

dr hab. Adriana Schetz, prof. US  
Instytut Filozofii i Kognitywistyki  
Uniwersytet Szczeciński  
ul. Krakowska 71-79  
71-017 Szczecin  
tel. 667 534 554

**Ocena rozprawy doktorskiej mgra Krzysztofa Krenca pt.  
*Can Machines have Free Will? Analysis of the Concept of Free Will in relation  
to the Psychophysical Problem***

**Streszczenie**

Tematem przedłożonej mi do recenzji pracy jest zagadnienie niemal tak stare, jak sama filozofia, mianowicie problem istnienia wolnej woli. Autor na wstępie zaznacza, że nie tyle interesuje go pojęcie wolnej woli pojawiające się w różnych ujęciach z zakresu filozofii umysłu, co pojęcie wolnej woli niejako implikowane postawami, jakie filozofowie obierają broniąc określonych poglądów na naturę relacji psychofizycznej (s. 5). Jest to interesujące podejście, które zapowiada pewną świeżość ujęcia tego starego problemu. Co więcej, pojęcie wolnej woli analizowane jest przez Autora dysertacji w kontekście idei myślących maszyn, a dokładniej, w kontekście sztucznych sieci neuronowych (s. 86-87). Doprecyzowując zatem temat dysertacji przygotowanej przez mgra Krenca można by powiedzieć, że brzmi on: *Can artificial neural networks operate on the basis of free will?*, albo jakoś podobnie. Autor dołożył starań by uzasadnić aktualność podjętego problemu oraz jego wagę dla współczesnej nauki odwołując się do niewątpliwej popularności ujęcia naturalistycznego w najnowszych rozważaniach filozoficznych jako takich. Dodaje również, że osiągnięte przez niego wyniki przyczynią się do rozstrzygnięcia takich kwestii jak to, czy należy postrzegać rozwój AI jako zagrażający ludzkości oraz czy odpowiednio zaawansowane, sztuczne umysły powinny mieć prawa moralne zbliżone do ludzkich (s. 5-6). Wydaje mi się, że są to wystarczające powody by uznać dysertację za wartościową merytorycznie, jednakże chciałabym podkreślić, że pojawiające się jakby nieco mimochodem, ostatnie zdanie części wprowadzającej, wzmacnia to poczucie i zachęca do dalszej lektury zdecydowanie bardziej, niż te przytoczone racje. Mgr

Krenc formułuje bowiem w owym zdaniu intrygującą i atrakcyjną z punktu widzenia coraz bardziej interdyscyplinarnej filozofii umysłu myśl, że „pewne wyniki jego rozprawy silnie wspierają ideę, w myśl której różnica między ludźmi a maszynami nie jest wyłącznie ilościowa” (s. 6). Myślę, że każdy czytelnik po takiej zapowiedzi czuje się zachęcony do sięgnięcia po dalszą lekturę tej pracy.

Od strony konstrukcji wywodu opracowanie jest kompletne, tj. składa się ze spisu treści, korpusu pracy oraz bibliografii. Cztery rozdziały kolejno traktują o: głównych rozwiązaniach współczesnych tzw. (jak określa to Autor) „problemu wolnej woli” (cudzysłów Autora); wolnej woli i dualizmie; wolnej woli i materializmie oraz wolnej woli i idealizmie transcendentálním. Całość podsumowuje część zatytułowana *Conclusion*. Praca w całości napisana jest w poprawnym i bardzo klarownym języku angielskim.

