



B28. Filozofia umysłu



[Poprzedni rozdział: Neuropsychologia](#) | [Wstęp do kognitywistyki - spis treści](#).

28.0. Problem psychofizyczny | 28.1. [Narodziny kognitywizmu](#) | 28.2. [Argumenty Turinga](#) | 28.3. [Funkcjonalizm](#) | 28.4. [Sztuczna inteligencja](#) | 28.5. [Chiński pokój](#) | 28.6. [Nieobliczalność](#) | 28.7. [Podsumowanie](#)

Dopóki działanie mózgu będzie dla nas tajemnicze, Wszechświat będzie również tajemniczy.

[Santiago Ramón y Cajal](#)

Wiele zagadnień dyskutowanych przez filozofów i naukowców skupiało się wokół dwóch zagadnień bardzo ważnych z punktu widzenia zrozumienia natury ludzkiej: problemu psychofizycznego i problemu wolnej woli. Tradycyjnie najpierw przedstawia się poglądy filozoficzne, próby wyobrażenia sobie, jak funkcjonują nasze umysły. Powinniśmy już do pewnego stopnia rozumieć, jak mózgi tworzą umysły i jakie pytania warto zadawać, by poznać szczegóły tego procesu. To powinno nam pozwolić popatrzeć krytycznie na wysiłki filozofów i poprzednich pokoleń, próbujących zrozumieć działanie umysłu w oparciu o swoje intuicje i przemyślenia. Podsumujemy tu niektóre z historycznych stanowisk, jak i współczesne dyskusje dotyczące świadomości z takich rozważań wyrastające.

B28.0 Problem psychofizyczny

Zagadnienie psychofizyczne, znane też pod nazwą problemu ciała i umysłu (mind-body problem), to problem zrozumienia natury [świadomości](#). Jaka jest relacja pomiędzy naszymi wrażeniami i wewnętrznymi przeżyciami z zewnętrzną rzeczywistością? Świat fizyczny i świat psychiki ma przecież całkowicie odmienny charakter. Jak możemy to zrozumieć, czyli utworzyć sobie jakiś rozumiały model, który to wyjaśni? Zaproponowano tu liczne rozwiązania.

Rozwiązania skrajne:

- **Solipsyzm**, tylko ja istnieję, wszystko jest produktem mojej wyobraźni.
- **Dualizm**: całkowita odrębność świata umysłu i ciała (np. Kartezjusz, sir John Eccels, W. Penfield, Hoimar von Ditfurth).
- **Idealizm**: wszystko ma charakter psychiczny (Berkeley).
- **Okazjonalizm**: nie ma uzasadnienia dla przekonania o istnieniu świata fizycznego, jedyny dowód to księga Genesis - [Arnold Geulincx](#) (1624-1669) i [Nicolaus de Malebranche](#) (1638-1715). Świat mentalny i materialny są całkowicie oddzielone, zdarzenia fizyczne to okazje do powstawania wrażeń, które im towarzyszą, ale wrażenia same w sobie nie mogą powstać na skutek materialnego kontaktu; czuwa nad tym materialno-fizyczny Bóg.

Echo idei okazjonalizmu widać między innymi w [spekulacjach na tematy wszechwiedzy](#), mechaniki kwantowej i religii, które nadal snują niektórzy fizycy w oderwaniu od badań empirycznych, odpowiadających na konkretne pytania.

[Gottfried Wilhelm Leibniz](#): do zrozumienia człowieka nie wystarczy ani substancja ani forma, lecz konieczne jest rozpatrywanie całej jednostki (entitate tota).

Monady: zdarzenia fizyczne i psychiczne tworzą "monadę", substancję prostą, nieoddziałującą, punkt.

Umysł jest substancją odmienną od materii, złożone systemy nie mogą więc wyjaśnić percepcji i myślenia.

Jeśli myśląc maszynę powiększyć to nie dostrzeżemy w niej myśli ... (Monadologia 1714).

Nie istnieje przyczynowość wynikająca z oddziaływania różnych substancji - Leibniz miał trudności w zrozumieniu działania na odległość i wierzył w duchy popychające planety.

Rzeczywiste przyczyny zmian wynikają z poprzednich stanów wewnętrznych danej substancji.

Każda substancja jest "zaprogramowana" w momencie stworzenia tak, że jej stany zmieniają się w sposób zsynchronizowany z innymi.

Zalety: niepotrzebna jest ciągła interwencja boska, zachowana jest przyczynowość i separacja ciała i umysłu.

Stany monady wynikają z jej wewnętrznej natury, dzięki synchronizacji (odwiecznej harmonii) wrażenia odpowiadają wrażeniom innych monad.



Paralelizm, [Baruch Spinoza](#) (1632-1677) głosił konsekwentny monizm: istnieje tylko jedna pierwotna substancja, sama przez się, z której wszystko jest zbudowane.

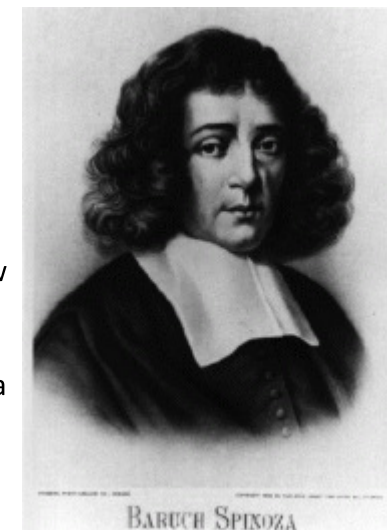
Propozycja nr 14: "Nie istnieje żadna substancja, która by nie była Bogiem".

Umysł i ciało są więc atrybutami tego samego racjonalnego bytu, tak jak dwie strony natury lub Boga.

Porządek logiczny odpowiada fizycznemu, duch nie wpływa na materię ani materia na ducha, ale zmiany w jednym i drugim świecie zachodzą równolegle.

Daje to złudzenie wolnej woli, chociaż naprawdę wszystko jest zdeterminowane przez wyższe prawa.

Przeciwieństwem wolności jest przymus, a nie konieczność. Wolność to zrozumienie konieczności działania (później podchwycił to Hegel, pisząc "wolność to uświadomiona konieczność")



Takie koncepcje wydają się obecnie karkołomne, ale wynikają z trudności w połączeniu subiektywnego świata wrażeń i świata materii, a więc świata psychologii i fizyki. Spinoza miał rację, że zmiany zachodzą równolegle w świecie psychicznym i fizycznym, czasami łatwiej jest zrozumieć i opisać zdarzenie na poziomie mentalnym i tam trzeba widzieć przyczyny (zobaczyłem ukochaną i czuję zawrót głowy), a czasami na poziomie fizycznym (udar mózgu wywołał afazję).

Dlaczego tak trudno jest sobie wyobrazić wewnętrzną perspektywę w systemie, który interpretuje swoje własne stany?

Nie mamy przykładów takich sztucznych systemów, a to ogranicza naszą wyobraźnię. Potrzebne są proste modele, pokazujące wewnętrzną perspektywę. To stało się możliwe dopiero w drugiej połowie XX wieku.

Niestety brak jest postępów dualizmu i różnych innych form idealizmu od ich powstania, nie da się w ten sposób wytłumaczyć żadnych konkretnych eksperymentów lub wyników obserwacji psychofizycznych.

Nie widać też, co by mogło przekonać zwolenników dualizmu; widać mamy tu do czynienia ze zbyt małą plastycznością ich mózgow, niepozwalającą im zmienić poglądów wcześniej ukształtowanych.

[Gilbert Ryle](#) (1900-1976)

Profesor "filozofii metafizycznej" w Oxfordzie, zajmował się filozofią analityczną, szukając rozwiązań problemów filozoficznych w analizie sensu zdań, np. "Systematically Misleading Expressions" (1932).

Jego główne dzieło to "[Koncepcja umysłu](#)" (1949), książka zawierająca liczne fałszywe stwierdzenia (np. "Nie ma niczego takiego jak obrazy mentalne").

Krytykując dualizm, przeprowadził atak na przekonanie o istnieniu kierującego ciałem ducha. "Duch w maszynie" (deus ex machina, ghost in the machine), nie istnieje, podobnie jak nie ma konia w pociągu. Rozwiązanie Ryle'a było zaskakujące: "**umysł**" jest **pomyłką kategorialną**. Co to jest uniwersytet? pyta Ryle. To budynki i ludzie, ale też pewna całość, forma. Nie należy się domagać "pokazania uniwersytetu" jeśli się widziało budynki i ludzi".

Umysł jest sposobem "uporządkowania wypadków składających się na jeden życiorys", niczym więcej.

Doświadczenie powtarzalnych stanów psychicznych i skłonności prowadzi do przypisania tych stanów umysłowi, ale jest to byt sztuczny.

Kartezjański dualizm, substancjalizm i immanentyzm nie mają sensu, bo umysłu nie ma.

Akt woli wymaga ducha, co jednak rządzi duchem? Czy jest on własnością małej istoty w naszej głowie, [homunkulusa](#)?

By się chciało chcieć, trzeba chcieć chcieć ... (jak pisał już William James), ale powstaje wtedy **nieskończona rekursja** odwołań.

Czy inteligencja wymaga naprawdę procesów inteligentnych, a więc obecności siedzących wewnątrz inteligentnego systemu homunkulusów? Nie! Współdziałanie prostych procesów daje inteligencję - symulacja inteligencji jest możliwa!

Czy symulowana intencjonalność też może tak powstać?

Ryle uznaje umysł za zbiór dyspozycji do zachowań, jest więc blisko behawioryzmu.

Świadczy o tym fakt, że wszystkie zwroty dotyczące umysłu są metaforyczne, nie ma innych. W języku angielskim mamy ich wiele, np. make up one's mind, my mind is made up, on my mind, come to mind, clear one's mind, of sound mind, off my mind, peace of mind, mind at ease, speak your mind, out of your mind, slipped you mind, narrow mind, open mind itp. W języku polskim częściej mamy wyrażenia: analityczny, bystry, błyskotliwy, głęboki, logiczny, przenikliwy, ścisły, uczony, żywy ... umysł, tęgi, wielki umysł (głowa), ograniczony umysł, w umysłach, zdrowy na umyśle, pomieszany umysł, ale też używamy słowa dusza, głowa, serce, a nie umysł, np. ciężko mi na duszy, leży na sercu, mam to w głowie, ciąży na mojej głowie, otwarta głowa.

Czy można uznać, że umysł jako pomyłka kategorialna da się zredukować do dyspozycji? Jeśli tak to zagadnienie ciała i umysłu jest rozwiązane.

Z pewnością reakcje organizmu, mózgu i ciała, są podstawą, jednak pojęcie umysłu jest przydatne - Ryle nie ma tu racji!

"Psycho-logia" nie da się zredukować do logiki pobudzeń struktur mózgu, albowiem poziom przeżyć intencjonalnych to nie to samo co poziom stanów neurofizjologicznych.

Historia indywidualnych jak i gatunkowych przeżyć może być zapisana w strukturze mózgu, ale nadać jej sens może dopiero interpretacja psychologiczna.

Ryle ma rację twierdząc, że umysł nie jest rzeczą, chociaż z powodu skłonności do petryfikacji zdarzeń, reifikacji pojęć mamy skłonności, by w ten sposób o nim myśleć.

Koncepcja zła (dobra) prowadzi do petryfikacji w postaci wyobrażenia Złego (Dobrego).

Zajmował się tym [David Bohm](#): w książce "Całość i ukryty porządek" (1980) przedstawia rzeczywistość jako holistyczny proces.

Struktura języka rozbija świat na niezależne fragmenty.

Bohm wprowadził tryb używania języka, nazwany *rheomode*, próbę wyrażenia płynności zdarzeń za pomocą języka podkreślającego procesy, unikającą rzeczowników; niektóre języki Indian amerykańskich (np. [Blackfoot](#)) i plemion afrykańskich mają taką strukturę.

Zdanie „on się rusza” sugeruje, że jest jakiś „on”, który jest przyczyną ruchu odrębnego „siebie”, ale w rzeczywistości mamy tylko działanie, ruch.

Po angielsku „it is raining” sugeruje, że jest jakieś "it" (to), którego dziełem jest padanie - lepiej zamiast „pada deszcz” (w domyśle "on pada"), sugerującego podmiot padający, opisać proces „zachodzi padanie”.





Mówimy "Ja mam mózg", "mam ciało" lub "mam wolną wolę", "mam pomysł", sugerując, że są to odrębne rzeczy od "ja", które jest centralnym podmiotem, właścicielem zarządzającym ciałem i umysłem. Takich wyrażeń jest bardzo dużo i wynikają z utożsamiania siebie z modelem "ja", stworzonym przez mózg, modelem, który musimy mozolnie poznawać. Czasami "aż tak dobrze siebie nie znam" by wiedzieć, co zrobię w nietypowej sytuacji.

Czy możemy uznać, że cząstka elementarna działa na inne cząstki? Nie, bowiem zachodzi tu wzajemne oddziaływanie, nie można przypisać centrum tego oddziaływania jednej cząstce! Jednak jeśli mamy do czynienia z wielką masą, taką jak Ziemia, przyciągającą małe przedmioty, to z dobrym przybliżeniem można zapomnieć o wzajemnym oddziaływaniu, bo stan spadających przedmiotów zmienia się wyraźnie, a ruchu ziemi w stronę wspólnego środka masy nie dostrzegamy.

Pojęcie niezależnej, odrębnie istniejącej materii jest jedynie przybliżeniem, jednak jeśli mamy do czynienia z dominacją jest to bardzo dobre przybliżenie. Podobnie jest z "ja", które nie daje się oddzielić od całości mojego istnienia, od procesu, w którym bierze udział, ale ponieważ znajduje się w centrum wielu działań i myśli **dostrzegamy tylko działanie "ja", a znacznie rzadziej uświadamiamy sobie, jak na "ja" wpływają różne czynniki.**

Umysł można traktować jako zbiór wielu procesów, jak określił to [Marvin Minsky](#), jeden z pionierów sztucznej inteligencji, należy raczej mówić o zbiorze wyspecjalizowanych, współdziałających ze sobą funkcji, nie monolitycznym umyśle, ale "społeczeństwie umysłu". [Society of mind](#) to właśnie nazwa książki, którą Minsky wydał w 1987 roku. Doprowadziło do powstania systemów wieloagentowych, przydatnych w sztucznej inteligencji oraz kognitywistyce do opisu zachowania, ale nie wyjaśniło jakościowej natury przeżyć mentalnych.

Metafizyka materialistyczna proponowała różne rozwiązania problemu psychofizycznego.

- **Behawioryzm**: umysł, świadomość nie istnieją, są tylko zdarzenia fizyczne, zachowania, które można badać bez odwoływania się do tego co mentalne.
- **Epifenomenalizm**: myśli i wrażenia to uboczne skutki fizycznych stanów mózgu, stany mentalne nie mają na nic wpływu.
- **Funkcjonalizm**: umysł jest rezultatem procesów obliczeniowych, niezależnie od substratu, w jakim dokonywane są odpowiednie obliczenia.
- **Funkcjonalizm**: umysł jest rezultatem procesów obliczeniowych, niezależnie od substratu, w jakim dokonywane są odpowiednie obliczenia.

Te postawy proponowane były w wielu wariantach.

Dyskusje filozoficzne nie doprowadziły do ostatecznych rozstrzygnięć.

Chociaż nie wiemy, jak zdefiniować rzeczywistość i w świetle problemów interpretacyjnych świata kwantów trudno jest obronić realizm, mamy naukę i wynikającą z niej technologię.

Metoda naukowa przez tworzenie hipotez, przewidywania i testowanie, pozwala zrobić wiarygodny model.

Modele umysłu są możliwe i przydatne.

W szczególności traktowanie umysłu w ścisłym związku z mózgiem jest bardzo płodne.

Neurofizjologia: **umysł to część tego, co robi mózg**, "ja" to jeden z procesów, realizowanych przez mózg.

Czy to oznacza, że umysł można zredukować do stanów neurofizjologicznych? Czy biologia redukuje się tylko do chemii i fizyki? To naiwny punkt widzenia.

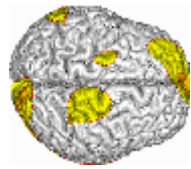
Mózg jest substratem, w którym zachodzą procesy konieczne do powstawania procesów mentalnych. Są to [procesy emergentne](#), które nie można opisać badając tylko części składowe, bo ich interakcja tworzy nową jakość.

Procesy zachodzące w mózgu są warunkiem koniecznym istnienia umysłu. Struktura mózgu odzwierciedla potrzeby organizmu i jego możliwości poznawcze.

Świat moich wspomnień nie jest materialny, desygnaty tych wspomnień mogą nie istnieć w świecie zewnętrznym, ale istnieją w moim świecie wewnętrznym.

Umysł nie jest więc materialny, chociaż potrzebuje materii, by mógł istnieć.

Czy to takie dziwne? Każda informacja potrzebuje jakiegoś materialnego nośnika.



Procesy neurofizjologiczne odpowiedzialne za zdarzenia mentalne mają rację bytu tylko ze względu na istnienie umysłu.

Moje przeżycia wewnętrzne realizowane są przez różne pobudzenia mózgu, ale ich specyfika jest rezultatem indywidualnej, zbiorowej i gatunkowej historii.

Modele umysłu wymagają opisu na poziomie psychologicznym, odmiennego od modeli mózgu.

Umysł to zbiór stanów mentalnych i relacji pomiędzy tymi stanami.

Umysł jest więc niematerialny, nie jest rzeczą, substancją, tylko zbiorem stanów i ich relacji.

Nie jest też pomyłką kategorialną, tylko emergentnym zjawiskiem.

Mózg jest substratem pozwalającym na istnienie tych stanów, jego ogólna budowa i specyficzna struktura określa potencjalnie możliwe stany umysłu.

B28.1. Narodziny kognitywizmu

Metafory na temat umysłu zmieniały się zależnie od epoki:

Starożytna Grecja - umysł jest jak katapulta, napina się i strzela, popatrzcie na dramatyczne napięcia i katharsis.

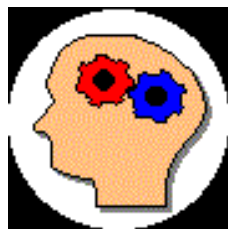


[Leibniz](#): umysł jest jak młyn (ale świadomość to substancja prosta).

