

FILOZOFIA I NAUKA
Studia filozoficzne i interdyscyplinarne
Tom 8, cz. 2, 2020

Barbara Trybulec

PODMIOT CZY AGENT? ROZUMIENIE PODMIOTOWOŚCI W ERZE ARTEFAKTÓW POZNAWCZYCH

10.37240/FiN.2020.8.2.6

STRESZCZENIE

Pokartezjańska tradycja filozoficzna określa podmiot działania jako istotę refleksyjną, z zasady świadomą swej podmiotowości, co czyni ją zdolną do podjęcia wolnej decyzji i wzięcia za nią odpowiedzialności. Tak rozumiana podmiotowość przypisywana jest, poza Bogiem i aniołami, wyłącznie ludziom, tylko ludzie bowiem rozumieją znaczenia języka pojęciowego, co stanowi konieczny warunek refleksyjności. Rozwój sztucznych wzmocnień poznawczych oraz coraz bardziej autonomicznej sztucznej inteligencji, który dokonał się w przeciągu ostatnich kilkudziesięciu lat, skłania jednak do refleksji, czy takie ujęcie podmiotu nie jest zbyt antropocentryczne. Na wątpliwość tą wskazują niektórzy zwolennicy rozszerzania procesów poznawczych i stanów mentalnych poza mózg, argumentując za potrzebą odpowiedniego rozszerzenia również ich podmiotu. Coraz częściej pojawiają się także propozycje nadawania podmiotowości samym artefaktom. W artykule oceniam próby nadawania podmiotowości zarówno rozszerzonym, jak i sztucznym systemom poznawczym podjęte przez Lambrosa Malafourisa, Richarda Menary'ego, Edwina Hutchinsa, Luciano Floridiego i J. W. Sandersa, oraz Samira Chopry i Laurence'a F. White'a. Ocenę tę wydaję z perspektywy dwóch koncepcji podmiotowości: podmiotu jako systemu refleksyjnego, oraz podmiotu jako przedmiotu nastawienia intencjonalnego. Reprezentantem pierwszego z tych ujęć czynię Donalda Davidsona, drugiego – Daniela Dennetta. Argument, którego bronię brzmi, że pojęcie podmiotu wypracowane w analitycznej filozofii działania nie może zostać zastosowane do rozszerzonych i sztucznych systemów poznawczych, a przyjęcie wobec nich strategii nastawienia intencjonalnego nie spełnia tej samej funkcji. Postuluję zatem, aby systemy rozszerzone i sztuczne określać za pomocą nazwy „agent” oznaczającej wszelki relatywnie autonomiczny system przetwarzający informacje w celu wykonania zadania poznawczego. Nazwę „podmiot działania” należy tymczasem zarezerwować dla systemów faktycznie spełniających odpowiednie kryteria.

Słowa kluczowe: podmiot, agent, sprawczość, teza umysłu rozszerzonego, rozszerzony system poznawczy, wzmocnienia poznawcze, sztuczny system poznawczy.

1. WPROWADZENIE

Osiągnięcia nauki i rozwój technologiczny na przestrzeni wieków zawsze silnie wpływały na treść rozważań dotyczących umysłu, wolności i odpowiedzialności, wiedzy i świadomości. Filozof przyjmujący te kategorie za obiekty swych badań, nie może przejść obojętnie obok faktu powstawania coraz bardziej zaawansowanych artefaktów wzmacniających ludzkie naturalne zdolności poznawcze. W prezentowanym artykule skupiam się na pojęciu podmiotu działania, które jest ściśle związane ze wspomnianymi powyżej kategoriami i silnie narażone na nadużycia w czasach upowszechniania się rozszerzonych i sztucznych systemów poznawczych. Teza umysłu rozszerzonego, wprowadzona przez Andy Clarka i Davida Chalmersa, uzasadniła możliwość istnienia rozszerzonych procesów poznawczych i stanów mentalnych przebiegających w takich systemach. Wciąż jednak pozostaje niejasna kwestia podmiotu owych procesów i stanów, którą można ująć w pytaniu, czy podmiot również się rozszerza wraz ze swoim umysłem, czy też nie. Istnienie rozszerzonych systemów poznawczych stanowi tym samym poważne wyzwanie dla filozoficznych i potocznych intuicji związanych z pojęciem podmiotu, wymaga bowiem zastosowania go na zupełnie nowym terenie.

Na gruncie badań kognitywistycznych prowadzonych w języku polskim, do oznaczenia systemu wykonującego zadania poznawcze używa się najczęściej nazwy „agent”. Nazwa „podmiot”, lub „sprawca” jest rzadko stosowana ze względu na jej filozoficzne obciążenie konotacjami ze świadomością, intencją, wolnym wyborem i wrażliwością na wartości – własnościami, których nie przejawia wiele systemów zdolnych do rozwiązywania zadań poznawczych. W kognitywistycznej literaturze anglojęzycznej również dominuje nazwa „agent”, lecz tutaj sytuacja znaczeniowa jest bardziej złożona. Podobnie jak w języku polskim jest ona najczęściej stosowana na określenie relatywnie złożonego systemu przetwarzającego informacje w celu wykonania zadania poznawczego, jednak, za sprawą analitycznej filozofii działania, jest ona filozoficznie obciążona w podobny sposób jak polski „podmiot”. Z tej racji, teoretycznie, angielska nazwa „agent” powinna funkcjonować wyłącznie jako określenie sprawcy – istoty refleksyjnej; w praktyce jej zakres jest jednak dużo szerszy, co może prowadzić do niejasności i nadużyć. Angielska nazwa „subject”, która mogłaby oznaczać system refleksyjny i odróżniać go od wszystkich innych agentów, jest rzadko stosowana ze względu na jej konotacje związane z subiektywnością i perspektywą pierwszoosobową. Sytuacja w kognitywistyce anglojęzycznej wydaje się zatem trudniejsza niż na gruncie języka polskiego. Stosowana jest tam bowiem właściwie tylko jedna nazwa – „agent” – do określania wszystkich systemów rozwiązujących zadania poznawcze. Kognitywistyka polskojęzyczna również zdaje się zmierzać w kierunku unifikacji wszystkich takich systemów pod nazwą „agent”, jest ona jednak w lepszej sytuacji, gdyż dysponuje nazwą „podmiot działania”, której

treść dobrze oddaje różnice pomiędzy pierwotną i wtórną sprawczością. Wyznaczenie wyraźnej granicy pomiędzy agentem a podmiotem działania (sprawcą) wprowadza porządek teoretyczny pozwalający rozjaśnić wiele kwestii praktycznych związanych nie tylko z podmiotowością epistemiczną, lecz również z moralną i prawną. Nazwę „podmiot poznający” stosuję w artykule na oznaczenie podmiotu procesów poznawczych. To niejasne sformułowanie wymaga precyzacji, która wykracza jednak poza rozważania tej pracy. W tym miejscu zakładam jedynie, że każdy podmiot działania jest podmiotem poznającym. Wynika to z koncepcji podmiotu wypracowanego na gruncie filozofii analitycznej, gdzie działania wymagają podjęcia decyzji na postawie danych będących efektem procesów poznawczych. Wstępnego wyjaśnienia wymaga również używany przeze mnie termin „system”. Badaniem jego desygnatu zajmują się różne dziedziny nauki, od fizyki i chemii przez biologię po nauki społeczne. Systemy stanowią zbiory różnych elementów powiązanych wzajemnymi relacjami, które razem biorą udział w realizowaniu procesów prowadzących do konkretnych efektów, np. wytworzenia pewnych zjawisk¹. W artykule, mówiąc o systemie poznawczym odwołuję się do pojęcia systemu dynamicznego rozwiniętego w ramach modelowania dynamicznego będącego częścią matematyki stosowanej i mającego na celu wyjaśnianie naturalnych zjawisk poprzez konstruowanie ich abstrakcyjnych, dynamicznych modeli. Teoria systemów dynamicznych jest z kolei gałęzią matematyki teoretycznej skupionej wokół analizowania geometrycznych i topologicznych własności możliwych zachowań, które owe abstrakcyjne modele mogą przejawiać w czasie. Pomijając techniczny opis działania systemu dynamicznego można opisać jego wyróżniającą cechę następująco: składa się on z podsystemu, jednego lub kilku, którego dane na wejściu pochodzą z informacji wyprodukowanych na wyjściu przez inny podsystem, z którym ten pierwszy wchodzi we wzajemne interakcje. System dynamiczny może tworzyć zatem dwa lub więcej mniejszych odpowiednio sprzężonych ze sobą systemów przetwarzania informacji, które w efekcie tego sprzężenia realizują cel poznawczy. Takie pojęcie systemu wykorzystują zwolennicy istnienia rozszerzonych systemów poznawczych składających się z człowieka i wykorzystywanego przez niego artefaktu, lub z wielu ludzi (Chemero 2009; Theiner, O'Connor 2010; Palermos 2014). Po tych wstępnych ustaleniach mogę przejść do określenia celu bieżących analiz.

Pierwszym celem prezentowanej pracy jest ocena wybranych prób nadania statusu podmiotu działania zarówno jednoosobowym rozszerzonym systemom poznawczym, jak i samym artefaktom stanowiącym taki system. Próby te oceniam z perspektywy rozumienia podmiotu jako istoty refleksyj-

¹ Pojęcie systemu analizowane jest szczegółowo na gruncie ogólnej teorii systemów zapoczątkowanej w biologii przez Ludwiga von Bertalanffyego (Bertalanffy 1984), rozwijanej początkowo głównie przez cybernetyków (Wiener 1971), a następnie zaadaptowanej przez wiele innych dziedzin nauki, w tym przez nauki społeczne (Mazur 1966).

nej, oraz z perspektywy strategii nastawienia intencjonalnego. Wybrałam te dwie koncepcje ze względu na ich precyzyjne stanowisko wobec pierwotnej i wtórnej sprawczości. Tezy w nich wyrażone są zgodne, mimo ich odmienności, z teoretycznymi i potocznymi intuicjami i z tego względu dominują w różnych dziedzinach filozofii jako określenia podmiotu działania. Drugim moim celem jest uzasadnienie potrzeby wyraźnego odróżnienia dwóch pojęć: podmiotu działania i agenta i odpowiedniego ich stosowania podczas analiz i badań dotyczących rozszerzonych i sztucznych systemów poznawczych.