Mam jednak pewne zastrzeżenia związane ze stroną merytoryczną konstrukcji dysertacji mgra Krenca, które zaznaczę od razu. Po pierwsze, nie jest dla mnie jasny zamysł umieszczenia w części dotyczącej związku wolnej woli z podejściem materialistycznym do podmiotu działania podpunktów odnoszących się do biologicznego naturalizmu oraz monizmu anomalnego. O ile stanowisko zaproponowane przez Donalda Davidsona może być pojmowane jako swego rodzaju nietypowy materializm, to już koncepcja rozwijana przez Johna Searle’a z samego założenia nie ma być ani materializmem, ani jego przeciwieństwem. Co prawda Autor dysertacji zdaje sobie sprawę z kontrowersyjności swojej decyzji i stara się wykazać jej zasadność, niemniej jednak nie czuję się przekonana, gdyż proponowane przez niego uzasadnienie jest zbyt zdawkowe. Po drugie, wbrew zapowiedziom we wstępie, brakuje wyraźnych, a przede wszystkim poszerzonych i pogłębionych, analiz dotyczących pojęcia wolnej woli u tytułowych maszyn. Ogromne zdziwienie wzbudził we mnie fakt, że praca nie zawiera części definiującej maszynę oraz nie ma dyskusji z podejściami klasycznymi (skoro takie Doktoranta interesują, czemu dał wyraz wyborem takich, a nie innych teorii relacji psychofizycznej), w których ucieleśniony umysł pierwotnie wiązany był z koncepcją maszyny, jak choćby u Francisca Vareli, Humberta Maturany, czy współcześnie Evana Thompsona. *De facto* zagadnieniu umysłu maszynowego, a ściśle umysłu w postaci sztucznej sieci neuronowej, poświęcono zaledwie 5 stron. Wrócę do tego zarzutu w części krytycznej recenzji.

Wracając jednak do zawartości poszczególnych części pracy zauważmy, że pierwszy rozdział poświęcony głównym, współczesnym rozwiązaniom problemu wolnej woli rozpoczyna omówienie samego problemu wolnej woli. Dowiadujemy się z niego, że przez wolną wolę Doktorant rozumie takie działanie podmiotu, które mogłoby nie mieć miejsca. Innymi słowy, jak określał to George E. Moore, podmiot działa w sposób wolny, jeśli mógłby

zrobić coś innego (s. 8). Z kolei determinizm rozumiany jest przez mgra Krenca Peterem za Van Inwagenem jako stanowisko, w myśl którego „przeszłość i prawa natury łącznie determinują w każdej chwili unikalną przyszłość” („*a unique future*”, *ibidem*). Rozumiejąc czym jest działanie podmiotu kierującego się wolną wolą oraz czym jest determinizm, nie powinniśmy oczekiwać, zdaniem Doktoranta, że problem istnienia i natury wolnej woli zostanie rozwiązany, gdy tylko uda się ustalić, czy rzeczywiście wszechświat jest deterministyczny. Jak czytamy na s. 9: „Przedstawiony problem nie jest problemem dotyczącym prawdziwości determinizmu – jest to problem relacji między wolną wolą i determinizmem, bez względu na to, czy determinizm jest prawdą czy nie. Innymi słowy, ktoś kto jest przekonany, że wszechświat nie jest deterministyczny mógłby nadal podtrzymywać ideę, że prawdziwość determinizmu nie musi pociągać za sobą odrzucenia istnienia wolnej woli”.

W kontekście tej uwagi szczególnie ważne staje się zagadnienie możliwych relacji pomiędzy wolną wolą i determinizmem. Autor odwołuje się do dwóch, dobrze znanych filozofom, rodzajów takiej relacji, mianowicie do kompatybilizmu i inkompatabilizmu. W związku z różnorodnością rozumień tych podejść w dysertacji został zaproponowany dalszy ich podział na superkompatybilizm oraz stanowisko, zgodnie z którym wolna wola pociąga za sobą determinizm oraz twarde, czy silny (*hard*) inkompatabilizm, twarde, czy silny determinizm oraz libertanizm. Stanowiska te są klarownie omówione na s. 11-36 i nie będę ich streszczać.

