

Jerzy Bobryk

Czy możliwa jest ingerencja wiedzy

Filozofia Nauki 1/1, 137-151

1993

Artykuł został zdigitalizowany i opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

Jerzy Bobryk

Czy możliwa jest integracja wiedzy

1. Wstęp

Jesteśmy świadkami powstania i rozwoju interdyscyplinarnego nurtu popularnie określanego mianem „kognitywizmu”. Kognitywizm jest złożonym fenomenem współczesnej kultury; jest on między innymi pewnym systemem wartości, pewną wizją człowieka i jego umysłu, jest także pewnym stylem uprawiania wielu dyscyplin naukowych. Kognitywizm jest jednocześnie silnie związany z wyłaniającą się właśnie nową dyscypliną wiedzy, określaną w języku angielskim terminem „*cognitive science*” (co w języku polskim bywa oddawane przez „nauka poznawcza”, „nauka o poznawaniu”), i jej najnowszym odgałęzieniem — neurologią kognitywną (*cognitive neuroscience*). Kognitywizm, w wersji proponowanej przez twórców «nauki o poznawaniu», nastawiony jest na zbadanie istoty poznania w ogóle — przez zintegrowanie wysiłków tak odmiennych dyscyplin, jak filozofia, psychologia, neurofizjologia, informatyka, lingwistyka, teoria kultury i wiele innych. W ramach tego nurtu, od samego początku jego powstania, czyli około roku 1975 [Bobrow, Collins], ukształtowała się specyficzna koncepcja umysłu. Koncepcja ta, wraz z towarzyszącymi jej metodami badań procesów poznawczych, została zdominowana — pomimo otwarcie deklarowanej interdyscyplinarności nurtu — przede wszystkim przez wstępne założenia i praktykę badawczą nauki o komputerach i teorii sztucznej inteligencji. Sytuacja ta budzi dyskusje na temat relacji pomiędzy metodami nauk formalnych i przyrodniczych z jednej, a humanistycznych — z drugiej strony. W dyskusjach tych biorą udział zarówno zwolennicy, jak i przeciwnicy nowej koncepcji umysłu. Rzecz sprowadza się w istocie do pytania, czy możemy przyjąć dyktowany przez paradygmat kognitywizmu kierunek integracji humanistyki, przyrodzoznawstwa i nauk formalnych.

Chciałbym w tym miejscu, nawiązując do sporów toczonych w ramach i wokół nowego paradygmatu badania umysłu, pokazać niezbędność oddzielenia obszarów przyrodznawstwa, humanistyki i nauk formalnych. Moim zdaniem, jasne oddzielenie obszarów tych trzech typów wiedzy nie tylko nie wyklucza, ale wręcz umożliwia integrację badań w tak odrębnych metodologicznie, a związanych przedmiotowo dziedzinach, jak psychologia, epistemologia, neurofizjologia i teoria sztucznej inteligencji.

Wygodną i użyteczną aparaturę pojęciową dla dokonania tego znalazłem w pracach Kazimierza Twardowskiego; przede wszystkim, choć nie tylko, w jego teorii czynności i wytworów.

1.1. Współczesna forma sporów o istotę myślenia. Cyborg Pylyshyna w chińskim pokoju Searle'a

Zwolennicy nurtu kognitywizmu, przedstawiając swoją wizję inteligencji, najczęściej powołują się na artykuł Alana M. Turinga „Computing Machinery and Intelligence” [Turing 1950], który zapoczątkował serię myślowych eksperymentów rozważających problem myślenia mechanicznego, i określających, jakie warunki są konieczne i wystarczające, aby określoną istotę lub urządzenie można było nazwać podmiotem myślącym. Turingowska odpowiedź na to pytanie oparta jest na sformułowanej w 1936 roku definicji „obliczania” i na koncepcji tak zwanej maszyny Turinga. Zadał on sobie wtedy pytanie, co naprawdę czyni ktoś, kto wykonuje obliczenia. Osoba taka postępuje według ciągu reguł, które są z góry ustalone i mechanicznie przestrzegane: utrwała pewne symbole na pewnym materialnym nośniku (np. taśmie papieru); jej kolejne kroki zależą od tego, co już utrzymała, i od przewidzianych reguł manipulowania symbolami. Ponieważ postępuje ona standardowo i automatycznie, działa w ustalony i dokładnie przewidziany sposób — działa jak «maszyna Turinga» — czynności jej mogą być zastąpione pracą urządzenia mechanicznego, np. komputera. Człowiek może obliczać i komputer może obliczać; pojawia się przeto chętnie przyjmowana w kręgu kognitywizmu hipoteza, że wszelkie procesy intelektualne, procesy myślenia, procesy poznawcze w ogóle, opierają się na manipulowaniu prostymi symbolami według z góry ustalonych reguł.

Przeciwnicy stawiania znaku równości, pomiędzy po Turingowsku rozumianym obliczaniem a szeroko pojmowanym myśleniem, podsuwają nam eksperymenty myślowe, które mają zachwiać wiarę w maszynową inteligencję. Wyobraźmy sobie — pisze John Searle [1984. 32-41] — że grono programistów napisało program mogący symulować rozumienie języka chińskiego. Kiedy zadamy komputerowi pytanie przy pomocy znaków z języka chińskiego, to wyprodukuje on odpowiedzi w języku chińskim. Wyobraźmy sobie dalej, że odpowiedzi komputera są tak doskonałe, jak odpowiedzi naturalnego użytkownika języka chińskiego. Czy faktycznie w takiej sytuacji komputerowe rozumienie języka chińskiego jest identyczne z rozumieniem Chińczyka? Przyjmijmy