Realizując te zadania wskazuję w pierwszej kolejności na rozumienie podmiotu jako istoty refleksyjnej, a w drugiej na jego określenie jako przedmiotu nastawienia intencjonalnego. Następnie, po krótkim wprowadzeniu w koncepcję poznania rozszerzonego, przedstawiam i oceniam argumentację Lambrosa Malafourisa, Richarda Menary'ego i Edwina Hutchinsa za ideą rozszerzania podmiotu działania na cały rozszerzony system poznawczy, którą nazywam tezą podmiotu rozszerzonego. W kolejnej części artykułu skupiam się na dwóch stanowiskach reprezentowanych przez Luciano Floridi'ego i J. W. Sandersa, oraz Samira Chopry i Laurence'a F. White'a, które przypisują podmiotowość moralną i prawną samym artefaktom poznawczym. Oceniam, jakie pojęcie podmiotu użyte jest w tych koncepcjach i jakie relacje łączą je z pojęciami podmiotu, na które wskazuję w pierwszej części. W ostatniej części pracy, przedstawiam propozycję rozróżnienia pomiędzy podmiotem działania a agentem. Wskazuję również na warunki, które muszą zostać spełnione, aby proces rozszerzania podmiotu działania można było uznać za dokonany, oraz zwracam uwagę na jego możliwe konsekwencje.

2. PODMIOT DZIAŁANIA JAKO ISTOTA REFLEKSYJNA

Do ugruntowania się w myśli nowożytnej i współczesnej idei podmiotu działania jako istoty refleksyjnej przyczyniła się w decydującym stopniu filozofia Kartezjusza. Podmiotem myślącym może być na jej gruncie, pomijając Boga i aniołów, tylko człowiek, bowiem tylko on spełnia konieczny warunek refleksyjności związanej ściśle z racjonalnością, samoświadomością, oraz zdolnością do działań moralnych opartą na wolności i odpowiedzialności (Descartes 2004/1647). Refleksyjność określana jest jako zdolność postrzegania siebie jako podmiotu zarówno doświadczenia zmysłowego jak i postaw propozycjonalnych. Owa samowiedza jest związana z posiadaniem jaźni rozumianej jako subiektywny punkt widzenia gwarantujący psychiczną tożsamość podmiotu w czasie (Bremer 2016). Podmiot posiadający jaźń stanowi osobę i wyróżnia się spośród wszystkich innych istot i systemów poznawczych zainteresowaniem własnymi reprezentacjami świata. Jest to bardzo wyrafinowana własność, gdyż zakłada rozumienie języka pojęciowego. Istota

refleksyjna, czyli świadoma własnej podmiotowości, musi dysponować pojęciem Ja i tworzyć metareprezentacje posiadające niepochodną treść na temat swoich stanów mentalnych. Umiejętność posługiwania się językiem pojęciowym leży również u podstaw wrażliwości na wartości – kolejnej koniecznej cechy podmiotu jako istoty refleksyjnej. Bez znajomości pojęć takich jak dobro, zło, godność, wolność, czy miłość nie jest bowiem możliwe planowanie działań, oraz refleksja nad zgodnością przekonań i pragnień z przyjętymi wartościami (Taylor 2001).² Każda osoba jest świadoma swojej osobistej historii, którą ocenia ze względu na owe wartości. Planuje ona swoje dalsze życie tak, aby je realizować. Nie byłoby to możliwe bez posiadania przez nią pojęć przeszłości i przyszłości, czyli bez dostępu do znaczeń. Wrażliwość na wartości, wraz z możliwością dokonania wolnego wyboru pomiędzy dostępnymi działaniami, decyduje o podmiotowości moralnej. Podmiot działania moralnego jest odpowiedzialny za swoje decyzje, można mu tym samym przypisać zasługę lub obarczyć go winą za dokonany wybór. Takie rozumienie podmiotu wskazuje wyraźnie na związek podmiotowości moralnej z poznawczą. Podmiot działania moralnego musi być podmiotem poznania znaczeń, wartości oraz własnych stanów mentalnych.

Idea podmiotu jako istoty refleksyjnej jest rozwinięta w XX-wiecznej analitycznej filozofii działania. Uznaję tu za jej reprezentanta Donalda Davidsona, gdyż jego stanowisko jest jednym z najbardziej wpływowych i dyskutowanych we współczesnej filozofii podmiotu (Davidson 1971; 1978). Analizy Davidsona dotyczą podmiotu działania, oznaczanego przez tego filozofa nazwą „agent”, co w dużym stopniu wpłynęło na rozwinięcie się na gruncie filozofii anglojęzycznej konotacji wiążących tę nazwę z istotą refleksyjną. Wśród różnych zjawisk zachodzących w świecie działanie wyróżnia się mianowicie tym, że jest pod pewnym opisem intencjonalne, czyli wykonane ze względu na racje posiadane przez podmiot, którymi są jego przekonania i pragnienia. Te stany mentalne składają się na intencję, która stanowi konieczny warunek działania.³ Podmiotowość przynależy zatem wyłącznie obiektowi (systemowi), który ma możliwość podjęcia decyzji o działaniu opierając się na posiadanych racjach. Aby zrealizować te wymagania, system musi posiadać stany wewnętrzne z odpowiednią treścią reprezentacyjną, innymi słowy, musi posługiwać się językiem pojęciowym. Bez tej umiejętności nie mógłby on używać stanów propozycyjalnych, takich jak przekonania i pragnienia w celu uzasadniania swych zachowań, które z tej przyczyny nie

² Wrażliwość na wartości jako cecha wyróżniająca osobę jest trafnie zilustrowana przez Annette Baier: „...spośród rzeczy, o których można powiedzieć, że są młode, tylko o osobach ludzkich (a są to jedyne osoby jakie znamy) można powiedzieć, że przeżywają swą młodość, marnują ją, stają się pełnoletnie lub mają najlepsze lata swojego życia za sobą. Osoby tworzą kalendarze, spisują swoje przemyślenia, świętują rocznice, rozpoznają momenty przejścia od jednej fazy życia do drugiej, odprawiają pogrzeby” (Baier 2001, s. 54).

³ W późniejszych pracach Davidson skłonny był uznawać intencje za osobne stany intencjonalne, które mogą przybrać postać nigdy niezrealizowanego zamierzenia (Davidson 1978).

uzyskałyby statusu działań (Davidson 1992). Nie każde działanie musi być jednak poprzedzone rozważaniem przez podmiot racji. Wystarczy, że jeśli byłby on ich świadomy, mógłby w oparciu o nie uzasadnić racjonalność swojego działania (Davidson 1978).

Zgodnie z propozycją Davidsona, system niedysponujący językiem nie może być podmiotem myślenia, gdyż myśl i język wzajemnie się warunkują. Aby myśleć, należy uczestniczyć w wymianie znaczeń językowych z innymi ludźmi, znać warunki prawdziwości i zasadności wypowiedzianych przez nich zdań. Myśl jest dyspozycją do określonego działania werbalnego w danych okolicznościach, polegającego na wyrażaniu odpowiedniej postawy wobec ich treści takiej jak, na przykład, asercja, wątpliwość, pragnienie jej prawdziwości itd. (Davidson 1980). Aby myśleć, system musi zatem dysponować pojęciem prawdy i fałszu; w innym przypadku nie mógłby przyjąć żadnej postawy wobec wybranego sądu, ani dokonać jej normatywnej oceny. Pojęcie obiektywnej prawdy kształtuje się natomiast tylko w wyniku komunikacji językowej z innymi ludźmi, która zakłada konieczność interpretacji ich wypowiedzi (Davidson 1992). System decydujący o swoim działaniu w oparciu o posiadane racje, będący innymi słowy podmiotem działania, musi zatem poruszać się w obiektywnej przestrzeni znaczeniowej stanowiącej podstawę do oceny jego wyborów i wypowiedzi innych ludzi pod względem prawdziwości, spójności i uzasadnienia.

Zdaniem Davidsona, tylko człowiek, spośród wszystkich zwierząt, może stanowić tak opisany podmiot, tylko on jest źródłem inicjatywy, elementem aktywnym, jednym słowem, sprawcą. Na gruncie polskiej filozofii, takie rozumienie podmiotu rozwija Stanisław Judycki. Jego zdaniem, podmiotowość powinna być przypisana wyłącznie systemowi, który posiada świadomość swojej autonomii i siły inicjującej działanie, który, innymi słowy ma poczucie sprawstwa i kontroli wobec swoich wyborów, intencji i działań. Judycki odmawia tych własności wszelkim teraźniejszym i przyszłym programom komputerowym niezależnie od tego, jak dokładnie będą imitować ludzką racjonalność i świadomość. Tylko człowiek, jego zdaniem, poznaje wartości, co jest niezbędne dla zaistnienia jakichkolwiek pragnień. Dodając do tego warunek wolności woli,⁴ który stanowi podstawę wszelkiej decyzji o działaniu, filozof ten konkluduje, że żaden nienadprzyrodzony system poznawczy poza człowiekiem nigdy nie będzie do niego zdolny (Judycki 2006).

Podsumowując, koncepcja podmiotu działania jako istoty refleksyjnej głosi, że podmiotowość przynależy wyłącznie systemowi, który ma poczucie sprawstwa i kontroli wobec swoich wyborów i działań. Zarówno Davidson jak i Judycki uznają, że żadne zwierzę poza człowiekiem, ani żaden sztuczny system nie spełnia tych wymagań. W trzeciej i czwartej części artykułu

⁴ Warunek ten Judycki, podobnie jak Frankfurt, definiuje jako możliwość refleksji nad własną hierarchią wartości i możliwość działania z nią niezgodnego (Judycki 2006), (Frankfurt 1971).

przedstawię argumenty filozofów, którzy twierdzą, że taka koncepcja podmiotu jest zbyt rygorystyczna i powinna zostać osłabiona w celu włączenia niektórych artefaktów do zbioru podmiotów działających.