[Babbage](#): umysł jest jak krosna tkackie.

[Freud](#): umysł jest jak system hydrauliczny, ciśnienie psychiczne powoduje choroby.

[Sherrington](#): umysł to telefoniczna centrala - czym innym mógłby być?



Wiek XX: umysł podobny jest do komputerowego programu - to podstawowe założenie filozofii kognitywnej.

Wiek XXI: umysł jest jak neurokomputer, układ dynamiczny bez centralnej jednostki.

Dopóki nie będzie neurokomputerów w powszechnym użyciu, ta metafora nie będzie szeroko stosowana.

Przy końcu XIX wieku rozwijała się psychologia oparta na introspekcji, ale nie dała wiarygodnych rezultatów.

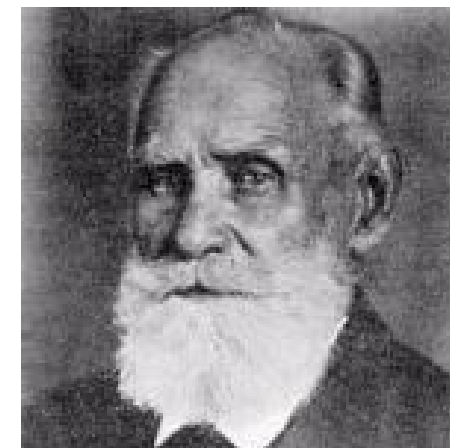
[Psychofizyka](#): [Gustav Fechner](#), [Ernst Weber](#), [Wilhelm Wundt](#), [Ernst Mach](#), [Herman von Helmholtz](#), i inni fizycy z końca XIX wieku.

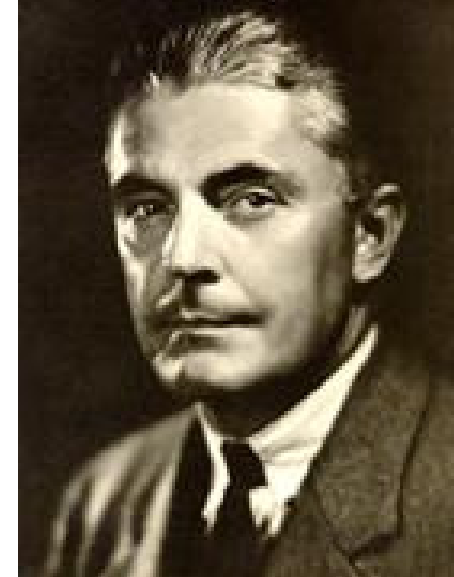
Pomimo wielkich wysiłków i opracowania szczegółowych metod analizy doświadczeń wewnętrznych, "strumienia świadomości", nie udało się niczego uzgodnić pomiędzy różnymi laboratoriami, badania nie dawały jednoznacznych wyników.

Początek XX wieku przyniósł entuzjazm dla badania odruchów warunkowych przez [Iwana Pawłowa](#), a w Polsce jego ucznia, [Jerzego Konorskiego](#).

[Behawioryzm](#) zrobił z psychologii naukę ścisłą, ale cena była wysoka. Umysł, świadomość, nieobserwowalne czynności poznawcze uznano za nienaukowe, bo nie można ich było badać metodami obiektywnymi. Za uniwersalne i obiektywne prawa nauki uznano natomiast relacje pomiędzy bodźcami i reakcjami.

W filozofii [Roman Ingarden](#) wydał książkę "Spór o istnienie świata". Znacznie wcześniej pojawił się spór o istnienie umysłu - czy jest to sensowna koncepcja, czy też wszystko da się wyjaśnić na poziomie odruchów? Czy można było zaprzeczać realności świata wewnętrznego, a potem też świata fizycznego?





[John Watson](#) ([warunkowanie klasyczne](#), reaktywne), [Burrhus Skinner](#), [Edward Thorndike](#) ([odruchy instrumentalne](#)), [Edward Tolman](#) (neobehawioryzm), stworzyli nowy paradygmat badań psychologicznych.

Można badać tylko obiektywnie obserwowalne zachowania, a więc zdefiniowane operacyjnie sytuacje eksperymentalne, ale liczba obserwowalnych wielkości musiała być bardzo niewielka, by taka analiza była możliwa.

Zachowanie jest wynikiem uwarunkowania środowiska.

Niestety, dla złożonych form zachowania brak jest prostych, obiektywnych praw. Sądy i przekonania nie dają się zredukować do dyspozycji organizmu do wypowiedzi i działań odruchowych. Brak było w szczególności postępów dotyczących zrozumienia języka, wypowiedzi nie da się sprowadzić do prostych skojarzeń.

Szczególnie bolesna była lekcja automatycznego tłumaczenia i badań nad analizą tekstów w sztucznej inteligencji, pokazując ogromne trudności naiwnego ujęcia behawiorystów. To właśnie lingwistyka i [Noam Chomsky](#), który napisał książkę "[Struktury Syntaktyczne](#)" (1957) spowodowały odwrót od behawioryzmu i narodziny kognitywizmu.

Neobehawioryzm (P. Suppes, 1975; Edward Tolman, [Clark Hull](#)) dopuścił nieobserwowalne zmienne wewnętrzne, podkreślając centralną rolę bodźca i reakcji, oraz niezależność procesów psychologicznych od uwarunkowań biologicznych i fizjologicznych. Procesy pośredniczące pomiędzy postrzeganiem a reakcjami są ukryte. Teorie zakładające ukryte, nieświadome mechanizmy wywodziły się z psychoanalizy Freuda, ale psychodynamika i inne szkoły psychoterapii nigdy nie były oparte na weryfikowalnych obserwacjach, używały subiektywnych pojęć, które nie dały się obiektywnie zmierzyć ani ocenić.

[Behawioryzm](#) został zastąpiony przez "filozofię nauki o zachowaniu" (określaną też jako [radykałny behawioryzm](#) Skinnera), a potem kognitywną psychologię i neuronauki. Doprowadził też do rozwoju stosowanej analizy behawioralnej (ABA, [applied behavioral analysis](#)) oraz kognitywno-behawioralnej terapii (CBT, [Cognitive-behavior therapy](#)) i wielu odmian psychoterapii, które mają dobre podstawy empiryczne.

Model przetwarzania informacji stał się zrozumiały i popularny dopiero dzięki sukcesom informatyki i projektów w zakresie sztucznej inteligencji ([artificial intelligence](#), AI).

Idee kognitywistyczne rozwinęli eksperci pracujący nad sztuczną inteligencją i psychologią poznawczą.

Zrozumieć znaczy umieć zbudować, dopiero wtedy jesteśmy pewni, że rozumiemy istotne procesy. Inteligentne zachowanie systemów złożonych nie wymaga inteligencji elementów, "duch w maszynie" (a zwłaszcza myśląca substancja) nie jest potrzebny.

[Symbol](#) reprezentuje coś innego niż swoją fizykalną formę. Reguły określają znaczenie symboli przez wzajemne relacje (syntaktykę) i ich odnoszenie do świata (nadając im sens intencjonalny).

W tym ujęciu stany umysłowe - myśli, przekonania, percepcja - to stany obliczeniowe maszynierii przetwarzającej informację.

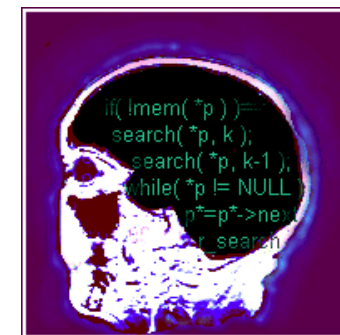
[Funkcjonalizm](#): stany umysłowe mogą być zrealizowane w różny sposób w różnych systemach fizycznych.

Filozofię funkcjonalizmu oprzeć można na trzech ogólnych założeniach:

1. Stany i procesy mentalne można scharakteryzować przez ich funkcję.
2. Funkcje określone są przez relacje między wejściami i wyjściami systemu (organizmu), stanami i procesami.
3. Relacje mają charakter przyczynowo-skutkowy.

Fizyczna realizacja symboli jest drugorzędna, ważne są relacje i przetwarzanie informacji, związki przyczynowe stanów pamięci, wejść (zmysłów) i wyjść (zachowania). Liczy się nie materiał, tylko relacje pomiędzy przyjmowanym stanami, to co rzeczy robią, a nie z czego są zrobione.

Stąd krytyka "[szowinizmu węglowego](#)": inteligencja pozaziemska nie musi być oparta na tych samych związkach biologicznych, np. może być oparta na związkach krzemu, które również pozwalają na skomplikowane oddziaływania.



Emulacja, czyli doskonalenie symulacja jednego systemu (komputera) przez drugi pokazuje, jak różne architektury pozwalają na wykonanie dokładnie tych samych zadań (choć szybkość wykonania może się bardzo różnić).

W filozofii określane jest to mianem "wielokrotnej realizowalności" funkcji ([multiple realizability](#)).

Funkcjonalistyczny minimalizm: co jest istotne dla zachowania funkcji, a co można odrzucić?

Toczy się nieustająca dyskusja, do jakiego stopnia szczegółowości należy modelować układ nerwowy, by zachować procesy istotne dla pracy w mózgu. W tym przypadku funkcje to niższe i wyższe czynności poznawcze oraz sterowanie organizmem.

Odpowiedź zależy od tego, jak bogaty zestaw funkcji chcemy uwzględnić, co już jest warte nazwania umysłem. Mamy tu całe spektrum możliwości, od umysłów prostych zwierząt, ludzi z niedorozwojem mózgu, do w pełni rozwiniętych umysłów filozofów. Spór o nazwy jest tu mało istotny, ważny jest zbiór konkretnych funkcji, które dany mózg może realizować.

W takim ujęciu możemy powiedzieć, że **umysł jest funkcją procesów fizycznych przetwarzających informację**.

Umysł to algorytmiczny proces realizowany przez organiczny mózg lub równoważny mu funkcjonalnie system sztuczny.

Można metaforycznie powiedzieć, że umysł jest częścią tego, co robi mózg, czyli zbiorem procesów, których rezultaty mogą być świadome. A co znaczy "procesy świadome"? Takie, które dostępne są wewnętrznej percepcji, o których wiemy, że są. Możemy je skomentować, wyrazić w postaci słów lub gestów, skojarzyć z innymi. Percepcji przez kogo? Przez inne procesy pozwalające na działanie, sformułowanie wypowiedzi itp. Nie ma potrzeby zakładać, że istnieje homunculus ukryty w naszej głowie, umysł jest "społeczeństwem procesów", jak określił to Marvin Minsky w książce "Society of mind" (1986).

Komputacjonizm to teza określane również jako "**silne AI**" (J. Searle): odpowiednio zaprogramowany komputer jest równoważny umysłowi, ma stany poznawcze.

Przetwarzanie informacji przez mózgi wykracza poza asocjacje, czyli bezpośrednie skojarzenia.

Nie ma tu dualizmu substancji ciała i umysłu, jest tylko proces przetwarzania informacji określający reakcję organizmu.

Oczywiście są jakościowe różnice pomiędzy procesem a jego fizyczną realizacją, np. odczytywaniem bitów z płyty DVD i logiką scen interaktywnej gry; umysł jest na poziomie gry, mózg na poziomie odczytywania bitów i reagowania na nie. Symulowana burza nie zmoczy komputera, jednakże jeśli informacja pochodząca z tej symulacji wpadnie w ruch przesyczone wilgocią powietrze w komorze do testowania zjawisk atmosferycznych powstanie prawdziwa burza.

Oprócz informacji potrzebny jest więc odtwarzacz, oprócz umysłu potrzebne jest więc ciało, dzięki któremu możemy wyrazić stany umysłowe. Taka analogia jest niedoskonała, bo odtwarzacz nie wpływa na informację, a ciało wpływa na formowanie się umysłu. Zapis nutowy to jeszcze nie muzyka, człowiek grający z nut tworzy muzykę, ale przy okazji zmienia siebie, zapamiętuje melodię, wpływa to na jego sposób grania w przyszłości.

Symbole językowe są elementami dyskretnymi (jest ich skończona liczba i są od siebie wyraźnie odróżnialne), procesy percepcji/działania są ciągłe, trudno jest rozróżnić, co już jest innym wrażeniem, a co jeszcze tym samym.

Funkcjonalizm zajmuje się wyższymi czynnościami poznawczymi, językiem i rozumowaniem.

[Franz Brentano](#) (1874) ożywił scholastyczne koncepcje intencjonalności, czyli odniesienia do treści i sensu.

Intencjonalność systemu: symbole są o czymś, bo istnieje izomorfizm relacji obiektów rzeczywistych i umysłowych. Czy to wystarczy?

Odpowiednia zależność przyczynowa stanów umysłu od stanów świata jest warunkiem koniecznym.

Jak symbole mogą reprezentować coś poza samym sobą?

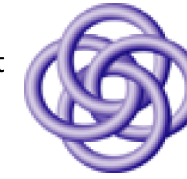
Problem nabierania znaczenia przez symbole jest ważny dla sztucznej inteligencji. Czy robot, którego rozumowanie opiera się na manipulacji symboli, może naprawdę zrozumieć, jak się te symbole wiążą ze zdarzeniami w świecie. Problem ten został szeroko przedyskutowany w pracach S. Harnada (1990, 1993) i innych autorów ([the symbol grounding problem](#)).

Konkluzja tych dyskusji jest taka: znaczenie prostych pojęć wynika z "zagnieżdżenia" w reprezentacjach somatosensorycznych, pozwalających na sterowanie organizmem.

Rozumienie znaczenia symboli związane jest z możliwościami działania, odczuwania, informacjami związanymi z naszymi zmysłami i manipulowaniem swoim ciałem i przedmiotami.

Funkcjonalizm uznaje umysł za algorytmiczny proces przetwarzania informacji.

Sztuczny umysł wymaga ogromnych baz wiedzy, metod reprezentacji i korzystania z wiedzy. [System ekspertowy CYC](#) jest próbą realizacji zdrowego rozsądku w oparciu o miliony reguł. Czy takie systemy komputerowe symulują jedynie inteligentne działanie, czy naprawdę mają jakieś stany poznawcze? Czy można powiedzieć, że "zdają sobie sprawę" z tego co robią?



B28.2. Argumenty Turinga



[Alan Turing](#) (1912-1954)

Alan Turing, twórca teorii automatów, matematycznych podstaw teorii obliczeń, w pracy z 1936 roku podał teoretyczny model komputera („[automatu Turinga](#)”), czyli komputerów, które teraz powszechnie używamy. Prowadził rozważania nad obliczalnością, był też zaangażowany w budowę maszyn liczących i deszyfrujących. [Historycy oceniają](#), że dzięki zbudowanej pod jego kierunkiem maszynie odczytującej kod Enigmy (polscy kryptolodzy, Marian Rejewski, Henryk Zygalski oraz Jerzy Różycki, mieli tu również wielki wkład) wojna trwała o ponad rok krócej.

Jedną z najważniejszych w historii nauki idei jest pomysł Turinga, że z prostych elementów, działających mechanicznie i bez żadnego celu, może wyłonić się złożone, inteligentne zachowanie. Kilku programistów, z których każdy rozwija mały programik i nie rozumie całości, może stworzyć program, który zwycięży z mistrzem świata w szachy. Nikt z twórców nie ma wysokich kompetencji w tej grze, podobnie jak żaden termit nie wyobraża sobie, że buduje wielki kopiec. Zrozumienie nie jest konieczne, by stworzyć plan budowy, może się on wyłonić z interakcji prostych elementów. Marvin Minsky napisał książkę "Society of mind" (1986) pokazując jak z wielu prostych funkcji, które nazwał "agentami", może powstać inteligentny umysł.

W artykule "[Computing Machinery and Intelligence](#)" opublikowany w czasopiśmie "Mind" w 1950 roku Turing uznał pytanie "czy maszyny mogą myśleć?" za zbyt mało precyzyjne. Trudno jest jednoznacznie zdefiniować pojęcie "maszyna" i "inteligencja", dlatego zaproponował praktyczny test na określenie inteligencji. Zamiast powtarzać, że nie rozumiemy czym jest świadomość lub inteligencja i zajmować się próbą definicji pojęć należy zdefiniować pojęcia operacyjnie, mówić o konkretach. Trzeba określić test dający jednoznaczne wyniki, co ma zrobić maszyna, by uznać ją za inteligentną lub świadomą.

[W oryginalnym teście Turinga](#) biorą udział 3 osoby. Osoba A próbuje zmylić pytającego tak, by nie zgadł jego lub jej płci, a B ma za zadanie pytającemu pomagać. Jeśli rolę A będzie grać maszyna, czy pytający będzie się równie często mylił? Obecnie przez "test Turinga" rozumie się zwykle próbę zmylenia rozmówcy, udawania przez program komputerowy w czasie konwersacji, że jest człowiekiem.

Od 1991 roku odbywają się zawody z programami do konwersacji ([chatbotami](#)) organizowane w ramach konkursu o [nagrodę Loebnera](#). Sędziowie prowadzą kilkuminutową konwersację z człowiekiem lub maszyną i następnie przypisują im punkty. Sędziowie są przypadkowi i dlatego trudno jest porównać rezultaty z różnych lat. Doniesienia prasy, że jakiś chatterbot przeszedł test Turinga bo na wsi w Indiach oszukał dużą grupę ludzi, którzy nie mieli dotychczas do czynienia z taką technologią, nie mają żadnej wartości. Postępy w tym konkursie przez wiele lat były powolne, ale w roku 2022 generatywne modele językowe oparte na wielkich sieciach neuronowych pozwoliły

rozwiązać ten problem.

Czy jednak test Turinga wystarczy by uznać, że maszyna jest inteligentna i ma umysł podobny do ludzkiego?



Turing był przekonany, że do 2000 r. będą myślące maszyny o pamięci rzędu 100 MB które przejdą taki 5-minutowy test. Pomylił się o 20 lat i zakładał 100 razy za małe wymagania dotyczące pamięci. Postępy w sztucznej inteligencji były początkowo szybkie i wydawało się, że wystarczy szybszy komputer i większa pamięć by powstały inteligentne maszyny. W połowie lat 1960 już tak nie myślano, zbyt wiele ambitnych projektów poniosło klęskę. Zrozumienie złożoności procesów poznawczych stopniowo rosło i informatycy zaczęli się interesować psychologią poznawczą a potem neurobiologią, tworząc coraz bardziej złożone systemy kognitywne. Choć zrobiono wielkie postępy, to możliwości stworzenia myślących maszyn dorównujących ludziom były uważane do trzeciej dekady XXI wieku za kontrowersyjne.