teraz, że jakiś człowiek zamknięty jest w pokoju wypełnionym koszykami z symbolami chińskiego języka. Nie zna on chińskiego, ma jednak napisaną w znanym mu języku książkę z instrukcjami manipulowania chińskimi znakami. Reguły te są czysto formalne, określające syntaktykę, nie zaś — semantykę. Dana reguła może mieć postać: weź symbol *X* z koszyka nr 1 i zapisz go za symbolem *Y* z koszyka nr 2. Do pokoju są przesyłane z zewnątrz symbole chińskie, a zamknięta w nim osoba ma — posługując się książką reguł — wysyłać na zewnątrz ciągi innych symboli. Symbole pojawiające się w pokoju są w istocie zadanymi w języku chińskim pytaniami, zaś symbole wysyłane są odpowiedziami na nie: są to odpowiedzi zadowolające, jeśli książka instrukcji jest dobrze napisana. Z punktu widzenia zewnętrznego obserwatora osoba w «chińskim pokoju» zachowuje się tak, jakby znała język chiński (pomijamy tu oczywiście długość czasu potrzebnego na przeczytanie instrukcji i udzielenie odpowiedzi). W istocie zna ona — podobnie jak komputer symulujący rozumienie języka chińskiego — co najwyżej reguły łączenia znaków chińskich, zasady manipulowania nimi, czyli syntaktykę. Faktycznie język chiński znają ci, którzy napisali książkę instrukcji, oraz ci, którzy są w stanie ocenić, czy wychodzące z pokoju znaki są poprawnymi i sensownymi odpowiedziami na przesłane wcześniej pytania.

Zwolennicy obliczeniowej teorii umysłu najczęściej komentowali argumenty Searle'a, stwierdzając, że wprawdzie osoba z «chińskiego pokoju» nie rozumie języka, rozumie go jednak cały «system», czyli osoba, pokój z symbolami i książka instrukcji. Takie rozszerzanie podmiotu rozumienia nie zmienia jednak niczego. Jak długo nie uwzględnimy autora książki instrukcji i zewnętrznego interpretatora poprawności odpowiedzi, tak długo nie będziemy mogli znaleźć przejścia od syntaktyki do semantyki. Znaczenie istnieje tylko w momencie układania książki instrukcji i w momencie interpretowania znaków chińskich; poza tym mamy wyłącznie do czynienia z mechanicznym manipulowaniem oderwanymi od znaczenia symbolami. Mówiąc inaczej, to nie komputer, lecz komputer-wraz-z-użytkownikiem może być potraktowany jako odpowiednik ludzkiego umysłu [Bobryk 1987]. Jeżeli odrzucimy Turingowską komputerową metaforę ludzkiego umysłu, a na jej miejsce zaakceptujemy metaforę komputera-wraz-z-użytkownikiem, znajdziemy się (pozornie tylko, jak to postaram się pokazać) w punkcie wyjścia, bo nie wiemy dalej, jakie to specyficznie ludzkie cechy zapewniają zdolność do pełnego rozumienia języka, czy do myślenia w ogóle.

Zdaniem Searle'a, prawdziwe i pełne rozumienie związane jest z intencjonalnością aktów poznawczych. Wyjaśnieniu swojego rozumienia intencjonalności oraz pokazaniu nieintencjonalnego charakteru procesów maszynowych poświęcił Searle szereg tekstów [np. Searle 1980, 1984, 1984a]; stwierdzając tam [np. 1984a.89] m.in.: *Intencjonalne zjawiska umysłowe są częścią naszej biologicznej biografii. Bycie spragnionym, doznawanie wrażeń wzrokowych, posiadanie pożądań, obaw czy oczekiwań jest w tym samym stopniu aspektem osobistej biografii biologicznej, co oddychanie, trawienie czy spanie.*

Zjawiska intencjonalne, podobnie jak inne zjawiska biologiczne, są realnymi wewnętrznymi cechami określonych biologicznych organizmów, w taki sam sposób, jak mitoza i mejoza oraz wydzielanie żółci są realnymi wewnętrznymi cechami biologicznych organizmów. Wprawdzie Searle rozumie intencjonalność w sposób zgodny z tradycją fenomenologiczną, to jednakże zacytowane wyżej stwierdzenie, że intencjonalność jest produktem pracy ludzkiego mózgu, podobnie jak wydzielanie żółci jest produktem działania woreczka żółciowego, jest oczywistym odstępstwem od Husserlo-wskiego antynaturalistycznego rozumienia podmiotu poznania. Sugeruje ono także, że autor cytatu całkowicie utożsamia czynności psychiczne z procesami neurofizjologicznymi.

To wyrażane przez Searle'a przekonanie o silnym związku intencjonalności z działaniem mózgu wykorzystał Zenon Pylyshyn do rozprawienia się z sugestiami zawartymi w idei «chińskiego pokoju». Załóżmy, pisze Pylyshyn [1980], że osoba znajdującą się w «chińskim pokoju», lecz faktycznie znająca język chiński, a więc nie tylko jego syntaktykę, ale także semantykę, będzie stopniowo poddawana zabiegowi cyborgizacji, czyli zastępowania fragmentów jej biologicznego mózgu, komórka po komórce, jakimiś elektronicznymi mikroukładami, których rezultaty pracy będą identyczne z rezultatami pracy organicznych neuronów. Znaczący to, że każda taka sztuczna komórka, odbierając od naturalnej pewne sygnały na wejściu, będzie na wyjściach przekazywać innym naturalnym komórkom takie same sygnały, jakie przekazywałaby komórka naturalna.