3. PODMIOT DZIAŁANIA JAKO PRZEDMIOT NASTAWIENIA INTENCJONALNEGO

Refleksyjności opartej na samoświadomości i perspektywie pierwszej osoby zagraża brak możliwości jej naukowego badania ze względu na jej subiektywny charakter. Aby uczynić ją przedmiotem obiektywnego opisu należy poddać ją naturalizacji, czego podejmuje się Daniel Dennett. Uważa on, że nazwy oznaczające stany mentalne są niezbędne w dyskursie naukowym, bowiem nie sposób bez nich przewidywać i wyjaśniać zachowania. Z drugiej strony, naukowcy nie mają dostępu do pragnień i przekonań stanowiących jego przyczyny. Z tego względu należy założyć, że każdy obiekt, którego zachowanie można uznać za racjonalne, posiada wyjaśniające je intencjonalne stany propozycjonalne, a człowiek nie jest wyróżnionym elementem tego zbioru. Innymi słowy, podmiotowość systemu jest uzależniona od postawy jaką obserwator wobec niego przyjmuje. System nie jest traktowany jako podmiot postaw propozycjonalnych dlatego, że faktycznie je posiada, przeciwnie, staje się on ich podmiotem dlatego, że tak jest traktowany (Dennett 1976). Taka strategia, nazwana nastawieniem intencjonalnym, jest stosowana głównie wtedy, gdy obserwator nie zna mechanizmu działania racjonalnie zachowującego się systemu. Aby wyjaśniać i przewidywać jego zachowanie, przyjmuje on zatem, że rządzą nim odpowiednie stany mentalne w postaci przekonań i pragnień.⁵

Stanowisko Dennetta wobec podmiotowości może być interpretowane instrumentalistycznie, gdyż jej przypisywanie nie opiera się na rozpoznaniu rzeczywistych cech systemu, lecz ma służyć jako użyteczne narzędzie wyjaśniania i przewidywania jego zachowania. Wydaje się zatem, że pytanie o świadomość, wrażliwość na wartości, wolność i odpowiedzialność nie ma na gruncie tego stanowiska sensu; refleksyjność można bowiem nadać systemowi, któremu nie jesteśmy skłonni przypisać tych cech, jeśli tylko uznamy taką strategię za wyjaśniająco użyteczną. W swych późniejszych pracach Dennett przyznaje postawom propozycjonalnym realność, mimo że nadal są one rozpoznawalne wyłącznie z perspektywy nastawienia intencjonalnego (Dennett 1981). Obserwator mianowicie przyjmując to nastawienie widzi pewien aspekt systemu, jego emergentną własność, realny wzorzec, który

⁵ Gdy znany jest mechanizm funkcjonowania systemu, można go opisać w terminach funkcjonalnych, bez konieczności stosowania terminów intencjonalnych. Nastawienie takie nazywa Dennett projektowym. Do wyjaśniania jeszcze prostszych zachowań, na przykład ruchu obiektów fizycznych, służy natomiast nastawienie fizykalne (Dennett 1971).

z innych perspektyw jest niedostrzegalny, a który system obiektywnie posiada. Różnica pomiędzy stanowiskiem Dennetta i Davidsona dotyczy przede wszystkim perspektywy z jakiej rozpatrywana jest problematyka podmiotowości. Pierwszego z tych filozofów interesuje mianowicie głównie kwestia metodologiczna: jak rozpoznać wśród różnych systemów poznawczych podmiot intencjonalnych stanów mentalnych. Davidson i inni przedstawiciele analitycznej filozofii podmiotu skupiają się natomiast przede wszystkim na wskazaniu koniecznych cech takiego podmiotu, bez precyzowania, jak owe cechy mają być rozpoznawane w praktyce i jaki status ontologiczny należy im nadać.

4. TEZA POZNANIA ROZSZERZONEGO

Rozumienie podmiotu działania jako istoty refleksyjnej stanowi punkt wyjścia dla ujęć przypisujących podmiotowość systemom rozszerzonym oraz artefaktom. Przed przystąpieniem do ich analizy przybliżę główną ideę tezy poznania i umysłu rozszerzonego. Obie tezy zostały wprowadzone przez Andy'ego Clarka i Davida Chalmersa w artykule *Umysł rozszerzony* (Clark, Chalmers 1998), a następnie rozwinięte przez Clarka w jego kolejnych pracach. Rozróżnienie pomiędzy nimi oraz rozumowanie, które doprowadziło autorów do ich sformułowania przedstawiłam w innym miejscu (Trybulec 2017), dla potrzeb tej pracy, przytoczę jedynie konkluzję argumentacji Clarka i Chalmersa. Teza poznania rozszerzonego głosi, że ludzkie procesy poznawcze mogą być częściowo realizowane przez procesy zachodzące w zewnętrznych wobec człowieka obiektach takich jak na przykład komputer. Jej autorzy uzasadniają ją w sposób funkcjonalistyczny wskazując, że miejsce realizacji własności poznawczych nie ma znaczenia, liczy się jedynie równoważność funkcjonalna struktury organizacyjnej realizatora. Wyjaśniając powiązanie procesów wewnętrznych wobec organizmu podmiotu z zewnętrznymi, Clark przywołuje pojęcie sprzężenia zwrotnego rozwiniętego na gruncie teorii systemów dynamicznych. Relacja ta polega na ciągłym, wzajemnym przyczynowaniu (*continuous reciprocal causation*) zachodzącym pomiędzy różnymi elementami dynamicznego systemu wykonującego dane zadanie poznawcze. Lokalizacja owych elementów i ich fizyczna budowa nie ma znaczenia dla funkcji, którą pełnią w systemie, liczy się jedynie ich wzajemne przyczynowe powiązanie tworzące odpowiednią, dla osiągnięcia danego celu, strukturę organizacyjną. Złożony, dynamiczny system może zatem składać się z procesów przebiegających w samym mózgu organizmu, w innych częściach jego ciała, a także poza nim w jego zewnętrznym środowisku i stanowić tym samym system rozszerzony (Clark 2008). O granicach systemu poznawczego decydują, jednym słowem relacje pomiędzy procesami realizowanymi przez różne fizyczne nośniki.

Teza umysłu rozszerzonego, zilustrowana znanym eksperymentem myślowym z Ottonem i Ingą w roli głównej, jest ideą dużo bardziej radykalną od omówionej powyżej. Nie tylko mianowicie procesy poznawcze, lecz również stany mentalne, takie jak przekonania dyspozycyjne, mogą być częściowo konstytuowane przez czynniki istniejące poza ciałem podmiotu, na przykład przez zapisy w notatniku (Clark, Chalmers 1998). Teza ta jest zagrożona tak zwanym rozděciem poznawczym, którego Clark i Chalmers próbują uniknąć, niezbyt skutecznie, zdaniem krytyków.⁶ W wyniku tego rozděcia stany mentalne rozprzestrzeniają się w świecie w niekontrolowany sposób prowadząc do absurdalnych scenariuszy. Najbardziej efektywną metodą zawężenia systemów rozszerzonych do ich właściwych przypadków jest przestrzeganie warunku obustronnego sprzężenia pomiędzy różnymi zachodzącymi w jego ramach procesami.⁷ W przypadku zaistnienia takich wzajemnych relacji przyczynowych nie ma możliwości odróżnienia procesów wewnętrznych i zewnętrznych wobec organizmu, wchodzą one bowiem w tak zwane ciągłe pętle sprzężenia zwrotnego (*ongoing feedback loops*) (Clark 2009, Chemero 2009). W wyniku tych oddziaływań powstają nowe emergentne własności takie jak, na przykład, postawy propozycyjalne, skutkujące odpowiednim zachowaniem, którego podmiotem jest cały rozszerzony system (Clark 2008).⁸ Czy to jednak wystarczy, aby systemowi takiemu, jako całości, nadać status podmiotu działającego?

5. TEZA PODMIOTU ROZSZERZONEGO

Szybki rozwój technologiczny wzmocnił przekonania wielu kognitywistów do sprzeciwiania się filozoficznej antropomorfizacji podmiotu działania. Podejmowane są próby rozszerzania podmiotowości na artefakty, które albo są w odpowiedni sposób sprzężone z ludzkim systemem poznawczym, albo same wykazują wymagane cechy. W tej części artykułu wskażę na rozwiązania pierwszego typu.

Pierwsze, na które chciałabym zwrócić uwagę, zaproponował Lambros Malafouris (Malafouris 2008). Autor ten ilustruje swoją argumentację przykładem doświadczonego garncarza, który wytwarza naczynie niemal bezre-

⁶ Błąd rozděcia poznawczego zarzucają Clarkowi między innymi Kenneth Aizawa i Frederic Adams (Adams, Aizawa 2010), oraz Brie Gertler (Gertler 2007).

⁷ Myśl tę rozwinęłam w książce *Wiedza i jej podmiot w szerokich systemach poznawczych* (Trybulec 2017). Argumentuję w niej, że kryterium rozszerzonego procesu poznawczego zaproponowane przez teorię systemów dynamicznych jest jaśniejsze i mniej problematyczne od propozycji Clarka i Chalmersa zawartej w czterech kryteriach rozszerzonego systemu poznawczego przedstawionych w artykule „Umysł rozszerzony”. Przede wszystkim stanowi ono warunek konieczny i wystarczający takiego systemu, którego nie dostarcza propozycja Clarka i Chalmersa.

⁸ Należy zaznaczyć, że Clark ogranicza się w rozważaniu rozszerzonych stanów mentalnych, do nieświadomych przekonań dyspozycyjnych, stany świadome w jego mniemaniu wciąż pozostają realizowane neuronalnie.

fleksyjnie, bez potrzeby świadomego decydowania o ruchach swych dłoni. Człowiek ten wykorzystuje w swej pracy ucieleśnioną umiejętność rządzoną zasadami, których sam nie potrafiłby jasno sformułować. W takiej sytuacji, zdaniem Malafourisa, podmiotem działania nie może być wyłącznie garncarz, gdyż procesy decyzyjne kierujące działalnością wytwarzania naczyń nie są realizowane wyłącznie przez procesy mózgowy. Są one mianowicie konstytuowane przez określone normami relacje manipulacji zachodzące pomiędzy człowiekiem i biorącym udział w działaniu artefaktem. Podmiotem zatem jest cały szeroki system złożony z garncarza, wszystkich przyrządów i materiałów, których on używa, oraz procesów zachodzących pomiędzy nimi. Stanowisko takie wymyka się intuicjom, wydaje się bowiem oczywiste, że to stany mentalne garncarza są przyczyną podjęcia działania i to on jest odpowiedzialny za jego efekt. Malafouris zauważa jednak, że intuicje te są osadzone w nie zawsze adekwatnym poczuciu sprawstwa, którego człowiek doznaje działając, mimo iż faktycznie nie on inicjuje działanie, a przynajmniej nie tylko. To prawda, że tylko człowiek jest w stanie uświadomić sobie poczucie autorstwa danej aktywności, nie ma to jednak znaczenia dla określenia faktycznego jej podmiotu (Malafouris 2008). Podmiotowość zatem nie jest istotnie związana ze zdolnością do refleksji i władania językiem pojęciowym – w tym miejscu stanowisko Malafourisa odchodzi od koncepcji Davidsona. Podmiot czynności wytwarzania naczyń jest konstytuowany przez sieć sprzężonych ze sobą procesów zachodzących zarówno w mózgu garncarza, w jego mięśniach, w jego systemach percepcyjnych (wizualnym, dotykowym), jak również przez własności koła garncarskiego decydujących o jego afordancjach, przez własności gliny, a także przez ogólny kontekst społeczny, w którym pojawia się owa aktywność (Malafouris 2008).

