

<http://dx.doi.org/10.17951/kw.2024.37.61-81>

Etyczny behawioryzm i zwrot relacyjny jako dwa przejawy antyantropocentryzmu¹

Maciej Musiał

 <https://orcid.org/0000-0002-9519-7429>

W artykule zostają omówione trzy stanowiska funkcjonujące w ramach dyskusji o statusie moralnym robotów. Pierwsze z nich, określane mianem „ortodoksyjnego”, wyraża wciąż chyba jeszcze dominujący we współczesnej kulturze zachodniej pogląd antropocentryczny, przypisujący człowiekowi wyróżniony, centralny status, ze względu na jego szczególne właściwości takie jak świadomość, emocje, wolna wola itp. Dwa stanowiska alternatywne wobec ortodoksyjnego, etyczny behawioryzm i zwrot relacyjny, proponują odmienne niż wspomniane wyżej właściwości kryteria przypisywania statusu moralnego i zostają zinterpretowane jako przejawy tendencji do negowania wyjątkowego statusu ontologicznego i aksjologicznego człowieka. W przypadku

MACIEJ MUSIAŁ, dr hab., profesor uczelni, Uniwersytet im. Adama Mickiewicza; adres do korespondencji: Wydział Filozoficzny, ul. Szamarzewskiego 89, 60-568 Poznań; e-mail: m.musial@amu.edu.pl

¹ Niniejsze rozważania stanowią kontynuację mojego namysłu nad problemem statusu moralnego robotów w kontekście szerszych tendencji kulturowych i filozoficznych, w szczególności w odniesieniu do etycznego behawioryzmu i zwrotu relacyjnego, które interpretuję z perspektywy społeczno-regulacyjnej koncepcji kultury jako przejawy desymbolizacji w postaci odczarowywania i ponownego zaczarowywania świata. Tym samym, niniejszy test stanowi kontynuację i rozwinięcie ustaleń poczynionych w następującej publikacji: Maciej Musiał, „Stanowiska w sporze o status moralny robotów jako przejaw tendencji desymbolizacyjnych”, w: *Obyczaj – kultura – nauka. Księga jubileuszowa dedykowana Profesorowi Janowi Gradowi*, red. Jacek Sójka i in. (Poznań: Wydawnictwo Nauk Społecznych i Humanistycznych UAM, 2023), 185–194.

etycznego behawioryzmu, człowiek zostaje „uzwyczajniony” do poziomu innych bytów na sposób redukcjonistyczno-naturalistyczny, podczas gdy w zwrocie relacyjnym mamy do czynienia „niezwykłym” tego, co nieludzkie poprzez relegitymizację myślenia magicznego w duchu posthumanistycznym. Tym samym, artykuł ilustruje w jaki sposób w dyskusji nad statusem moralnym robotów ujawniają się obecne we współczesnej kulturze tendencje do dwojakiej eliminacji wyjątkowości człowieka na gruncie redukcjonistycznego naturalizmu oraz magicznego posthumanizmu.

Słowa kluczowe: roboty, status moralny, etyczny behawioryzm, zwrot relacyjny, antropocentryzm

Problem statusu moralnego robotów – jak również wszelkich innych bytów – w największym skrócie dotyczy tego, czy roboty zasługują na uznanie ich za byty posiadające autoteliczną wartość moralną, którą należy mieć na uwadze mając z nimi do czynienia. Idzie tu zatem m.in. o to, czy posiadamy wobec robota jakieś powinności ze względu na niego samego (z perspektywy etyki deontologicznej), lub, czy możemy robotowi uczynić coś, co będzie dobre/przyjemne lub złe/nieprzyjemne – dla niego samego (z perspektywy etyki konsekwencjalistycznej)². Spośród wielu meandrów związanych z przypisywaniem statusu moralnego, tym, który jest najżywiej dyskutowany w odniesieniu do robotów, i który stanowić będzie główny przedmiot zainteresowania w niniejszym tekście, jest zagadnienie kryteriów przyznawania statusu moralnego. Innymi słowy, idzie o to, jakie warunki są konieczne i/lub wystarczające, by danemu robotowi przyznać status moralny.

Niniejszy tekst omawia trzy stanowiska próbujące rozwiązać ten problem: jedno stanowisko określane mianem ortodoksyjnego oraz dwa stanowiska alternatywne: etyczny behawioryzm i zwrot relacyjny. Stanowisko ortodoksyjne zostanie omówione skrótowo, głównie pod kątem kierowanych do niego zarzutów, stanowiących punkt wyjścia ujęć alternatywnych, które zostaną omówione bardziej szczegółowo. Warto zastrzec, że nie próbujemy tu rozstrzygać, które z tych stanowisk jest słuszne. Zamiast tego, koncentrujemy się na uwypukleniu swoistości każ-

² Więcej na temat pojęcia statusu moralnego zob. m.in. Agnieszka Jaworska i Julie Tannenbaum, „The grounds of moral status”, w: *Stanford Encyclopedia of Philosophy*, red. Edward N. Zalta, <https://plato.stanford.edu/entries/grounds-moral-status/> (dostęp: 10.12.2023).

dego z nich, a także na zinterpretowaniu obu stanowisk alternatywnych jako specyficznych form antyantropocentryzmu oraz wpisaniu ich w szersze tendencje filozoficzne i kulturowe.

Ortodoksyjne stanowisko wobec problemu statusu moralnego robotów i jego krytyka

Najbardziej standardowa odpowiedź na pytanie o kryteria przyznawania statusu moralnego robotom, która jest często określana mianem „ortodoksyjnej”, głosi, że kryterium takie stanowi posiadanie przez danego robota odpowiednich właściwości, takich jak świadomość, samoświadomość, wolna wola, emocje itp. Zasadniczo twierdzi się, że aktualnie istniejące roboty i inne artefakty techniczne nie posiadają rzeczonych właściwości, a zatem nie przysługuje im status moralny. W sposób szczególnie wyraźny takie stanowisko zajmują np. Joanna J. Bryson³, Sven Nyholm i Lily Frank⁴ oraz Vincent C. Müller⁵, jednakże jest ono dość często przyjmowane *implicite* jako zbieżne z punktem widzenia tzw. zdrowego rozsądku (przynajmniej tego typowego dla współczesnej kultury zachodniej). Zgodnie z wcześniejszą deklaracją, nie będziemy szczegółowo omawiać meandrów stanowiska ortodoksyjnego, skupimy się natomiast na jego krytyce, którą podzielić można na krytykę o charakterze epistemologicznym/poznawczym, oraz aksjologicznym/etycznym⁶.

³ Joanna J. Bryson, „Robots should be slaves”, w: *Close engagements with artificial companions: Key social, psychological, ethical and design issues*, red. Yorick Wilks (Amsterdam: John Benjamins, 2010), 63–74; teźże, „Patience is not a virtue: The design of intelligent systems and systems of ethics”. *Ethics and Information Technology* 2018, nr 20: 15–26, <https://doi.org/10.1007/s10676-018-9448-6>

⁴ Sven Nyholm, Lily Frank, „From sex robots to love robots: Is mutual love with a robot possible?”, w: *Robot sex: Social and ethical implications*, red. John Danaher i Neil McArthur (Cambridge, MA: MIT Press, 2017), 219–244.