Rodzi się pytanie, w którym momencie cyborgizowany człowiek utraci intencjonalność swoich aktów poznawczych. Jeśli, jak chce Searle, intencjonalność związana jest z organicznością struktur przetwarzających informacje musimy uznać, że wymiana struktur organicznych na nieorganiczne pozbawia człowieka tej cechy. Czy jednak utrata intencjonalności nastąpi w momencie wymiany jakiegokolwiek pierwszej komórki? Może trzeba w tym celu wymieniać ich znacznie więcej, może 25%, może 51%? A może mózg ludzki zachowuje intencjonalność tak długo, jak długo zostanie w nim przynajmniej jedna naturalna komórka? Przy takim postawieniu problemu każda odpowiedź będzie mało przekonująca.

1.2. Na czym polegają różnice w stanowiskach oponentów

W powyższym ścieraniu dwu eksperymentów myślowych uchwycić można kilka (paradoksalnie — więcej niż dwie) nie do końca wyartykułowanych koncepcji umysłu. W stanowisku Pylyshyna, będącym kontynuacją poglądów Turinga, mamy osobliwe połączenie podejścia behawiorystycznego z podejściem strukturalno-funkcjonalnym. Z jednej strony, autorzy mają tendencję do rozpoznawania inteligencji, przede wszystkim, lub wyłącznie, po efektach działania jakiegoś układu. Stwierdzają zatem, że skoro sztuczne urządzenie potrafi sprostać dokładnie tym samym — a uznanym za wyma-

gające cechy inteligencji — zadaniom, co człowiek, to nie ma istotnej różnicy pomiędzy inteligencją ludzką i maszynową. Z drugiej strony jednakże, autorzy ci porzucają konsekwentnie behaworystyczny punkt widzenia, «zaglądają» do wnętrza inteligentnie działającej «czarnej skrzynki», pokazują nam, jakie i jak uporządkowane operacje gwarantują inteligentne zachowanie systemu. (Interesują się tak zwanymi programami działania systemu.) W tym strukturalno-funkcjonalnym podejściu abstrahuje się jedynie od materialnego podłoża inteligencji: określa, że jest sprawą bez znaczenia, jakie to struktury, organiczne, mechaniczne, elektroniczne czy jakiegokolwiek inne realizują funkcje intelektualne.

Na stanowisku Searle'a patrzy się na inteligencję od strony fenomenologicznej. (Nie jest to jednak tradycyjna Husserlowska transcendentna fenomenologia, jest to raczej stanowisko zbliżone do tzw. fenomenologii egzystencjalnej [Dreyfus w: Dreyfus, Hall 1984].) Mówi się tam, że ani odpowiednie uporządkowanie operacji składających się na procesy umysłowe, ani nawet efekty działania tych procesów, nie wystarczą do uczynienia systemu prawdziwie myślącym. Liczy się przede wszystkim «intencjonalność» owych procesów operacji czy «intencjonalność» działania systemu. Jednocześnie sam Searle, jako oponent funkcjonalnego podejścia Pylyshyna, dość stanowczo stwierdza, że intencjonalność w sposób konieczny związana jest z działaniem substancji organicznej mózgu, co przynajmniej pozornie zbliża go do teorii identyczności lub tożsamości stanów umysłowych i stanów ośrodkowego układu nerwowego. Wprawdzie Searle nie podpisuje się w istocie rzeczy pod poglądem, że stany umysłowe nie są niczym innym niż fizycznymi stanami mózgu, a więc pod teorią J.J.C. Smarta czy jego współczesnego następcy Davida M. Armstronga [1982], jednak jego oponent Pylyshyn rozprawia się z jego poglądami tak, jakby w istocie nie było różnic pomiędzy Searlem i Armstrongiem. Różnice te istnieją jednak i zawierają się w Searle'owskim pojęciu intencjonalności. Ma ono u Searle'a i jego poprzedników szereg złożonych i subtelnych treści (niektóre z nich będą dalej analizował), pośród których najważniejsze jest to, że procesy intencjonalne zawsze mają jakieś znaczenie (*meaning*) i odniesienie (*refer-*

ence), zawsze też ustalają stosunek poznającego podmiotu do czegoś poza nim. Tymczasem autorzy, reprezentujący stanowisko, jakiego broni Pylyshyn, definiują procesy umysłowe jako „procesy manipulowania fizycznie ucieleśnionymi znakami według pewnych reguł formalnych”, i przyjmują zasadę tzw. metodologicznego solipsyzmu [Fodor 1981/1984] nakazującą badanie umysłu (maszynowego czy ludzkiego) tak, jakby istniał jedynie on. Tym samym pomijają oni, jak to podkreśla Searle, problemy semantyczne i sferę intencjonalności¹.

2. Czynności umysłowe, czynności fizyczne, czynności posługiwania się znakami z języków formalnych

Właściwe postawienie pytań o istotę tego, co umysłowe, musi być poprzedzone jasnym ustaleniem relacji pomiędzy zewnętrznymi obserwowalnymi zachowaniami, czynnościami umysłowymi, a czynnościami posługiwania się znakami w ogóle, znakami zaś języków formalnych w szczególności. Przy tej okazji pojawi się pytanie o stosunek czynności umysłowych do nieobserwowalnych czynności «wewnętrznych», między innymi do czynności ośrodkowego układu nerwowego. Rozwikłanie tego typu problemów to w istocie częściowe przynajmniej rozwikłanie stosunków pomiędzy takimi dyscyplinami, jak nauki o zachowaniu, introspekcjonistyczna psychologia, lingwistyka lub semiotyka, logika i neurofizjologia.