Malafouris zdaje sobie sprawę, jak fundamentalną zmianę wprowadza wykluczając z warunków koniecznych podmiotowości intencję jako stan mentalny będący przyczyną działania. Nie chcąc zupełnie zrezygnować z tego pojęcia zapożycza od Johna Searle'a kategorię intencji w działaniu (Searle 1995), która charakteryzuje codzienne, rutynowe czynności nie wymagające podejmowania świadomych decyzji. Tak rozumiana intencja jest, zdaniem Malafourisa nieodróżnialna od samego działania, którego podmiotem jest cały szeroki system. Ta szczególna własność, tworząca trzon podmiotowości, nie powinna być zatem rozumiana jako wewnętrzny stan mentalny, lecz jako rozproszone, emergentne zjawisko będące wytworem omawianego systemu. Wraz z intencją rozszerzona zostaje również sprawczość, autorstwo działania oraz odpowiedzialność za nie, co w oczywisty sposób prowadzi do sprzecznych z intuicją scenariuszy. Malafouris przekonuje jednak, że warto podjąć wysiłek porzucenia dualistycznych intuicji dzielących świat na „ja” i „wszystko inne”, gdyż tylko wtedy ujawni nam się prawdziwa natura podmiotowości (*agency*), która polega na działaniu w świecie, a nie na planowaniu owego działania „w głowie”.

Stanowisko Malafourisa wybrałam jako reprezentacyjne dla tezy rozszerzonego podmiotu, gdyż autor ten w oryginalny sposób stara się zmodyfikować warunki wypracowane na gruncie analitycznej filozofii działania tak, aby mogły być one spełnione przez jednoosobowy rozszerzony system poznawczy. Nie odbiera on podmiotowi istotnej intencjonalności, nie zamienia pojęcia podmiotu działania na pojęcie nieintencjonalnego agenta,⁹ nie stosuje również strategii nastawienia intencjonalnego. Malafouris przeprowadza jednak pewną zmianę pojęcia intencji, która pozwala mu na włączenie w zbiór podmiotów działania systemów rozszerzonych. Istotna dla działania intencja nie stanowi, jak u Davidsona, stanu mentalnego, który poprzedza je jako jego przyczyna. Zgodnie z tą koncepcją, w systemie złożonym z człowieka manipulującego artefaktem, działanie i jego intencja są tak silnie ze sobą sprzężone, że nie można ich odizolować, stają się właściwie nieodróżnialne. Granica pomiędzy sferą mentalną, a fizycznym artefaktem zanika, artefakt staje się częścią umysłu sprawcy. Należy jednak pamiętać, że intencja w działaniu jest wciąż stanem mentalnym z tą różnicą, że nie pojawia się przed aktywnością, lecz równocześnie z nią. Nawet jeśli stan ten jest konstytuowany, tak jak wskazuje Malafouris, przez wiele składających się na system elementów, to człowiek stanowi najważniejszy z nich i konieczny do jego zaistnienia. Zdaje się zatem, że człowiek pełni w takim systemie rolę dominującą chociażby dlatego, że ponosi każdego rodzaju odpowiedzialność za wykonane działanie (epistemiczną, moralną, prawną). Trudno również odmówić mu sprawczości w sensie podjęcia decyzji o przeprowadzeniu działania. Nawiasem mówiąc próby, które podejmuje Malafouris przywołując badania Libeta (Libet 1985), których wyniki mają pokazywać, że mózg podejmuje decyzję o działaniu bez świadomości podmiotu, nie mogą być zastosowane do złożonych działań, takich jak wyrób garnka. Podsumowując, za podmiot działania opisanego przez Malafourisa należy uznać człowieka, mimo że procesy poznawcze kierujące samym wykonaniem tej czynności mogą być rozproszone na cały rozszerzony system.

Przekonanie o konieczności poszerzenia zakresu nazwy „podmiot działania” doprowadzają do skrajności zwolennicy teorii aktora – sieci (ANT – *actor-network theory*), której twórcami są Bruno Latour, Michel Callon i John Law. Podmiotowość rozumiana jest przez tych badaczy jako rozproszona na sieć złożoną z ludzi, artefaktów i interakcji pomiędzy nimi, w której każdy element jest równie ważny, a człowiek nie jest wyróżniony tylko ze względu na swoje człowieczeństwo (Law 1999). Podmiotem (*agent*), nazywanym również aktorem lub aktantem, może być zatem element sieci lub jej emergentny wytwór. Jest nim, jednym słowem, wszystko, co działa, a człowiek w tym zbiorze nie posiada żadnej wyróżnionej pozycji. Połączeni

⁹ W tym przypadku angielskim odpowiednikiem „podmiotu działania” jest „agent”, natomiast odpowiednikiem „agenta” może być „relatywnie autonomiczny system przetwarzający informacje”.

dynamicznymi relacjami aktorzy wytwarzają nowy działający obiekt, czyli nowego aktora – sieć (Abriszewski 2008). Pytanie, które w tym miejscu się narzuca, brzmi, jak jest rozumiane w tej teorii działanie (*agency*). Standardowo, działanie jest czynnością spowodowaną przez intencję podmiotu, gdzie intencja jest stanem mentalnym posiadającym treść. Działający jest zatem sprawcą, sprawia swoją wolą, że coś się dzieje. W ramach ANT działanie jest przypisywane dużo szerszej grupie obiektów. Wszystko, co wchodzi w relacje z innymi obiektami, co modyfikuje i kształtuje działanie innych aktorów, samo stanowi podmiot działania (Latour 2013). Działanie nie wymaga zatem refleksji, aktorzy do niej zdolni nie mają przewagi nad innymi obiektami organizującymi przestrzeń wokół siebie w taki sposób, aby efektywnie realizować cele. Podmiot nie musi być zatem sprawcą, wystarczy, że jest podmiotem czynu, zachowania, ruchu, lecz niekoniecznie działania. Bruno Latour zwraca uwagę, że technologia tak silnie kształtuje nasze decyzje i wszelką działalność w świecie, że bezowocne staje się wyróżnianie człowieka jako podmiotu owych decyzji i intencji działania. Istnieją po prostu aktorzy, z których część jest ludzka, a część nie, własność człowieczeństwa jest przy tym nieistotna, liczy się wyłącznie sieć, w której dokonuje się działanie poprzez rozproszone umiejętności, procesy poznawcze i cele (Latour 1992). Intencja, konieczna dla działania w stanowisku Malafourisa, chociaż rozproszona na cały szeroki system, traci w ANT znaczenie. Malafouris rozważa system, który zawiera w sobie człowieka i bez którego intencja, mimo, że rozproszona i konstytuowana przez wiele innych elementów, nie mogłaby się pojawić. Człowiek w takim systemie wydaje się być koniecznym, chociaż niewystarczającym warunkiem podmiotowości poznawczej. W ANT człowiek nie pełni już tak ważnej roli, gdyż działanie nie wymaga żadnej specyficznie ludzkiej cechy. Teoria ta stanowi, moim zdaniem, najbardziej radykalne odejście od rozumienia podmiotu wypracowanego w analitycznej filozofii działania, jak również od potocznych intuicji.

Dwa kolejne głosy za potrzebą rozszerzenia podmiotowości poza człowieka w niektórych systemach rozszerzonych stanowią komentarz i modyfikację myśli Clarka. Richard Menary i Edwin Hutchins zauważają, że autor ten niesłusznie przyznaje rolę podmiotu działania w takich systemach ludzkiemu organizmowi, jego stanowisko jest, innymi słowy zbyt internalistyczne. Pierwszy tych krytyków, zarzuca Clarkowi, że nie wziął pod uwagę dwóch ważnych cech rozszerzonych procesów poznawczych. Po pierwsze, procesy takie przekształcają ludzkie naturalne zdolności poznawcze w tak dużym stopniu, że ich wyróżnianie jako wewnętrznych i istotnie poznawczych jest arbitralne i bezowocne. Po drugie, manipulowanie artefaktami, nazywane przez Menary'ego praktykami poznawczymi, ma ważny wymiar normatywny związany z koniecznością przestrzegania odpowiednich reguł. Rozszerzony proces nie jest zatem poznawczy dlatego, że jego elementy organiczne i środowiskowe są odpowiednio sprzężone, lecz dlatego, że stanowi normatywną

praktykę polegającą na manipulowaniu artefaktem i prowadzącą do rozwiązania problemu poznawczego. Jest on rozproszony na cały szeroki system, który tworzy niepodzielną całość stanowiącą podmiot podejmowanej przez niego aktywności poznawczej (Menary 2007, 50–64). Należy zatem porzucić założenie, że istnieje cieleśnie ograniczony podmiot, którego wewnętrzne procesy poznawcze rozszerzają się, lub nie poprzez tworzenie i wykorzystywanie przez niego zewnętrznych reprezentacji. Pojęcie umysłu rozszerzonego jest, zdaniem Menary'ego, zwodnicze i niepoprawnie określa omawiane przez niego zjawisko. Umysł podmiotu nie rozszerza się w niektórych sytuacjach poznawczych, gdyż nigdy nie był on nierozszerzony. Podmiot działania, stany mentalne odpowiedzialne za podmiotowość zawsze są konstytuowane przez praktyki poznawcze polegające na manipulowaniu zewnętrznymi reprezentacjami, gdyż zawsze rządzą nimi normy zakorzenione w kulturze i stosunkach społecznych (Menary 2012, 151–152).