⁵ Vincent C. Müller, „Is it time for robot rights? Moral status in artificial entities”, *Ethics and Information Technology* 23, nr 4(2021): 579–587, <https://doi.org/10.1007/s10676-021-09596-w>

⁶ Ograniczam się tutaj do przytoczenia krytyk stanowiska ortodoksyjnego sformułowanych na gruncie omówionych niżej ujęć alternatywnych.

Jeśli idzie o krytykę epistemologiczną/poznawczą, to na pierwszy plan wysuwają się trzy zarzuty. Pierwszy dotyczy tego, że nie ma – i nie zanoszą się, by w najbliższym czasie ta sytuacja miała się zmienić – zgody co do tego, jak właściwie zdefiniować wzmiankowane wyżej właściwości mające stanowić kryteria przyznawania statusu moralnego, czyli świadomość, emocje itd. Drugi zarzut zwraca uwagę, że nawet gdybyśmy potrafili zdefiniować te właściwości, to nie ma zgody co do tego, które z nich winny być brane pod uwagę podczas ustalania statusu moralnego, czyli, które z nich powinien posiadać robot, by uzyskać status moralny. Wreszcie, trzeci zarzut stanowi, że nawet gdybyśmy zdefiniowali rzeczony właściwości i zgodzili się, które z nich stanowią konieczne i/lub wystarczające kryteria przyznawania statusu moralnego robotom, to nie mamy jasności co do tego, jak stwierdzić obecność tych właściwości w robotach (dodajmy, że kwestia ustalenia obecności tych właściwości w ludziach również jest – mówiąc najogólniej – daleka od oczywistości). Innymi słowy, nie wiemy i nie zanoszą się, byśmy mieli w najbliższym czasie dowiedzieć się, jak uzyskać bezpośredni dostęp poznawczy do świadomości, emocji itp., w związku z czym stwierdzanie ich obecności jest wysoce problematyczne. Widać to choćby w przypadku badań nad zwierzętami, jednakże sprawa dodatkowo komplikuje się w przypadku robotów, w oczywisty sposób posiadających odmienny ustrój ontologiczny niż organizmy żywe. Idzie tu w szczególności o fakt, że – ujmując sprawę w dużym skrócie i uproszczeniu – uważamy, że świadomość, emocje itp. w przypadku ludzi są możliwe dzięki posiadaniu określonego rodzaju organizmu. Nie wiemy jednak, co dokładnie w naszych organizmach sprawia, że posiadamy wspomniane wyżej właściwości. Tym bardziej trudno nam stwierdzić, czy inne byty, w szczególności zwierzęta uposażone w inne organizmy, tudzież – w przypadku robotów – odmienny *hardware*, również posiadają rzeczony właściwości. Zatem krytyka o charakterze epistemologicznym/poznawczym wskazuje na fakt, że w przypadku kryteriów postulowanych przez ujęcie ortodoksyjne mamy do czynienia z tyłoma znakami zapytania, że w praktyce nie da się tych kryteriów stosować.

Krytyka aksjologiczna/etyczna wskazuje natomiast na fakt, że kryteria postulowane przez ujęcie ortodoksyjne w nieuzasadniony sposób uprzywilejowują człowieka i (na co kładzie się z reguły mniejszy nacisk) organizmy żywe w ogóle. W szczególności zwraca się uwagę na to, że właściwości postulowane przez stanowisko ortodoksyjne jako kryteria przyznawania statusu moralnego – czyli świado-

mość, wolna wola, emocje itp. – są cechami, których posiadanie w pierwszej kolejności przypisujemy ludziom, a zatem postulowanie tych właściwości jako kryteriów uprzywilejowuje ludzi, nadając im status bytów wyjątkowych, „wzorcowych”, a tym samym wymaga od jakichkolwiek bytów mających posiadać status moralny określonego stopnia podobieństwa do ludzi właśnie – krytycy postrzegają takie podejście jako wynikające z arbitralnego przekonania o wyjątkowej wartości człowieka, które określić można mianem antropocentryzmu aksjologicznego.

Podsumowując, krytycy ortodoksyjnego stanowiska w kwestii kryteriów przyznawania statusu moralnego zwracają uwagę, że kryteria te – czyli posiadanie właściwości takich jak świadomość, wolna wola, emocje itp. – wiążą się z szeregiem problemów poznawczych, uniemożliwiających ich praktyczną aplikację, a zarazem opierają się na arbitralnych założeniach w sposób nieuprawniony uprzywilejowujących człowieka (i ewentualnie inne istoty żywe), a tym samym czyniących z człowieka byt wyjątkowy i „wzorcowy”. W związku z tym krytycy stanowiska ortodoksyjnego proponują alternatywne stanowiska mające na celu sformułowanie odmiennych kryteriów przyznawania statusu moralnego, pozwalających uniknąć wzmiankowanych wyżej problemów. Spośród nich na pierwszy plan wysuwa się etyczny behawioryzm oraz zwrot relacyjny.

Etyczny behawioryzm jako aksjologiczny antropocentryzm redukcjonistyczno-naturalistyczny

Etyczny behawioryzm to stanowisko zaproponowane i najpełniej wyartykułowane przez Johna Danahera w serii artykułów⁷, choć – jak przyznaje sam Danaher – rozmaite elementy jego koncepcji są obecne we wcześniejszych stanowiskach innych myślicieli i myślicielek, takich np. jak David Levy czy Erica Neely.

⁷ John Danaher, „Welcoming robots into the moral circle: A defence of ethical behaviourism”, *Science and Engineering Ethics* 26, nr 4(2019): 2023–2049, <https://doi.org/10.1007/s11948-019-00119-x>; tenże, „Robot Betrayal: a guide to the ethics of robotic deception”, *Ethics and Information Technology* 22, nr 2(2020): 117–128, <https://doi.org/10.1007/s10676-019-09520-3>; tenże, „What Matters for Moral Status: Behavioral or Cognitive Equivalence?”, *Cambridge Quarterly of Healthcare Ethics* 30, nr 3(2021): 472–478, <https://doi.org/10.1017/S0963180120001024>

W niniejszych rozważaniach skupiam się wszelako na ujęciu autorstwa Danahera. Chociaż pod wpływem krytyki stanowisko Danahera uległo pewnej ewolucji, jego rdzeń pozostał niezmienny⁸. Stanowi go przekonanie, że wystarczającym warunkiem przyznania robotowi (lub jakimkolwiek innemu bytowi) statusu moralnego jest przejawianie zachowań analogicznych do bytu, który posiada już status moralny. W sposób najbardziej wyrazisty Danaher artykułuje to przekonanie w ramach – nigdy przez siebie nieodwołanej i niezmodyfikowanej – zasadzie porównawczej etycznego behawioryzmu, która stanowi: „Jeśli byt X wykazuje lub prezentuje w przybliżeniu równoważne wzorce zachowań (P1...Pn) jak jednostka Y, i jeśli jesteśmy przekonani, że te wzorce uzasadniają nasze przypisywanie statusu moralnego jednostce Y, to albo (a) ten sam status moralny musi być przypisany X, albo (b) używanie P1...Pn do uzasadniania statusu moralnego Y musi być ponownie ocenione”⁹.