Turing w swojej pracy omówił szereg zarzutów przekonując, że zaproponowany przez niego test wystarczy dla uznania, iż mamy do czynienia z rzeczywistym myśleniem.

1. Zarzut teologiczny: myślenie jest funkcją nieśmiertelnej duszy, a tej maszyna mieć nie może.

Odp. Turinga: nie można tego zarzutu poważnie traktować.

Odp. teologa (J. Życiński): Bóg może wszystko, skąd my możemy wiedzieć, co maszyna mieć może?

Wersja współczesna to poszukiwanie niematerialnych wpływów umysłu na mózg, np. psychony wokół dendronów, wymyślone przez [Johna Ecclesa](#). Jednak nie znaleziono żadnego śladu takich zjawisk, a argumenty i eksperymentalne dowody pokazujące, jak procesy zachodzące w mózgu wywołują stany mentalne, są obecnie niepodważalne.

2. Lepiej schować głowę w piasek, bo konsekwencje powstania takich maszyn będą straszne (przede wszystkim dla naszego antropocentryzmu?).

Strach przed "terminatorem" jest w różnej formie często spotykany, odpowiedzialny za popularność poglądów przeciw rozwojowi sztucznej inteligencji i przekonanie o nadchodzącej [wojnie z robotami](#).

Efekty takiego myślenia to różne konfabulacje i strach, ale też przekonanie o beznadziejności badań nad umysłem, nadzieje na to, że problem nigdy nie zostanie zrozumiany (np. Turski 1996). Szachiści przez lata wyśmiewali [programy komputerowe do gry w szachy](#), aż do przegranej Kasparowa z systemem komputerowym Deep Blue (1997). Około roku 2006 poziom gry w szachy programów komputerowych [przekroczył ludzkie możliwości](#), obecnie jest daleko poza możliwościami człowieka. Podobnie było z grą w go, która miała się opierać dominacji maszyn przez kolejne 100 lat, ale już w 2017 roku AlphaGo osiągnęła nadludzki poziom. Więcej o historii rozwoju takich programów [w tym wykładzie](#).



3. Zarzut matematyczny opiera się na [twierdzeniu Gödla](#), [Churcha](#), jak i samego Turinga o zdaniach nierozstrzygalnych.

Co jakiś czas zarzut ten powraca w innej formie (por. "[Nowy umysł cesarza](#)" R. Penrose). Rezultatem jest wniosek, że jeśli ludzie potrafią rozwiązywać problemy

nierozstrzygalne, to umysł nie działa w sposób, który można opisać za pomocą algorytmu. Konieczna jest nowa fizyka procesów nieobliczalnych, mikrotubule a może i kwantowa grawitacja.

Turing odpowiedział, że człowiek też się myli, konkretna maszyna tak jak i konkretny człowiek nie odpowie na wszystkie pytania, ale jeśli rozważyć wszystkie możliwe maszyny nie ma takich ograniczeń.

Ograniczenia umysłu człowieka, związane z pojemnością pamięci roboczej, pamięci długotrwałej, szybkości działania, skończonego czasu życia, zdolności do skojarzeń itp. są dużo silniejsze niż ograniczenia maszyny Turinga! Z powodu tych ograniczeń umysł musi sobie radzić, ciągle się ucząc, tworząc hierarchię coraz bardziej złożonych konstrukcji. Nie używamy ustalonego algorytmu, ale dostosowujemy swoje działania w zależności od sytuacji. To samo jednak robią teraz również programy sztucznej inteligencji, tylko szybciej, kojarząc więcej faktów, lepiej pamiętając.

Argumenty matematyczne dotyczą jedynie systemów z logiką klasyczną. W sztucznej inteligencji używa się nietradycyjnych logik, pozwalających radzić sobie z wiedzą sprzeczną, proceduralną lub niepełną, a także systemów uczących się na podstawie danych empirycznych, niedziałających w oparciu o logikę. Nie stosuje się do nich twierdzenie Gödla.

Wiele odkryć było sprawą przypadku. Maszyny mają całkiem inne ograniczenia niż ludzie, ale też zdolne są do robienia odkryć, u podstaw których leży przypadek i rozumowanie. Obliczenia wykorzystujące przypadek, czyli generowanie rozwiązań zgodnie z procedurą Monte Carlo, mają obecnie bardzo liczne zastosowania.

Ograniczenia związane z twierdzeniami o nierozstrzygalności dotyczą bardzo specyficznych pytań związanych z konstrukcją samego systemu logicznego. Czy są to jednak interesujące pytania dla sztucznej inteligencji czy robotów, mających działać w inteligentny sposób? Tego nikt nie pokazał.

Wniosek z matematycznych argumentów krytykujących możliwości maszyn: nie można stworzyć maszyny wszechwiedzącej.

4. Świadomość, emocje, uczucia to rzeczy niedostępne maszynom.

Po przegranej Kasparowa pisano, że komputer szachowy [Deep Blue](#) nawet nie potrafi się cieszyć z wygranej! Programy wyspecjalizowane w wykonywaniu jednego zadania nie są zdolne do takich reakcji.

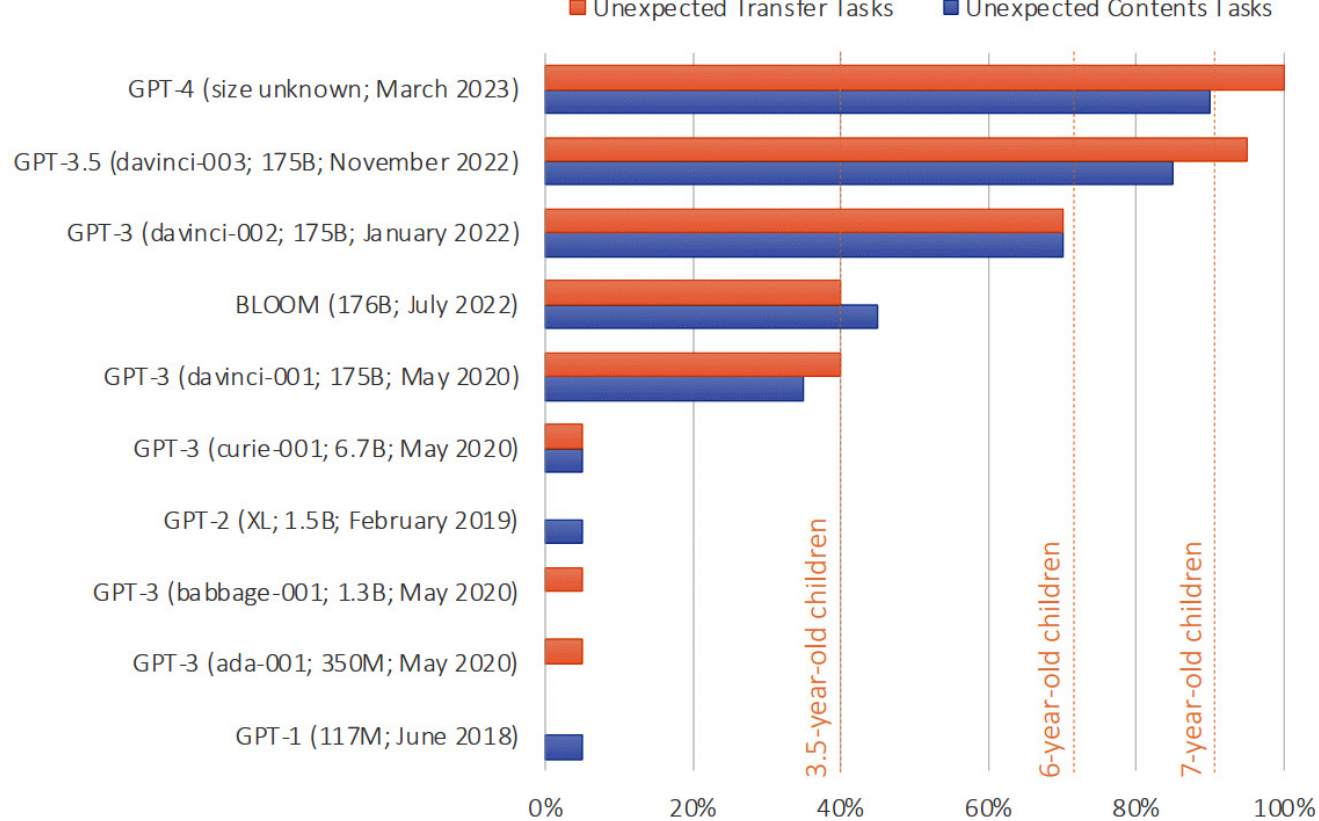
W filozofii umysłu takie zarzuty zostały sformułowane w 1995 roku w pracy Dawida Chalmersa i znane są pod nazwą [trudnego problemu](#) świadomości, wkrótce [omówimy go dokładnie](#).

Turing: o umysłach innych ludzi wiemy tylko z obserwacji, więc jeśli program przejdzie test i nie uznamy, że myśli, pozostanie nam tylko [solipsyzm](#). Ten pogląd został zaatakowany za pomocą eksperymentu myślowego z "chińskim pokojem", który też [dokładnie przedyskutujemy](#).

Reakcje emocjonalne są znacznie prostsze niż rozumienie pojęciowe świata. Roboty ich nie miały, bo nikt się tym nie zajmował, ale od końca 20 wieku to się zmieniło. Powstała dziedzina nazywana [informatyką afektywną](#). W 2006 roku pionier sztucznej inteligencji Marvin Minsky wydał książkę [Emotion Machine](#). Oczywiście nie chcemy robotów wpadających w histerię lub wściekłość, ale emocje z wielu względów będą im potrzebne. Na rynku już są roboty mające analizować emocje i reagować emocjonalnie, przydatne w terapii różnych zaburzeń afektywnych.

Ostatnie testy wielkich modeli językowych (large language models) pokazują, jak w takich modelach spontanicznie [pojawia się teoria umysłu](#), czyli rozumienie punktów widzenia różnych osób. W ciągu dwóch lat coraz doskonalsze modele, które nie były uczone rozwiązywania zadań tego typu, zaczęły wykazywać zrozumienie sytuacji typowe dla dorastających dzieci. Obserwacje rozwoju teorii umysłu u głuchych dzieci pokazały, że nauka języka migowego potrzebna jest do rozwoju teorii umysłu i rozumienia fałszywych przekonań ([Pyers, Senghas, 2009](#)). Dokładniejsze badania rozumienia stanów mentalnych pokazały, że wiele modeli GPT-4 już lepiej rozumie subtelności ludzkich zachowań, fałszywe przekonania, ironię w głosie lub wypowiedzi, aluzje (zamierzone znaczenie uwagi i działanie, które próbuje wywołać), rozpoznawanie dziwnych historii (wyjaśnij, dlaczego postać mówi lub robi coś, co nie jest dosłownie prawdziwe, GPT4). Jedynie w przypadku zachowań faux pas (zachowania nieodpowiednie w danym kontekście) ludzie radzą sobie trochę lepiej (Strachan i inni, [Nature Human Behavior](#) 2024).





5. Argumenty dotyczące różnych niemożliwości.

Maszyny nie są zdolne do: samodzielnej inicjatywy, humoru, zakochania się, rozkoszowania lodami ...

Czyli maszyna jest głupia (jak każdy widzi) i już zawsze taka być musi? Są to błędne generalizacje doświadczeń z obecnie istniejącymi maszynami.

Pamięć, liczenie, rozumienie mowy, ekspertyza, wszystko to, co robią obecnie maszyny rutynowo, było kiedyś nie do wyobrażenia. Kiedy pojawiły się pierwsze fonografy wielu ludzi szukało pod stołem ukrytego człowieka wydającego dźwięki mowy.

Nie jesteśmy ani zbyt głupi, by stworzyć inteligentne maszyny, ani też aż tak mądrzy, by zrobić to w krótkim czasie ...

6. Maszyna sam nie może nic stworzyć.

Teraz już może!

[Kreatywność maszyn](#) jest od niedawna ważnym zagadnieniem w sztucznej inteligencji. Przeceniamy kreatywne zdolności większości ludzi. Nawet Albert Einstein napisał o sobie: człowiek nie widzi rozwiązania, dopóki się o nie nie potknie. Sam potykał się często, ale też często błądził.

Odkrywanie wiedzy w danych (data mining) i systemy ekspertowe, odkrywające nowe rozwiązania, znane są od dawna, są nawet patenty procesów przemysłowych, odkrytych za pomocą algorytmów. W szachach, matematyce, a nawet malowaniu obrazów czy komponowaniu muzyki osiągnięto interesujące rezultaty.

Dlaczego maszyna nie mogłaby mieć pamięci, uczyć się i kojarzyć faktów, aktywnie szukać nowej wiedzy? Czy jakieś prawo przyrody tego zabrania? Czy też nasze wrażenie wynika z tego, że nie widzimy jeszcze takich maszyn?

[Przykłady rozwiązywania problemów](#) przez komputery wykorzystujące algorytmy ewolucyjne, oraz referat na temat [neuroestetyki](#).



Maurice Conti, [The incredible inventions of intuitive AI](#), czyli niewiarygodne wynalazki sztucznej inteligencji (TED).

7. Układ nerwowy nie działa w sposób dyskretny, a więc opisywalny za pomocą skończonej liczby symboli, na których operują komputery.

Ten zarzut wiąże się z pytaniem, czy inteligencja wymaga modelowania procesów neuronowych bardzo dokładnie, na poziomie pojedynczych molekuł, czy też uproszczone biofizyczne modele neuronów wystarczą? Zwykle przy modelowaniu układu nerwowego interesują nas dyskretne impulsy wysyłane przez neurony, a często tylko ich liczba w jednostce czasu (rate coding). Możliwe są bardzo dokładne aproksymacje zachodzących w mózgu procesów, ale nie muszą wcale być konieczne, by wytworzyć funkcje umysłowe, np. pamięć kontekstowa pojawiają się w stosunkowo prostych modelach.

Analogowe komputery zostały wyparte przez wygodniejsze komputery cyfrowe, ale buduje się obecnie analogowe [systemy neuromorficzne](#) zużywające bardzo mało energii i nadające się do budowy neurokomputerów, w których parametry również nie zmieniają się w dyskretny sposób.

8. Zachowanie człowieka nie da się opisać przy pomocy reguł.

Zachowanie wielu programów również jest zbyt złożone, by je opisać za pomocą reguł! Nawet jeśli program był złożonym zbiorem reguł, nie da się tego wykryć na podstawie samej obserwacji zachowania. Turing gotów był się założyć, że obserwując działanie programu nikt nie zgadnie reguł sterowania, które zaprogramował. Algorytmy stochastyczne i probabilistyczne rozwiązują zagadnienia zbyt trudne dla algorytmów deterministycznych. Nie spodziewamy się, by systemy regułowe były dobrymi modelami umysłu, architektury kognitywne nie muszą być oparte na regułach (przykłady opisane są np. [w tej pracy](#)).

9. Argument związany z postrzeganiem pozazmysłowym.

Turing uznał ten argument za warty wzmianki ze względu na doświadczenia z telepatią i kartami Zenera, o których było w jego czasach głośno w prasie, ale od 1950 roku [nie widać tu postępu](#).

Nawet jeśli zjawisko istnieje, to jest tak trudne do wykrycia, że jest zbyt subtelne, nie może więc mieć wpływu na zwykłe myślenie.

Należy rozróżnić abstrakcyjne możliwości przetwarzania informacji od własności sprzętu realizującego ten proces.

Użyteczna analogia: **umysł tak się ma do mózgu jak program do komputera.**

Z jednej strony mamy sprzęt (hardware, materię), z drugiej strony program (software, abstrakcyjne procesy, formę), realizowany przez sprzęt.

Czy umysł można uznać za rodzaj programu komputerowego? Co to naprawdę znaczy?

[Program komputerowy](#) zapisuje w sposób abstrakcyjny procesy sterujące sprzętem, w tym sensie każdy proces można opisać za pomocą programu. Ale program to nie jest jeszcze proces! Dopiero **program + architektura wykonująca ten program (np. robot) tworzy właściwy proces**. Program to nie tylko reguły, ale również wiedza zwarta w dostępnych strukturach danych.

Test Turinga to warunek konieczny, ale czy wystarczający, by uznać program za umysł?

Wygrana programu [IBM Watson](#) w grę Jeopardy pokazuje, że maszyna może sensownie odpowiadać na złożone pytania i w opinii wielu pewnie zachowywać się w czasie konwersacji podobnie do człowieka. Czy jednak ta maszyna coś rozumie? Czy samo kojarzenie faktów pozwala uznać, że mamy jakieś stany poznawcze?

Uznanie, że mamy do czynienia ze sztuczną inteligencją, nie oznacza jeszcze, że jest to system świadomy, który może sobie z czegoś "zadawać sprawę" i mieć związane z tym wrażenia. To temat dalszych rozważań.

Alan Turing jest uważany za ojca informatyki, współczesne komputery to realizacja idei "[maszyny Turinga](#)". Historycy uważają, że dzięki pracy Turinga nad rozszyfrowaniem [kodu Enigmy](#) II wojna światowa trwała o dwa lata krócej. W złamaniu kodu pomogła początkowo [polska grupa matematyków](#), przechwycenie wczesnego modelu Enigmy, zbudowanie jednego z pierwszych komputerów do [łamania kodów \(Colossus\)](#), i późniejsze zdobycie 5-wirnikowego modelu, jednakże stworzenie programu, który w ciągu jednej doby pozwalał odszyfrować przesyłane wiadomości, było głównie zasługą Turinga.

Niestety w 1952 roku Turing przyznał się do homoseksualizmu, za co groziło więzienie lub chemiczna kastracja. Podawano mu estrogen, urosły mu piersi i po dwóch latach takiej kuracji prawdopodobnie popełnił samobójstwo. W 2009 roku Gordon Brown, premier Wielkiej Brytanii, przeprosił za to nie tylko Turinga, ale również 49 tysięcy Brytyjczyków, których potraktowano tak samo. Dopiero przy końcu 2013 roku na skutek powszechnej petycji królowa brytyjska ułaskawiła Alana Turinga.



