilość informacji jaka może dotrzeć do niego nie przewyższa ogólnej ilości zastosowanych nagród i kar. W czasie uczenia się na pamięć „Casablanki" dziecko prawdopodobnie byłoby bardzo rozdrażnione, gdyby tekst można było poznawać tylko przy pomocy metody „dwudziestu pytań", przy czym każde „NIE" byłoby ciosem. Dlatego niezbędne jest posiadanie jakichś innych „nieemocjonalnych" kanałów komunikacji. Jeśli będą one dostępne, to będzie można nauczyć maszynę metoda kar i nagród, słuchania rozkazów, wydanych w jakimś języku np. języku symbolicznym. Te rozkazy będą przesyłane przez „nieemocjonalne" kanały. Stosowanie tego języka zmniejszy znacznie ilość potrzebnych kar i nagród.

Zapatrywania dotyczące odpowiedniej złożoności automatu-dziecka mogą się zmieniać. Można by spróbować zrobić go tak prostym jak to jest tylko możliwe zgodnie z ogólnymi zasadami. Albo też można by do niego „wbudować" kompletny system logicznego wnioskowania. W tym ostatnim wypadku pamięć byłaby przeważnie zajęta przez definicje i

założenia. Założeniami byłyby np. dobrze ustalone fakty, przypuszczenia, twierdzenia matematyczne udowodnione, wypowiedzi podane przez autorytet, wyrażenia o postaci zdań logicznych, ale bez wartościowania. Pewne założenia można określić jako „imperatywy". Maszyna powinna być tak zbudowana, aby natychmiast po stwierdzeniu, że imperatyw jest „dobrze ustalony" automatycznie odbywało się odpowiednie działanie. Aby to zilustrować załóżmy, że nauczyciel mówi do maszyny: „Odrób teraz swoją pracę domową". Może to spowodować, że „Nauczyciel mówi: „Odrób teraz swoją pracę domową"" zostanie zaliczone do dobrze ustalonych faktów. Innym takim faktem mogło by być „Wszystko, co mówi nauczyciel jest prawdą". Powiązanie tych faktów może w końcu doprowadzić do zaliczenia imperatywu „Odrób teraz swoją pracę domową" do dobrze ustalonych faktów i będzie to, dzięki konstrukcji maszyny znaczyło, że praca domowa faktycznie rozpoczyna się, a jej efekt jest istotnie zadowalający. Stosowane przez maszynę procesy wnioskowania nie muszą spełniać wymagań stawianych przez najbardziej wymagających logików. Na przykład mogłoby nie być hierarchii typów. Ale nie musi to znaczyć, że fałszywe rozumowania będą zdarzały się częściej, niż grożący nam spadek z nieogrodzonego urwiska. Odpowiednie imperatywy (wyrażone w systemach, a nie stanowiące części reguł systemu) takie jak: „Nie stosuj klasy, chyba, że jest ona podklasą takiej klasy, którą wymienił nauczyciel", mogą wywierać podobny skutek co „Nie podchodź za blisko krawędzi".

Imperatywy, które może wykonać maszyna nie posiadająca kończyn, muszą posiadać raczej intelektualny charakter, tak jak w podanym wyżej przykładzie (odrabianie pracy domowej). Wśród tego rodzaju imperatywów ważne będą takie imperatywy, które ustalają kolejność stosowania reguł odnośnego systemu logicznego. W każdym stadium stosowania systemu logicznego istnieje duża ilość alternatywnych kroków, z których każdy można zastosować, o ile zachowane jest posłuszeństwo regułom systemu logicznego. Zależnie od dokonanych wyborów otrzymuje się różnice takie jakie występują między znakomitym i nędznym argumentatorem, ale nie różnice, występujące między logicznym i nielogicznym argumentatorem. Zdania, prowadzące do tego rodzaju imperatywów mogłyby być następujące: „Gdy wymieni się Sokratesa, to zastosuje sylogizm Barbara" albo „Jeśli udowodniono, że jedna metoda jest szybsza od drugiej, to nie stosuj powolniejsze metody". Niektóre z nich mogą być „dane na mocy autorytetu", ale inne może wytwarzać sama maszyna na drodze naukowej indukcji.

Pewnym czytelnikom idea uczącej się maszyny może wydawać się paradoksalna. Jak mogą zmieniać się reguły działania maszyny? One powinny całkowicie opisywać działanie maszyny, bez względu na jej historię oraz zmiany jakim mogłaby być poddana. Wobec tego, reguły te są zupełnie niezmienne w czasie. Jest to prawda. Wyjaśnienie paradoksu polega na tym, że reguły, które zmieniają się w procesie uczenia się nie roszczą sobie tak dużych pretensji, wymagając jedynie efemerycznej słuszności. Czytelnik może porównać to z Konstytucją Stanów Zjednoczonych.