Zauważmy, że Danaher proponuje zachowania i ich podobieństwo do innych zachowań jako jedyne kryterium przyznawania statusu moralnego. Nie interesują go „kryjące” się (lub nie) za zachowaniami – jak to określa – „metafizyczne” właściwości, które stanowią kryterium w ramach ujęcia ortodoksyjnego. Zdaniem Danahera, są one niedostępne dla bezpośredniego poznania, a zatem nie warto brać ich pod uwagę¹⁰. Ponadto Danaher przekonuje, że nie ma znaczenia ontologia danego bytu, tudzież *hardware*, a zatem to, z czego i w jaki sposób dany byt jest zbudowany – liczą się bowiem jedynie zachowania, natomiast ich ontologiczne „podłoże” nie powinno odgrywać żadnej roli. Dla przykładu, Danaher uważa, że jesteśmy w coraz większym stopniu „cyborgami”, w tym sensie, że posiadamy rozruszniki serca i inne niebiologiczne komponenty, i przekonuje, że za-

⁸ Zob. Danaher, „What Matters for Moral Status”, gdzie Danaher odpowiada na krytykę sformułowaną w następujących tekstach: Jilles Smids, „Danaher’s Ethical Behaviourism: An Adequate Guide to Assessing the Moral Status of a Robot?”, *Science and Engineering Ethics* 26, nr 5(2020): 2849–2866, <https://doi.org/10.1007/s11948-020-00230-4>; Sven Nyholm, *Humans and robots: Ethics, agency, and anthropomorphism* (London, New York: Rowman & Littlefield, 2020), 116–119; Henry Shevlin, „How Could We Know When a Robot was a Moral Patient?”, *Cambridge Quarterly of Healthcare Ethics* 30, nr 3(2021): 459–471, <https://doi.org/10.1017/S0963180120001012>

⁹ Danaher, „Welcoming”: 2030, tłum. własne.

¹⁰ Tamże: 2025, 2029.

stępowanie biologicznych organów nieorganicznymi odpowiednikami nie powinno prowadzić do odebrania ludziom (lub zwierzętom) statusu moralnego, a zatem ontologia nie powinna mieć znaczenia dla przypisywania statusu moralnego¹¹. Wreszcie, Danaher sądzi, że żadnego znaczenia nie ma fakt, iż roboty są zaprogramowane przez ludzi, a nawet gdyby miał on znaczenie, to i tak nie odróżnia on robotów od ludzi, ponieważ ludzie również są zaprogramowani przez ewolucję (w wariancie ateistycznym) lub przez Boga (w wariancie teistycznym). Również w tym przypadku Danaher twierdzi, że gdyby dzięki osiągnięciom genetyki możliwe było – w jakiejś mierze – prenatalne intencjonalne programowanie istoty ludzkiej, to nie powinno to negatywnie wpływać na należny tej istocie status moralny, a zatem fakt programowania nie ma znaczenia w przypisywaniu statusu moralnego¹². Podkreślmy zatem raz jeszcze, dla Danahera liczą się tylko zachowania, a nie to, co się za nimi „kryje” (metafizyczne właściwości), ich podłoże (ontologia danego bytu) oraz geneza (proces powstania danego bytu na płaszczyźnie biologicznej i społeczno-kulturowej, zarówno w sensie filogenetycznym jak ontogenetycznym).

Warto zaznaczyć, że Danaher rozumie zachowanie bardzo szeroko, szerzej niż znaczna część tradycyjnego behawioryzmu, zaliczając do nich np. obserwowalne empirycznie aktywności mózgu. Dodajmy również, że Danaher – przynajmniej deklaratywnie – nie twierdzi, iż świadomość, emocje, wolna wola i inne metafizyczne właściwości nie istnieją, a jedynie, że nie mamy do nich bezpośredniego dostępu poznawczego w związku z czym nie mogą one stanowić kryterium orzekania o statusie moralnym. Jego behawioryzm nie ma zatem charakteru ontologicznego, lecz metodologiczny, nie przesądza bowiem o tym, co istnieje, a jedynie o tym, co należy wziąć pod uwagę¹³.

Zdaniem Danahera, uznanie zachowania za jedyne kryterium orzekania o statusie moralnym pozwala uniknąć problemów epistemologicznych, które dotyczą kryteriów postulowanych przez ujęcie ortodoksyjne. W szczególności idzie tu o fakt, że Danaher nie musi rozstrzygać, czym są poszczególne właściwości, które nazywa „metafizycznymi”, takie jak świadomość, wolna wola itd., które z nich są konieczne dla uzyskania statusu moralnego oraz jak uzyskać do nich dostęp poznawczy, ponieważ to nie one stanowią dla niego kryterium orzekania

¹¹ Tamże: 2031–3032.

¹² Tamże: 2032–2035.

¹³ Tamże: 2027.

o statusie moralnym, lecz zachowania, które można empirycznie zaobserwować. Zatem Danaher proponuje porównywać zachowania z innymi zachowaniami i powstrzymać się od pytania, co się za tymi zachowaniami „kryje”, a więc jak te zachowania można interpretować lub wyjaśniać. W ramach swojego stanowiska Danaher zdaje się dążyć do „uzewnętrznienia” i zobiektywizowania kryteriów, co ujawnia się szczególnie w jego stosunku do metafizycznych właściwości, które – za Svenem Nyholmem i Lily Frank¹⁴ – określa mianem „tego, co dzieje się wewnątrz”, a co w jego przekonaniu pozostaje „głęboko ukryte”, niedostępne dla bezpośredniego, obiektywnego poznania i jest legitymizowane jedynie intersubiektywnym konsensusem, w przeciwieństwie do dostępnego dla bezpośredniego i obiektywnego poznania „zewnątrznego” zachowania. Tym samym Danaher sądzi, że proponowane przez niego oparcie się na zachowaniu jest bardziej pewnym i wiarygodnym kryterium orzekania o statusie moralnym, albowiem pozwala kierować się obiektywnymi, obserwowalnymi zachowaniami, a nie spekulacjami na temat metafizycznych właściwości¹⁵.

¹⁴ Nyholm, Frank, „From sex robots to love robots”, wszędzie.

¹⁵ Koncepcja Danahera budzi liczne wątpliwości. W szczególności wydaje się, że Danaherowi nie udaje się – wbrew jego deklaracjom – ustalić pewnych i wiarygodnych kryteriów, które nie budziłyby żadnych poznawczych wątpliwości. Zauważmy bowiem, że Danaher nie proponuje, by kryterium były zachowania robotów identyczne z zachowaniami ludzi lub innych bytów posiadających już status moralny, lecz zachowania „w przybliżeniu równoważne” (w oryginale: „*roughly equivalent*”). Oznacza to, że roboty nie muszą przejawiać wszystkich zachowań charakterystycznych dla ludzi, a jedynie niektóre, a ponadto, że nawet w przypadku tych, które będą porównywane z ludzkimi, nie jest konieczna ich identyczność. Nieuchronnie prowadzi to do pytania: które zachowania i jakie ich aspekty będą brane pod uwagę, a także: w jaki sposób rozstrzygnąć te kwestie? Jak wskazują krytycy, jedynym sposobem odpowiedzi na te pytania jest porównywanie nie tylko samych zachowań, ale również tego, co się za nimi „kryje”, czyli branie po uwagę ich interpretacji i wyjaśnień. Korzystając z metafory Danahera, który twierdzi, że metafizyczne właściwości nie mówią własnym głosem, lecz poprzez zachowania, należy stwierdzić, że również zachowania nie mówią własnym głosem, lecz poprzez swoje wyjaśnienia i interpretacje – tylko one pozwalają nam stwierdzić, które z zachowań, i które z aspektów tych zachowań są istotne dla ustalania statusu moralnego, a które nie. Zatem, dla przykładu, gdy Danaher pisze, że jego zdaniem kluczowymi zachowaniami mającymi znaczenie dla nadawania statusu moralnego są zachowania poznawcze/kognitywne, to na gruncie samego zachowania nie jest on w stanie ani wyselekcjonować tych, które są poznawcze/kognitywne, ani też uzasadnić dlaczego to właśnie one – a nie inne zachowania – miałyby odgrywać kluczową rolę. Danaher przyznaje się

