

Uniwersytet im. Adama Mickiewicza w Poznaniu

Wydział Psychologii i Kognitywistyki

Kierunek: Kognitywistyka



**Neuroetyka jako neurokognitywistyczne spojrzenie
na zagadnienie moralności i etyki**

Neuroethics as cognitive neuroscience perspective on
morality and ethics

Ewelina Gajewska

445841

Praca licencjacka napisana pod kierunkiem
dr Łukasza Przybylskiego

Poznań 2021

Abstrakt

Od niespełna dwóch dekad rozwija się program badawczy neuroetyki, mający aspiracje do stania się samodzielną dyscypliną nauki. Neuroetyka pyta o „moralność mózgu/umysłu” oraz etyczne kontrowersje związane z wdrażaniem osiągnięć neurokognitywistyki w życiu społecznym. Tym samym obszar jej zainteresowań podzielić można na dwie główne „tradycje”, tj. neuronaukę etyki i etykę neuronauk. Celem pierwszej z nich jest opisanie sposobu funkcjonowania zmysłu moralnego w mózgu i umyśle człowieka. Idea ta wpisuje się w ogólniejszą tendencję nauk do naturalizacji zagadnień klasycznie przypisywanych filozofii. Z kolei koncepcja i założenia etyki neuronauk nawiązują do bioetyki, jednakże ze względu na złożoność i wyjątkowość przedmiotu jej zainteresowań, tj. mózgu/umysłu człowieka, zdecydowano o jej odrębności względem wspomnianej dyscypliny. Bowiem rozwój neurokognitywistyki, a więc i wiedzy na temat budowy oraz sposobu funkcjonowania układu nerwowego i poznawczego wiąże się z prowadzeniem nowatorskich i często ryzykownych badań naukowych oraz powstawaniem nowych neuro-technologii. Zatem osoby zaangażowane w etykę neuronauk dbają o poprawność wniosków wyciąganych z tych badań i sposobów wykorzystania narzędzi technologicznych – pytają, w jaki sposób wpłyną one na obecnie przyjmowane sposoby rozumienia natury człowieka, tego czym jest świadomość, osoba czy wolna wola.

Słowa kluczowe: emocje społeczne, moralność, neuroetyka, obrazowanie mózgu

Neuroethics as cognitive neuroscience perspective on morality and ethics

Abstract

Neuroethics' research program has started about two decades ago and today it aspires to become an independent academic discipline. This area of study asks about roles of the brain and mind in human morality as well as ethical issues related to applications of cognitive neuroscience research and neurotechnologies in public life. Therefore neuroethics can be divided into two subfields or „traditions”, i.e. neuroscience of ethics and ethics of neurosciences. The former focuses on the study of a moral sense – a feeling of the rightness or wrongness of an action – and neural bases of behavior, moral responsibility, consciousness, free will and personal identity. Whereas the latter, often compared to bioethics, deals with ethical and social problems raised by advances in psychopharmacology, neuroimaging, brain-computer interfaces, brain stimulation as well as new research techniques and emerging technologies. Therefore neuroethicists ask how growing knowledge in the field of cognitive neuroscience about our brains and minds will influence and change definitions related to human nature.

Key words: brain imaging, morality, neuroethics, social emotions

Spis treści

| | |
|---|-----------|
| 1. Koncepcja neuroetyki | 4 |
| 2. Neuronauka etyki | 8 |
| 2.1 Moralność z perspektywy filozofii | 11 |
| 2.2 Moralność z perspektywy psychologii | 15 |
| 2.3 Moralność z perspektywy neurokognitywistyki | 27 |
| 3. Etyka neuronauk | 40 |
| 3.1 Medycyna | 44 |
| 3.2 Wojsko | 48 |
| 3.3 Ekonomia i marketing | 50 |
| 3.4 Prawo | 52 |
| 3.5 Sztuczna inteligencja | 54 |
| 3.6 Edukacja | 58 |
| 4. Podsumowanie | 62 |
| Bibliografia | 66 |