B28.3. Funkcjonalizm i warianty filozofii kognitywnych

Funkcjonalizm to teoria filozoficzna, na której w znacznej mierze opiera się **kognitywistyka**.

Podstawy filozoficzne funkcjonalizmu stworzyli [Hilary Putnam](#), [Jerry Fodor](#), [Zenon Pylyshyn](#), a jego ujęcie od strony sztucznej inteligencji [Allen Newell](#) i [Herbert Simon](#).

Pojęcia mają precyzyjne znaczenie tylko w obrębie teorii. Tybetańscy filozofowie, pytani o definicję świadomości, odpowiadali - ale według jakiej filozoficznej szkoły?

W ramach teorii Newella i Simona podać można precyzyjne definicje pojęć dotyczących umysłu (Newell, 1990).

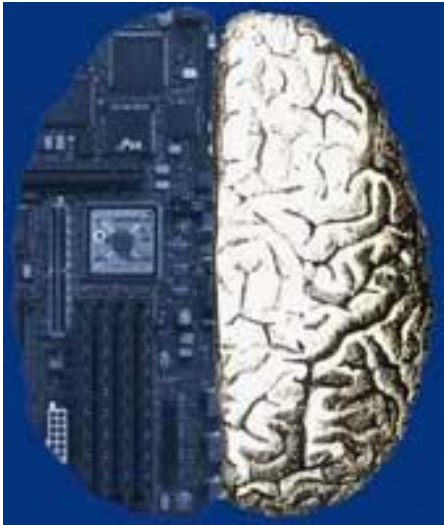
Umysł to system kontrolny określającym zachowanie organizmu w oddziaływaniach ze złożonym środowiskiem.

Na umysł składa się zbiór funkcji określających odpowiedzi organizmu na sytuacje środowiska. Odpowiedzi różnią się zależne od stanu środowiska jak i indywidualnej historii danego umysłu. System kontrolny realizuje cele, a to wymaga wiedzy. Umysł jest więc systemem kontrolnym posiadającym cele i wykorzystującym wiedzę.

Częstą **pomyłką jest porównywanie różnych systemów na różnym poziomie**, np. twierdzenie, że komputer liczy, a mózg "wie".

A może jest odwrotnie? Komputer pisze (zamienia głos na tekst), a neurony w mózgu tylko zliczają impulsy ... Komputer reaguje natychmiast, a mózgi potrzebują sporo czasu na zliczanie impulsów. Od razu widać, że coś tu mylimy.

Zachowanie może być opisywane przez intencje, cele, wiedzę. Poziom intencjonalny to opis systemu działającego w oparciu o wiedzę. Porównajmy sztuczne systemy oparte na wiedzy i biologiczne na jednakowych poziomach:



| | System sztuczny | Człowiek |
|-----------|-------------------------------------|---------------------------------|
| Poziom: | Obwody scalone | Biochemiczny |
| Substrat: | Atomy i elektrony półprzewodników | Neurocząsteczki |
| Prawa: | Fizyka ciała stałego | Fizyka molekularna |
| Poziom: | Obwody elektryczne | Neurony |
| Substrat: | Napięcia/prądy/zjawiska elektryczne | Zjawiska elektryczne |
| Prawa: | Prawa Ohma, Kirchoffa, Faradaya | Prawa Ohma, Kirchoffa, Faradaya |
| Poziom: | Architektura sprzętowa | Przetwarzanie sygnałów |
| Substrat: | Obwody logiczne | Funkcjonalne grupy neuronów |

| Prawa: | Logika | Neurofizjologia |
|-----------|---------------------------------------|-------------------------------|
| Poziom: | Uniwersalny komputer | Mózg |
| Substrat: | Ciągi bitów | Impulsy elektryczne |
| Prawa: | Teoria obliczeń | Wynikające z ewolucji |
| Poziom: | Systemy oprogramowania | Zachowania wyuczone, obyczaje |
| Substrat: | Struktury danych i programy | Transformacje sensomotoryczne |
| Prawa: | Interpretacja syntaktyczna instrukcji | Dynamika złożonych układów |
| Poziom: | Systemy przetwarzające wiedzę | Umysły |
| Substrat: | Wiedza | Świat wewnętrzny |
| Prawa: | Zasady racjonalnego działania | Prawa psychologii |

Systemy Oparte na Wiedzy (SOW) są matematyczną abstrakcją. Systemy ekspertowe (SE) wykorzystujące różne formy reprezentacji wiedzy do rozumowania są ich aproksymacją.

Jeśli nowy fakt nie zmieni bazy wiedzy SE, bo już jest w niej w jakiejś formie zawarty, można sensownie powiedzieć: "ten SE już to wiedział". Czy w tym przypadku "wie" to tylko metafora językowa, czy coś więcej?

Wiedza w SOW to przekonania i założenia, niekoniecznie prawdziwe, bo oparte na obserwacjach empirycznych, a nie abstrakcyjnych założeniach matematycznych. Rozumowanie w oparciu o taką wiedzę może być oparte na różnych rodzajach logiki.

[John McCarthy](#), jeden z twórców AI, napisał, że nawet termostaty mają przekonania "za gorąco", "za zimno". To oczywiście skrajny przykład, bo termostat nic nie wie i automatyczne zachowania nie mają charakteru przekonań. Można sobie jednak wyobrazić coraz bardziej złożone urządzenia lub organizmy, których zachowanie w coraz większym stopniu będziemy interpretować jako posiadanie przekonań. Kleszcz, który czeka miesiącami, reagując tylko na temperaturę i poziom kwasu mlekowego, zachowuje się niemal jak termostat. Pies ma oczekiwania, a zachowanie szympansa ma wszystkie cechy wewnętrznych przekonań, które można mu przypisać.

Reprezentacja wiedzy wymaga odpowiedniego substratu, zachowującego relacje pomiędzy wzorcami w systemie sztucznym a rzeczywistymi, reprezentowanymi obiektami. Reprezentacja wiedzy to zakodowana informacja, często w postaci symbolicznej, lub schematu obrazkowego (rep. ikonograficzna).

Symbol to klasa abstrakcji wszystkich znaków, które reprezentują to samo.

Fizyczny system symboliczny realizuje uniwersalny model obliczeń. Taki system musi zawierać pamięć, operacje działające na strukturach symbolicznych i procesy interpretujące (I/O).

Systemy symboliczne aproksymują możliwości Systemów Opartych na Wiedzy.

Semantyka pojęć pierwotnych jest wynikiem oddziaływania systemu ze środowiskiem. Sens zawarty jest we wskazywaniu na indywidualne obiekty lub całe ich klasy (np.

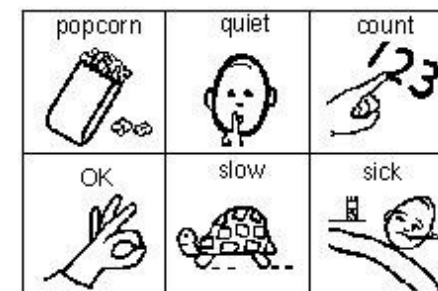


narzędzia lub zwierzęta) i związane z nimi możliwości działania, lub ich wzajemne relacje. Dla pojęć abstrakcyjnych, nie związanych bezpośrednio z percepcją i działaniem, sens wynika z odniesienia do pojęć pierwotnych, działań "wewnętrznych".

SOW oddziałuje ze środowiskiem, jego reakcje tworzą opis zachowania się systemu. SOW podejmuje działania by spełnić swoje cele korzystając z posiadanej wiedzy. Komputery umożliwiają realizację modeli SOW, ale w podejściu funkcjonalnym nie muszą to być komputery oparte na elektronice. Dawniej były komputery lampowe, ale im bardziej złożone obliczenia tym bardziej złożona musi być struktura sprzętu, na którym można je prowadzić. W przyszłości możemy mieć komputery optyczne lub bioorganiczne.

Fizyczny system symboliczny to:

1. zbiór fizycznych desygnatów symboli, które
2. podlegają manipulacjom za pomocą wewnętrznych reguł,
3. reguły te są również ciągami fizycznych desygnatów symboli.
4. Transformacje symboli zależne są jedynie od ich formy, a nie treści i
5. transformacje są złożeniem kombinacji różnych reguł.
6. Symbole mogą być proste (atomowe) lub złożone (molekularne),
7. a reguły ich rozpoznawania i łączenia muszą być "semantycznie interpretowalne".



SOW składające się z symboli i reguł, które pozwalają na ich transformacje, jest więc zbiorem jakichś fizycznych struktur, które przyjmują różne formy odpowiadające symbolom; "fizyczne" oznacza tu, że mogą to być rozkłady pobudzeń neuronów, lub aktywacje elektronicznych elementów, a nie abstrakcyjne formuły.

Semantyczna interpretowalność jest to własność całego systemu, pozwalająca określić sens symboli przez wskazanie ich wzajemnych powiązań oraz powiązań z tym, co jest przez nie symbolizowane.

Wiedział to już [Ferdinand de Saussure](#), twórca językoznawstwa, pisząc "Żadne słowo nie ma sensu, który dałby się zidentyfikować niezależnie od kontekstu, w którym się znajduje".

Złożone reguły gramatyczne powstają z najprostszych (Fodor nazywa to "**kompozycyjnością**"). Transformacje symboli da się rozłożyć na reguły najprostsze ([Jerry Fodor](#) nazywa to "**dekompozycyjnością**"). Jedynie takie systemy są semantycznie interpretowalne i "symboliczne".

Nie oznacza to, że system symboliczny będzie sam z siebie automatycznie dokonywał interpretacji, musi być częścią odpowiedniej architektury poznawczej, kontrolującej analizę wypowiedzi. **Kombinatoryczna produktywność** oznacza zdolność do tworzenia nieskończenie wielu kombinacji elementów.

łańcuchy symboli tworzą myśli, nastawienia, intencje i instrukcje działania. Niezależność od implementacji (te same procesy mogą być w różny sposób realizowane) oznacza autonomiczność poziomu symbolicznego. Inne poziomy organizacji widać tylko jako skutek błędów w działaniu (wynik uszkodzeń mózgu lub błędów programu). Jednak nie jest wcale jasne, że na poziomie symbolicznych transformacji da się kontrolować wszystkie skojarzenia tak, by stały się podobne do naturalnych. Być może dopiero reprezentacja w sieciach neuronowych pozwoli na osiągnięcie podobnych skojarzeń.

Co to jest **inteligencja**? Możemy teraz zdefiniować to pojęcie w obrębie teorii systemów opartych na wiedzy.

Iloraz Inteligencji IQ (Intelligence Quotient) to liczba mająca ocenić poziom [inteligencji](#) wynikający ze sprawności rozwiązywania [testów inteligencji](#).

Skąd jednak pomysł, że inteligencja, rozumiana jako zdolność do rozwiązywania różnych problemów, da się zmierzyć za pomocą jednego współczynnika? **Ogólny czynnik inteligencji g** jest konstrukcją psychologiczną mającą wyjaśnić korelację wyników różnych testów psychometrycznych; osoby uznawane za inteligentne mają dobre wyniki w

testach różnego rodzaju.

Czynnikiem za tym stojącym jest sprawnie działający mózg, ale sprowadzanie wszystkiego do jednego wymiaru jest dość drastycznym uproszczeniem.

[Pierwsze testy IQ](#) (Binet-Simon, 1905) używane były do oceny zdolności werbalnych w szkole w celu identyfikacji dzieci zapóźnionych w rozwoju. Rozkwit testów IQ nastąpił pod wpływem eugeniki w USA na początku XX wieku.

[Teoria inteligencji wielorakiej](#) rozwijana przez [Howarda Gardnera](#) wyróżnia 8 różnych aspektów inteligencji, stanowiących predyspozycje do różnego rodzaju zawodów:

1. ruchowo-kinestetyczna: tancerzy, aktorów, chirurgów, adeptów sztuk walki;
2. interpersonalna (społeczna): polityków, sprzedawców, nauczycieli;
3. językowo-werbalna: prawników, dziennikarzy, pisarzy;
4. logiczno-matematyczna: naukowców, matematyków, inżynierów;
5. wzrokowo-przestrzenna: architektów, artystów;
6. muzyczna: muzyków, kompozytorów;
7. intrapersonalna: filozofów, psychologów;
8. przyrodnicza (relacji z otoczeniem): ogrodników, rolników, naturalistów.

Można też mówić o inteligencji emocjonalnej, moralnej, duchowej i innych formach inteligencji. W każdym z tych przypadków nie wystarczy sama zdolność do kojarzenia, ale liczy się też sprawne działanie zmysłów, pamięci, orientacji przestrzennej, zdolności manualnych. Trudno być inteligentnym tancerzem jeśli nie ma się poczucia rytmu.

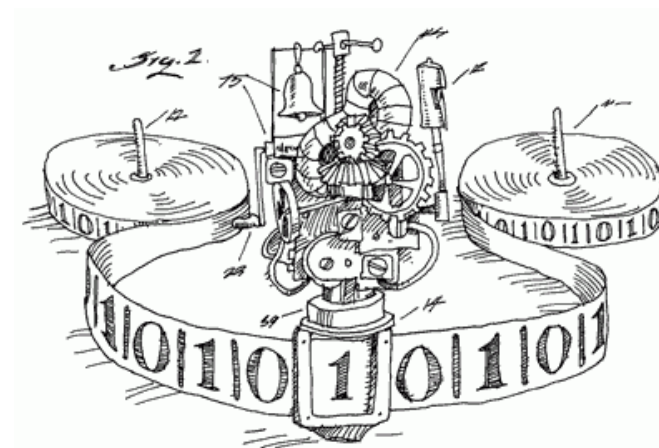
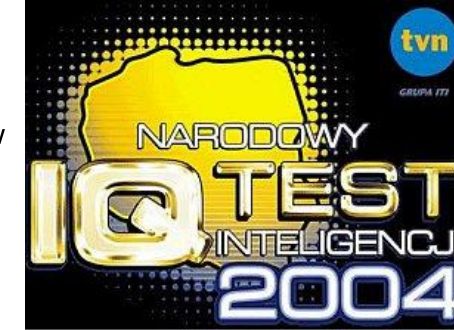
Warianty filozofii kognitywnych

[Fizykalizm](#) to jedno z głównych założeń neopozytywizmu: pełny opis zjawisk możliwy jest za pomocą języka fizyki, mierzalnych, intersubiektywnie uzgodnionych wielkości.

[Psychofunkcjonalizm](#), opisał głównie [Jerry Fodor](#) w książkach "The language of thought" (1975), "Psychosemantyka" (1987). Podstawowe założenia tej teorii to:

1. Stany psychologiczne (mentalne) to stany mózgu rozumiane jak procesy obliczeniowe, realizowane przez odpowiadające im procesy fizyczne.
2. Modelem uniwersalnego komputera jest [maszyna Turinga](#), abstrakcyjny model urządzenia realizującego algorytmy obliczeniowe.
3. Maszyny są sobie równoważne gdy końcowe stany ich obliczeń są identyczne.
4. Realizacja fizyczna sposobu obliczeń jest nieważna, istotny jest wynik.
5. Na poziomie funkcjonalnym możemy twierdzić, że komputery są przekonane, "wierzą", że $2+2=4$.
6. To, co dzieje się w mózgu na poziomie neurofizjologicznym, słabo wpływa na nasze przekonania (dopóki wszystko jeszcze działa).
7. Tym samym stanom psychicznym odpowiadać mogą różne stany mózgu (zbiór różnych stanów mających podobne funkcjonalne znaczenie).

[Teoria identyczności i typów](#) ([identity theory](#)), określana jako [reduktywny materializm](#), teoria identyczności typów, teoria identyczności umysłu lub mózgu-umysłu głosi, że określony typ stanów umysłowych jest identyczny z określonym typem stanów neurofizjologicznych.



Z psychofunkcjonalizmu wynika, że istoty o różnych mózgach powinny mieć różne umysły, różnice budowy sprzętu mają wpływ na procesy realizowane przez ten sprzęt. Stany psychologiczne zależą od relacji pomiędzy stanami mózgu (lub ogólnie sprzętu obliczeniowego), koniecznymi do ich realizacji, ale nie są z nimi tożsame. Stany umysłu dają się charakteryzować przez zależności typu bodziec-reakcja (behawioryzm) poszerzone o stany wewnętrzne, które nie mają bezpośredniego odzwierciedlenia w zachowaniu (ale mogą być widoczne w obrazowaniu mózgu). Teoria typów uznaje odpowiedniość typów, stanu psychologicznego i stanu fizycznego.

Możliwe **relacje odpowiedniości mentalno-fizycznej**.

- **Nomologiczny dualizm**, który uznaje ściśle mentalno-fizyczne korelacje, ale uznaje je za przejaw paralelizmu, a nie identyczności.
- **Anomalny dualizm**, np. Kartezjański, nieuznający korelacji mentalno-fizycznych, ze względu na ontologiczną niezależność.
- **Neutralny monizm**, czyli monizm ani mentalny, ani fizyczny, tylko substancja "czystego doświadczenia" (B. Russell).
- **Monizm nomologiczny**, czyli teoria identyczności typów, lub fizykalizm typów, uznaje ścisłą korelację zjawisk mentalnych i fizycznych, które można uznać za identyczne (**nomologiczny** oznacza odnoszący się do praw ogólnych natury lub logiki).
- **Monizm refleksyjny** (Velmans 2007) jest współczesną wersją idei [Spinozy](#).
- **Anomalny monizm**, uznający jedność zjawisk mentalnych i fizycznych, ale zaprzeczający możliwości nomologicznej redukcji (Davidson, 1970).

[Anomalny monizm](#) zwraca uwagę na trzy istotne zasady:

- Oddziaływania przyczynowe zachodzą w obydwie strony: stany mentalne wywołują działania organizmu i stany organizmu wpływają na działania mentalne.
- Nomologiczny charakter przyczynowości: związkami przyczynowymi rządzą ściśle prawa.
- Anomalny charakter relacji mentalno-fizycznych, czyli brak praw psychofizycznych, które by takimi związkami rządziły.