Danaher uważa również, że jego stanowisko unika aksjologicznego/etycznego problemu kryteriów proponowanych przez ujęcie ortodoksyjne w postaci arbitralnego uprzywilejowania człowieka jako swoistego wzorca, czyli aksjologicznego antropocentryzmu; zatem jego stanowisko można określić mianem aksjologicznego antyantropocentryzmu. Zdaniem irlandzkiego filozofa, ograniczenie się do zachowań jako wystarczającego kryterium czyni przedsięwzięcie mniej antropocentrycznym i bardziej inkluzywnym, albowiem: 1) jakkolwiek rolę przestaje odgrywać konstytucja ontologiczna, czyli w szczególności to, z czego i w jaki sposób zbudowany jest dany byt (co oznacza, że roboty nie muszą być pod tym względem podobne do człowieka); 2) przestaje mieć znaczenie fakt, że roboty są zaprogramowane przez ludzi (co również w swoisty sposób obniża pozycję ludzi w stosunku do robotów); 3) tracą znaczenie metafizyczne właściwości mające zaświadczać o wyjątkowości człowieka (świadomość, wolna wola, emocje itp.) – co prawda, jak zauważyliśmy wcześniej, Danaher nie twierdzi, że one nie istnieją, lecz ze względu na brak bezpośredniego dostępu poznawczego w gruncie rzeczy odbiera im jakkolwiek rolę w wyznaczaniu statusu moralnego robotów, ludzi i jakichkolwiek innych bytów. Oznacza to, że cechy traktowane – w ramach podejść antropocentrycznych – jako zaświadczone o wyjątkowości i wyższości ludzi i – w ramach podejścia ortodoksyjnego – wymagane w celu uzyskania statusu moralnego, przestają mieć jakiegokolwiek znaczenie, w związku z czym wyjątkowy status człowieka i jego pozycja „wzorca” ulegają daleko posuniętej redukcji: z punktu widzenia Danahera, jedyną istotną różnicą pomiędzy ludźmi i robotami są ich zachowania, a jeśli zachowania tych pierwszych w wystarczającym stopniu upodobnią się do tych drugich, to nie sposób będzie mówić o dalszym występowaniu istotnych aksjologicznie różnic.

Przejdźmy do wyłuszczenia, w jakim sensie aksjologiczny antyantropocentryzm Danahera ma charakter redukcyjno-naturalistyczny. Uzasadniając tę tezę, nie chcemy wikłać się w roztrząsanie meandrów rozmaitych ujęć naturalizmu i redukcjonizmu. Zamiast tego, odwołując się do dość powszechnych przekonań i in-

do tego w odpowiedzi na krytykę Jillesa Smidsa, Svena Nyholma i Henry'ego Shevlina, nie zmienia jednak brzmienia swojej zasady porównawczej. Szczegółową krytykę stanowiska Danahera w powyższym duchu przeprowadzam w: Maciej Musiał, „Criticizing Danaher's Approach to Superficial State Deception”, *Science and Engineering Ethics* 29, nr 5(2023): 1–15, <https://doi.org/10.1007/s11948-023-00452-2>

tuicji dotyczących sedna obu tych pojęć, pragniemy zwrócić uwagę na fakt, że Danaher pozbawia człowieka wyjątkowego statusu poprzez zredukowanie jego metafizycznych właściwości do zbiorów zachowań. Powiadając, że Danaher dokonuje wspomnianej wyżej redukcji, mamy na myśli w szczególności redukcję metodologiczną, w tym sensie, że postuluje on, by badanie metafizycznych właściwości – do których w jego przekonaniu nie mamy poznawczego dostępu – zastąpić badaniem zachowań. Zauważmy, że ta redukcja ma charakter naturalistyczny w tym sensie, że proponuje zastąpienie antynaturalistycznych metod badawczych (takich np. jak interpretacja, którą można by stosować do zachowań) metodami nauk przyrodniczych, w szczególności np. zwróceniem uwagi na ilościowe parametry porównywanych zachowań. Dodajmy, że określanie poglądów Danahera mianem naturalistycznych i redukcjonistycznych trudno uznać za efekt imputacji, szczególnie w obliczu tego, w jaki sposób on sam opisuje swoje zapatrywania – fakt, że zdecydował się on na nazwanie swojego stanowiska mianem „behawioryzmu” dość wyraźnie sygnalizuje, iż pozycje naturalistyczne są mu bliskie, natomiast jego deklaracje, że jego zamierzeniem jest „zredukować etykę do epistemologii”¹⁶ sygnalizują, iż nie jest nadużyciem traktowanie jego stanowiska jako odchodzącego od antynaturalistycznego odnoszenia faktów do wartości na rzecz zredukowania metod badawczych do skupiających się na porównywaniu nieinterpretowanych faktów (w postaci zachowań).

Warto zaznaczyć, że redukcjonistyczno-naturalistyczny charakter koncepcji Danahera sam w sobie nie jest szczególnie zaskakujący. Tym, co wydaje się natomiast specyficzne dla etycznego behawioryzmu jest fakt, że jego redukcjonistyczno-naturalistyczna proveniencja idzie w parze z daleko posuniętym antyantropocentryzmem aksjologicznym. Zwracając uwagę jedynie na zachowania, etyczny behawioryzm deprecjonuje wyjątkowość człowieka, eliminując jego swoiste cechy, poczynawszy od „metafizycznych” właściwości, które „kryją się” za zachowaniami, a których Danaher nie chce brać pod uwagę, poprzez specyficzny biologiczny *hardware* człowieka, a skończywszy na odrzuceniu zaświadczających na korzyść człowieka interpretacji faktu, że to właśnie człowiek tworzy i programuje roboty. Tym samym, odrzucając wszystko poza zachowaniami, etyczny behawioryzm staje się stanowiskiem eliminującym różnice jakościowe pomiędzy

¹⁶ Danaher, „Welcoming”: 2028.

człowiekiem a robotem, czego konsekwencją jest m.in. stanowisko Danahera, głoszące, że przyjaźń człowieka z robotem jest równie wartościowa jak przyjaźń człowieka z człowiekiem¹⁷.

Zwrot relacyjny jako aksjologiczny antyantropocentryzm magiczno-posthumanistyczny

David Gunkel oraz Mark Coeckelbergh, najpierw niezależnie od siebie¹⁸, a później przez pewien czas wspólnie¹⁹ wypracowali stanowisko określane mianem zwrotu relacyjnego²⁰; w ostatnim czasie jego interesujące rozwinięcie zaprezentowali m.in. Nancy S. Jecker, Caesar A. Atiure i Martin Odei Ajei²¹. Istotą tego poglądu – wspólną dla jego rozmaitych wariantów – jest przekonanie, że o statusie moralnym danego bytu decyduje sposób wchodzenia z nim w relacje, a w szczególności to, w jaki sposób jest on doświadczany w ramach tych relacji. Nie idzie zatem o to, by ustalić, czym jest dany byt i dopasować do niego swoje z nim relacje, ale raczej o to, by to dopiero relacja i pojawiające się w jej ramach doświadczenie ustaliły, czym dany byt jest. Gunkel twierdzi, że właściwością są „aposteriorycznym produktem [...] relacji społecznych”²²; Coeckelbergh sądzi, że „relacje są pierwotne wobec obiektów, które ta relacja łączy”²³; natomiast Jecker, Atiure i Ajei powiadają, że „to, co sprawia, że osoby są osobami to ich wzajemne relacje z innymi i ze

¹⁷ Tenże, „The philosophical case for robot friendship”, *Journal of Posthuman Studies* 3, nr 1(2019): 5–24.