2.1. Czynności a wytwory

Myślenie, obliczanie, posługiwanie się znakami — to przykłady czynności ludzkich. Każda czynność to mniej lub bardziej złożony proces ukierunkowany na powstanie wytworu [Tomaszewski 1963, Twardowski 1912/1965]. Wytworem czynności wyobrażania sobie jest wyobrażenie, wytworem czynności przemawiania jest przemowa, wytworem czynności malowania jest malowidło, a wytworem czynności pisania

1) Powyższe polemiki streszczałem tak, jakby nie było różnicy pomiędzy używaniem zwrotów typu „myślenie”, „poznanie”, „działanie intelektu”, „procesy umysłowe”. Cytowani polemicy mają tendencję do utożsamiania niektórych spośród tych terminów; ich rozróżnienie nie wydaje się zresztą konieczne, gdy stawiamy tak ogólny problem, jak podobieństwo sztucznych i naturalnych mózgów. Można mimo wszystko odnieść wrażenie, że spory tego typu mają charakter czysto terminologiczny, i mogą zostać przecięte odpowiednimi definicjami. Można np. „umysł” albo „intelekt” lub „podmiot myślenia”, zdefiniować jako „system posługujący się znakami...” lub jako „coś, co posiada organiczny mózg...”. Nie będzie wtedy żadnych wątpliwości, co jest, a co nie jest podmiotem myślącym. Jednakże w cytowanych sporach, nie chodzi o uznanie całkowitej równości (aktualnie konstruowanych i programowanych) sztucznych i naturalnych intelektów. Chodzi w nich o odpowiednie postawienie problemów, o zrozumienie fenomenu ludzkiej inteligencji. Zarzut mój polega na tym, że komputerowa metafora ludzkiego umysłu nie zawsze jest rozumiana właściwie. Chcę też pokazać przyczyny tego niewłaściwego rozumienia.

jest pismo. Mamy nietrwałe wytwory czynności: wtedy wytwór zlewa się z czynnością i istnieje tak długo, jak długo istnieje czynność (spacerowanie-spacer, mówienie-mowa); mamy też wytwory trwale istniejące po zakończeniu samej czynności (rysowanie-rysunek). Zdaniem twórcy teorii czynności i wytworów [Twardowski 1912/1965.229]², wszystkie wytwory czynności psychicznych są nietrwałe, wyobrażenie istnieje tylko tak długo jak długo sobie coś wyobrażam, a każda niezapisana myśl istnieje tylko wtedy, kiedy myślę. To kontrowersyjne (w sytuacji kiedy ani psychologia, ani inna dyscyplina nie ustaliły ostatecznie charakteru relacji pomiędzy procesami psychicznymi i procesami neurofizjologicznymi) twierdzenie Twardowskiego o nietrwałości wytworów czynności psychicznych zostało powtórzone przeszło pół wieku później przez Neissera [1967], który uznał, że w ludzkim umyśle istnieją nie tyle gotowe i wielokrotnie przywoływane idee, pojęcia lub wyobrażenia, co ślady wykonywanych poprzednio czynności umysłowych, ślady powodujące, że później podejmowane czynności umysłowe przebiegają podobnie i tworzą podobne wytwory. Przekonanie o trwałości wytworów czynności psychicznych wynika z tego, że czynnościom takim towarzyszą wielokrotnie czynności fizyczne, prowadzące do trwałych — a więc istniejących jeszcze po zakończeniu obu czynności — wytworów, zwanych przez Twardowskiego „psychofizycznymi”, których funkcja polega na przechowywaniu treści psychicznych. Najbanalniejszym przykładem będzie tu (psychiczna) czynność myślenia, której towarzyszy (fizyczna) czynność zapisywania powstających myśli. Mamy zatem do czynienia z psychofizyczną czynnością myślenia i notowania, w wyniku której powstaje trwały psychofizyczny wytwór — notatka, tekst *etc.* Taki wytwór psychofizyczny jest znakiem (ekspresją) pewnych treści psychicznych; wytwór ten utrwalony i oglądany później będzie budził (w tym samym lub innym podmiocie) podobne treści psychiczne, jakie istniały przy zapisywaniu. Znaczy to, że pewne wytwory psychiczne istnieją potencjalnie w wytworach psychofizycznych (w swoich znakach) [Twardowski 1912/1965.234]. Nietrwały wytwór psychiczny może także znaleźć swój wyraz w równie nietrwałym wytworze psychofizycznym, czego przykładem jest ekspresja myśli w mowie.

Ten zaproponowany przez Twardowskiego podział czynności i wytworów pokazuje wyraźnie funkcje znaków w kształtowaniu czynności umysłowych. Znaki odrywają wytwory od czynności, utrwalają nietrwałe wytwory i w ten sposób ułatwiają ponowne powstanie pewnych wytworów, i organizują przebieg czynności umysłowych. Ponieważ znaki, jako wytwory trwałe, a już z pewnością zewnętrzne, przekazywane są między podmiotami, są też wspólnie tworzone i doskonalone, a także zbiorowo organizowane w pewne systemy (systemy semiotyczne). Kolejnym banalnym (bo powszech-

2) Pracę Twardowskiego *O czynnościach i wytworach* uważam za naturalną kontynuację pracy *O treści i przedmiocie przedstawień*. Mówiąc inaczej, teoria czynności i wytworów [Twardowski 1912/1965] jest kontynuacją teorii intencjonalności [Twardowski 1984/1965].

nie znanym, choć nie zawsze do końca przeanalizowanym) przykładem jest poprawianie pisemnej lub ustnej wypowiedzi ucznia przez nauczyciela, co oczywiście ma wpływ nie tylko na doskonałość wytworu psychofizycznego (wypowiedzi, tekstu), ale także i na przebieg procesów umysłowych ucznia. Także pojedynczy podmiot nauczony utrwalania wytworów własnych czynności umysłowych, korzysta z tej szansy nie tylko, by zapamiętać (utrwalić), ale także by uchwycić własne rodzące się myśli. Wystarczy tutaj przypomnieć rolę notatek, szkiców, schematów, ideogramów.