Ważną cechą maszyny uczącej się jest to, że jej nauczyciel często będzie w bardzo dużym stopniu nieświadomy tego, co dzieje się w niej, chociaż, mimo to, może do pewnego stopnia przewidzieć zachowanie się swojego ucznia. Powinno to dotyczyć głównie późniejszej edukacji maszyny, powstającej z maszyny-dziecka o dobrze wypróbowanym projekcie (lub programie). To wyraźnie kontrastuje z normalną procedurą stosowania maszyny do wykonywania obliczeń: wtedy zależy nam na posiadaniu wyraźnego umysłowego obrazu stanu maszyny w każdej chwili liczenia. Cel ten można osiągnąć jedynie z trudem. Wobec tego pogląd, że „maszyna może wykonać tylko to, co wiemy w jaki sposób zlecić jej do

wykonania”[1] wydaje się dziwny. W rezultacie działania większości programów, które możemy wprowadzić do maszyny, maszyna zachowuje się w sposób bezsensowny lub też przypadkowy. Inteligentne zachowanie się przypuszczalnie polega na odstępstwie od całkowicie zdyscyplinowanego zachowania się, wymaganego przy liczeniu, ale dosyć nieznacznym, takim, które nie powoduje przypadkowego zachowania się lub banalnych pętli repetycyjnych. Innym ważnym rezultatem przygotowania naszej maszyny do udziału w grze w naśladownictwo na drodze procesu nauczania i uczenia się jest to, że prawdopodobnie „ludzka omylność” zostanie usunięta w dosyć naturalny sposób, to znaczy bez specjalnego „trenowania”. (Czytelnik powinien pogodzić to z punktem widzenia, przedstawionym na stronach 12-13). Procesy nauczone nie dają stuprocentowej pewności wyniku; gdyby ją dawały, to nie podlegałyby zapominaniu.

Zapewne rozsądne jest wprowadzenie elementu przypadkowego do uczącej się maszyny. Element przypadkowy jest dosyć użyteczny, gdy szukamy rozwiązania jakiegoś problemu. Przypuśćmy na przykład, że chcielibyśmy znaleźć liczbę zawartą między 50 i 200, równą kwadratowi sumy swoich cyfr. Moglibyśmy na przykład najpierw wypróbować liczbę 51, następnie 52 i potem następne liczby, aż do otrzymania liczby, spełniającej powyższy warunek. Bądź też moglibyśmy wybierać liczby na chybił trafił, aż do otrzymamy dobrą. Zaletą tej metody jest to, że nie potrzeba śledzić wypróbowanych wartości natychmiast; wadą to, że nie można dwukrotnie wypróbowywać tę samą wartość ale nie jest to zbyt istotne, jeśli istnieje kilka rozwiązań. Metoda systematyczna ma tę ujemną stronę, że w obszarze, który ma być badany w pierwszej kolejności może występować ogromny blok bez żadnych rozwiązań. Obecnie uważa się, że proces uczenia się polega na szukaniu takiej postaci zachowania się, która spełni wymagania nauczyciela (lub jakieś inne kryterium). Ponieważ prawdopodobnie istnieje duża liczba zadawalających rozwiązań, wydaje się, że metoda przypadkowa jest lepsza od metody systematycznej. Zauważmy, że metoda ta jest stosowana w procesie analogicznym – ewolucji. Ale tam stosowanie metody systematycznej nie jest możliwe. W jaki sposób można by śledzić różne wypróbowane kombinacje genetyczne, tak aby uniknąć ponownego ich wypróbowywania?

Możemy mieć nadzieję. Że maszyny będą współzawodniczyć z ludźmi we wszystkich czysto intelektualnych dziedzinach. Ale od których z nich należałoby zacząć? Trudno nawet to przesądzić. Wielu ludzi myśli, że najlepsza byłaby bardzo abstrakcyjna działalność w rodzaju gry w szachy. Można również twierdzić, że najlepiej dostarczyć maszynie najlepsze organy zmysłowe i następnie nauczyć ją rozumieć i mówić po angielsku. Ten proces mógłby naśladować normalne nauczanie dziecka. Maszynie pokazywałoby się rzeczy i nazywało je itd. Znowu nie wiem jaka jest prawidłowa odpowiedź, ale myślę, że należałoby wypróbować obydwa podejścia. Widzimy tylko mały odcinek drogi przed nami, ale możemy dostrzec tam mnóstwo rzeczy do zrobienia.

[1] Porównaj wypowiedź lady Lovelace, która nie zawiera słowa „tylko”. Przyp. autora.