Takie holistyczne ujęcie podmiotu działania budzi jednak wątpliwość ze względu na pojęcie manipulacji, wokół którego Menary buduje swoją wersję tezy poznania rozszerzonego. Wydaje się, że w systemie złożonym z człowieka manipulującego artefaktem, część podmiotowa jest jednoznacznie określona, gdyż relacja manipulacji zakłada istnienie dwóch elementów – sprawczego, który manipuluje oraz pasywnego, manipulowanego. Można zgodzić się z omawianym autorem, że treść stanów mentalnych odpowiedzialnych za decyzję o takim a nie innym działaniu jest zdeterminowana przez praktyki kultowe, stany te jednak należą do podmiotu, którym jest człowiek, to on bowiem stanowi element aktywny i inicjujący w rozszerzonym systemie.

Hutchins również zauważa u Clarka brak konsekwencji w zacieraniu różnic funkcjonalnych pomiędzy procesami zachodzącymi w ciele człowieka a tymi, które poza niego wykraczają (Clark 2008). Traktując organizm ludzki jako centrum dowodzenia, autor *Supersizing the Mind* izoluje go ze środowiska społeczno-kulturowego i ignoruje moc, z jaką kształtuje ono ludzką działalność poznawczą. Hutchins tymczasem jest zwolennikiem idei podmiotu rozproszonego, którą w książce *Cognition in the Wild* ilustruje przykładem działania polegającego na nawigacji statku podczas jego cumowania w porcie (Hutchins 1995). Cały proces nawigacji jest tak skomplikowany, że musi być przeprowadzany przez szeroki system poznawczy złożony z ludzi i artefaktów połączonych siecią wzajemnych relacji. Kim jest w takiej sytuacji podmiot owego działania? Czy jest tyle podmiotów, ile ludzi zaangażowanych w to zadanie, czy też jest jeden podmiot – szeroki system, w którym proces poznawczy prowadzący do wykonania zadania jest rozproszony na wszystkie jego części? Wybierając drugą odpowiedź Hutchins przekonuje, że podmiotowość w tym przypadku nie jest rozproszona tylko na ludzi biorących udział w tym zadaniu, lecz również na artefakty, którymi manipulują, a także na kontekst społeczny, który ową manipulacją rządzi. Wszystkie te elementy składają się na szeroki system poznawczy, w którym procesy kon-

trone i decyzyjne odpowiedzialne za podmiotowość są zdecentralizowane i rozproszone (Hutchins 2011). Analiza tej sytuacji jako przejawu działania jednego podmiotu wykracza jednak poza tematykę tej pracy, gdyż dotyczy podmiotu grupowego, na który składa się wiele podmiotów intencjonalnych. Problematykę podmiotowości tego typu systemów podejmują w innych moich pracach (Trybulec 2017a; 2018), w tym miejscu natomiast pragnę skupić się na rozszerzonym systemie jednoosobowym. Zdaniem Hutchinsa, jeśli chodzi o przypisywanie podmiotowości, nie zachodzi znacząca różnica pomiędzy rozszerzonym podmiotem jednoosobowym, a wieloosobowym. Komentując przedstawioną przez Clarka sytuację Ottona autor ten wskazuje, że umiejętność posługiwania się notatnikiem jest ukształtowana przez praktyki kulturowe, które konstytuują takie stany mentalne jak intencja działania i poczucie kontroli stanowiące o podmiotowości rozszerzonego systemu. Człowiek nie powinien zatem być traktowany jako wyizolowany ze środowiska społecznego aktywny element szerokiego systemu manipulujący innym również wyizolowanym, ale tym razem biernym elementem. Oba obiekty są wytworem praktyk kulturowych i oba przejawiają aktywność; stanowią bowiem sprzężoną, dynamiczną jedność (Hutchins 2011).

Modyfikacja pierwotnej tezy umysłu rozszerzonego, wprowadzona przez Hutchinsa, budzi podobne wątpliwości jak propozycja Menary'ego. Argumentując przeciwko internalizmowi Hutchins wskazuje, że procesy mózgowie nie są wystarczające dla realizacji procesów poznawczych, decydującą rolę odgrywają w niej bowiem również interakcje z artefaktami i innymi ludźmi, którymi rządzą normatywne praktyki kulturowe (Hutchins 2011). Warto jednak zauważyć, że wskazywanie na procesy neurologiczne i interakcje społeczne jako współtworzące realizator procesów poznawczych jest ryzykowne, gdyż grozi błędem kategoriałnym. Praktyki kulturowe są mianowicie realizowane przez stany mentalne i procesy poznawcze indywidualnych ludzi oraz przez relacje zachodzące pomiędzy nimi, gdy współdziałają oni ze sobą. Procesy mentalne, z kolei, są realizowane przez procesy neurologiczne przynajmniej częściowo, jeśli nie w całości. Elementy, które mają wspólnie – według Hutchinsa – realizować procesy poznawcze należą zatem do różnych poziomów rzeczywistości, co wprowadza niejasności w dyskusję nad podmiotowością systemu. Można zgodzić się z omawianym autorem, że człowiek nie stanowi jedyne aktywnego elementu w szerokim systemie poznawczym, jeśli ma się na myśli to, że inne elementy systemu kształtują stany mentalne decydujące o podejmowanym działaniu. Jeśli jednak rozumie się aktywność jako sprawczość i odpowiedzialność, to przynależy ona wyłącznie elementowi refleksyjnemu systemu, w tym wypadku człowiekowi.

Na gruncie polskiej filozofii działania, na rolę czynników społecznych w konstytuowaniu podmiotowości systemu poznawczego zwraca uwagę Katarzyna Paprzycka. Chociaż autorka ta nie głosi tezy podmiotu rozszerzonego, w jej argumentacji można odnaleźć wątek wskazujący na możliwość jego

zaistnienia. Podkreśla ona, że podmiot działający jest zanurzony w relacje z innymi ludźmi, a czynnikiem, który często istotnie wpływa na podejmowane przez niego decyzje, są ich normatywne oczekiwania. Gdy decyzją człowieka kieruje zobowiązanie zewnętrzne, którego źródłem nie jest on sam, lecz normy społeczne rządzące relacjami międzyludzkimi, należy uznać inicjatywę sprawczą za rozszerzoną poza jego organizm (Paprzycka 2010). Jako przykład takich sytuacji podaje Paprzycka zaniechania, które zachodzą wtedy, gdy podmiot nie spełnia oczekiwań innych ludzi. W przeciwieństwie do działań, które są skutkiem intencji danej jednostki, zaniechania są spowodowane przez stany mentalne innych podmiotów. Paprzycka przekonuje, że do wyjaśnienia takich przypadków konieczne jest odrzucenie kartezjańskiego rozumienia podmiotu jako autonomicznej, oderwanej od świata substancji i ujęcie go jako zanurzonego w relacje z innymi podmiotami, których normatywne oczekiwania determinują jego działania i wpływają na ich ocenę (Paprzycka 2010).¹⁰ Zauważyć należy jednak, że takie działanie bez intencji, którym jest zaniechanie, podjąć może jedynie system poznawczy, który ma możliwość uświadomienia sobie tego, że nie spełnia oczekiwań mimo, że leży to w jego mocy. Jeden z elementów takiego systemu musi zatem wykazywać się refleksyjnością, co czyni z człowieka konieczny warunek podmiotowości systemu.

Menary i Hutchins argumentują, że ludzkie stany mentalne decydujące o podmiotowości rozszerzonego systemu są realizowane nie tylko przez procesy mózgowo, lecz również przez czynniki zachodzące poza organizmem człowieka. Mimo że procesy neurologiczne są warunkiem koniecznym podmiotowości, nie stanowią one warunku wystarczającego. Na gruncie przytoczonych koncepcji wystarczy to, aby status podmiotu przypisać całemu systemowi, a nie tylko jednej z jego części. Wstępną ocenę tej tezy już tu przedstawiłam; uściślię ją i podsumuję w końcowej części artykułu. Teraz natomiast przedstawię dwie propozycje podważające konieczność czynnika ludzkiego dla ustanowienia podmiotowości i nadające status podmiotu działania samemu artefaktom.

6. PODMIOTOWOŚĆ SYSTEMÓW SZTUCZNYCH

Filozofowie, których stanowiska przywołuję w tej części artykułu, traktują niektóre artefakty będące systemami jako podmioty moralności oraz odpowiedzialności prawnej, czyli jako specjalnego rodzaju podmioty działania. Zdawałoby się zatem, że przypisują im cechy, które są swoiste dla istoty refleksyjnej, przede wszystkim świadomość wolnego wyboru zgodnego (bądź nie) z posiadaną hierarchią wartości. Tak jednak nie jest. Luciano Floridi

¹⁰ Przykład zaniechania, do którego odwołuje się Paprzycka, pojawia się w sytuacji wymyślonej przez Joshua Knobe'a w celu przedstawienia tzw. efektu Knobe'a (Knobe 2002).

i J. W. Sanders uznają za podmiot działania każdy system, któremu z odpowiedniego poziomu abstrakcji można przypisać trzy cechy: interaktywność, autonomię i adaptacyjność (Floridi, Sanders 2004). Pierwsza z nich oznacza możliwość zmiany stanu systemu w reakcji na bodźce, druga, możliwość takiej zmiany bez konieczności stymulacji, a trzecia – zdolność do zmiany reguł zachodzenia zmian. System przejawiający wszystkie te cechy ma dostęp do własnych zasad funkcjonowania i może wprowadzać w nich zmiany w odpowiedzi na warunki środowiskowe (Floridi, Sanders 2004, 357–358). Sugeruje to pewną formę refleksyjności rozumianą jednak nie jako świadomość treści własnych stanów mentalnych, lecz jako możliwość ich oceny i modyfikacji pod kątem ich zgodności z zasadami racjonalności. Floridi i Sanders podkreślają, że system aspirujący do miana podmiotu działania przejawia niezbędną ku temu autonomiczność nie w sposób obiektywny, lecz tylko na odpowiednim poziomie abstrakcji, na którym stworzony jest jego model zakładający istnienie tej własności. Ocena danego systemu pod względem podmiotowości zakłada zatem wcześniejszy wybór odpowiedniego jego modelu (Floridi, Sanders 2004, 360–362). Przypomina to postulowaną przez Dennetta strategię nastawienia intencjonalnego, która umożliwi nadanie podmiotowości każdemu wystarczająco złożonemu systemowi. Nawet taka cecha jak wrażliwość na wartości, która ze względu na konieczność posługiwania się językiem pojęciowym przysługuje, według Davidsona tylko człowiekowi, może, zgodnie z omawianą koncepcją zostać przypisana artefaktowi, jeśli stworzy się odpowiedni jego model. Wystarczy mianowicie, że opisuje on system jako spełniający trzy warunki podmiotowości, a jego działanie jako prowadzące do moralnego dobra lub zła.¹¹