2.2. Artefakty

Aby do końca oddać rolę zewnętrznych i zbiorowo tworzonych znaków dla kształtowania przebiegu indywidualnych i wewnętrznych czynności umysłowych, trzeba przypomnieć jeszcze jedno wprowadzone przez Twardowskiego [1912/1965.237] pojęcie. Istnieje pewna szczególna klasa wytworów, zwana przez niego „artefaktami”. Odcisk stopy w glinie może być zrobiony sztucznie, choć może niczym nie różnić się od odcisku prawdziwej stopy: aktor może zachować się dokładnie tak, jak zachowuje się człowiek, przeżywający w danym momencie autentyczne uczucie; przemawiający może wyrazić myśli rodzące się w jego głowie, ale może też odtworzyć mechanicznie zapamiętane subtelne rozumowanie (swoje własne z przeszłości lub rozumowanie innego), może wreszcie odczytać coś z kartki. Autentyczność, czyli nieartefaktowość wytworu, jest w pewnych wypadkach sprawą bez znaczenia (drukarnik nie musi mieć umysłowości Husserla, by przekazać nam jego *Idee*); bardzo często liczy się nie to, jak powstaje wytwór, lecz to, jaki on jest. Sami zresztą także nie musimy powtarzać czynności psychicznych prowadzących do powstania określonych treści, możemy sobie przypomnieć lub odczytać z notatek słowa te treści wyrażające. Ze szczególnego rodzaju artefaktów korzysta logika [Twardowski 1912/1965. 238]. Przykład logiki pokazuje coś więcej. Znaki z systemów formalnych nie tylko utrwalają i zastępują pewne, psychicznie istniejące, pojęcia, sądy i operacje intelektualne; pozwalają nam ponadto na posługiwanie się tymi znakami, w oderwaniu od ich treści, bez konieczności tworzenia jakichś wytworów psychicznych, w których te treści istnieją. *Tego rodzaju wytwory zastępcze przedstawiają najskrajniejszy przypadek niezależnienia wytworów psychicznych od czynności, dzięki którym istnieć mogą naprawdę, czyli aktualnie. Operując takimi wytworami zastępczymi i w nauce i w życiu potocznym nie odczuwamy też żadnej trudności w operowaniu niezastępczymi wytworami niezależnie od wytwarzających je czynności, tym bardziej, że niezmiernie często wytwory rzeczywiste i zastępcze zjawiają się na przemian...* [Twardowski 1912/1965.239].

2.3. Myśli a środki wyrazu myśli

Tak więc jeszcze przed powstaniem idei maszyny Turinga, przed jej ucieleśnieniem w wynalazku komputera, wielu autorów nie miało wątpliwości, że manipulowanie

symbolami według sztywno ustalonych reguł nie jest istotą myślenia; jest co najwyżej jego częścią lub jego szczególną postacią. Postać ta istnieje dopiero wtedy, gdy zostaną już wynalezione pewne szczególne formy utrwalania wytworów czynności psychicznych, narzędzia oddzielania czynności od wytworów, sposoby uwalniania człowieka od części jego pracy intelektualnej, posługiwania się artefaktami w miejsce oryginalnych wytworów. Tego typu wzmacniacze naturalnych intelektów zmieniają ludzkie nawyki intelektualne w sposób podobny do tego, w jaki maszyny, narzędzia i silniki zmieniają charakter ludzkiej aktywności fizycznej. Podobnie też myślenie w języku logicznego rachunku zdań (a więc przy pomocy specjalnych symboli i specjalnych zasad posługiwania się nimi) jest nową formą myślenia. Myślenie przy pomocy komputera jest nową formą działalności intelektualnej. Z drugiej strony, trzeba pamiętać, że np. komputerowa symulacja zjawisk meteorologicznych jest tylko formą ich analizy (być może ich przewidywania), nie zaś formą tworzenia wiatru, deszczu i opadów. Trzeba zatem pamiętać o podwójnej funkcji tak zwanej sztucznej inteligencji. Może ona być wzmacniaczem, uzupełnieniem, lub protezą ludzkiej działalności poznawczej; może też być narzędziem badania ludzkiego intelektu i jako komputerowa symulacja może służyć do analizy ludzkiego umysłu. Symulacje procesów poznawczych nie są przy tym tworzeniem myśli, podobnie jak symulacje procesów meteorologicznych nie są tworzeniem deszczu i śniegu.

Zajmijmy się jednak w tym miejscu tylko pierwszą funkcją sztucznej inteligencji, a więc przykładem ogólniejszego (i wcześniejszego od wynalazku komputera) zjawiska uzewnętrzniania i automatyzowania ludzkich czynności psychicznych i utrwalania ich wytworów. Jak to już zostało powiedziane, do tej kategorii zjawisk należy np. rachowanie przy pomocy kartki i ołówka, czy wnioskowanie w obrębie logicznego rachunku zdań. Przyjęta tutaj za Kazimierzem Twardowskim perspektywa pozwala na jasne ujęcie stosunku przedmiotu logiki do przedmiotu psychologii procesów poznawczych. (Oczywiście «jasne ujęcie» jest tylko początkiem, a nie końcem rozwikłania problemu zależności pomiędzy tymi naukami.) Z tej perspektywy logika nie jest nauką o myśleniu w ogóle, ani nauką o ukrytych («głębokich») aspektach ludzkiego intelektu. Jest ona nauką o specjalnym typie ludzkiej działalności intelektualnej, działalności nie mającej charakteru czysto wewnętrznego (niezależnie od tego, czy przez procesy «wewnętrzne» rozumiemy procesy dostępne introspekcyjnie, czy procesy mózgowy, czy też jedno i drugie razem) — działalności przy użyciu zewnętrznych oraz kulturowo i społecznie wypracowanych narzędzi. Takimi narzędziami są systemy notacji symbolicznej i zasady posługiwania się tą notacją. Sam Twardowski widział przedmiot logiki i nauk formalnych w ogóle chyba nieco wężiej, utrzymywał bowiem, że logika zajmuje się przede wszystkim pewnego typu wytworami (głównie artefaktami) [Twardowski 1912 /1965. 238-240], nie zaś czynnościami intelektualnymi w ogóle. Najistotniejsze jest w każdym razie stwierdzenie artefaktowości obszaru dociekań logiki i sztuczności