¹⁸ Mark Coeckelbergh, *Growing moral relations: Critique of moral status ascription* (New York: Palgrave MacMillan, 2012); David Gunkel, *The machine question* (Cambridge, MA: MIT Press, 2012).

¹⁹ Mark Coeckelbergh, David Gunkel, „Facing animals: A relational, other-oriented approach to moral standing”, *Journal of Agricultural and Environmental Ethics* 27, nr 5(2014): 715–733, <https://doi.org/10.1007/s10806-013-9486-3>

²⁰ Anne Gerdes, „The Issue of Moral Consideration in Robot Ethics”, *ACM SIGCAS Computers & Society* 45, nr 3(2015): 274–280, <https://dl.acm.org/doi/10.1145/2874239.2874278>

²¹ Nancy S. Jecker, Caesar A. Atiure i Martin Odei Ajei, „The Moral Standing of Social Robots: Untapped Insights from Africa”, *Philosophy & Technology* 35 (2022): 34, <https://doi.org/10.1007/s13347-022-00531-5>

²² David Gunkel, *Robot rights* (Cambridge, MA: MIT Press, 2018), 200.

²³ Coeckelbergh, *Growing moral relations*, 45.

społecznością”²⁴. Tym samym, to właśnie rzeczzone relacje i pojawiające się w ich ramach doświadczenia stanowią kryterium orzekania o statusie moralnym. Można zatem powiedzieć, że o ile Danaher stara się zobjektywizować i „uwnętrznić” kryterium proponowane przez stanowisko ortodoksyjne, to relacjonisci usiłują te kryteria – nomen omen – „urelacyjnić” i zsubiektywizować.

Z punktu widzenia zwolenników zwrotu relacyjnego proponowane przez nich stanowisko unika problemów epistemologicznych charakterystycznych dla stanowiska ortodoksyjnego. W szczególności problem zdefiniowania, skatalogowania i zweryfikowania obecności właściwości – takich jak świadomość, emocje itp. – przestaje stanowić istotne wyzwanie, ponieważ o obecności tych i innych właściwości decydują dokonujące się w ramach relacji subiektywne doświadczenia. Z punktu widzenia relacjonistów, nie ma zatem – podobnie jak w przypadku zachowań u Danahera – sensu pytać, co „tak naprawdę” „kryje się” za tymi doświadczeniami, ponieważ właściwości obiektów, o których jest sens mówić konstytuują się dopiero w ramach relacji, a zatem subiektywne doświadczenia można porównywać jedynie z innymi subiektywnymi doświadczeniami. Tym samym, zwolennicy zwrotu relacyjnego legitymizują właściwie wszelkie doświadczenia wynikające z relacji z robotami²⁵.

Zwolennicy zwrotu relacyjnego sądzą również, że ich stanowisko pozwala uniknąć antropocentryzmu aksjologicznego, a zatem stanowi przejaw antyantropocentryzmu aksjologicznego, ponieważ w jego ramach obiekt, który miałby otrzymać status moralny, w ogóle nie musi być podobny do człowieka; zatem człowiek nie jest wzorcowym bytem, do którego inne muszą być podobne. Warto zauważyć, że zmiana statusu człowieka dokonuje się tu nie tylko poprzez rezygnację z jego „wzorcowej” roli, ale również poprzez zrównanie człowieka z innymi bytami. Odbyna się ono jednak inaczej niż w przypadku etycznego behawioryzmu,

²⁴ Jecker, Atiure, Ajei, „The moral standing”: 34.

²⁵ Poznawcza płaszczyzna zwrotu relacyjnego budzi szereg wątpliwości, spośród których na pierwszy plan wysuwa się relatywizm czy wręcz idiografizm i skrajny subiektywizm. Doświadczenia robotów – i jakichkolwiek innych bytów – różnią się bowiem nie tylko w zależności od kultury, ale także w ramach poszczególnych kultur i wreszcie w przypadku konkretnych osób, które mogą zwyczajnie zmieniać swój sposób doświadczania danego robota. Jeśli te wszystkie doświadczenia są równie autentyczne i nie można ich porównywać z niczym „głębszym”, a jedynie z innymi doświadczeniami, to pojawia się problem uzgodnienia jakiegokolwiek intersubiektywnej i względnie stałej odpowiedzi na pytanie o status moralny czegokolwiek.

który – przypomnijmy – eliminował znaczenie wyjątkowych właściwości człowieka poprzez zastąpienie ich zachowaniami. W przypadku zwrotu relacyjnego nie tyle odbiera się tym ludzkim właściwościom ich znaczenie, co raczej odbiera się im czysto ludzki charakter. Innymi słowy, relacjoniści sądzą, że jeśli w ramach naszej relacji z robotem nasze subiektywne doświadczenie sugeruje nam, że ten robot posiada świadomość, odczuwa emocje itd., to należy uznać, że tak właśnie jest. O ile zatem w przypadku etycznego behawioryzmu mieliśmy do czynienia ze zrównywaniem człowieka i robota poprzez odebranie znaczenia wyjątkowym cechom człowieka, o tyle w przypadku zwrotu relacyjnego owo zrównywanie dokonuje się poprzez nadanie tych właściwości robotom.

Aksjologiczny antyantropocentryzm w przypadku zwrotu relacyjnego można zinterpretować jako posiadający charakter magiczno-posthumanistyczny w tym sensie, że: 1) odbiera człowiekowi wyjątkowy status poprzez legitymizację myślenia magicznego pozwalającego na zrównanie bytów nieludzkich z człowiekiem; 2) to z kolei upodabnia go do myśli posthumanistycznej, która również posiada wyżej wymienioną cechę. Pochylmy się najpierw nad pierwszym z tych zagadnień.

Myślenie magiczne, zwane też m.in. myśleniem magiczno-mitycznym, mitem lub magią, oznacza sposób myślenia najbardziej charakterystyczny dla ludów tubylczych, badanych w szczególności przez antropologię kulturową²⁶, małych dzieci, badanych w szczególności przez psychologię rozwojową²⁷ oraz osób z niektórymi schorzeniami psychicznymi, które bada w szczególności psychiatria²⁸ – wszelako występowanie myślenia magicznego nie ogranicza się jedynie do tych trzech grup, lecz jedynie występuje w ich ramach – jak wspomnieliśmy – w sposób

²⁶ Zob. m.in.: James George Frazer, *Złota gałąź*, tłum. Jan Lutyński (Warszawa: PIW, 1962 [1890]); Susan Greenwood, *The Anthropology of Magic* (Oxford, New York: Bloomsbury Academic, 2009).

²⁷ Zob. m.in.: Jean Piaget, *Jak sobie dziecko wyobraża świat*, tłum. Maria Gawlik (Warszawa: PWN, 2006 [1926]); Eugene Subbotsky, *Magic and the Mind: Mechanisms, Functions, and Development of Magical Thinking and Behavior* (Oxford, New York: Oxford University Press, 2010).