Najpoważniejszy zarzut wobec stanowiska Floridiego i Sandersa dotyczy braku realnej intencjonalności, czyli rzeczywistych stanów mentalnych będących przyczynami wolnej decyzji o działaniu. Omawiani autorzy argumentują jednak, że określanie, które systemy faktycznie posiadają stany mentalne, a które nie, jest filozoficzną spekulacją, gdyż nie istnieje możliwość obiektywnego rozstrzygnięcia tego problemu. Dostęp do faktycznych stanów mentalnych jest wyłącznie pierwszoosobowy, decyzja o przypisaniu systemowi podmiotowości powinna być jednak oparta na obserwowalnych faktach, z tego też względu należy skupić się wyłącznie na jego zachowaniu. Podmiotem działań moralnych, czyli takich, które mają konsekwencje oceniane jako moralnie dobre lub złe, jest zdaniem Floridiego i Sandersa ten system, który bierze udział w tak zwanej grze w moralność (*moral game*), niezależnie czy jest tego świadomy, czy nie. Zdaje się jednak, że świadomość owej „gry” jest niezbędna w celu obarczenia systemu odpowiedzialnością,

¹¹ Przykładem takiego systemu może być program stworzony w celu diagnozowania chorób, planowania terapii, oraz monitorowania i modyfikowania jej przebiegu. Jego działanie prowadzi do moralnego dobra (zdrowia pacjenta), lub zła (dalszej choroby lub śmierci) (Floridi, Sanders 2004, 364).

która jest nierozzerwalnie złączona z podmiotowością moralną. Funkcjonowanie sztucznego systemu może prowadzić do skutków ocenianych z perspektywy moralnej, nie znaczy to jednak, że on sam jest podmiotem działania moralnego. Stałby się nim dopiero wtedy, gdyby mógł zostać obciążony odpowiedzialnością za ten skutek. Odpowiedzialny za swoje działanie może być jednak tylko artefakt, który ma możliwość popełnienia błędu, za który on sam ponosi winę (przynajmniej w przeważającej części). Wysiłek, aby go nie popełnić, bywa ceniony jako osiągnięcie i nagradzany, a jego popełnienie, czy to z premedytacją, czy z zaniechania, skutkuje porażką i naganą. To, czy działanie systemu prowadzi do moralnego dobra czy zła, jest uzależnione od hierarchii wartości, którą się przyjmuje do jego oceny. Działanie moralnie naganne może przeprowadzić tylko system, który realizuje wartość niezgodną z hierarchią obowiązującą w danej społeczności lub ze swoją własną, i ponadto powinien wiedzieć o tym odstępstwie (ma taką poznawczą możliwość). Wszystkie te warunki może spełnić tylko system o bardzo złożonych stanach reprezentacyjnych, wśród których znajdują się postawy propozycyjne nie tylko odpowiadające przekonaniom, lecz również pragnieniom, wątpliwościom, oczekiwaniom, czy lękom. Bez rozumienia znaczeń języka pojęciowego, a być może również bez możliwości przeżywania emocji, system nie jest w stanie spełnić warunków podmiotowości moralnej.

Aby utrzymać koncepcję sztucznych systemów jako podmiotów moralności, Floridi i Sanders rozrywają ścisły związek moralności z odpowiedzialnością. Wskazują przy tym na praktyki traktowania dzieci, zwierząt, lub niektórych artefaktów jako systemów pomnażających dobro lub zło, mimo że nie odpowiadają one za swoje czyny (Floridi, Sanders 2004, 369). Omawiani autorzy mierzą się również z powszechną intuicją skłaniającą do obarczania odpowiedzialnością za działanie artefaktu jego twórcę. Ich zdaniem, nie nadąża ona za współczesnym rozwojem technologicznym, gdzie bardzo wiele czynników wpływających na funkcjonowanie, na przykład oprogramowania, leży poza kontrolą pojedynczego człowieka a nawet grupy ludzi. Problem odpowiedzialności za efekty działania złożonych programów staje się współcześnie coraz poważniejszy i wymaga porzucenia antropocentrycznych przekonań na temat podmiotowości (Floridi, Sanders 2004, 371–372). Kierowanie się takim zaleceniem pozwala dostrzec sprawcę działania poza człowiekiem, sprawcę w znaczeniu odpowiedzialnej jednostki uczestniczącej w życiu społecznym na równych człowiekowi prawach.

Na potrzebę dostrzeżenia podmiotowości poza człowiekiem zwracają również uwagę badacze skupieni na aspekcie prawnym funkcjonowania artefaktów. Samir Chopra i Laurence F. White postulują nadanie niektórym programom komputerowym osobowości prawnej, gdyż w niektórych sytuacjach nie ma możliwości obciążenia odpowiedzialnością człowieka lub grupy ludzi (Chopra, White 2011). Stosują przy tym bardzo szerokie pojęcie podmiotu działania jako systemu autonomicznie przetwarzającego informacje

w celu wykonania danego zadania.¹² Autonomiczność rozumieją przy tym jako zdolność do funkcjonowania bez bezpośredniej kontroli człowieka lub innych systemów, oraz do samodzielnego kontrolowania własnych działań i stanów. Cecha ta związana jest również ze społeczną zdolnością do wchodzenia w relacje z innymi podmiotami, a także z możliwością inicjowania działania nakierowanego na określony cel nawet, jeśli jest on zlecony przez programistę (Chopra, White 2011, 10). Podobnie jak na gruncie stanowiska Floridego i Sandersa, omawiana propozycja nie zakłada konieczności posiadania przez system, uznany za podmiot działania, intencji poprzedzonej rozumowaniem i wolną decyzją. Taka sprawczość bez intencji, charakterystyczna dla nauk komputerowych i ekonomicznych, przeczy zarówno potocznym jak i filozoficznym intuicjom. Zdaniem Chopra i White'a, jest to jednak cena konieczna do zapłacenia wzięwszy pod uwagę rozwój technologiczny. Aby rozjaśnić potrzebę zmiany pojęcia podmiotu działania na bardziej liberalną (mniej antropocentryczną), omawiani autorzy podają konkretny przykład. Pytają mianowicie, kto stanowi stronę kontraktu w umowach kupna-sprzedaży za pośrednictwem sklepu internetowego w przypadku, gdy żadna osoba po stronie sprzedającego nie zna jego dokładnych warunków i nie może świadomie wyrazić na nie zgody (Chopra, White 2011, 30). Dostępność towaru, jego cena, proponowana obniżka i inne bonusy są często dostosowywane do konkretnego klienta przez program w sposób będący poza bezpośrednią kontrolą człowieka. Prognozuje się, że w przyszłości będą istnieć programy negocjujące na bieżąco warunki sprzedaży. Procesem tym rządzi oczywiście odpowiedni algorytm stworzony przez programistę, jednak ani ów programista, ani jego pracodawca nie są w stanie kontrolować warunków proponowanych przez program każdemu klientowi. W takim wypadku jednak nie mogą oni stanowić strony ważnej umowy, nie można im bowiem przypisać intencji jej zawarcia. Z kim w takim razie klient zawiera umowę? W praktyce, rzecz jasna odpowiedzialnością za wszystkie zawierane transakcje obarczane są konkretne osoby, które w pełni się na to zgadzają, zjawisko to jednak nie ma teoretycznego uzasadnienia, gdyż nie podjęli oni w przypadku każdej transakcji decyzji, za którą można byłoby im przypisać zasługę lub obarczyć ich winą. Teoretycznie stroną umowy powinien stanowić sztuczny system określający jej warunki, a nie człowiek zlecający mu to zadanie i określający ogólne zasady jego wykonania. Pozostaje jednak problem intencji zawarcia kontraktu, która jest konieczna dla jego ważności, ponieważ tylko wtedy mamy do czynienia z podmiotem, którego można obarczyć odpowiedzialnością, w tym wypadku prawną.

¹² Przykładem takiego programu jest oprogramowanie sklepu internetowego oferujące elastyczne warunki sprzedaży, lub program przeprowadzający aukcje internetowe i pośredniczący w zawarciu umowy pomiędzy stronami.

Chopra i White zdają sobie sprawę, że aby zyskać status podmiotu działania, sztuczny system poznawczy musi spełnić kryterium intencjonalności sformułowane między innymi przez Davidsona. Interpretują jednak to kryterium następująco: “If artificial agents are to be viewed as possessing agency in the intentional sense, then it should be possible *to view them as taking actions whose causes are their own beliefs and desires*” (Chopra, White 2011, 12). Podkreśliłam w tym cytacie sformułowanie, które wyraźnie wskazuje na zastosowanie przez autorów strategii nastawienia intencjonalnego. W kolejnym zdaniu swojej pracy autorzy ci odwołują się już bezpośrednio do zalecenia Dennetta, aby traktować sztuczny system jako podmiot intencjonalny, jeśli tylko takie jego rozumienie ułatwia wyjaśnianie i przewidywanie jego zachowania (Dennett 1987). Im bardziej zachowanie artefaktu jest złożone, im trudniej jest wyjaśnić je w odwołaniu do funkcjonowania jego mechanizmu, tym bardziej rozsądne jest przypisywanie mu stanów mentalnych. Strategia ta jest szczególnie użyteczna w chronionych prawem umowach, gdzie jedną ze stron stanowi artefakt (Chopra, White, 39–43).