tak zwanych języków formalnych (jako języków skonstruowanych przez ludzi). Warto może dodać, że wątek ten rozwija np. Ludwig Wittgenstein w *Dociekaniach filozoficznych* [1972], a za nim niektórzy przedstawiciele tzw. filozofii postmodernistycznej [np. Rorty 1989 i 1991].

2.4. Procesy neurofizjologiczne, ruchy fizyczne, a procesy intencjonalne

Jeśli przyjmiemy zasadę tzw. ekstrakortykalnej organizacji wyższych czynności psychicznych [Łuria 1976, Wygotski 1971], czyli teorię uznającą, że dla wykonania pewnych operacji intelektualnych musimy wspierać nasze biologiczne mózgi sztucznymi narzędziami i społecznie wypracowanymi zasadami ich użycia, jeśli uznamy postulowaną przez Twardowskiego nietrwałość wytworów czynności psychicznych i powszechność używania artefaktów, to musimy odrzucić teorię identyczności procesów mózgowych i procesów umysłowych. Znaczy to, że żadna teoria, usiłująca opisać ludzkie procesy poznawcze, nie może być zastąpiona fizjologią ośrodkowego układu nerwowego. Istnieje dodatkowy i być może istotniejszy powód odrzucenia neurofizjologicznego redukcjonizmu. Przyjmuje się powszechnie, że procesy poznawcze mają charakter procesów intencjonalnych. Procesy intencjonalne (zwane najczęściej „aktami” lub „czynnościami”) wykazują szczególne właściwości, których nie wykazują inne zjawiska i procesy. Specyfika procesów intencjonalnych nie polega wyłącznie na tym, że zawsze są one «o czymś», czyli mają przedmioty i treści. Procesy te podlegają, lub zdają się podlegać, dosyć szczególnemu rodzajowi przyczynowego uwarunkowania. Faktycznie lub pozornie przebiegają one mianowicie tak, że nie można ich w sposób satysfakcjonujący opisać i nie można wyjaśnić ich przebiegu używając kategorii przyczyny sprawczej. Dlatego współczesna humanistyka [Searle 1984] wprowadza pojęcie „przyczynowości intencjonalnej” (*intentional causation*). Z powodu intencjonalności ludzkich procesów poznawczych nie można też ludzkich czynności zredukować do «reakcji» i opisać w taki sposób, jak spodziewała się tego psychologia behawiorystyczna. Reakcje, w przeciwieństwie do czynności, zdeterminowane są przyczynami sprawczymi — bodźcami. O ile czynności (i akty intencjonalne) zdeterminowane są tym, co w wyniku ich pojawienia się nastąpi, o tyle reakcje zdeterminowane są tym, co je poprzedza; na tym zresztą polega ich specyfika.

Czynności ludzkie nie są zewnętrznymi zachowaniami, nie dadzą się też opisać jako reakcje, którym towarzyszą, lub które poprzedzają jakies «wewnętrzne» umysłowe procesy. Różnorodne «umysłowe» aspekty ludzkiego działania, takie jak np. plany tego działania, jego potocznie rozumiane intencje, treści *etc.*, zawarte są bardzo często w samym działaniu, są w nim immanentnie obecne, nieoddzielalne od niego. John Searle [1984] wprowadził pojęcie „uprzywilejowanego opisu ludzkiego działania”. Opis taki obejmuje informacje, które muszą się pojawić, jeśli chcemy wiedzieć, jakiego typu działania są właśnie realizowane. Informacje te pojawią się w świadomości podmiotu,

gdy spytamy go, co robi (choć wcale nie muszą pojawia się w jego świadomości, gdy nie będziemy go pytać, a działanie nie natrafi na przeszkody). Dla aktów wyobrażenia sobie, charakter «uprzywilejowanej informacji» ma informacja o treści wyobrażenia; dla opisu ludzkich pragnień i obaw najważniejszą jest informacja, czego pragniemy lub czego się obawiamy³.

Akty intencjonalne opierają się na materialnych strukturach i zjawiskach lub procesach nie mających charakteru intencjonalnego. Nie można myśleć, nie mając mózgu, w którym zachodzą złożone procesy biologiczne, chemiczne i fizyczne, podobnie nie można malować, nie mając farb. Można jednak, mając mózg, nie myśleć, a mając farby, nie malować. Fizjologia układu nerwowego może być nauką pomocniczą dla psychologii i epistemologii, podobnie jak optyka i chemia mogą być naukami pomocniczymi dla teorii sztuki. Zresztą to właśnie neurofizjologia uczy nas, że pewne funkcje psychiczne mają zmienną i dynamiczną lokalizację w mózgu, co między innymi znaczy, że dla zrealizowania danej czynności umysłowej w różnych warunkach (chodzi zarówno o warunki zewnętrzne jak i warunki podmiotowe) muszą być uaktywnione za każdym razem nieco inne partie mózgu [Łuria 1976 i 1976a]. Mówiąc inaczej, tożsamość funkcji (czynności umysłowej) nie oznacza tożsamości przebiegu procesów nerwowych. Dynamiczność mózgowej lokalizacji funkcji umysłowych to drugi, poza ich ekstrakortykalną organizacją, powód odrzucenia programu neurofizjologicznego redukcjonizmu.