Drugi rozdział dysertacji, równie kompetentny co pierwszy, prezentuje główne stanowiska oraz argumentacje i definicje wymagane do prowadzenia dalszych analiz zapowiedzianych przez Autora. Tym razem zatrzymujemy się przy zagadnieniu wolnej woli i dualizmu, powinniśmy zatem dowiedzieć się czym jest dualizm, jakie są jego rodzaje, jak argumentuje się na rzecz tego stanowiska oraz czy jest ono zgodne z ideą wolnej woli. Większość z tych wymaganych informacji udało się Doktorantowi przedstawić w sposób jasny i kompetentny. W tym miejscu pragnę zauważyć natomiast, że część poświęcona innym niż kartezjańskie rozwiązaniom dualistycznym jest bardzo skromna. Stanowi ją zaledwie jeden podrozdział o objętości 6 stron. Tymczasem w rozprawie doktorskiej bardziej liczyłam na dyskusję z tymi rozwiązaniami, niż z podejściem kartezjańskim, które stanowi przedmiot omówienia już chyba niezliczonych prac dyplomowych z wszelkich możliwych poziomów akademickich.

To, co nieco łagodzi niedosyt we wspomnianym wyżej względzie, mógł przynieść rozdział trzeci poświęcony zagadnieniu dualizmu w kontekście materializmu, a tu przede

wszystkim materializmów nieredukcyjnych Johna Searle'a oraz Donalda Davidsona oraz redukcyjnego, czy wręcz eliminacyjnego, prezentowanych przez Paula i Patricję Churchlandów. Niemniej jednak ta część rozprawy doktorskiej Mgra Krenca niestety również nieco rozczarowuje swoją „szkolnością”. Pomijając jednak to, czy biologiczny naturalizm i monizm anomalny są w ogóle stanowiskami, które należało włączać do rozważań nad materializmem, zauważalna jest przewaga objętościowa streszczenia na dość ogólnym, choć zarazem poprawnym merytorycznie, poziomie tych stanowisk, a dużo bardziej interesującą częścią dysertacji, jaką stanowi własna propozycja Doktoranta rozwiązania impasu w dyskusjach nad wolną wolą. W rozdziale na temat materialistycznego podejścia do problemu istnienia wolnej woli dowiadujemy się zasadniczo jedynie tyle, że nie daje ono podstaw do rozstrzygnięcia spornej kwestii. Poruszony jest jednak (nareszcie!, s. 84) problem implementacji procesu podejmowania decyzji opartego na działaniach wolnej woli w sztucznych sieciach neuronowych. Doktorant zauważa, że ani materializm – także w wersji koneksjonistycznej angażującej model sieci neuronowych – ani dualizm kartezjański nie przyczyniają się do zrozumienia mechanizmu działania wolnej woli w świecie fizycznym. Nieco anachroniczne wydało mi się wyciągnięcie tego wniosku bazując wyłącznie na koncepcji Ramsey'a, Sticha i Garona (s. 86). Na ten temat więcej powiem w części poświęconej dyskusji z poglądami Mgra Krenca.

Najbardziej dla mnie interesujący rozdział czwarty dysertacji poświęcono próbie ukazania idealizmu transcendentального, jako stanowiska filozoficznego, które niesie ze sobą, by tak rzec, potencjał rozstrzygnięcia tytułowego problemu, który – powtórzmy – polega na trudności w wyznaczeniu relacji pomiędzy zagadnieniem wolnej woli a rozstrzygnięciami problemu psychofizycznego. Zdaniem Autora dysertacji połączenie materializmu eliminacyjnego z idealizmem transcendentальnym Kanta daje pewną nadzieję na znalezienie miejsca na rozważania o wolnej woli we współczesnych dyskusjach filozoficznych, w których króluje naturalizm. Jest to nadzieja wątpliwa, jak mniemam, gdyż sam Autor przyznaje, że kantyzm nie musi być stanowiskiem, które w pełni go satysfakcjonuje jako rozwiązanie problemu wolnej woli. Niemniej jednak dzięki kantowskiemu rozróżnieniu na wyglądy, czy to, co się jawi oraz na rzeczy same w sobie, Mgr Krenc dostrzega możliwość takiego ujęcia podmiotu działania, który na poziomie tego, jak się jawi może zostać w pełni opisany jako deterministycznie działający obiekt materialny. Natomiast rozpatrywany jako rzecz sama w sobie byłby ujmowany jako podmiot działań moralnych, a zatem, zdaniem Doktoranta, jako osoba obdarzona wolną wolą.