Analizy przeprowadzone w tym i w poprzednim paragrafie wyraźnie wskazują na istnienie w dyskusji nad podmiotowością rozszerzonych systemów poznawczych dwóch pojęć sprawczości, które niekiedy są ze sobą mylone. Pierwsze zostało wypracowane na gruncie analitycznej filozofii podmiotu i określa system posiadający reprezentacje o niepochođnej treści tworzące złożoną sieć postaw propozycjonalnych będących faktyczną przyczyną konkretnych działań. Drugie pojęcie sprawczości określa systemy, które mogą być w pewnych kontekstach traktowane jako podmioty działania, a w innych nie, w zależności od tego z jakiej perspektywy się je spostrzeżga, lub jaki ich model aktualnie się przyjmuje. Aby jasno odróżniać systemy, do których stosowane są te dwa rodzaje sprawczości, należy konsekwentnie stosować do ich oznaczania odpowiednie nazwy. „Podmiot działania” mianowicie to system faktycznie spełniający warunki podmiotowości Davidsona, system, który może stać się „pełnokrwistym” podmiotem każdego rodzaju odpowiedzialności. „Agent” natomiast to każdy relatywnie autonomiczny obiekt przetwarzający informację w celu wykonania zadania poznawczego. Ta druga nazwa jest coraz częściej stosowana na gruncie polskojęzycznej kognitywistyki ze względu na szybko rozwijającą się technologię wzmocnień poznawczych i systemów sztucznej inteligencji. Już teraz istnieją artefakty wykonujące złożone czynności poznawcze prowadzące do skutków znaczących pod względem moralnym i prawnym, i z pewnością będą one coraz intensywniej wchodzić w interakcje z człowiekiem. Przeprowadzając refleksję nad ich podmiotowością należy przede wszystkim ocenić, czy przysługują im cechy umożliwiające obarczenie ich odpowiedniego rodzaju odpowiedzialnością, a co za tym idzie przypisanie im konkretnych praw i obowiązków.

6. WNIOSKI

Propozycje rozszerzania podmiotowości poza człowieka, którym poświęciłam ten artykuł, są motywowane rozwojem artefaktów stanowiących wzmocnienia poznawcze. Zarówno systemy, których funkcjonowanie wymaga ciągłego sprzężenia z człowiekiem, jak i takie, które posiadają pewien stopień autonomii, wkraczają coraz wyraźniej na arenę życia społecznego niosąc ze sobą wiele teoretycznych oraz praktycznych pytań. W artykule przedstawiłam wybrane próby zastosowania pojęcia podmiotu działania do nowego zjawiska, jakim są rozszerzone i sztuczne systemy poznawcze. Jako pierwszy cel wyznaczyłam sobie ocenę tych prób z perspektywy dwóch dominujących we współczesnej filozofii ujęć podmiotu działania: jako istoty refleksyjnej oraz jako przedmiotu nastawienia intencjonalnego. Reprezentantem pierwszego stanowiska uczyniłam Davidsona, zaś drugiego – Dennetta. Z jednej strony, koncepcje te znacznie się od siebie różnią, z drugiej jednak, nie jest łatwo wskazać na to, co odróżnia podmiot istotnie intencjonalny od tylko postrzeganego jako taki z odpowiedniej perspektywy. Jaki system mianowicie uznać za refleksyjny, a jaki nie, skoro własności o tym decydujące są widoczne tylko z pierwszoosobowego punktu widzenia? Za Floridim i Sandersem można stwierdzić, że każda decyzja w tej sprawie oparta jest wyłącznie na filozoficznej spekulacji, a nie na obserwowalnych faktach. Ta teoretyczna trudność uzasadnienia specyfiki ludzkiej sprawczości, oraz coraz wyraźniej dostrzegalna obecność artefaktów wykonujących zadania poznawcze skłoniła autorów, wskazanych przez mnie w punkcie piątym i szóstym, do postawienia, odpowiednio, tez podmiotu rozszerzonego i sztucznego. Należy zdecydowanie podkreślić, że są to tezy o odmiennej treści prowadzące do różnych teoretycznych i praktycznych konsekwencji. Teza podmiotu rozszerzonego, którą przypisałam Malafouris'owi, Menary'emu i Hutchinsowi, wyrasta z przekonania o konstytutywnym wpływie artefaktów i innych elementów środowiskowych, które można zebrać pod pojęciem praktyk kulturowych (w tym poznawczych), na własności systemu odpowiedzialne za inicjatywę sprawczą. Własności te, a chodzi tu o postawy propozycyjne takie jak przekonania, pragnienia i intencje, realizowane są, innymi słowy nie tylko przez procesy zachodzące w organizmie, lecz również przez takie, które poza niego wykraczają. Skoro tak, to dlaczego ograniczać podmiotowość tylko do człowieka, jeśli tworzą ją procesy rozproszone na cały szeroki system? Jest tak dlatego, brzmi moja odpowiedź, że z działaniem wiąże się w sposób nierozdzielny odpowiedzialność. W zależności od tego, jaką wartość dane działanie ma realizować (epistemiczną, moralną, estetyczną, itd.), takim rodzajem odpowiedzialności obarczony jest jego podmiot. Działanie jest normatywne, gdyż zawsze stanowi próbę realizacji celu, który jest wartościowy dla jego podmiotu, i które może się zakończyć nagradzonym osiągnięciem lub ganioną porażką. Odpowiedzialny za swoje

funkcjonowanie może być jednak tylko system refleksyjny posiadający reprezentacje o zrozumiałej dla niego treści, nazywaną w filozofii treścią nie-pochodną (Adams, Aizawa 2005).

W mojej ocenie, żaden z przywołanych w tym tekście autorów koncepcji podmiotu rozszerzonego nie wykazuje, że podmiotem działania może być rozszerzony system poznawczy. Ocenę tę uzasadniam koniecznym powiązaniem podmiotowości z odpowiedzialnością, lecz ten praktyczny wymiar dyskusji nad sprawczością systemów rozszerzonych i sztucznych ma również oparcie w rozwiązaniach teoretycznych. Rozproszenie realizatora odpowiedzialnych za działanie stanów intencjonalnych na cały rozszerzony system nie prowadzi mianowicie do rozproszenia ich podmiotu. Tak, jak już wspomniałam, podmiotem działania, a tym samym odpowiedzialności może być jedynie system refleksyjny. Czy jest jednak możliwe, aby realizator podmiotowości był rozproszony, a sama podmiotowość nie? W wyjaśnieniu tej kwestii z pomocą przychodzi rozwiązanie zaproponowane przez Lynne R. Baker (Baker 2009). Dotyczy ono co prawda rozumienia jaźni w rozszerzonych systemach poznawczych, jednak nie dokonam zbyt dużego nadużycia, jeśli zastosuję tę propozycję do systemu refleksyjnego. Baker odwołuje się mianowicie do podziału rzeczywistości na poziomy dokonanego na gruncie nieredukcyjnego fizykalizmu. Stany mentalne, zgodnie z tym ujęciem należą do własności wyższego rzędu systemu poznawczego i powstają z niższego rzędu własności fizycznych. Oba rodzaje własności posiadają wzajemnie różne charakterystyki. Własności fizyczne, w przeciwieństwie do mentalnych, zachodzą w przestrzeni – w organizmie lub poza nim. Własności systemowe, takie jak podmiotowość, nie zajmują natomiast miejsca, nie można zatem określić, czy znajdują się w organizmie, czy poza nim. Zdaniem Baker, fakt, że środowisko społeczne, językowe i fizyczne odgrywa istotną rolę w kształtowaniu decyzji o działaniu, nie oznacza, że ich podmiot rozszerza się w jakikolwiek sposób. Innymi słowy, subosobowe stany podmiotu działania mogą częściowo składać się z pozabiologicznych elementów, które wchodząc ze sobą w złożone relacje przyczynowe wytwarzają własności systemowe takie jak przekonania, pragnienia i inne stany mentalne. Fizyczny realizator podmiotowości jest w takim wypadku rozszerzony, lecz sam podmiot nie, określenie to stosuje się bowiem wyłącznie do własności fizycznych. Wniosek ten zdaje się podważać samą tezę umysłu rozszerzonego Clarka i Chalmersa, tego tematu nie będę jednak w tym miejscu podejmować.

Podmiotowość systemów całkowicie sztucznych jest obecnie tematem dużo częściej podejmowanym przez filozofów i kognitywistów. Jak pokazałam w punkcie szóstym, nie tylko podawane są konkretne przykłady takich systemów, lecz również padają argumenty za teoretyczną możliwością i praktyczną korzyścią przypisywania im statusu podmiotu wraz ze związanymi z tym prawami i obowiązkami. Teoretycznym uzasadnieniem możliwości zaistnienia sztucznych podmiotów działania jest zdaniem omówionych auto-

ród strategią nastawienia intencjonalnego. W mojej ocenie nie spełnia ona jednak tego zadania. Mimo, że stanowi efektywne narzędzie obiektywnego wyjaśniania i przewidywania zachowań złożonych artefaktów, nie uzasadnia możliwości obarczenia ich odpowiedzialnością. Należy zauważyć, że aby uznać artefakt za system refleksyjny trzeba przypisać mu dużo bardziej złożoną umysłowość niż posiadanie stanów intencjonalnych wyjaśniających jego zachowanie. Osoba stosująca wobec takiego systemu strategię nastawienia intencjonalnego musiałaby obdarzyć go umiejętnością oceny owych stanów pod względem racjonalności i zgodności z przyjętą przez niego hierarchią wartości, a także możliwością porównania hierarchii przyjętej przez ten system z panującą w danej społeczności i refleksji nad ryzykiem ewentualnego jej naruszenia. Co za tym idzie, system taki musiałby być postrzegany jako dokonujący wyboru pomiędzy różnymi możliwymi działaniami na podstawie posiadanych racji, a zatem jako posiadający dostęp do treści swoich stanów intencjonalnych, oraz jako świadomy tego, że jest ich podmiotem i sprawuje nad nimi kontrolę. Zdaje się, że na obecnym etapie rozwoju technologicznego byłoby to poważne nadużycie narzędzia zaproponowanego przez Dennetta, mogące prowadzić do zbyt dużego rozděcia sprawczego. Traktowanie artefaktów jako sztucznych podmiotów działania nie jest uzasadnione podobieństwem do ich naturalnych pierwowzorów; systemy te jeszcze nie przypominają bowiem człowieka ani zachowaniem, ani złożonością funkcjonalną realizatora procesów przetwarzania informacyjnego.