3. Podsumowanie. Obszary badań różnych dyscyplin naukowych w świetle teorii czynności i wytworów

Najczęściej rozróżniamy dyscypliny naukowe na podstawie przedmiotu ich badań lub na podstawie stosowanych metod badawczych. Obydwa kryteria zawodzą w wypadku teorii wiedzy (*cognitive science*). Mamy tu do czynienia z pozorną jednolitością przedmiotu integrowanych w jej ramach dyscyplin. Wszystkie one wszak zajmują się «poznaniem». Ta pozorna jednolitość przedmiotu narzuca nam jednolitość metod badawczych (komputerowe symulacje ludzkich procesów poznawczych w wypadku teorii wiedzy, i komputerowe symulacje procesów neurofizjologicznych w wypadku neurologii kognitywnej). Wszystko to stwarza wrażenie, jakoby humanistyczne, przyrodnicze i formalne dyscypliny wiedzy współpracujące w badaniach nad poznaniem i umysłem mogły automatycznie przejmować od siebie podstawowe założenia i metody badawcze. W rzeczywistości dyscypliny badające «poznanie» nie dotyczą tego samego przedmiotu. Widać to wyraźnie, gdy ujmemy takie zjawiska, jak procesy poznawcze, czy czynności poznawcze, w kategorii pojęciowe zaproponowane przez Kazimierza

3) Zauważmy, że obawy i pragnienia mają swoje treści i przedmioty nawet wtedy, gdy nie są uświadomione. To jednak, czego pragnę lub czego się obawiam, będzie wstępnie określać moje działanie, jak inne akty intencjonalne; będzie ustalać mój stosunek do otoczenia.

Twardowskiego. Pewne z tych dyscyplin (np. filozofia i humanistycznie zorientowana psychologia) zajmują się procesami poznawczymi od strony ich intencjonalności; badają one nie tyle procesy, co czynności poznawcze. Inne dyscypliny (np. neurofizjologia i przyrodniczo zorientowana psychologia) mogą badać nieintencjonalne podstawy procesów intencjonalnych. Ta druga grupa dyscyplin, w przeciwieństwie do pierwszej, może opisywać badane przez siebie procesy w kategoriach przyczyn sprawczych.

3.1. Niemożliwość zredukowania psychologii procesów poznawczych do neurofizjologii

Procesy zachodzące w mózgu, niezależnie od tego, czy rozważamy je na poziomie procesów elektrycznych, chemicznych czy neurofizjologicznych, same w sobie nie są «o czymś», nie są intencjonalne. To, co się dzieje w mózgu, jedynie umożliwia przebieg procesów intencjonalnych. Nawet doskonała znajomość procesów fizjologicznych będących podstawą mojego aktu wyobrażenia sobie czegoś, nie powie nam nic ani o przedmiocie, ani o treści wyobrażenia, tak jak chemiczna analiza atramentu nie powie nam nic o treści pisma. Wiemy, co wiąże określone dźwięki lub plamy atramentu na papierze — czyli materialny nośnik znaku — z jego treścią; są to tzw. społeczne konwencje. Odpowiedź na pytanie, co decyduje o tym, jak poszczególne ciągi procesów mózgowych (np. te wszystkie procesy mózgowie, które zachodzą w moim mózgu, gdy myślę lub piszę o intencjonalności) powiązane są z pewnymi przedmiotami, jest czymś bardziej złożonym. Niemniej jednak obydwa pytania są do siebie bardzo podobne. Tymczasem oczywisty udział konwencji w tym, że określone dźwięki lub ślady na papierze stają się znakami, nieczęsto prowadzi do przypuszczenia, że w sposób analogiczny o semantyczności pewnych procesów mózgowych decydują nie tylko naturalne czynniki. Żadna z nauk przyrodniczych nie jest w stanie badać «sensu», «znaczenia»; nie może tego czynić ani fizyka, ani chemia, ani biologia, ani też neurofizjologia. Jeżeli może istnieć neurologia kognitywna, to nie może być ona w całości nauką przyrodniczą, nie może być też uprawiana według wzoru nauk przyrodniczych. Wprawdzie procesy mózgowie możemy porównywać do procesów trawienia, jednakże zarówno pierwsze, jak i drugie nie są procesami intencjonalnymi.

3.2. Odrębność przedmiotu psychologii procesów poznawczych i nauk formalnych

Stwierdziliśmy powyżej, że procesy obliczania w świetle teorii czynności i wytworów są procesami psychofizycznymi, nie mogą być zatem modelem czysto psychicznych (wewnętrznych) procesów poznawczych. Musimy więc odrzucić albo «obliczeniową» teorię umysłu, albo metodologiczny solipsyzm; tymczasem propagatoryzy kognitywizmu [Fodor 1981/1984] każą nam przyjąć jedno i drugie. Sytuacja wygląda jednak tak, że w istocie nie możemy przyjąć ani jednego, ani drugiego.

Po pierwsze, nie możemy postawić znaku równości pomiędzy procesami obliczania a procesami «wewnętrzny» (niezależnie od tego, czy będziemy te ostatnie rozumieć jako procesy dostępne introspekcyjnie, czy jako procesy mózgowy), bowiem dla realizacji niektórych aktów poznawczych ludzie uciekają się także do zewnętrznych środków i zachowań.