Część ostatnią dysertacji stanowi rzeczowe i merytoryczne streszczenie głównych wniosków, do których Doktorant doszedł w trakcie prowadzonych analiz.

## Dyskusja

Pierwsze co przyszło mi na myśl po zakończeniu lektury pracy doktorskiej Mgra Krenca to pytanie, gdzie jest mowa o wolnej woli maszyn? Krótka wzmianka na temat sieci neuronowych nie może dawać pełnej satysfakcji, skoro pojęcie woli maszyn znalazło się w tytule dysertacji. Ponadto analizy przeprowadzone w tym miejscu nie są w pełni satysfakcjonujące również dlatego, że bazują na, co prawda, doniosłej koncepcji rozwijanej przez trójkę autorów: Sticha, Ramsey'a i Garona, niemniej jednak dość wąsko charakteryzującej rolę sieci neuronowych, czego świadomość zresztą Doktorant ma (s. 86). Mgr Krenc dostrzega bowiem przydatność tej koncepcji głównie w wyjaśnianiu pewnych konkretnych sposobów posługiwania się językiem.

W rzeczywistości istnieje interesujące powiązanie koneksjonizmu z problemem wolnej woli, a dokładniej rzecz ujmując, z możliwością opisu i wyjaśnienia zjawiska podejmowania pewnych działań określanych jako niedookreślone, czy niezdeteminowane. Chodzi o odniesienie do wspomnianych już przeze mnie Vareli i Maturany, Thompsona, a także Eleanor Rosh, tym bardziej że są to autorzy bazujący na początkowo klasycznym, a następnie zmodyfikowanym modelu koneksjonistycznym umysłu, a w tym działania. Bardzo interesująco mogłaby przebiegać dyskusja Doktoranta z tym stanowiskiem, gdyż zważywszy na zaprezentowany w pracy doktorskiej warsztat metodologiczny i umiejętności analitycznego myślenia Doktoranta, bez wątpienia potrafiłby On nadać dużo subtelniejszy i bardziej wyrafinowany charakter dyskusji na temat wolnej woli w kontekście materializmu, aniżeli pozwalają na to koncepcje zaproponowane przez wybranych przez niego autorów. Varela i Maturana, a obecnie głównie Thompson, ukazują koneksjonizm jako wersję materializmu nieredukcyjnego w postaci emergentyzmu. Jest to zabieg bardzo przydatny, gdy chce się mówić o wolnej woli. Pozwala wyjść poza taką interpretację reprezentacji decyzji podmiotu na poziomie aktywności sieci neuronowej, która byłaby li tylko stosunkowo prostym wzorem lokalnego jej pobudzenia, (na które składałyby się wyłącznie odpowiednio zidentyfikowane wagi, połączenia, warstwy itp.). W ujęciu tych autorów sieć neuronowa reprezentuje proces podejmowania decyzji jako zjawisko rozproszonej po niej aktywności, która to aktywność ostatecznie generuje emergentny stan całościowy. Próżno mówić w tym miejscu o „zwykłym” czy „klasyczne rozumianym” determinizmie. Należy też wyraźnie podkreślić, że zarówno

Varela i Maturana, jak i Thompson nie byli zwolennikami koneksjonizmu. Proponowali inne ujęcie, rodzaj enektywizmu, który dziś jest wciąż bardzo popularnym rodzajem materializmu, albo używając bezpieczniejszego terminu, choć może mniej klarownego, rodzajem naturalizmu. Szkoda, że do tego podejścia Doktorant również się nie ustosunkował. Jestem przekonana, że miałby interesujące uwagi, a jego spostrzeżenia wzbogaciłyby dysertację o dyskusje bliższe najnowszym sporom na temat architektury umysłu w kontekście problemu wolnej woli.