Drugi cel tej pracy dotyczył uzasadnienia potrzeby wyraźnego odróżnienia podmiotu działania od agenta w analizach i badaniach rozszerzonych i sztucznych systemów poznawczych. Potrzeba ta wynika z różnic pomiędzy podmiotem działania opisanym przez Davidsona a systemem stanowiącym przedmiot nastawienia intencjonalnego. Na gruncie zarówno polskojęzycznej jak i anglojęzycznej kognitywistyki stosuje się często nazwę „agent” na oznaczenie relatywnie autonomicznego artefaktu przetwarzającego informacje w celu rozwiązania zadania poznawczego. Mimo, że nazwa ta w języku polskim nie ma tak silnych konotacji ze sprawczością jak w języku angielskim, należy zdecydowanie odróżnić jej desygnat od desygnatu nazwy „podmiot działania”, której znaczenie wskazuje na istotną, „pełnokrwistą” intencjonalność związaną z wrażliwością na wartości i odpowiedzialnością. Każdy podmiot działania jest agentem, lecz nie odwrotnie. Jako odpowiednik polskiej nazwy „agent” w języku angielskim można zaproponować nazwę „relatywnie autonomiczny system przetwarzający informację w celu wykonania zadania poznawczego”, odpowiednikiem nazwy „podmiot działania” jest, rzecz jasna, angielska nazwa „agent”. Nazwa „agent” stosowana w języku polskim jest bardzo użyteczna, a wręcz niezbędna w dobie rozwoju technologicznych wzmocnień poznawczych i coraz bardziej autonomicznej sztucznej inteligencji. Trudno odmówić jej desygnatom pewnego rodzaju sprawczości, jednak jest to inna sprawczość od tej, która cechuje podmioty

działania. Świadomość istnienia tych dwóch rodzajów sprawczości wprowadza w badaniach kognitywistycznych porządek pojęciowy, który może przeciwdziałać nadużyciom. Rozróżnienie to pozwala również określić różnice pomiędzy człowiekiem a innymi systemami poznawczymi, co stanowi niezbędny punkt startu dla prób stworzenia sztucznych podmiotów działania.

BIBLIOGRAFIA

- K. Abriszewski, *Poznanie, zbiorowość, polityka. Analiza Teorii aktora-sieci Bruno Latoura*, Universitas, Kraków 2008.
- F. Adams, K. Aizawa, *Defending Non-derived Content*, *Philosophical Psychology*, 18, 2005, 661–669.
- F. Adams, K. Aizawa, *The Bounds of Cognition*, Wiley-Blackwell, Oxford 2010.
- L. R. Baker, *Instrumental Intentionality*, *Philosophy of Science*, 56 (2), 1989, 303–316.
- L. R. Baker, *Persons and the Extended Mind Thesis*, *Zygon*, 44 (3), 2009, 642–658.
- L. von Bertalanffy, *Ogólna teoria systemów. Podstawy, rozwój, zastosowania*, PWN, Warszawa 1984.
- J. Bremer, *Naukowy i potoczny obraz osoby w kognitywistyce*, Wydawnictwo WAM, Kraków 2016.
- A. Chemero, *Radical Embodied Cognitive Science*, MIT, Cambridge (Mass.) 2009.
- R. Chisholm, *Human Freedom and the Self. The Lindley Lecture*, Department of Philosophy, University of Kansas 1964.
- A. Clark, D. Chalmers, *Umysł rozszerzony*, w: M. Miłkowski, R. Poczobut (red.) *Analityczna metafizyka umysłu. Najnowsze kontrowersje*, Wydawnictwo IFiS PAN, Warszawa 2008, 342–357. Oryginał: A. Clark, D. Chalmers, *Extended Mind*, *Analysis*, 58, 1998, 7–19.
- A. Clark, *Supersizing the Mind. Embodiment, Action, and Cognitive Extension*, Oxford University Press, Oxford 2008.
- S. Chopra, L.F. White, *A Legal Theory for Autonomous Artificial Agents*, The University of Michigan Press 2011.
- D. Davidson, *Intending*, *Philosophy of History and Action*, 11, 1978, 41–60.
- _____, *Zwierzęta racjonalne*, przeł. C. Cieśliński, w: D. Davidson, *Eseje o prawdzie, języku i umyśle*, PWN, Warszawa, 1992, 234–250.
- _____, *Agency*, w: D. Davidson, *Essays on Action and Events*, Clarendon Press, Oxford 2002/1971, 43–61.
- D. Dennett, *Intentional Systems*, *Journal of Philosophy* 68, 1971, 87–106.
- _____, *Conditions of Personhood*, w: A. Oksenberg Rorty (red.), *The Identities of Persons*, University of California Press, 1976, 175–196.
- _____, *True Believers*, w: J. Haugeland (red.), *Mind Design II*, MIT, 1981, 57–79.
- R. Descartes, *Medytacje o pierwszej filozofii*, Wydawnictwo Zielona Sowa, Kraków, 2004/1647.
- L. Floridi, J.W. Sanders, *On the Morality of Artificial Agents*, *Minds and Machines*, 14, 2004, 349–379.
- H. G. Frankfurt, *Freedom of the Will and the Concept of a Person*, *The Journal of Philosophy*, 68 (1), 1971, 5–20.
- B. Gertler, *Overextending the Mind*, w: *Arguing about the Mind*, B. Gertler, L. Shapiro (red.), Routledge, 2007, 192–206.
- E. Hutchins, *Enculturating the Supersized Mind*, *Philosophical Studies* 152 (3), 2011, 437–446.
- S. Judycki, *Zachowanie i działanie*, *Diametros* 7, 2006, 82–97.
- J. Knobe, *Intentional Action and Side Effects in Ordinary Language*, *Analysis* 63 (3), 2003, 190–194.
- B. Latour, *Where Are the Missing Masses? Sociology of a Few Mundane Artefacts*, w: *Shaping Technology, Building Society. Studies in Sociotechnical Change*, W. Bijker, J. Law (red.), MIT Press, Cambridge (Mass.) 1992, 225–258.

- B. Latour, *Nadzieja Pandory. Eseje o rzeczywistości w studiach nad nauką*, przeł. K. Abri-szewski, A. Derra, M. Smoczyński, M. Wróblewski, M. Zuber, WN UMK, Toruń 2013.
- J., Law, *After ANT: Complexity, Naming and Topology*, w: J. Law and J. Hassard (red.), *Actor Network Theory and After*, Blackwell, Oxford, 1999, s. 1–14.
- B. Libet, *Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action*, *Behavioral and Brain Sciences* 8, 1985, 529–566.
- L. Malafouris, *At the Potter's Wheel: An Argument for Material Agency*, w: *Material Agency. Towards a Non-Anthropocentric Approach*, C. Knappet, L. Malafouris (red.), Springer 2008, 19–36.
- M. Mazur, *Cybernetyczna teoria układów samodzielnych*, PWN, Warszawa 1966.
- R. Menary, *Cognitive Integration: Mind and Cognition Unbounded*, Palgrave Macmillan, Basingstoke 2007, 62–64.
- O. Palermos, *Loops, Constitution and Cognitive Extension*, *Cognitive Systems Research*, 27, 25–41.
- K. Paprzycka, *Problem indywidualizacji i internalizacji działań*, *Analiza i egzystencja* 11, 2010, 291–315.
- M. Scheler, *Stanowisko człowieka w kosmosie*, w: M. Scheler, *Pisma z antropologii filozoficznej i teorii wiedzy*, przeł. S. Czerniak, A. Węgrzecki, PWN, Warszawa 1987.
- J.R. Searle, *Umysł, mózg, nauka*, przeł. J. Bobryk, PWN, Warszawa 1995, 52–63.
- Ch. Taylor, *Pojęcie osoby*, w: *Filozofia podmiotu. Fragmenty filozofii analitycznej*, przeł. R. Wieczorek, Aletheia, Warszawa, 2001, 401–420.
- Ch. Taylor, *Źródła podmiotowości. Narodziny tożsamości nowoczesnej*, przeł. M. Gruszczyński, O. Latek, A. Lipszyc, A. Michalak, A. Rostkowska, M. Rychter, Ł. Sommer, PWN, Warszawa 2001.
- G. Theiner, T. O'Connor, *The Emergence of Group Cognition*, w: *Emergence in Science and Philosophy*, A. Corradini, T. O'Connor (red.), Routledge, 2010, s. 6–78.
- B. Trybulec, *Wiedza i jej podmiot w szerokich systemach poznawczych*, Wyd. IFiS PAN, Warszawa 2017.
- _____, *Group as a Distributed Subject of Knowledge. Between Radicalism and Triviality*, *Avant*, VIII, 2017a, 183–207.
- _____, *It Takes Effort to Be (Collectively Rational): Group as a Reasoning Agent*, w: *Rationality and Decision Making*, M. Hetmański (red.), seria: *Poznań Studies in the Philosophy of the Sciences and the Humanities*, t. 111, 2018, 54–70.
- N. Wiener, *Cybernetyka, czyli sterowanie i komunikacja w zwierzęciu i maszynie*, PWN, Warszawa 1971.
- K. Wojtyła, *Osoba i czyn, oraz inne studia antropologiczne*, Towarzystwo Naukowe KUL, Lublin 1994.

**THE SUBJECT OR AGENT? UNDERSTANDING SUBJECTIVITY
IN THE COGNITIVE ARTEFACTS ERA**

ABSTRACT

The philosophical tradition defines the subject as a reflective being, in principle aware of its agency which makes it capable of making free decisions and taking responsibility for them. Agency, understood in this way, is clearly attributed only to people. However, the technological development of artificial cognitive enhancements and of increasingly autonomous artificial intelligence, that has been taken place in last few decades, casts doubts whether such an approach is not too anthropocentric. This doubt is indicated by some proponents of extending cognitive processes beyond the human brain; they argue for the need of appropriate extension of

the subject as well. Moreover, there is an increasing number of proposals attributing agency to artifacts. In the first part of the article, I refer to the two most commonly used philosophical criteria distinguishing the subject of cognition from all information processing systems: being a reflective system, and being the subject of intentional stance. Next, I assess, from such a perspective, the attempts to attribute agency to both one-person extended cognitive systems and artificial systems, such as relatively autonomous computer programs. I argue that the gap between conceptions of the extended subject and the artificial subject, and the standard approach incline toward the usage of the term “agent” designating this phenomenon. The term is already widely used in cognitive science to designate any relatively autonomous information processing system performing a cognitive task. The need of the clear distinction between “the subject” (“subjectivity”) and “the agent” (“agency”) is especially noticeable in Polish, where the difference in meanings of these concepts is not so evident as in English. The awareness of the applying in cognitive science these two different notions of agency prevents against a conceptual misuse which could lead to erroneous explanations and predictions.

Keywords: subject, agent, agency, extended mind thesis, extended cognitive system, cognitive enhancement, artificial cognitive system.

O AUTORCE — dr, Uniwersytet Marii Curie-Skłodowskiej, Instytut Filozofii, pl. Marii Curie-Skłodowskiej 4, 20-031 Lublin.

E-mail: Barbara Trybulec <barbara.trybulec1@gmail.com>