²⁸ Zob. m.in.: Geza Roheim, *Magic and Schizophrenia* (Bloomington: Indiana University Press, 1955); Artur Dobosz, *Myślenie magiczno-mityczne a schizofrenia* (Bydgoszcz: Oficyna Wydawnicza Epigram, 2013).

najbardziej wyrazisty. Nie sposób tu w pełni scharakteryzować myślenia magicznego oraz związanych z nim kontrowersji, ograniczymy się zatem do wskazania na te jego cechy, które są istotne dla rozważanego tu zagadnienia.

Po pierwsze, zakładając, że subiektywne, relacyjne doświadczenia konstytuują rzeczywistość, za którą nie „kryją się” żadne stałe esencje, relacjoniści w istocie proponują „płaską” ontologię, w ramach której nie ma zbyt wiele miejsca na hierarchie typu *rzeczywistość – pozór*, którą zresztą relacjoniści otwarcie kontestują. Jest to również cecha myślenia magicznego, które w swej najbardziej wyrazistej postaci implikuje ontologię jednowymiarową, pozbawioną pionowej hierarchii, w ramach której wszystko jest równie realne (np. sen i jawa mają taką samą moc sprawczą i taki sam status ontologiczny)²⁹.

Po drugie, uznając, że to subiektywne, relacyjne doświadczenie wyznacza cechy danego bytu i akceptując, że doświadczenie to może zmieniać się w czasie, relacjoniści w gruncie rzeczy akceptują możliwość – charakterystycznego dla magii – metamorficznego przekształcania się rzeczywistości wraz z przekształcaniem się jej doświadczenia, jako że za tym doświadczeniem „nie kryje się” żadna „głębocka” stała esencja, a zatem ulegając zmianie „na powierzchni” obiekty w istocie ulegają zmianie całościowej (bo pod „powierzchnią” nic nie ma), a zatem zmianie metamorficznej. Tym samym, można dodać, że ontologia będąca konsekwencją propozycji relacjonistów, jest nie tylko „płaska” i jednowymiarowa, ale również dynamiczna i procesualna, co również stanowi cechę myślenia magicznego³⁰.

Po trzecie, relacjoniści – a spośród nich najbardziej wyraźnie Mark Coeckelbergh – uznają, że w subiektywnych, relacyjnych doświadczeniach obecny jest synkretyczny splot emocji, działania i intelektu. Nie sądzą zatem, że intelekt należy oddzielić od działania i emocji oraz że to ta pierwsza władza poznawcza dostarcza nam bardziej wiarygodnych informacji – relacjoniści postulują, że w naszym subiektywnym doświadczeniu robotów istotne jest nie tylko to, co myślimy o tych

²⁹ Zob. *Czy metamorfoza magiczna rekompensuje brak symbolu*, red. Jerzy Kmita (Poznań: UAM, 2001).

³⁰ Ernst Cassirer, *The philosophy of symbolic form. Volume 2, Mythical thought*, tłum. Charles Hendel (New Haven: Yale University Press 1960 [1925]); Emile Durkheim i Marcel Mauss, *Primitive Classification*, tłum. Rodney Needham (Chicago: University of Chicago Press, 1963 [1902]).

robotach, ale też jakie emocje względem nich odczuwamy i jakie działania podejmujemy wobec nich. Taki synkretyczny splot emocji, działania i intelektu również jest cechą myślenia magicznego³¹.

Po czwarte wreszcie, wśród rozmaitych subiektywnych doświadczeń wchodzenia w subiektywne relacje z robotami możemy zaobserwować również takie, które polegają na przypisaniu rozmaitym robotom – począwszy od odkurzaczy Roomba³², poprzez robopieski AIBO³³, a skończywszy na militarnych Packbotach³⁴ – cech istot żywych, takich jak posiadanie osobowości i uczuć, a także silne emocjonalne przywiązanie do tych technicznych artefaktów. Warto dodać, że istnieją również badania wskazujące na fakt przypisywania przez ludzi moralnej podmiotowości (którą można uznać za pojęcie w znacznej mierze pokrywające się z pojęciem statusu moralnego) rozmaitym nieludzkim bytom, w tym również robotom³⁵. Tego typu stosunek do robotów zawiera cechę myślenia magicznego, tj. animizm, czyli traktowanie bytów nieożywionych, jak gdyby były ożywione. Innymi słowy, traktując roboty jak ożywione byty podobne do ludzi zachowujemy się analogicznie do dzieci przypisujących takie stany swoim zabawkom lub ludy tubylcze projektujące takie cechy na przyrodę nieożywioną. Relacjoniści legitymizują takie magiczne doświadczenia robotów, nie poddają ich krytyce jako niewiarogodnych, lecz przeciwnie, postulują otwarcie się na tego typu doświadczenia poprzez naśladowanie sposobów doświadczania świata przez kultury niezachodnie, często dość mocno zakorzenionych w myśleniu magicznym. Najlepszym tego przykładem jest zaproponowana przez Jecker, Atiure i Ajei symbioza relacjonizmu z afrykańską filozofią ubuntu, która jest u swoich fundamentów radykalnie

³¹ Lucien Levy-Bruhl, *Czynności umysłowe w społeczeństwach pierwotnych*, tłum. Bella Szwarman-Czarnota (Warszawa: PWN, 1992[1910]), 55–57.

³² Ja-Young Sung i in., „«My Roomba Is Rambo»: Intimate Home Appliances”, w: *Proceedings of UbiComp 2007: Ubiquitous Computing*, red. John Krumm i in. (Berlin, Heidelberg: Springer-Verlag, 2007), 145–162, https://doi.org/10.1007/978-3-540-74853-3_9

³³ Sherry Turkle, „A Nascent Robotics Culture: New Complicities for Companionship”, AAI Technical Report Series, 2006, https://sherryturkle.mit.edu/sites/default/files/images/ST_Nascent%20Robotics%20Culture.pdf (dostęp: 22.11.2023).

³⁴ Julie Carpenter, *Culture and Human-Robot Interaction in Militarized Spaces: A War Story* (Surrey, Burlington: Ashgate, 2016).

³⁵ Aleksandra Lukaszewicz, Paweł Fortuna, „Towards Turing Test 2.0 – Attribution of Moral Status and Personhood to Human and Non-Human Agents”, *Postdigital Science and Education* 4, nr 1(2022): 860–876, <https://doi.org/10.1007/s42438-022-00303-6>

relacjonistyczna, zakładając, że byty uzyskują status moralny nie ze względu na swoje właściwości (w szczególności nie ze względu na właściwości upodabiające do ludzi), lecz ze względu na bycie akceptowanymi jako członkowie i członkinie określonej społeczności, co można zawrzeć we frazie: posiadam „ja”, ponieważ jestem częścią „my”, a tym samym dopuszcza rozmaite elementy myślenia magicznego, m.in. animizm³⁶.