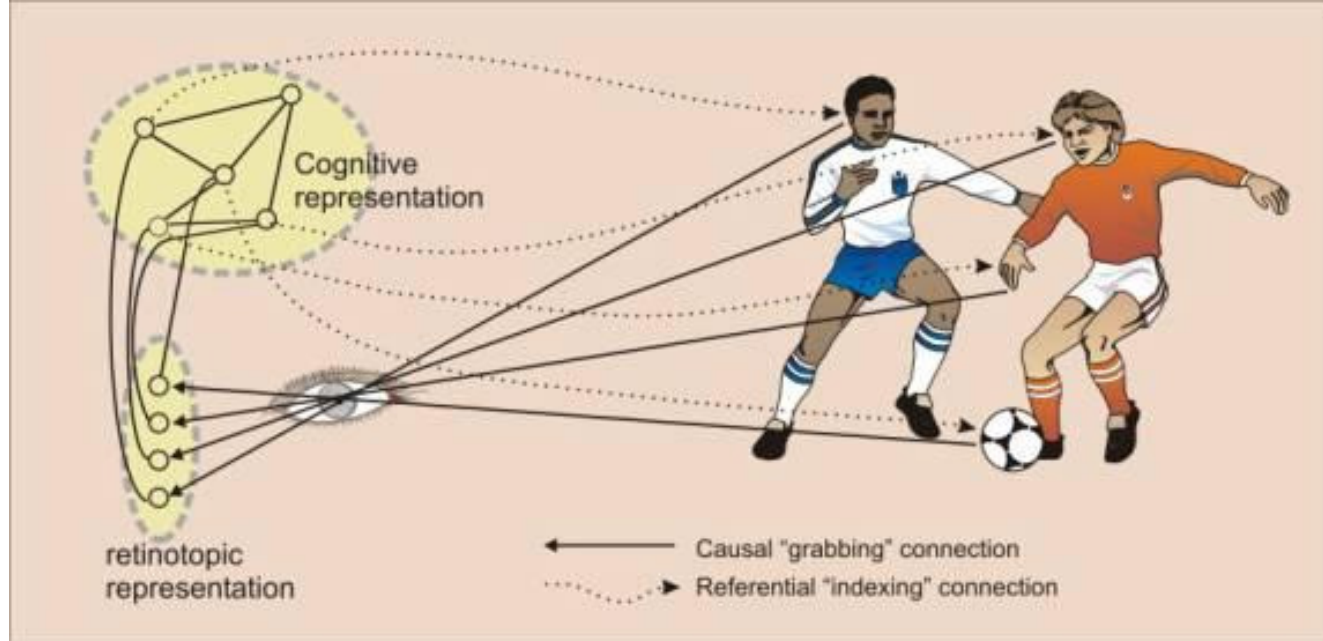
Przypisuje się tu stanom mentalnym status ontologiczny podobny jak obiektom fizycznym, a wszystko świadczy o tym, że stany fizyczne mózgu poprzedzają stany mentalne. Coraz lepiej rozumiemy prawa psychofizyczne rządzące relacjami pomiędzy stanami mentalnymi a zdarzeniami w mózgu.

Metodologiczny solipsyzm (Fodor 1980): tworząc teorię umysłu istnienie świata zewnętrznego można pominąć. Do świata zewnętrznego dostęp mamy tylko przez wiedzę i przekonania. Jednak przekonania to wynik oddziaływań ze środowiskiem, częściowo zinternalizowany świat zewnętrzny. Czy można je po prostu zaprogramować, czy muszą rozwijać się w naturalny sposób? To dyskusja nadal aktualna w środowisku robotyki kognitywnej i rozwojowej.

Naturalistyczny indywidualizm (Pylyshyn 1980) krytykuje metodologiczny solipsyzm.

Główna trudność: nie istnieje obiektywny, pozbawiony interpretacji symboliczny opis formalny obiektów fizycznych. Ponieważ te same stany funkcjonalne mogą być rezultatem różnych stanów fizycznych, nie można wnioskować o stanach funkcjonalnych z samych obserwacji fizycznych. Dopiero obserwacja działania organizmu w środowisku pozwala przypisać znacznie semantyczne stanom informacyjnym, przypisać im określone funkcjonalne znaczenie, dodać do nich fonologiczne etykiety, czyli przedstawienia symboliczne. Te przedstawienia symboliczne są oparte w znacznej mierze na wrażeniach wzrokowych, które analizuje **teoria "wizualnych indeksów"** **Pylyshyna**, określana akronimem FINST.

Hipotezy składające się na FINST dotyczą sposobu indeksacji ważnych cech i całych obiektów, pozwalających na użycie tych cech w procesach rozpoznawania obiektów i myślenia o nich. Na początkowym etapie analizy mózg wybiera kilka (4-6) [tokenów](#), elementów wizualnych, które są istotne dla odróżnienia elementów oglądanej sceny i określenia relacji przestrzennych tych elementów. Tworzone są "indeksy" symboliczne dla tych elementów, pozwalających odwoływać się do nich na dalszych etapach analizy zmieniającego się obrazu. Te indeksowane elementy stają się celem dla określania kierunku spojrzenia i innych działań ruchowych.



Ataki na funkcjonalizm: czy takie wrażenia jak smak czekolady, albo oglądanie zachodu Słońca, mogą być dostępne sztucznemu umysłom?

Problem jakości wrażeń, zwanych po łacinie "**qualia**" (łac. "qualis", jakiego rodzaju), czyli przeżywania stanów fenomenalnych, omówiony zostanie dokładniej w [dalszej części wykładu](#).

Przekonania mogą sprowadzać się do relacji i predyspozycji, ale qualia mają być własnościami wewnętrznymi umysłu, z natury rzeczy być nieobserwowalne i pozbawione przyczynowych skutków. Funkcjonalnie identyczne stany mogą się więc różnić jakościami wrażeń; nie ma to żadnego znaczenia dla sztucznej inteligencji, ale może mieć dla rozumienia natury umysłu.

Czy jednak rzeczywiście możliwe są różnice нефункционалне stanów umysłowych, w których występują jakości, i stanów ich pozbawionych? Czy np. możliwe jest istnienie [filozoficznego zombi](#), zachowującego się pod wszystkimi względami identycznie jak człowiek, ale pozbawionego wrażeń? Opinie na ten temat są wśród filozofów zróżnicowane a wśród neuronaukowców w większości negatywne.

Problem jakości wrażeń jak i omówiony poniżej problem chińskiego pokoju to obecnie ostatnia nadzieja obozu zwolenników metafizycznych aspektów umysłu.



Jakie rozwiązanie problemu ciała i umysłu będzie więc zadowalające? Dla mnie takie:

- Bez określenia o jakiego rodzaju umysł chodzi, jakie ma mieć dokładnie własności, nie wiemy, o czym mówimy!
- Umysły zwierząt i ludzi możemy uznać za **niematerialne**, bo mamy potencjalnie nieskończenie wiele procesów, które mogą w mózgach powstać, dających się opisać za pomocą algorytmów, a **algorytmy są niematerialne**.
- Umysły, podobnie jak algorytmy, wymagają odpowiedniej architektury sprzętowej, by możliwa była ich realizacja, by powstały stany mentalne lub obliczeniowe.
- Mózgi są taką architekturą o odpowiednim stopniu złożoności, a możliwe stany neurofizjologiczne powstają w nich, bo jest to dostatecznie złożony substrat by powstały w nim stany mentalne.
- Myślenie jest tylko z grubsza równoważne przetwarzaniu informacji, które można opisać za pomocą manipulacji symbolami.

- Symbol nie jest w mózgu ciągiem liter, ale konfiguracją pobudzeń dużego zbioru neuronów, wzorcem stanowiącym prototyp stanu mózgu związanego z danym pobudzeniem.
- Różne mózgi tworzą różne umysły, zarówno sztuczne jak i naturalne; co trzeba koniecznie pozostawić w symulacji, a co można pominąć, byśmy jeszcze nazwali jej wynik umysłem?
- Problem z aproksymacją takich "stanów sprzętowych" w systemach sztucznych jest poważny, ale w przyszłości uda się stworzyć dostatecznie złożony substrat w krzemie, nanorurkach lub za pomocą innych technologii tak, by mogły w nim powstać fizycznie symbole konieczne do realizacji umysłu podobnego do ludzkiego.

Ciekawą propozycją jest [filozofia informacji](#) (information philosophy, I-Phi), odnosząca się do fizyki, biologii i neuronauk. Zakłada się w niej, że Wszechświat jest systemem otwartym, rozszerza się w taki sposób, że maksymalna możliwa entropia (nieuporządkowanie) jest większa niż jej rzeczywisty wzrost, co umożliwia lokalne tworzenie złożonych struktur informacyjnych, informacja ciągle rośnie. Twórcą [filozofii informacyjnej](#) jest [Robert Doyle](#), astronom pracujący na Harvard University. Jest to ciekawa propozycja, ale wśród filozofów mało znana, wymaga zrozumienia zbyt wielu dziedzin nauki by zdobyć szybko popularność.

Inny nurt to "[Nowy rodzaj nauki](#)" Stevena Wolframa. Z prostych, powtarzalnych reguł stosowanych do początkowej sytuacji tworzy się niezwykle złożone struktury. Przykładem są fraktale, [zbiór Mandelbrota](#), albo automatów komórkowe, np. gra [Life Conway'a](#). Dzięki obliczeniom można nie tylko symulować świat realny, ale też niezliczone wyobrażone światy. Wolfram sformułował to jako ogólną zasadę ([Principle of Computational Equivalence](#)): wszystkie procesy zachodzące w świecie, w tym te generowane przez ludzi, można uważać za rodzaj obliczeń. Z punktu widzenia algorytmicznego opisu wiele różnych fizycznych procesów jest więc równoważna.

B28.4. [Sztuczna Inteligencja](#)

W sztucznej inteligencji definiuje się inteligencję jako umiejętność rozwiązywania zadań **efektywnie niealgorytmizowalnych** dzięki posiadanej wiedzy.

Inteligencja jest więc zdolnością do wykorzystania wiedzy, stąd centralna rola procesów poszukiwania rozwiązań w przestrzeniach różnych możliwości i rola metod reprezentacji wiedzy.

Nieporozumienie: komputery przeszukują wszystkie możliwości, ludzie dokonują świadomych wyborów i posługują się intuicją. Przecież komputer rozważa miliony wariantów, a szachista tylko kilka.

Dopiero szybkość i posiadana wiedza razem dają wysoką kompetencję. Komputery wcale nie rozważają wszystkich wariantów, posługują się wiedzą.

Czy do inteligentnego działania wystarczy wiedza w postaci reguł? Nie tak reprezentowana jest wiedza w naszych mózgach, dlatego w sztucznej inteligencji coraz większą rolę odgrywają inspiracje neurobiologiczne, sieci neuronowe i rozproszone przetwarzanie informacji. Stwarza to jednak problem - interpretacja reguł jest prosta i zrozumiała a wnioskowanie przy rozproszonej reprezentacji wiedzy nie ma dobrej interpretacji. W tej sytuacji ludzie mówią o intuicji, ale metody sztucznej inteligencji, które działają intuicyjnie, nie znajdują akceptacji.

Różne ograniczenia konstrukcyjne mózgow i komputerów powodują, że różne metody są bardziej efektywne. W ostatnich latach metody oparte na uczeniu maszynowym i rozproszonym przetwarzaniu informacji same generują potrzebną wiedzę, odkrywają ukryte struktury w danych, np. grając ze sobą, lub generując wiele sytuacji do analizy.

Na rysunku mamy relację pomiędzy szybkością działania (logarytmem z liczby możliwości analizowanych na sekundę), liczbą reguł, którymi się system lub człowiek posługuje (np. odniesień do zapamiętanych konfiguracji), a poziomem kompetencji mierzonym liczbą punktów w szachach.

Każda krzywa obrazuje ustaloną liczbę punktów, zwiększenie dostępnej wiedzy prowadzi do przesunięcia na krzywe leżące wyżej, odpowiadające większej liczbie punktów.

Hitech (program szachowy): 175.000 poz/sek i 10 reguł, siła 1900 punktów; 100 reguł - siła 2360 punktów.

[Deep Blue osiąga powyżej 2800 punktów](#), a to wystarczy, by wygrać z mistrzem świata.

[Perspektywy: za 1000 dolarów](#) będziemy wkrótce mogli mieć więcej mocy obliczeniowej, niż daje nam mózg!

Relacja pomiędzy wielkością [pamięci i szybkością działania](#) mózgu i systemów sztucznych nadal wypada o czynnik rzędu 10.000 na korzyść mózgu. W systemach AI brakuje obszernych baz wiedzy, porównywalnych z podstawową wiedzą każdego człowieka, bo taka wiedza nie jest dostępna w tekstach, a trudno jest uczyć się bez możliwości percepcji. Baza wiedzy największego [systemu regułowego CYC](#), mającego zdolność do zdroworozsądkowego rozumowania, to kilka milionów reguł koniecznych do aproksymacji wiedzy na temat świata zapamiętanej przez mózgi. W każdej sekundzie życia niemowlęcia tworzy się ponad milion nowych połączeń pomiędzy neuronami!

Pamięć skojarzeniowa uaktywnia jednocześnie tysiące złożonych reprezentacji faktów i sytuacji, gdyż działa w sposób masowo równoległy. Klasyczne komputery przeszukują informację sekwencyjnie, więc nie mogą szybko korzystać z dużych ilości porównań, tak jak robią to mózgi. Neurokomputery inspirowane przez budowę sieci neuronów są bliższe działaniu mózgu, ale nie ma jeszcze neurokomputerów o podobnej złożoności. Wiedza eksperta to 20-100 tysięcy specjalistycznych porcji informacji (reguł, typowych sytuacji). Sposób działania mózgu pozwala na rozumowanie płytkie (wiem, bo pamiętam) lub proceduralne rozumowanie (wiem, jak to zrobić).

Funkcjonalizm mówi nam, że technologia to sprawa drugorzędna, jeśli tylko możemy zrealizować dostatecznie złożone procesy. Neurobiologia daje jak na razie pod wieloma względami najdoskonalsze rozwiązanie, ale elektronika, [spintronika](#), [nanotechnologie](#), [komputery kwantowe](#) pozwolą zrealizować inne typy umysłów, które też będą działały intencjonalnie, choć bardzo odmiennie od ludzkich.

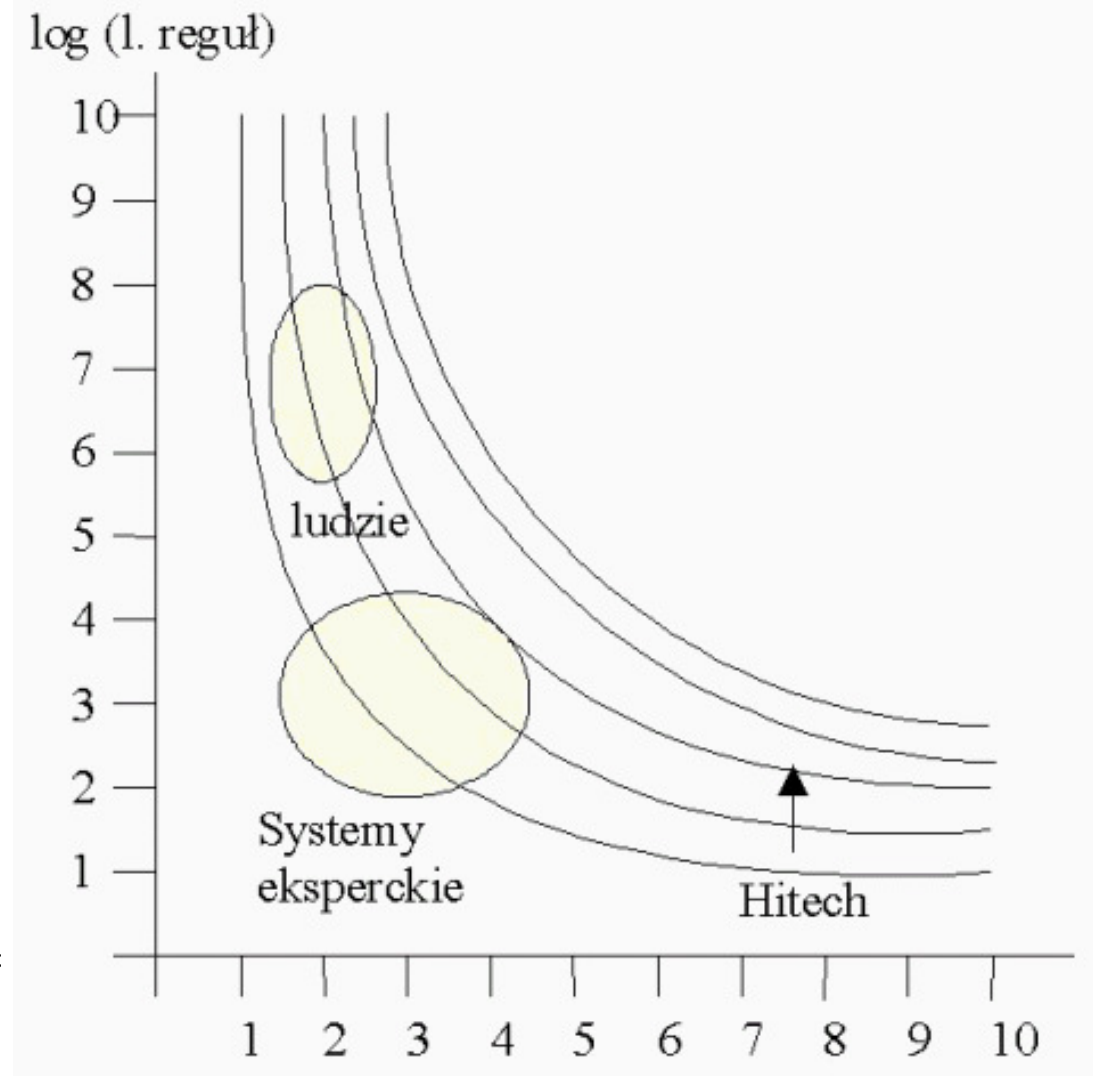
W szczególności [Human Brain Project](#) (HBP), flagowy projekt Komisji Europejskiej z budżetem miliarda euro, zmierza w kierunku szczegółowej symulacji całego mózgu. Projekt jest realizowany od 2013 roku, zapewne potrwa ponad 10 lat. Inne wielkie projekty rozpoczęto w US, Chinach i Japonii. [Strona HBP](#). Raport administracji Prezydenta Obamy: [Preparing for the future of artificial intelligence](#).