Po drugie, procesy obliczeniowe (np. procesy przebiegające w maszynach cyfrowych), jako procesy manipulowania znakami według mechanicznie ustalonych reguł, nie mogą być, z dwu powodów, modelem ludzkich procesów poznawczych. Procesy obliczeniowe same w sobie pozbawione są aspektu semantycznego (nie są intencjonalne), są czysto syntaktyczne. To, co dzieje się w komputerze, wymaga interpretacji semantycznej, a ową interpretację semantyczną zapewniają użytkownicy komputerów, czyli ludzie. Dalej, maszynowe obliczanie, jako manipulowanie znakami według zawsze tych samych i z góry ustalonych reguł, jest co najwyżej odpowiednikiem niektórych ludzkich czynności poznawczych⁴.

3.3. Czy *cognitive science* może zastąpić epistemologię

Interdyscyplinarny ruch kognitywizmu przyjmuje aktualnie postać specyficznego redukcjonizmu. Próbuje się w nim zamiast procesów poznawczych (albo procesów neurofizjologicznych) badać procesy przebiegające w maszynach cyfrowych. Przy takim ujęciu mamy do czynienia nie tyle z integracją różnych dyscyplin naukowych, co raczej z próbą redukcji ich przedmiotu badań do badań procesów zachodzących w maszynach cyfrowych.

Niektórzy filozofowie głoszą konieczność zastąpienia refleksji epistemologicznej badaniami różnych szczegółowych dyscyplin naukowych [Niżnik 1992]. Program kognitywizmu zdaje się wychodzić naprzeciw tym filozoficznym tendencjom. Wizja zastąpienia spekulacji filozoficznych badaniami nad sztuczną inteligencją, możliwość uprawiania «twardej nauki» zamiast «martwej» filozofii jest kusząca. Rezygnacja z filozofii zdaje się jednak oznaczać rezygnację z jakiegokolwiek refleksji metanaukowej, rezygnację z badania odrębności i specyfiki poszczególnych dyscyplin naukowych. W konsekwencji prowadzi to do redukcjonizmu, do mieszania procesów psychicznych z psychofizycznymi, artefaktów z oryginalnymi wytworami, aktów intencjonalnych z procesami fizycznymi; na koniec — mieszania różnych rodzajów przyczynowości.

4) Nie chciałbym rozstrzygać, czym jest przedmiot nauk formalnych. Po prostu akceptuję stanowisko antypsychologiczne, stwierdzające m.in., że przedmiotem badań logiki nie są ani psychologiczne procesy „zachodzące w głowie konkretnej jednostki, która właśnie myśli”, ani tym bardziej procesy mózgowy. Akceptacja antypsychologizmu nie oznacza automatycznie akceptacji twierdzenia Twardowskiego, że logika jest nauką o szczególnym typie artefaktów, ani twierdzenia, że logika i matematyka zajmują się wyłącznie po Turingowsku rozumianym obliczaniem. Skłaniałbym się ku uznaniu obu tych twierdzeń; nie wydają mi się one jednak tak bezdyskusyjne, jak ogólne założenia antypsychologizmu.

Bibliografia

- Armstrong, David M.
 1982 - *Materialistyczna teoria umysłu*, Warszawa, PWN.
 1980 - „What is consciousness ?”, [w:] *The Nature of Mind*, Queensland, University of Queensland Press.
- Bobryk, Jerzy
 1987 - *Locus umysłu*, Wrocław, Ossolineum.
- Bobrow, Daniel G. ; Collins, A.
 1975 - *Representation and Understanding. Studies in Cognitive Science*, New York, Academic Press.
- Dreyfys, Hubert L.
 1984 - „Husserl's Perceptual Noema” [w:] H.L.Dreyfus, H.Hall (red), *Husserl, Intentionality, and Cognitive Science*, Cambridge (Mass.), The MIT Press.
- Fodor, Jerry A.
 1981/1984 - „Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology”, [w:] H.L.Dreyfus, H.Hall (red), *Husserl, Intentionality, and Cognitive Science*, Cambridge (Mass.), The MIT Press.
- Luria, Aleksander R.
 1976 - *Podstawy neuropsychologii*, Warszawa, PZWL.
 1976a - *Problemy neuropsychologii*, Warszawa, PWN.
- Neisser, Ulrich
 1967- *Cognitive Psychology*, Englewood Cliffs, Prentice-Hall.
- Niżnik, Józef
 1992 - „Pogranicza epistemologii”, [w:] J.Niżnik (red.) *Pogranicza epistemologii*, Warszawa, Polska Akademia Nauk, Instytut Filozofii i Socjologii .
- Pylyshyn, Zenon W.
 1980 - „The causal power of machines”, *The Behavioral and Brain Sciences*, 3, s.442-444.
- Rorty, Richard
 1989 - *Contingency, Irony and Solidarity*, Cambridge, Cambridge University Press.
 1991 - *Essays on Heidegger and Others*, Cambridge, Cambridge University Press.
- Searle, John R.
 1980 - „Minds, Brains and Programs”, *Behavioral and Brain Sciences*, 3, s.417-424.
 1984 - *Minds, Brains and Science*, Cambridge (Mass.) , Harvard University Press.
 1984a - „Intentionality and Its Place in Nature”, *Dialectica*, vol. 38, no 2/3, s.87-99.
- Tomaszewski, Tadeusz

1963 - *Wstęp do psychologii*, Warszawa, PWN.

Turing, Allan M

1950 - „Computing Machinery and Intelligence”, *Mind*, vol. LIX, no 236, s.433-460.

Twardowski, Kazimierz

1912/1965 - „O czynnościach i wytworach”, [w:] *Rozprawy i artykuły filozoficzne*, Warszawa, PWN.

1894/1965 - „O treści i przedmiocie przedstawień”, [w:] *Rozprawy i artykuły filozoficzne*, Warszawa, PWN.

Wittgenstein, Ludwig

1972 - *Dociekania filozoficzne*, Warszawa, PWN.

Wygotski, Lew S.

1934/1971 - *Wybrane prace psychologiczne*, Warszawa, PWN.