Jeśli idzie o posthumanistyczny aspekt aksjologicznego antyantropocentryzmu relacjonistów, to wskaźmy kilka powodów, ze względu na które można uznać, że propozycja relacjonistów ma stosunkowo wiele wspólnego z niektórymi – również tymi dominującymi – nurtami posthumanizmu. Zacznijmy od tego, że pierwszym symptomem posthumanistycznego charakteru zwrotu relacyjnego jest fakt, że relacjoniści stosunkowo często i aprobatywnie cytują luminarzy tego nurtu, takich jak Bruno Latour, Donna Haraway, Karen Barad czy Rosi Braidotti. Wynika to rzecz jasna z wielorakich podobieństw pomiędzy relacjonizmem w wydaniu Gunkela, Coeckelbergha, Jecker, Atiure i Ajei a posthumanizmem. Pierwsze, najbardziej oczywiste podobieństwo dotyczy wspólnego celu, którym jest propagowanie aksjologicznego antyantropocentryzmu. Drugie, nieco mniej oczywiste, dotyczy środków stosowanych do realizacji powyższego celu – w postaci reaktywacji myślenia magicznego. Nie sposób tu dokonać szczegółowej analizy tego, w jaki sposób posthumanizm osiąga swe cele poprzez – podobnie jak ma to miejsce w przypadku zwrotu relacyjnego – relegitymizację myślenia magicznego. Zauważmy jednak, że zasadniczo w ramach wiodących nurtów posthumanizmu da się odnaleźć wszystkie cechy myślenia magicznego, które wymieniliśmy kilka akapitów wyżej. Dla przykładu, wspomnijmy dyskusję na temat neoanimizmu, w której udział wziął m.in. Tim Ingold. Jej główną konkluzję – podzielaną również przez Ingolda – stanowiło przekonanie, że o animizmie nie należy myśleć jako „prymitywnym” i „irracjonalnym” relikcie myślowym charakterystycznym dla ludów tubylczych, lecz jako o perspektywie przyglądania się światu pod wieloma względami lepszej od tego, co można określić mianem „racjonalności” i „zdrowego rozsądku”, którą należy upowszechniać i kultywować³⁷. Ta swoista relegitymizacja animizmu stanowi przykład jednej z bardziej wyrazistych form „przywra-

³⁶ Jecker, Atiure i Ajei, „The moral standing”: 23–24.

³⁷ Tim Ingold, „Rethinking the Animate, Re-Animating Thought”, *Ethnos* 71, nr 1 (2006): 9–20.

kania do łask” elementów myślenia magicznego przez posthumanizm. Szczegółową argumentację dotyczącą magicznego charakteru posthumanizmu, w szczególności w koncepcjach Bruno Latoura, Tima Ingolda i Bjornara Olsena, przedstawia Michał Rydlewski, formułując jedną ze swych głównych tez następująco: „posthumanizm w wersji, którą określam mianem antropologii animistycznej [...], jest próbą skonstruowania i upowszechnienia nowego typu doświadczenia kulturowego o charakterze magiczno-metamorficznym”³⁸.

Podsumowanie

Omówione w niniejszym tekście stanowiska można potraktować jako ilustrujące pewne szersze tendencje obecne we współczesnej kulturze w odniesieniu do sposobów określania ontologicznego i aksjologicznego statusu bytów ludzkich i nieludzkich, a w szczególności kwestii statusu człowieka wobec bytów nieludzkich. Stanowisko ortodoksyjne wyraża – wciąż chyba jeszcze dominujący we współczesnej kulturze zachodniej – pogląd antropocentryczny przypisujący człowiekowi wyróżniony, centralny status ze względu na jego szczególne właściwości. Etyczny behawioryzm deprecjonuje wyróżniony status człowieka, w praktyce unieważniając (choć w teorii nie eliminując) wspomniane wyżej wyjątkowe właściwości, a zatem – w duchu redukcjonistyczno-naturalistycznym – „uzwyczajniając” człowieka do poziomu innych bytów. Wreszcie, zwrot relacyjny również kontestuje wyjątkowość człowieka, ale czyni to na drodze charakterystycznej dla myślenia magicznego i posthumanizmu „uniezwyklania” tego, co nieludzkie, a zatem nie tyle odbierając człowiekowi jego wyjątkowe cechy, co raczej odbierając tym cechom wyjątkowość i przypisując je innym bytom.

³⁸ Michał Rydlewski, „Posthumanizm a myślenie magiczno-metamorficzne: w poszukiwaniu materialno-cieleśnej intersubiektywności (część I)”, *Rocznik Antropologii Historii* 10, nr 13 (2020): 275–352, <https://doi.org/10.25945/rah.2020.13.011>

Bibliografia

- Bryson, Joanna J. „Patience is not a virtue: The design of intelligent systems and systems of ethics”. *Ethics and Information Technology* 2018, nr 20: 15–26. <https://doi.org/10.1007/s10676-018-9448-6>
- Bryson, Joanna J. „Robots should be slaves”. W: *Close engagements with artificial companions: Key social, psychological, ethical and design issues*, red. Wilks, Yorick, 63–74. Amsterdam: John Benjamins, 2010.
- Carpenter, Julie. *Culture and Human-Robot Interaction in Militarized Spaces: A War Story*. Surrey, Burlington: Ashgate, 2016.
- Cassirer, Ernst. *The philosophy of symbolic form. Volume 2, Mythical thought*. Tłum. Charles Hendel, New Haven: Yale University Press 1960 [1925].
- Coeckelbergh, Mark. *Growing moral relations: Critique of moral status ascription*. New York: Palgrave MacMillan, 2012.
- Coeckelbergh, Mark i David Gunkel. „Facing animals: A relational, other-oriented approach to moral standing”. *Journal of Agricultural and Environmental Ethics* 27, nr 5(2014): 715–733. <https://doi.org/10.1007/s10806-013-9486-3>
- Czy metamorfoza magiczna rekompensuje brak symbol*. Red. Jerzy Kmita. Poznań: UAM, 2001.
- Danaher, John. „Welcoming robots into the moral circle: A defence of ethical behaviourism”. *Science and Engineering Ethics* 26, nr 4(2019): 2023–2049. <https://doi.org/10.1007/s11948-019-00119-x>
- Danaher, John. „Robot Betrayal: a guide to the ethics of robotic deception”. *Ethics and Information Technology* 22, nr 2(2020): 117–128. <https://doi.org/10.1007/s10676-019-09520-3>
- Danaher, John. „What Matters for Moral Status: Behavioral or Cognitive Equivalence?”. *Cambridge Quarterly of Healthcare Ethics* 30, nr 3(2021): 472–478. <https://doi.org/10.1017/S0963180120001024>
- Dobosz, Artur. *Myślenie magiczno-mityczne a schizofrenia*. Bydgoszcz: Oficyna Wydawnicza Epigram, 2013.
- Durkheim, Emile i Marcel Mauss. *Primitive Classification*. Tłum. Rodney Needham. Chicago: University of Chicago Press, 1963 [1902].
- Danaher, John. „The philosophical case for robot friendship”. *Journal of Posthuman Studies* 3, 1(2019): 5–24. <https://doi.org/10.5325/jpoststud.3.1.0005>
- Frazer, James George. *Złota gałąź*. Tłum. Jan Lutyński (Warszawa: PIW, 1962 [1890]).
- Gerdes, Anne. „The Issue of Moral Consideration in Robot Ethics”. *ACM SIGCAS Computers & Society* 45, 3(2015): 274–280. <https://doi.org/10.1145/2874239.2874278>
- Greenwood, Susan. *The Anthropology of Magic*. Oxford, New York: Bloomsbury Academic, 2009.
- Gunkel, David. *The machine question*. Cambridge, MA: MIT Press, 2012.
- Gunkel, David. *Robot rights*. Cambridge, MA: MIT Press, 2018.