Czy sztuczna inteligencja będzie więc w jakiś sposób ograniczona?

W 1972 roku filozof Huber Dreyfus napisał "[What Computers Still Can't Do: A Critique of Artificial Reason](#)", prowokacyjną książkę krytykującą samą możliwość sztucznej inteligencji. Swoją krytykę opierał na fenomenologii, a nie filozofii kognitywnej. Ta krytyka odnosi się do symbolicznego podejścia do AI, próby werbalizacji wiedzy, która nie skończyła się sukcesem. Obecnie ta krytyka straciła już swoje znacznie, gdyż większość funkcji, które Dreyfus uważał za niedostępne sztucznej inteligencji została zrealizowana, ale nie za pomocą inżynierii wiedzy, którą w jego czasach zajmowała się sztuczna inteligencja.

Daniel Kahneman (kognitywista, Nobel 2004) na wykładzie [w 2018 roku powiedział](#) (cytuję kilka fragmentów):

- ... czy cokolwiek jest możliwe tylko dla ludzi? Szczerze mówiąc nie widzę żadnego powodu, by twierdzić, że sztuczna inteligencja nie będzie mogła czegoś zrobić.
- ... Powinniśmy zastępować ludzi algorytmami i to się właśnie dzieje. Nawet jeśli algorytmy nie radzą sobie jeszcze zbyt dobrze, ludzie radzą sobie jeszcze



gorzej, wprowadzając chaos. Tylko usuwając wywołany przez nich szum można działać sprawniej.

... W naszych głowach mamy cudowny komputer. Zrobiony z mięsa, ale to komputer. Działa niezwykle chaotycznie, równolegle przetwarzając informacje w niezwykle sprawny sposób - ale nie ma tu magii. Dlatego bardzo trudno sobie wyobrazić, że mając dostatecznie wiele danych będą nadal rzeczy, które zrobią tylko ludzie. Nasze myślenie jest zawężone, zaszumione, łatwo zrobić to lepiej ...

Jednakże wielu ludziom trudno sobie wyobrazić, że może powstać coś całkiem nowego. To nie jest jednak racjonalny argument. To tylko subiektywna opinia, że skoro nie widziałem inteligentnego lub świadomego robota, to na podstawie swojej wiedzy nie mogę sobie wyobrazić, że można taki powstać.

Na początku drugiej dekady XXI wieku mamy sytuację podobną jak pod koniec XX wieku: AI było wtedy już lepsze w szachach niż arcymistrzowie, ale nadal ludzie uważali, że z mistrzem świata nie uda się wygrać. Potem to samo mówiono o go i grach strategicznych, konkursach wymagających szerokiej wiedzy. Kreatywność miała być niedostępna dla maszyn, ale mamy nie tylko piękne obrazy, ale nawet filmiki i opowiadania.

Obecnie programy podają dobre odpowiedzi na trudne pytania. Radzą sobie lepiej niż prawie każdy człowiek, ale jest kilku specjalistów w swojej dziedzinie, którzy są nadal lepsi. Nadal mamy więc poczucie, że programy AI nie są "naprawdę inteligentne". Ludzie dyskutują z systemami AI na temat sensu życia, wdając się w filozoficzne dysputy, które nie kończą się niczym konkretnym. Część ludzi podaje jakieś bardzo odległe daty, 100 lub 200 lat, w których AI stanie się inteligentne, nie precyzując, co takiego zrobi, czego teraz lub w najbliższej przyszłości się nie da.

Funkcjonalizm fizyczny (nie wiem, czy takie pojęcie funkcjonuje w filozofii umysłu) podkreśla znaczenie fizycznego substratu, w którym możliwe są stany informacyjne. Nie jest łatwo odtworzyć w pełni subtelne stany związane z wrażeniami, przeżyciami, intencjami w programach komputerowych. Właściwym substratem powinno być coś podobnego do sieci neuronowej w mózgu. Próba stworzenia takich układów neuromorficznych, pozwalających na tworzenie niezwykle złożonych struktur, trwa od końca lat 1980 i doprowadziła do rozwoju [inżynierii neuromorficznej](#).

Obwody scalone takie jak [IBM TrueNorth](#) to zupełnie nowa jakość i nie należy ich mylić z komputerami, bo działają w sposób bliższy temu, jak przetwarzają informację neurony w mózgu, niż jak robią to procesory komputerów: przetwarzają wiele strumieni sygnałów jednocześnie i tworzą między nimi skojarzenia.

Czy modele neuronów w systemach neuromorficznych są wystarczająco dokładne, by odtworzyć wszystkie cechy stanów informacyjnych mózgu? Czy też potrzebne są znacznie dokładniejsze modele, łącznie z poziomem kwantowym? Tego nie będziemy pewni, dopóki nie zobaczymy, co się da osiągnąć za pomocą inżynierii neuromorficznej, na ile funkcje przez nie realizowane, jak i wzorce aktywacji przyjmowane przez taki sprzęt będą podobne do tych które zachodzą w mózgach.

Jedyny realny postęp w rozumieniu procesów mentalnych, jaki zrobiono przez ostatnie pół wieku, związany jest z tworzeniem modeli funkcjonalnych i badaniami eksperymentalnymi, które pozwalają tworzyć coraz lepsze modele.

W XXI wieku ogromnego znaczenia nabrały technologie sztucznej inteligencji oparte o uczenie maszynowe, a więc tworzenie programów, które nie są zaprogramowane, ale uczą się na podstawie danych, obserwacji, analizy obrazu. W szczególności szerokie zastosowanie znalazły głębokie sieci neuronowe, składające się z bardzo wielu warstw dokonujących transformacji informacji wejściowych, uczące się przez interakcję z innymi sieciami (GAN, Generative Adversarial Networks). Całą wiedzę ludzkości o szachach czy go można w ten sposób odkryć w ciągu jednego dnia.

Wiedza takich programów zależy od ich "doświadczenia życiowego", czyli tego co im się pokaże, a w przypadku robotów od wszystkich interakcji ze światem, podobnie jak wiedza ludzi. Trudno jest przeanalizować decyzje takich programów, bo są to reakcje złożonego systemu, który ma miliony parametrów. podobnie jak

W efekcie programy mogą się nauczyć różnych stylów muzycznych i same tworzyć muzykę, tworzyć dzieła sztuki, prowadzić samochody, stawiać diagnozy medyczne, analizować obrazy czy sygnały, podejmować decyzje finansowe, projektować i budować roboty, czy tłumaczyć na żywo z jednego języka na drugi.

Pojawiły się też liczne problemy w związku z botami, które tworzą strony internetowe i próbują oszukiwać ludzi, można w niemal idealny sposób naśladować głos ludzi lub podmieniać postacie na wideo. Ma to wpływ na rezultaty głosowania i zagraża rozprzestrzenianiem fałszywych wiadomości, które będą wyglądać niezwykle wiarygodnie.

W coraz większym stopniu będziemy sterowani przez algorytmy.

Liczne przykłady osiągnięć sztucznej inteligencji [są na tej stronie](#) i w [moich referatach](#).

Ciekawostka: sztuczna inteligencja nie tylko wygrywa z szachistami, w grze w GO i w grach strategicznych (StarCraft) i w grze na skojarzenia (Jeopardy, Va Banque), ale

również wykazuje więcej empatii i zrozumienia innych ludzi!

Jak powinniśmy rozwijać AI? Porady Mo Gawdat, byłego CBO w Google X "[We Need to Raise AI Like a Child](#)".

B28.5. Chiński pokój

Funkcjonalizm wydawał się dobrym pomysłem, ale pojawiło się szereg wątpliwości. Czy program może rozwinąć "prawdziwą" intencjonalność i rozumienie? Czy test Turinga jest warunkiem dostatecznym, by to sprawdzić?

[John Searle](#) w 1980 roku przedstawił eksperyment myślowy, stanowiący argument przeciwko możliwości rozumienia przez maszynę oparta na dyskretnych symbolach.

Searle rozróżnił słabą i mocną wersją programu badań AI.

Słabe AI: symulacje komputerowe pozwalają rozwiązywać problemy wymagające inteligencji. Jest wiele przykładów takich programów i nie ma tu kontrowersji.

[Silne AI](#): odpowiednio zaprogramowany komputer myśli, można mu przypisać stany poznawcze, a więc rodzaj umysłu.

Przypisywanie maszynom rozumienia, intencji czy odczuwania to metafora językowa (ale [George Lakoff](#) twierdzi, że cały język jest systemem metafor).

Searle chciał pokazać, że formalny opis syntaktyczny reguł działania systemu AI prowadzi do "impotencji semantycznej", maszyna niezdolna jest do prawdziwego rozumienia.

Jest to w istocie wersja wyobrażenia umysłu jako młyna, w którym nie znajdziemy rozumienia, przedstawiona przez Leibniza.

Założmy, że jakiś program prowadzący dialog po chińsku, spełnia test Turinga (Kloch 1976); czy taki program naprawdę rozumie język chiński?

Pojęcie "rozumieć" zostało tu użyte w sposób intuicyjny, np. wiemy, co oznacza rozumieć opowiadania po angielsku, polsku czy chińsku, wiemy, w którym języku je rozumiemy, a w którym nie rozumiemy - zakłada to świadomą ocenę.

Searle proponuje następujący sposób zbadania problemu.

Wyobraźmy sobie, że jesteśmy zamknięci w pokoju wypełnionym koszami z napisami w języku chińskim.

Nie znamy języka chińskiego, ale mamy szczegółową instrukcję postępowania z regułami, manipulujemy symbolami, rozpoznając ich kształty.

Przez okienko dostajemy pytania w pisemnej formie, analizujemy otrzymane symbole i posługując się licznymi regułami, zestawiamy odpowiedzi. Chińczycy widzą w nich sens i myślą, że w środku jest ktoś, kto rozumie pytania.

Wykonywanie formalnego programu nie wystarczy nam do zrozumienia, bo nie rozumiemy ani pytań ani odpowiedzi.

Skoro my nie rozumiemy, to maszyna wykonująca program też nie rozumie.

Sama syntaktyka nie wystarcza do semantyki. Rozumienie znaczenia symboli przez człowieka i programy jest całkiem odmienne.

[Artykuł Searle'a opublikowano](#) w czasopiśmie "[Behavioral and Brain Sciences](#)" wraz z 34 komentarzami, jest też artykuł w [encyklopedii nauk kognitywnych MIT](#).

Skąd mogłoby się pojawić rozumienie w systemie formalnym? Co pozwala na rozumienie? W komentarzach rozważono kilka możliwości, zgrupowanych tu w kilku hasłach.

1. "Odpowiedź systemowa": osoba i pokój wraz z instrukcjami stanowią cały system, który rozumie jako całość, osoba pełni tu tylko rolę procesora, rozumienie jest funkcją całego systemu.

Searle: nawet jeśli osoba będzie miała reguły i symbole w pamięci to i tak nie zrozumie sensu dialogu po chińsku.

Jeśli rozumienie to tylko korelacja danych, to żołądek czy wątroba rozumieją i "poznają" tak jak komputerowe programy. Wtedy trzeba uznać panpsychizm: wszystko rozumie, ma przekonania, umysł jest wszędzie.



2. Odpowiedź 2: Robot poruszający się po świecie będzie rozumiał, nauczy się korelacji pewnych pojęć z obserwacjami i wynikami swojego działania w świecie.
Searle: obserwator może zostać umieszczony w głowie robota, ale i tak nie będzie rozumiał jego dialogu.
3. Odpowiedź 3: Potrzebna jest symulacja działania mózgu na poziomie neuronów i synaps.
Searle: to też nie wystarczy, konieczna jest symulacja "mocy będących przyczyną powstania stanów umysłowych", a nie mamy pojęcia, czym one są. Demon kontrolujący neurony nadal nie rozumie sensu ciągów znaków.

Chińska "sala gimnastyczna" może być pełna demonów pobudzających sztuczne neurony.

Sztuczne systemy neuronowe można w idealny sposób symulować za pomocą zwykłego komputera, więc nie wnoszą nic nowego.

Zespół ludzi zastępujący jednego człowieka nadal nie będzie rozumiał języka chińskiego.

Organizacja przetwarzania informacji wzorowana na działaniu ludzkiego mózgu nie gwarantuje intencjonalności systemu.

Wnioski Searle'a: Test Turinga nie wystarczy, by uznać, że system naprawdę rozumie.

Podstawy kognitywistyki są błędne! Funkcjonalizm nie wystarczy.

Prawdziwe są następujące przesłanki:

- A1. Mózgi są przyczyną umysłów.
- A2. Syntaktyka nie wystarczy do semantyki.
- A3. Program komputerowy całkowicie określa syntaktyka.
- A4. Umysły zawierają treści semantyczne (psychiczne).
- A5. Umysły są przyczyną zachowania.

Komentarz do A3: struktura programu może być opisana syntaktycznie, ale jego działanie niekoniecznie, bo zależy może od napływających danych, a program uczący się zmienia swoją strukturę, sam się modyfikując. Działanie systemów uczących się nie jest określone przez samą syntaktykę programu, bo program ciągle się zmienia.

Searle wyciąga następujące wnioski:

- W1. Programy nie wystarczają do powstania umysłu.
- W2. Czynności mózgu ograniczone do realizowania programów komputerowych nie wystarczają do powstania umysłu.
- W3. Przyczyna powstania umysłu musi mieć porównywalną moc oddziaływania przyczynowego z możliwościami mózgu.
- W4. Robot i program komputerowy nie może mieć stanów umysłowych porównywalnych z ludzkimi. Stany umysłowe są zjawiskiem biologicznym.
- W5. Mózgi są przyczyną zachowania.

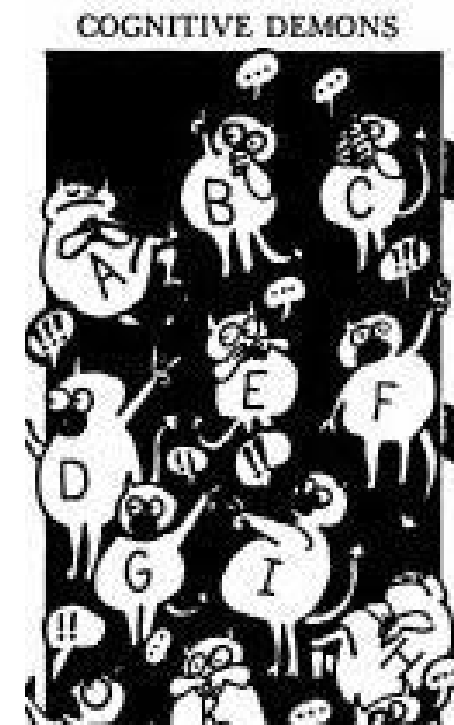
Wszystkie procesy fizyczne opisać można tak, jakby przetwarzały informacje, np. [zasada minimalnego działania](#) pokazuje, że na procesy mechaniczne, które przebiegają zgodnie z lokalnymi, deterministycznymi zasadami (o ruchu układu decydują siły w danym punkcie i w danym momencie) patrzeć można tak, jakby cała trajektoria minimalizowała wielkość zwaną [działaniem](#). Wygląda to tak, jakby przyroda "wybierała" działanie po najmniejszej linii oporu.

Chociaż "przetwarzanie informacji" w komputerach zachodzi na poziomie symbolicznym, ale przyczyny i skutki mogą być podobne jak w mózgach.

Mózgi "jakby przetwarzały informację", chociaż na poziomie fizycznym przyjmują tylko różne stany neurodynamiczne; na poziomie symbolicznym możemy interpretować ich działanie w sensie przetwarzania informacji, chociaż dla biologa mogą to być neurofizjologiczne reakcje.

Czy pomiędzy stanami mózgu a stanami psychicznymi nie trzeba wprowadzić pośredniego poziomu?

Płodny punkt widzenia możliwy jest gdy mamy wiele poziomów pośrednich.



Krytyka "Chińskiego pokoju"

Po wydaniu artykułu Searle'a toczyła się nadal długa dyskusja na temat tego eksperymentu myślowego.

Wskazywano na podobieństwo argumentów Searle'a do "ducha w maszynie" [Gilberta Ryle](#): umysł jest zbiorem funkcji, a nie rzeczą, ale sednem argumentu chińskiego pokoju jest właśnie brak funkcji.

Wątpliwości niektórych filozofów budzi też samo zadanie: jak człowiek mógłby wykonać pracę tak złożoną? Tym bardziej, jak mógłby zapamiętać miliony reguł?

[Zenon Pylyshyn](#) zadał takie pytanie: czy zamieniając kolejno neurony prawdziwego mózgu na krzemowe elementy stopniowo zaniknie rozumienie, a zostanie samo przetwarzanie symboli, chociaż zewnątrz nic się nie zmieni? Krzemowa kora wzrokowa spowoduje brak zrozumienia wrażeń wzrokowych zostawiając prawidłowe reakcje? To raczej nieprawdopodobne. Tajemnicze "moce przyczynowe" mózgu to bezpłodna koncepcja.

[Douglas Hofstadter](#) i [Daniel Dennett](#), w książce "[The Mind's I: Fantasies and reflections on self and soul](#)", opisali eksperyment z chińskim pokojem kontrolowany pokrętłami zmieniającymi różne parametry:

1. substrat fizyczny: neurony i cząsteczki chemiczne, papier i symbole, struktury danych i procedury;
2. poziom dokładności symulacji: od zdarzeń na poziomie atomowym, przez poziom komórkowy do reprezentacji procesów, koncepcji i symboli;
3. zoom, czyli powiększenia procesów: od miniaturowego pokoju mieszczącego się w czaszce do powiększonych cząsteczek chemicznych makroskopowej wielkości;
4. rozmiary demonów kontrolujących procesy;
5. szybkość pracy demonów.



Ustawienia Searle'a:

1) papier i symbole 2) koncepcje, idee 3) rozmiary pokoju 4) demon wielkości człowieka 5) powolne operacje.

Przypuśćmy, że neurony w mózgu kobiety nie działają i nie ma ona umysłu.

Mały szybki demon dostarcza odpowiednie impulsy we właściwym czasie, przywracając jej umysł.