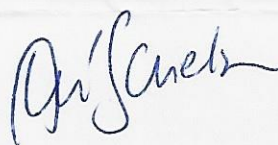
Druga moja uwaga mająca charakter polemiczny z rozprawą przedłożoną mi do recenzji wiąże się z deklaracją Mgra Krenca na temat głównego celu jego rozprawy doktorskiej. Celem tym ma być, jak przytoczyłam to w części omawiającej strukturę i zwartość merytoryczną, ukazanie powiązania postaw autorów koncepcji relacji psychofizycznej z ich zapatrywaniami na zagadnienie wolnej woli oraz, niejako przy okazji, odpowiedź na pytanie, czy rozwój sztucznej inteligencji niesie ze sobą zagrożenia dla ludzkości i czy przynajmniej niektóre wytwory określane mianem AI powinny zostać objęte normami moralnymi podobnymi do tych, które obejmują podmioty ludzkie. Z pierwszego zadania Doktorant wywiązał się częściowo, gdyż znacznie zawęził obszar poszukiwań koncepcji, a zatem i postaw filozoficznych ich twórców, co byłoby zrozumiałe, gdyby jakoś uzasadnił ten swój bardzo kostyczny wybór. Przydałoby się solidne podbudowanie argumentacyjne decyzji o zignorowaniu równie popularnych co wybrane przez Doktoranta, a także najnowszych, dyskusji wokół problemu relacji umysł-ciało oraz wokół zagadnienia wolnej woli. Co do zadania drugiego i trzeciego jest nieco podobnie. Nie można powiedzieć, że czytelnik nie znajduje odpowiedzi na postawione tu dwa pytania, niemniej jednak może być nieco rozczarowany zbyt słabym ich uwypukleniem oraz brakiem szerszego kontekstu (może „soczystych”, czyli obrazowych przykładów), który mógłby niejako podbić, wyakcentować i ponownie wyostrzyć najciekawszy – według mnie – wątek dysertacji, mianowicie tezę, w myśl której różnica między umysłem działającym wedle praw opisywanych przez koneksjonistów, a umysłem człowieka jest jakościowa.

Mimo tych uwag polemicznych, które wynikają bardziej z faktu, że praca przedłożona mi do recenzji jest w zamierzeniu ambitna oraz merytorycznie wartościowa, nie mam wątpliwości, że Doktorant jest biegle zorientowany w metodologii, terminologii, zakresie i funkcji tych analiz, a zatem i samej filozofii oraz nauk humanistycznych jako takich. Nie mam też wątpliwości, że Mgr Krenc znakomicie potrafi oceniać, analizować, porównywać, a także integrować informacje z wykorzystaniem rozmaitych źródeł oraz warsztatu naukowego właściwego dla filozofa, a nawet dla przedstawiciela interdyscyplinarnych nauk o poznaniu i komunikacji, do jakich należy kognitywistyka (z uwagi na temat rozprawy oraz wybór sposobu

jego analizy). Struktura dysertacji pokazuje także, że Doktorant jest w pełni kompetentny w określaniu priorytetów służących realizacji narzuconych sobie zadań cząstkowych w postaci pytań problemowych stanowiących punkty węzłowe rozprawy doktorskiej.

### **Konkluzja**

Pomimo wskazanych problemów oraz uwag polemicznych, które – podkreślam ponownie – mają charakter bardziej dyskusji akademickiej, aniżeli krytyki podyktowanej istotnymi brakami, rażącymi błędami, czy ubogim warszatem naukowym, pragnę podkreślić, że **przedłożona mi do recenzji rozprawa doktorska pt. *Can machines Have Free Will? Analysis of the Concept of Free Will in Relation to the Psychophysical Problem* spełnia ustawowe warunki określone dla tego rodzaju prac i w związku z tym wnoszę o procedowanie dalszych etapów postępowania zmierzającego do nadania Magistrowi Krzysztofowi Michałowi Krencowi stopnia doktora nauk humanistycznych w zakresie filozofii.**



Szczecin, 20.06.2022