- Jaworska, Agnieszka i Julie J. Tannenbaum. „The grounds of moral status”. W: *Stanford Encyclopedia of Philosophy*. Red. Edward N. Zalta. <https://plato.stanford.edu/entries/grounds-moral-status/> (dostęp: 10.12.2023).
- Ingold, Tim. „Rethinking the Animate, Re-Animating Thought”. *Ethnos* 71, nr 1(2006): 9–20.
- Jecker, Nancy S., Caesar A. Atiure i Martin Odei Ajei. „The Moral Standing of Social Robots: Untapped Insights from Africa”. *Philosophy & Technology* 35, (2022): 1–22. <https://doi.org/10.1007/s13347-022-00531-5>
- Levy-Bruhl, Lucien. *Czynności umysłowe w społeczeństwach pierwotnych*. Tłum. Bella Szwarzman-Czarnota. Warszawa: PWN, 1992 [1910].
- Lukaszewicz Aleksandra, Paweł Fortuna. „Towards Turing Test 2.0 – Attribution of Moral Status and Personhood to Human and Non-Human Agents”, *Postdigital Science and Education* 4, nr 1(2022): 860–876. <https://doi.org/10.1007/s42438-022-00303-6>
- Musiał, Maciej. „Criticizing Danaher’s Approach to Superficial State Deception”. *Science and Engineering Ethics* 29, 5(2023): 1–15. <https://doi.org/10.1007/s11948-023-00452-2>
- Musiał, Maciej. „Stanowiska w sporze o status moralny robotów jako przejaw tendencji desymbolizacyjnych”. W: *Obyczaj – kultura – nauka. Księga jubileuszowa dedykowana Profesorowi Janowi Gradowi*, red. Jacek Sójka, Agata Skórzyńska i Andrzej Bełkot, 185–194. Poznań: Wydawnictwo Nauk Społecznych i Humanistycznych UAM, 2023.
- Müller, Vincent C. „Is it time for robot rights? Moral status in artificial entities”. *Ethics and Information Technology* 23, nr 4(2021): 579–587. <https://doi.org/10.1007/s10676-021-09596-w>
- Nyholm, Sven. *Humans and robots: Ethics, agency, and anthropomorphism*. London, New York: Rowman & Littlefield, 2020.
- Nyholm, Sven, i Lily Frank. „From sex robots to love robots: Is mutual love with a robot possible?”. W: *Robot sex: Social and ethical implications*, red. John Danaher i Neil McArthur, 219–244. Cambridge, MA: MIT Press, 2017.
- Piaget, Jean. *Jak sobie dziecko wyobraża świat*. Tłum. Maria Gawlik. Warszawa: PWN, 2006 [1926].
- Roheim, Geza. *Magic and Schizophrenia*. Bloomington: Indiana University Press, 1955.
- Rydlewski, Michał. „Posthumanizm a myślenie magiczno-metamorficzne: w poszukiwaniu materialno-cieleśnej intersubiektywności (część I)”. *Rocznik Antropologii Historii* 10, nr 13, (2020): 275–352. <https://doi.org/10.25945/rah.2020.13.011>
- Shevlin, Henry. „How Could We Know When a Robot was a Moral Patient?”. *Cambridge Quarterly of Healthcare Ethics* 30, nr 3(2021): 459–471. <https://doi.org/10.1017/S0963180120001012>
- Smids, Jilles. „Danaher’s Ethical Behaviourism: An Adequate Guide to Assessing the Moral Status of a Robot?”. *Science and Engineering Ethics* 26, nr 5(2020): 2849–2866. <https://doi.org/10.1007/s11948-020-00230-4>
- Subbotsky, Eugene. *Magic and the Mind: Mechanisms, Functions, and Development of Magical Thinking and Behavior*. Oxford, New York: Oxford University Press, 2010.

Sung, Ja-Young, Lan Guo, Rebecca E. Grinter i Henrik I. Christensen. „«My Roomba Is Rambo»: Intimate Home Appliances”. W: *Proceedings of UbiComp 2007: Ubiquitous Computing*, red. John Krumm, Gregory D. Abowd, Aruna P. Seneviratne i Thomas Strang, 145–162. Berlin, Heidelberg: Springer-Verlag, 2007. https://doi.org/10.1007/978-3-540-74853-3_9

Turkle, Sherry. „A Nascent Robotics Culture: New Complicities for Companionship”. AAAI Technical Report Series, 2006. https://sherryturkle.mit.edu/sites/default/files/images/ST_Nascent%20Robotics%20Culture.pdf (dostęp: 22.11.2023).

Summary

Ethical Behaviorism and Relational Turn as Two Manifestations of Anti-Anthropocentrism

This text discusses three positions within the debate on the moral status of robots. The first, referred to as the “orthodox” position, expresses the still-dominant anthropocentric view in contemporary Western culture, which assigns humans a privileged, central status due to their unique qualities such as consciousness, emotions, free will, etc. The two alternative positions to the orthodox view, ethical behaviorism and relational turn, propose different criteria for attributing moral status than the aforementioned qualities and are interpreted as manifestations of a tendency to deny the unique ontological and axiological status of humans. In the case of ethical behaviorism, humans are “trivialized” to the level of other entities in a reductionist-naturalistic manner, whereas the relational turn involves “extraordinarizing” the non-human by re-legitimizing magical thinking in a posthumanist spirit. Thus, the article illustrates how contemporary cultural tendencies towards dual elimination of human uniqueness, based on reductionist naturalism and magical posthumanism, are revealed in the discussion on the moral status of robots.

Keywords: robots, moral status, ethical behaviorism, relational turn, Anti-Anthropocentrism

Zusammenfassung

Ethischer Behaviorismus und die relationale Wende als zwei Erscheinungsformen des Anti-Anthropozentrismus

Ziel des Textes ist es, die Frage zu beantworten: Wie ist die Unterbestimmtheitstheese auf dem Boden der Technikwissenschaften zu verstehen? Was trägt diese These zur Erklärung bei?

Welche Schwierigkeiten sind zu erkennen, wenn sie angenommen wird? Ich stelle zunächst den Entstehungskontext der These und ihre erste Formulierung in der Technikphilosophie vor. Dann stelle ich ein Fallbeispiel vor, das den in der These rekonstruierten Mechanismus veranschaulicht und gleichzeitig für seine Übernahme plädiert. Sie dient auch als Musterbeispiel für die Analysen, die im Rahmen von SCOT (Social Construction of Technology) – der sozialen Konstruktion von Technik – durchgeführt werden. Als nächstes rekonstruiere ich eine alternative Formulierung der These – sie ist nicht unvereinbar mit Feenbergs Vorschlag, sondern kann meiner Meinung nach als eine Verfeinerung desselben angesehen werden. Im folgenden Abschnitt stelle ich die wichtigsten Konsequenzen vor, die sich aus der Annahme der fraglichen These ergeben.

Schlüsselwörter: Wissenschafts- und Technikforschung, Technikphilosophie, Andrew Feenberg, SCOT, Unterbestimmtheitstheorie, Antianthropozentrismus

Ins Deutsche übersetzt von Anna Pastuszka

Information about Author:

MACIEJ MUSIAŁ, PhD, Habilitation, Associate Professor, Adam Mickiewicz University in Poznań; address for correspondence: Faculty of Philosophy, ul. Szamarzewskiego 89, PL 60-568 Poznań; e-mail: m.musial@amu.edu.pl