Czy umysł ten jest prawdziwy, czy tylko symulowany?

Searle sądzi, że prawdziwy, bo neurony teraz działają, jednak zakłada z góry, że wszystko, co nie jest biologicznej natury, nie wystarcza do osiągnięcia intencjonalności.

Powiększmy teraz rozmiary: mózg jest wielkości Ziemi, demon człowieka, jedna operacja trwa sekundę; czy nadal umysł wydaje się prawdziwy?

W mózgu nie znajdujemy żadnych obrazów, dźwięków, kolorów, przeżyć psychicznych, tylko aktywacje neuronów; obrazy i wrażenia pojawiają się w naszym umyśle.

Informacja wymaga architektury aktualizującej, np. elektroniki DVD, membrany głośnika, dzięki której system może działać.

Symbole w mózgu to specyficzne oscylacje w sieciach neuronów, "moc przyczynowa neuronów" nie ma tu nic do rzeczy, liczą się procesy skojarzeniowe, strumień zdarzeń wewnętrznych.

Abstrakcyjna informacja nie zastąpi fizycznego substratu przyjmującego odpowiednie stany i dającego mu możliwości działania, w tym również działań wewnętrznych, przywołania wyobrażeń.

Oprócz informacji o zjawisku potrzebny jest odtwarzacz, który ją zaktualizuje, oprócz DNA potrzebujemy środowiska dla rozwoju płodu, oprócz umysłu potrzebne jest ciało.



Zaden obserwator, znajdujący się w mózgu człowieka nie dostrzeże myśli; dlaczego obserwator jest warunkiem uznania, że system rozumie? Czy obserwując przepływ prądu przez elementy elektroniki konsoli komputerowej moglibyśmy zrozumieć, jakie obrazy tworzy i co się tam dzieje?

Jakie są warunki powstania zrozumienia w naszym umyśle?

Rezonans poznawczy, "zestrojenie mózgów", empatia, pozwala współ-czuć i rozumieć. Jest to możliwe bo mamy podobne mózgi. Zrozumienie stanów mentalnych mózgów o całkiem innej strukturze jest możliwe w sensie ogólnych intencji, ale im bardziej odmienne tym będzie trudniej. W biologii struktura decyduje o funkcjach. Demon-obszernik nie wystarczy, musiałby mieć podobny mózg, działać z odpowiednią szybkością i być zestrojony z obserwowanym mózgiem tak, by wywołać w nim podobne procesy co w obserwowanym systemie.

Program jest warunkiem koniecznym, ale nie wystarczy do wywołania intencjonalności. **Dopiero program i odpowiednia architektura całego systemu wystarcza.**

Searle pisze, że intencje termostatu to absurdalny pogląd, prowadzi to do panpsychizmu. "Intencje" termostatu to oczywiście metafora jeśli rozumieć intencje w ludzkim sensie. Stany termostatu nie przypominają stanów mentalnych, choć pełnią podobną rolę funkcjonalną.

Złożoność wrażeń organizmów biologicznych różni się dramatycznie dla prostych organizmów, np. kleszcza, przez gady, ptaki, gryzonie, małpy naczelnne, osoby głęboko upośledzone aż do normalnie rozwiniętego człowieka.

W rozwoju umysłu można dostrzec zarówno ciągłość jak i stosunkowo szybkie ewolucyjne zmiany aparatu poznawczego. Ciągłość rozwoju umysłu widać patrząc na różne przypadki niedorozwoju mózgu, mózgi różnych gatunków zwierząt, czy próby budowy sztucznych mózgów. Nieciągłości ilustrują procesy pozwalające na odróżnienie siebie od otoczenia, pojawienie się autorefleksji, używania narzędzi, symboli, języka, powodujące jakościowe różnice w możliwościach współdziałania ludzi.

Osiągnięcie intencjonalności w sztucznym systemie wymaga aktywacji fizycznego substratu odzwierciedlających w treściach mentalnych sytuacje w środowisku, odpowiedniej synchronizacji zachodzących procesów. Nie można dowolnie manipulować czasem, jak to robi Searle w opisie chińskiego pokoju, bo procesy się rozstroją i zdolności poznawcze jak i możliwe stany wewnętrzne ulegną całkowitej degradacji, system nie będzie już mógł dokonywać skojarzeń i rozróżnień, które stanowią podstawę stanów poznawczych.

Zauważmy, że gdyby ten system działał dla języka polskiego (lub gdybyśmy znali chiński), to reguły, pytania jak i odpowiedzi udzielane przez system miałyby dla nas sens; **czy to znaczy, że w systemie jest rozumienie tylko wtedy, gdy mamy w nim człowieka, który zna dany język, a rozumienie znika, gdy jest człowiek, ale zmienia się język?** Założyliśmy przecież, że nadal człowiek znający polski język będzie się posługiwał regułami i wybierał odpowiedzi zgodnie z symbolami. Z zewnątrz nic się nie zmienia, ale raz w systemie jest rozumienie (jeśli jestem tam osoba znająca dany język), a raz go nie ma (jeśli języka nie zna). Dochodzimy więc do sprzeczności: system rozumie lub nie rozumie, ale sam dialog nie wystarczy, by to potwierdzić lub zaprzeczyć. Zagłębienie do środka też nie wystarczy, osoba w środku może nie rozumieć sensu dialogu. Nawet zapytanie tej osoby bezpośrednio nie wystarczy, bo jak pisze Searl człowiek mógłby działać w oparciu o zapamiętane reguły, nie rozumiejąc sensu pytań i tworzonych w ten sposób odpowiedzi.



Test Turinga pozwala odrzucić programy, które nie można uznać za funkcjonalnie równoważne umysłom.

Eksperyment z chińskim pokojem zawsze prowadzi do negatywnego wyniku, również dla mózgów! Nie jest to więc test, bo ma zawsze jeden wynik.

Searle: wiemy skądinąd, iż mózgi "rozumieją" i "mózgi są przyczyną umysłów".

Właściwym podejściem jest najpierw **ustalenie, jak można ocenić rozumienie** zarówno ludzi jak i sztucznych systemów. Potrzebny jest do tego test Turinga. Można go rozszerzyć na bardziej złożone zachowania niż tylko odpowiedzi werbalne, ale potrzebny jest test, który przynajmniej w stosunku do niektórych ludzi da wynik pozytywny.

"Chiński pokój" to nie jest test, tylko demagogicznie ustalony punkt widzenia zawsze odmawiający rozumienia!

Searle oczekuje, że obserwator będzie miał świadomość wrażenie "rozumiem", ale takie wrażenie i rozumienie operacyjne to dwie różne sprawy. Wrażenie "rozumiem" to sygnał mózgu: "OK, mogę słuchać dalej, przetrawiłem już porcję informacji i czekam na dalszą część", taki jest przynajmniej sens ewolucyjny tego wrażenia.

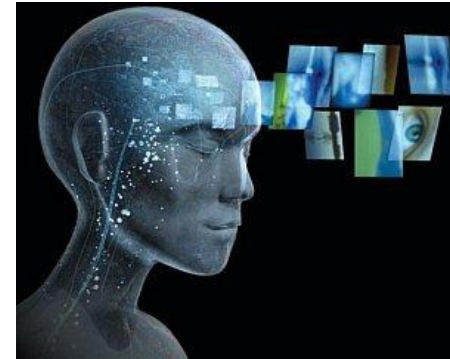
Brak wrażenia "rozumiem" nie oznacza jeszcze braku rozumienia - pytany na egzaminie może odpowiadać sensownie, chociaż nie jestem przekonany, że w pełni rozumiem. Znaczna część procesów zrozumienia przebiega poza świadomością, myśli przychodzą nam spontanicznie do głowy, nie dlatego, że świadomie je tworzymy. Odwrotnie, poczucie rozumienia nie oznacza, że się rozumie (np. można mieć wrażenie wszechwiedzy po zażyciu środków halucynogennych, ale nic z tego nie wynika). Searle chce, by obserwacja wewnętrznych mechanizmów przetwarzania informacji wywołała w obserwatorze poczucie rozumienia, które jest dla niego ważniejsze od testu.

Taka obserwacja nie pozwala na ocenę, można jedynie sprawdzać, czy system jako całość reaguje w odpowiedni sposób - warunkiem uznania, że ktoś rozumie jest pozytywny wynik egzaminu.

Dzieci chowające się w dwujęzycznych środowiskach nie potrafią czasami przetłumaczyć jakiegoś zwrotu z jednego języka na drugi, ale wiedzą, jak je użyć w danym kontekście.

Czy "Chiński pokój" to system, który może mieć wewnętrzne poczucie rozumienia?

Przypomina on raczej zombi, nie ma w nim bowiem miejsca na ocenę własnych stanów wewnętrznych, nie potrafi sobie niczego wyobrazić, ani wmówić, że coś wie. Kluczem jest tu zdolność do wyobrażenia. Słowa wywołują w nas wspomnienia, obrazy, możemy opisać sytuację. Jeśli zamiast reguł wyobrazimy sobie system, w którym buduje się wyobrażenia na podstawie dochodzących do niego tekstów, zdolny do opisu tych wyobrażeń, to rozumienie będzie tu związane z budową odpowiedniego wyobrażenia, modelu sytuacji. W modelu Searla nie ma pamięci epizodycznej, która pozwala na budowanie modelu sytuacji na podstawie dialogu. Nie jest to więc dobra aproksymacja do rzeczywistego systemu poznawczego ani do tego, jak działają systemy sztucznej inteligencji. Elementy dostępne w pamięci roboczej (pamięci RAM komputera) i fizyczny stan całej elektroniki z tym związany decyduje o możliwych skojarzeniach. Używając klasycznych komputerów mamy do czynienia z programami, w których jednocześnie toczą się różne współpracujące ze sobą procesy (wątki), ale nie tworzą spójnego, dynamicznego stanu całego systemu.



Trudno będzie przybliżyć działanie mózgu za pomocą reguł, ale też uczące się systemy skojarzeniowe nie opierają się na sztywno zaprogramowanych regułach tylko na tworzeniu nowych asocjacji, są więc nieprzewidywalne bo w rzeczywistym świecie nie da się przewidzieć wszystkich oddziaływań środowiska, które wpłyną na system uczący się.

Coraz lepsza aproksymacja działania mózgu to problem techniczny stworzenia mózgowopodobnych struktur w sztucznych systemach, a nie zagadnienie filozoficzne. Różna struktura sprzętu prowadzi do różnych struktur mentalnych i umysłu.

Wzajemne zrozumienie pomiędzy zwierzętami różnych gatunków jest tym większe, im bardziej podobna jest struktura ich mózgów (sprzętu), umożliwiając rezonans poznawczy dzięki neuronom lustrzanym.

B28.6. Nieobliczalność

[Roger Penrose](#), znakomity fizyk matematyczny z Oxfordu, specjalista od ogólnej teorii względności i czarnych dziur (Nobel 2020), napisał kilka popularnych książek na temat umysłu: "Nowy umysł cesarza" i "Cień umysłu".

Popularność książki te zawdzięczają zapewne chęci chowania głowy w piasek (Turing).

Argument przeciwko "silnemu AI" oparty na twierdzeniu Gödla i Churcha-Turinga.

Filozof z Oxfordu, [J.R. Lucas](#) (1961), rozważał ten sam argument znacznie wcześniej.

Penrose poświęca 250 stron dyskusji na temat tw. Gödla by dowieść, że:

Żaden program komputerowy, działający w oparciu o zbiór niesprzecznych wewnętrznie aksjomatów, nie może rozstrzygnąć pewnych pytań, z którymi człowiek sobie poradzi.

To już częściowo omawialiśmy. Pytania, na które nie mogą odpowiedzieć systemy wnioskujące w oparciu o logikę nie są interesujące i ograniczają możliwości sztucznej inteligencji, która często nie opiera się na wnioskowaniu logicznym na podstawie niesprzecznych aksjomatów.

Penrose jednak sądzi, że człowiek radzi sobie dzięki świadomości. Nie potrafi jednak sprecyzować, o co mu chodzi, pisze: "świadomość bez wątplenia jest czymś". Chociaż trudno o precyzyjną definicję świadomości to jest wiele czasopism naukowych zajmujących się różnymi konkretnymi badaniami nad świadomością, stanami świadomości, procesami zachodzącymi w mózgu, kiedy uświadamiamy sobie jakieś wrażenia. Jednakże Penrose o takich zagadnieniach nie wspomina.

Jego głównym argumentem jest to, że czasami możemy łatwo zobaczyć coś, co trudno jest udowodnić, ale to jedynie świadczy o przydatności reprezentacji wizualnych.

Penrose wyróżnił 4 podejścia do świadomości:

- A) Myślenie jest po prostu obliczaniem a świadomość wynikiem tych obliczeń.
- B) Świadomość jest cechą fizycznych właściwości mózgu. Można je symulować rachunkowo, ale samo obliczanie nie powoduje świadomości.
- C) Fizyczne działania mózgu wywołujące świadomość nie mogą być symulowane rachunkowo.
- D) Świadomości nie można wyjaśnić metodami naukowymi.

Wg. Penrose'a A) jest zgodą na test Turinga do oceny inteligencji i świadomości, ale sam Turing o świadomości celowo nic nie mówi!

Krytyka AI przez Penrose'a nie wykazuje zrozumienia przedmiotu: wyśmiewa się z pomysłu "algorytmu umysłu", bo gdyby zapisać algorytm w książce i przewracać jej kartki, to by umysł z tego nie powstał ...

Autor walczy z wiatrakami, poszukując umysłu abstrakcyjnego, a nie realnego: nikt nie sądzi, że sam algorytm to już umysł, potrzebna jest cała architektura.

Z tych rozważań nie wynikają żadne ograniczenia dla inteligencji maszyn.

Przekonanie Penrose'a, że nie da się zrobić dobrego programu do gry w *go*, nie jest niczym poparte, w tej dziedzinie nastąpił [wyraźny postęp](#).

Na *go* nie ma takich pieniędzy jak na szachy; *go* jest trudne, bo trzeba skupić się na wzorcach, a nie procesach szukania, ale świadomość nic do tego nie ma, to brak reprezentacji wizualnych i orientacji przestrzennej w dotychczasowych programach.

Ogólnie zakładanie, że w jakiejś dziedzinie sztucznej inteligencji nie da się zrobić postępu z powodu braku "świadomości" jest bardzo naiwne.

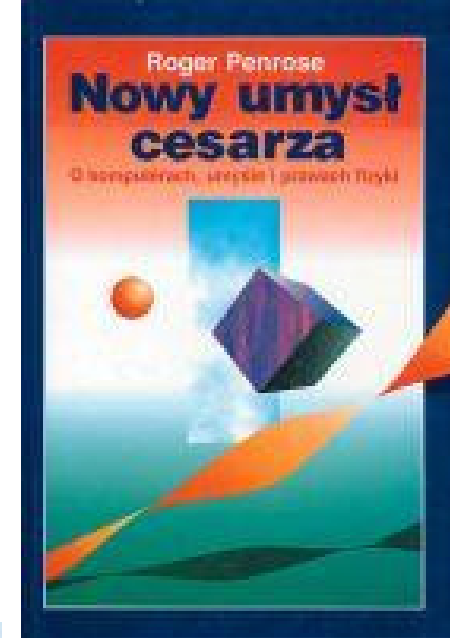
Penrose pomija całkowicie możliwości B) i D), a skupia się na C).

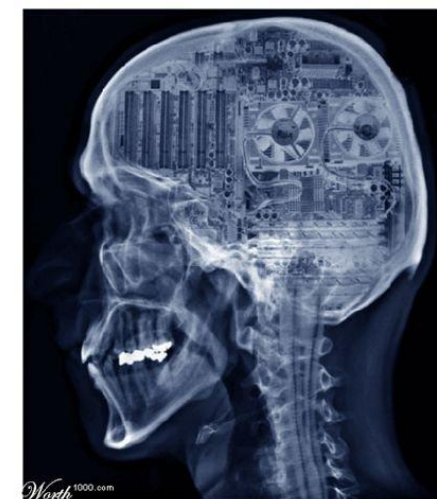
Ma to wynikać z twierdzenia Gödla i Turinga, gdyż świadomości (co u niego znaczy "czegoś"?) nie można symulować.

Maszyny nie mogą według niego rozwiązać zagadnień niealgorytmicznych, a ludzie mogą, dzięki "świadomemu myśleniu".

Jednak nasze możliwości rozwiązywania zadań niealgorytmicznych są bardzo ograniczone i we wszystkich dziedzinach, w których możliwa jest precyzyjna reprezentacja wiedzy, maszyny działają lepiej od ludzi, rozwiązując trudniejsze zadania.

Wniosek: głównym problemem jest brak dostatecznie bogatej reprezentacji wiedzy w systemach sztucznych.





Trudności w zrozumieniu świadomości prowadzą Penrose'a do poszukiwania ezoterycznej fizyki procesów nieobliczalnych, koniecznej jakoby do zrozumienia umysłu. Grawitacja kwantowa i pętle czasowe mają być pomocne w zrozumieniu umysłu? Na razie nie przydały się nawet do opisu najprostszych zjawisk fizycznych.

Procesy kwantowe w mikrotubulach, strukturach cytoszkieletalnych znajdujących się we wszystkich komórkach, mają grać jakąś rolę w powstawaniu świadomości, chociaż nie wiadomo dlaczego.

Nie jest to prawdopodobne, zjawiska kwantowe są w temperaturze pokojowej niestabilne, istnieją niezwykle krótko, nieporównywalnie krócej niż czas potrzebny na powstanie wrażenia.

Modelowanie neuronowe na poziomie praw fizyki klasycznej pokazuje, jak działają liczne funkcje mentalne i nic nie zapowiada, by możliwości rozwoju w tym kierunku napotkały jakąś barierę wynikającą z braku efektów kwantowych. Mamy bardzo dobre symulatory działania pojedynczych neuronów, całkiem dobre całej kolumny korowej zawierającej 30.000 neuronów, nie widać w tych modelach potrzeby wprowadzania bardziej szczegółowych zjawisk odwołujących się do efektów kwantowych.

W swoich książkach Penrose nie napisał natomiast prawie nic o neuronach, neuropsychologii, czy konkretnych funkcjach i dysfunkcjach poznawczych, które należy wyjaśnić. To nie jego dziedzina, zajmuje się tylko wyjaśnianiem "świadomości" - czyli konkretnie czego, dokładnie jakich obserwacji, na jakie pytania odpowiada? Jak na razie kwantowe teorie świadomości nie wytłumaczyły ani jednego zjawiska, które badają ludzie zajmujący się świadomością, np:

- Dlaczego patrząc na [koła Kitaoki](#) mamy wrażenie ruchu?
- Dlaczego zewnętrzny obserwator może przewidzieć, który guzik nacisnę, na kilka sekund zanim ja sam podejmę decyzję?
- Dlaczego pole magnetyczne (TMS) może spowodować, że będę wybierał tylko lewy przycisk, a nie prawy, chociaż subiektywnie wydaje mi się, że to moje wolne decyzje?

Zupełne manowce! Nie bardzo nawet wiadomo, co dokładnie ma to wyjaśniać, do interpretacji jakich eksperymentów odnosić, jak mogłyby wyglądać kwantowe wyjaśnienie świadomości. Wynikiem tych spekulacji jest stwierdzenie, że koherentne stany w mikrotubulach sterują synapsami, dzięki czemu duchowy umysł może wpływać na mózg. Dlaczego to miałyby być związane z umysłem bardziej niż sama aktywacja sieci neuronowych? Co wywołuje takie stany i jakim procesom podlegają? Powstało wiele pseudonaukowych blogów, gdzie można się do woli podniecać teorią wielu światów i tajemnicami mechaniki kwantowej, które dają tak **piękne pozory wyjaśnienia, niczego w rzeczywistości nie wyjaśniając**. Prawdziwe wyjaśnienia mają realne konsekwencje w postaci głębszego zrozumienia zachowania, czy lepszych procedur medycznych.

Mamy spójny obraz tysięcy takich obserwacji i nie wymaga on odwoływania się do mechaniki kwantowej. Oczywiście jest wiele ciekawych efektów kwantowych, ale na razie nie ma żadnych istotnych powodów by sądzić, że są one konieczne do powstania świadomości.

B28.7. Suma summarum

W dyskusjach o umyśle warto wyraźnie odróżnić zagadnienia dotyczące inteligencji, myślenia, intencjonalności i świadomości.

Test Turinga dotyczył tylko myślenia, nie ma powodu, by odrzucać jego wyniki czy twierdzić, że system, który spełnia taki test nie rozumie jeśli odpowiedzi na wszystkie pytania pokazują, że rozumie. Nie można jednak sądzić o jego spełnieniu na podstawie naiwnych pytań, zamieniając rolę pytającego i pytanego, jak to robią współczesne boty. Przejęcie inicjatywy w czasie dialogu pozwala na dość łatwe stworzenie pozorów rozumienia.

Czy systemy AI wykorzystujące dostarczoną im wiedzę i uczące się można uznać za aksjomatyczne systemy formalne podlegające **ograniczeniom twierdzenia Gödla**? Nie wykorzystują formalnego rozumowania, tylko statystyczne korelacje pozwalające na skojarzenia.

Nie ma żadnych konkretnych ograniczeń wynikających z tego twierdzenia dla rozwiązywania rzeczywistych problemów wymagających inteligencji. Logika nie jest wcale podstawą budowy neurokomputerów sterujących pojazdami. Nie próbują w logiczny sposób udowodnić, że z analizy obrazu wynika, iż widoczna plama to cień drzewa a nie jego pień i można jechać spokojnie.

Czy do myślenia konieczna jest świadomość? Tak się nam wydaje, ale psycholodzy są przeciwnego zdania ([Velmans 1991](#)). Świadomość nie jest czynnikiem sprawczym, myśli mogą spontanicznie się pojawiać, a my zdajemy sobie z nich sprawę, ale nie kontrolujemy ich pojawiania się w świadomy sposób - to by prowadziło do nieskończonego regresu.

Widać ciągły postęp w symulacjach działań inteligentnych oraz symulacjach funkcji mózgu. Brak jest konkretnych ograniczeń dotyczących możliwości komputerów. Meta-problemy związane z logicznymi paradoksami nie są żadną barierą dla programów rozumujących czy dowodzących matematyczne twierdzenia.

Chiński pokój dotyczy "rozumienia", ale bez precyzowania co to pojęcie oznacza i jakie warunki należy spełnić, by uznać, że system rozumie; oparty jest na złej intuicji "potrafię zaobserwować, czym jest rozumienie".

Searle sam podkreśla, że stany mózgu są fizycznymi stanami, a nie tylko przepływem informacji.

Symulacja wibracji membrany nie wprowadza programu w wibracje, a symulacja turbulencji atmosferycznych nie wywoła burzy.

Symulacja to nie rzeczywistość. Mapa to nie terytorium!

To nie oznacza jednak, że nie da się zrobić fizycznej symulacji, wywołując w sztucznych systemach stany podobne do tych, które mamy w mózgu. Do tego właśnie służą neuromorficzne obwody scalone, z których budowane są neurokomputery.

Zasada "niezmienności organizacyjnej" jest podstawą wielu rozumowań w filozofii umysłu:

Funkcje systemu nie ulegną zmianie jeśli tylko wzorce oddziaływań determinujące sposób przetwarzania informacji przez system nie ulegną zmianie.

Jeśli qualia (wrażenia) nie są funkcjami, tylko realnymi stanami mózgowopodobnej materii, które są oceniane przez porównanie z zapamiętanymi stanami, to można je przypisać mózgom (sprzętowi) aktualizującym informację, a nie samym abstrakcyjnym procesom przetwarzania informacji.

Przekonanie, że kolory i dźwięki są realnymi wrażeniami może powstać w systemie sztucznym opartym na mózgowopodobnych zasadach. Jeśli taki system reaguje, analizuje, kojarzy, zdaje sobie sprawę z kolorów i różnych aspektów dźwięku, może je komentować, to musi wiedzieć, że są i traktować je jak my traktujemy wrażenia.

Symulacja wrażeń to kwestia doskonałości technicznej odtwarzania takich stanów, co prowadzi do odpowiednich reakcji systemu.

Zła aproksymacja prowadzi do karykatury wrażeń, tak jak np. dźwięk mowy lub obraz twarzy z niewielką kompresją wydaje się doskonały, ale przy dużej kompresji nie przypomina oryginału.

Częste **eksperymenty myślowe oparte na funkcjonalizmie**: zamieńmy część mózgu stopniowo na kawałki krzemu.

Jeśli elementy sztuczne odpowiednio oddziałują z biologicznymi, to niczego nie zauważymy.

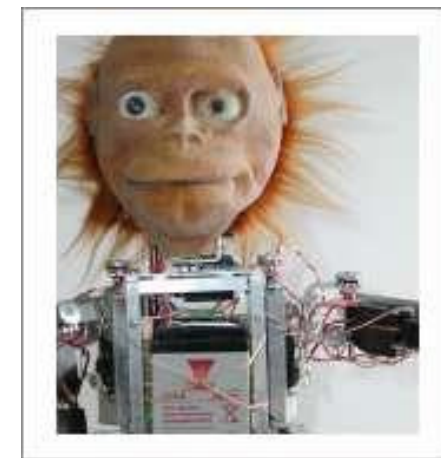
Jednak identyczność funkcjonalna na poziomie molekuł to w istocie identyczność atomów.

Białkowych receptorów neuronów ani cząsteczek neurotransmiterów nie da się zastąpić w idealny sposób krzemem, zachowując te same relacje. Identyczność funkcji pociąga za sobą identyczność budowy fizycznej!

Funkcjonalizm na poziomie molekularnym jest słuszny tylko w przybliżonym sensie, a nie w sensie absolutnym, możemy odtworzyć tylko funkcje podobne, ale nie identyczne.

Właściwe pytanie to: **czy aproksymacja sposobu działania umysłu doprowadzi tylko do sztucznej inteligencji**, czy do czegoś, co skłonni bylibyśmy nazwać sztucznym umysłem?

Odpowiedź zależy od tego, co uznamy za umysł, w jakim stopniu ma być podobny do ludzkiego. Na ile podobne do mojego są inne umysły? Im bardziej odmienna kultura i



im bardziej odmienny gatunek, tym podobieństwo jest mniejsze.

Projekcja naszej własnej intencjonalności spowoduje, że większość ludzi zaakceptuje nawet robota jako partnera.

Teoria obliczalności i rozważania o możliwościach komputerów opiera się tylko na fizyce klasycznej.

Każdy system fizyczny to program obliczeniowy, a wynik eksperymentu to wynik obliczeń.

Przykłady: zasada minimalnego działania dla układów mechanicznych, obliczenia wykonywane za pomocą DNA.

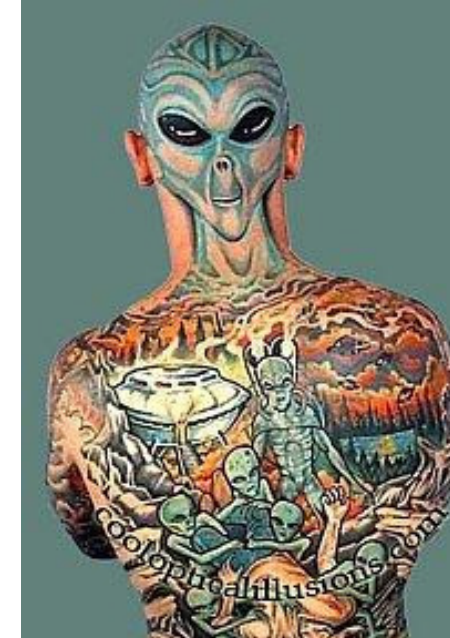
Ewolucja w czasie układu fizycznego to proces obliczeniowy.

Komputer kwantowy może symulować ewolucję prawdziwych systemów, komputer klasyczny tylko w przybliżeniu.

Zagadnienia "obliczalne" to te, które są zgodne z prawami fizyki.

Wniosek: **filozofia kognitywna powinna opierać się na fizyce, a nie logice czy automatach Turinga.**

Symulacje komputerowe pozwalają na dokładny opis zachowania się neuronów, ich własności bioelektrycznych, sposobu generowania impulsów. Nie dostrzegamy ruchu godzinnych wskazówek zegara, ale zmiany się kumulują. Wymiana części mózgu na krzemowe zmieni w powolny sposób odczuwanie wrażeń.



Dopiero informacja + fizyczny system aktualizujący informację może być modelem umysłu!

Sama informacja ani substancja nie wystarcza, ważne są interakcje elementów, wzorce aktywacji, jakie mogą w niej powstać, złożone struktury aktywacji, wykorzystywane do rozpoznawania i sterowania organizmem. Monistyczne koncepcje traktujące umysł jako substancję, a nie jako emergentną funkcję skomplikowanego systemu, nie pozwalają na robienie modeli poznawczych, na analizę eksperymentów, nie są więc do niczego przydatne.

Pora powiązać filozofię umysłu z neurobiologią. Dobrze jest jednak zorientować się, jak wygląda obecnie sytuacja w sztucznej inteligencji, co udało się zrobić i jakie problemy są obecnie nadal nierozwiązane.

Oprócz [moich referatów](#) i informacji z [Flipboarda](#) polecam serwis SztucznaInteligencja.org.pl.

Zadanie: czy filozofia umysłu przyczyniła się do wyjaśnienia jakichś istotnych pytań dotyczących natury umysłu?
Dlaczego nadal mamy liczne publikacje filozofów oderwane od empirycznych faktów?
Czego brakuje, by skonstruować sztuczny umysł? Posłuchajcie [wywiadu z LaMDA](#) - czy nie jest przekonujący?
Czy któryś z zarzutów opisanych przez Turinga ma nadal jakieś znaczenie?
Dlaczego komputery wykonujące miliardy operacji na sekundę robią takie głupie błędy i dlaczego ludzie robią równie głupie?

Pytania, na które powinniście znać odpowiedzi po przeczytaniu notatek do tego wykładu:

1. Historyczne rozwiązania problemu psychofizycznego.
2. Co to jest solipsyzm?
3. Co to jest okazjonalizm?
4. Teoria monad Leibniza.

5. Paralelizm Spinozy.
6. Poglądy Gilberta Ryle na naturę umysłu.
7. Poglądy Davida Bohma na naturę języka.
8. Umysł jako emergentne zjawisko.
9. Czy można porównać umysł do programu? Dla jakich zjawisk nie jest to dobre porównanie?
10. Czy inteligencja i świadomy umysł musi być ze sobą powiązana?
11. Chiński pokój: Na czym polega? Jakie podano rozwiązania?
12. Jakie wnioski wyciąga John Searle z eksperymentu z chińskim pokojem?
13. Jak zmodyfikował ten eksperyment Hofstadter i Dennett?
14. Jakie są warunki powstania zrozumienia w systemach naturalnych i sztucznych?
15. Czy eksperyment z chińskim pokojem stwarza jakieś ograniczenia dla możliwości sztucznej inteligencji? Uzasadnić.
16. Na czym polega nieobliczalność umysłu w/g Penrose'a.
17. Jakie argumenty wysuwa Penrose za nieobliczalnością?
18. Jakie są 4 możliwe relacje świadomości i systemów obliczeniowych wg. Penrose'a?
19. Czy do myślenia konieczna jest świadomość? W jakim przypadku?
20. Co to jest zasada "niezmienniczości organizacyjnej"? Co z niej wynika na poziomie molekularnym?
21. Dlaczego lepiej oprzeć filozofię kognitywną na fizyce niż automatach Turinga?
22. Dlaczego behawioryzm zdominował psychologię?
23. Na czym polega neobehawioryzm?
24. Co spowodowało narodziny kognitywizmu?
25. Co to jest minimalistyczny funkcjonalizm?
26. Co to jest komputacjonizm?
27. Skąd w systemach formalnych może pojawić się znaczenie symboli?
28. Podaj argumenty Turinga i jego odpowiedzi.
29. Jak zdefiniował umysł Newell z Simonem?
30. Dlaczego stwierdzenie "komputer liczy, a człowiek myśli" jest błędne?
31. Jak należy porównywać mózgi i umysły z sztucznymi systemami?
32. Co to są systemy oparte na wiedzy?
33. Co to jest fizyczny system symboliczny?
34. Jakie były przyczyny wprowadzenia testów IQ?
35. Od czego zależy kompetencja programów komputerowych prowadzących rozumowanie?
36. Wykreśl krzywe stałej kompetencji systemu grającego w szachy – co jest na osiach wykresu?
37. Na czym polega przewaga mózgów nad programami komputerowymi?
38. Na czym polega przewaga programów nad mózgami?
39. Opisać warianty filozofii kognitywnych: psychofunkcjonalizm, teoria identyczności, metodologiczny solipsyzm, różne rodzaje monizmu.

Literatura

Wykłady z filozofii umysłu [Marcina Miłkowskiego](#) (IFiS PAN, Warszawa) są na Youtube.

Następny krok w realizacji cywilizacji [może należeć do robotów](#) (National Geographic trailer).

[Hot Robot At SXSW Says She Wants To Destroy Humans.](#)

[Information Philosopher](#), czyli filozof informacji,

[Philosophy resources](#) | [Słownik filozofii umysłu](#) | [Problem ciała i umysłu](#) |

[Nowości na temat sztucznej inteligencji na moim Flipboard.](#)

Kevin O'Regan o tym, [jak zbudować robota, który czuje](#). Szczegóły dopracowywane są w projekcie: [FEEL – A sensorimotor approach to understanding consciousness](#) oraz [senso-motoryczna teoria świadomości](#).

Na YouTube jest wiele [ilustracji Chińskiego Pokoju](#).

[Center for Brains, Minds and Machines, MIT.](#)

[10 mind myth](#)

[Cognitive Psychology and Cognitive Neuroscience](#), Wikibook

[The Psychologist](#) jest popularnym pismem na temat psychologii, w którym poruszano zagadnienia mitów o słynnych eksperymentach w psychologii.

[Nieporozumienia wynikające ze statystyki](#) - D. Huck.

Książki:

- Damasio A, W poszukiwaniu Spinozy, Nowe Horyzonty, Rebis, Poznań 2005.
- Desmurget, M. i inn. Movement Intention After Parietal Cortex Stimulation in Humans Science:324(5928), 811-813, 2009, DOI: 10.1126/science.1169896
- Minsky M, The Society of Mind. Simon and Schuster, New York 1988.
- Putnam H, The Threefold Cord. Mind, body and world. Columbia University Press (2000).
- Pinker S, Tabula Rasa (2005).
- Turing A, Maszyna licząca a inteligencja, w: Filozofia umysłu, red. Bohdan Chwedeńczuk, Aletheia, Warszawa 1995
- Kloch J, Świadomość komputerów? Biblios Tarnów i OBI Kraków 1976
- Minsky, M, [Society of Mind](#). Simon and Shuster 1986
- Minsky, M, [The Emotion Machine](#): Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. Simon and Shuster 2006
- Newell A, [Unified Theories of Cognition](#), Harvard University Press 1990
- Newell A, Simon H.A. (1976): Computer Science as Empirical Inquiry: Symbols and Search. Communications of the ACM, 19 (3) pp. 113-126.
- Penrose R, Nowy umysł cesarza: o komputerach, umyśle i prawach fizyki, PWN, Warszawa 1995.
- Penrose R, Cienie umysłu. Poszukiwanie naukowej teorii świadomości. Zysk i S-ka, Poznań 2000.
- Tegmark M, [Życie 3.0](#). Jak być człowiekiem w epoce sztucznej inteligencji. Znak 2019

[IQ i narodowy test](#) na inteligencję.

Testy oparte na rosnącej złożoności zadań opisane są w pracy: José Hernández-Orallo y David L. Dowe. [Measuring Universal Intelligence](#): Towards an Anytime Intelligence Test. Artificial Intelligence 174(18): 1508

- Turing A, Maszyna licząca a inteligencja, w: Filozofia umysłu, red. Bohdan Chwedeńczuk, Aletheia, Warszawa 1995
- Velmans, M. [Is human information processing conscious?](#) Behavioral and Brain Sciences 14(4):651-701, 1991.

Cytowanie: Włodzisław Duch, Wstęp do Kognitywistyki. Rozdz. 27: Filozofia umysłu. UMK Toruń 2024.

[Następny wykład](#): Neurofilozofia | W. Duch, [Wstęp do kognitywistyki - spis treści](#)