1. Koncepcja neuroetyki

Za moment powstania neuroetyki uznaje się konferencję *Neuroethics: Mapping the Field*, która miała miejsce 13 – 14 maja 2002 r. w San Francisco (Niebrój, 2013, s. 131), choć sam termin pojawiał się już wcześniej – w 2001 r. w jednym z artykułów Williama Safire, de facto jednego z uczestników owej konferencji założycielskiej, w gazecie *New York Times* (Farah, 2012, s. 572), a nawet w latach siedemdziesiątych ubiegłego wieku (Racine i in., 2017, s. 328). Przez niespełna dwie dekady dyscyplina ta zdołała zyskać zainteresowanie nie tylko pojedynczych badaczy, ale także towarzystw i instytucji naukowych, jak *Society for Neuroscience* i *Federation of European Neuroscience Societies*. Powstało nowe czasopismo dedykowane neuroetyce – *Neuroethics* (Niebrój, 2013, s. 131). Pojawiło się także już kilka podręczników napisanych przez prekursorów tej dziedziny, jak Martha Farah (2010), Eric Racine (Racine, Aspler, 2017), Neil Levy (2007) czy Judy Illes (Illes, Sahakian, 2013). Definicja neuroetyki nie jest ścisła i najczęściej podawana na zasadzie enumeracji podejmowanych w jej ramach zagadnień. Bazując na podziale tematycznym dokonany na konferencji w 2002 r., Niebrój (2013, s. 132) wymienia następujące cztery obszary zainteresowania neuroetyki: 1) badania nad tożsamością człowieka jako podmiotu moralnego, 2) problemy etyczne związane z wdrażaniem osiągnięć neurokognitywistyki w życiu społecznym, 3) kontrowersje etyczne dotyczące badań naukowych mózgu i umysłu oraz 4) dialog z opinią publiczną nad osiągnięciami neuronauk. Ponadto w jej ramach wyróżnić można dwie oddzielne subdyscypliny, nazywane czasem też dwoma tradycjami, tj. neuronaukę etyki i etykę neuronauk (Roskies, 2002). Pierwsza z nich nazywana potocznie „moralnością mózgu” ma charakter bardziej empiryczny i obejmuje tematykę wskazaną w punkcie 1) w przedstawionym wyżej podziale. Tym samym neuronauka etyki odnosi się do badania neurokognitywnego podłoża (mechanizmów i lokalizacji) koncepcji etycznych oraz zagadnień: wolnej woli, odpowiedzialności moralnej, jaźni czy świadomości. Zatem przyczynia się do poszerzania wiedzy odnośnie tych pojęć i zjawisk, a tym samym lepszego ich zrozumienia. Druga jest dziedziną bardziej praktyczną, obejmującą pozostałe obszary (2-4) wskazanego wyżej podziału i poruszającą kwestie etyczne w projektowaniu i prowadzeniu badań nad mózgiem i umysłem oraz wpływu i skutków aplikacji wyników tych badań w życiu publicznym. Chociaż te dwie subdyscypliny są względem siebie niezależne, niemniej posiadają także wspólne obszary badań, jak w przypadku tzw. świadomej zgody osób badanych (Northoff, 2009, s. 565-566). Bowiem neuronauki starają się tu odpowiedzieć

na pytanie, jak w mózgu i umyśle przebiegają procesy przyswajania i przetwarzania informacji oraz podejmowania decyzji. Tym samym odpowiedzi te mogą pomóc wyznaczyć, jak ujmuje to Niebrój (2014, s. 134), neurologiczne kryteria zdolności do podejmowania decyzji. W tym kontekście również zagadnienie roli emocji w procesach poznawczych dotyczy obu tradycji neuroetyki. Bowiem, czy emocje są zaangażowane w wydawanie sądów moralnych i jakie mechanizmy mózgowe są tu aktywowane? Czy członkowie komisji etycznych powinni uwzględniać rolę emocji w formułowaniu norm etycznych i podejmowaniu przez nich decyzji? Czy osoby z zachowanymi funkcjami poznawczymi, lecz nieprawidłowościami w sferze emocjonalnej posiadają pełnię zdolności psychicznych, np. do podejmowania decyzji? Niemniej we wskazanym wyżej fragmencie artykułu Niebrój trafnie przestrzega przed zbyt skrajnym i zbyt uproszczonym redukcjonizmem koncepcji wolnej woli, świadomości czy moralności człowieka do ich neurobiologicznych korelatów. Należy bowiem pamiętać, iż są to złożone zjawiska, także zjawiska społeczne, uwzględniające nie tylko jednostkę je posiadającą, ale obejmujące również zależności i relacje interindywidualne. Podejmowane przez wielu myślicieli na przestrzeni setek czy tysięcy lat nadal nie posiadają jednak jednoznacznych definicji, a tym bardziej jasnego wyjaśnienia. Nowatorskie podejścia do badania tych zjawisk przez współczesnych filozofów i naukowców innych dyscyplin dają początek właśnie takim dziedzinom czy programom badawczym jak neuroetyka. Między innymi z tego względu warto przyjrzeć się tej nowej dyscyplinie nauki, jako że podejmuje „stare” tematy filozofii i psychologii moralności oraz wznowia istniejące już w bioetyce debaty, jednak czyni to z innej perspektywy. Perspektywy zyskujących z roku na rok zainteresowania i zasobów wiedzy nauk neurokognitywnych. Dodatkowo według Northoffa (2009, s. 568) tym, co pozwoli neuroetyce umocnić swoją pozycję jako autonomicznej dyscypliny nauki jest wypracowanie przez nią swoistej metodologii łączenia faktów empirycznych z wnioskami i twierdzeniami normatywnymi. Dotyczy to szczególnie neuronauki etyki, która ma aspiracje nie tylko do znalezienia neurokognitywnych podstaw moralności, ale i przełożenie ich na fakty preskryptywne i systemy normatywne (Racine, Aspler, 2017, s. 94-95). Dziedzina ta powinna więc wypracować sposób na rozwiązanie problemu gilotyny Hume’a. Zatem chociaż te dwie dyscypliny mają oddzielne przedmioty zainteresowań oraz różnić badacze są zaangażowani w ich studiowaniu, niemniej istnieją „punkty styku” między nimi i ustalenia w jednej z nich mogą informować badania i sposób ich prowadzenia w tej drugiej.

W tym miejscu warto również wspomnieć, iż okoliczności powstania neuroetyki są podobne do narodzin kognitywistyki jako nauki. Bowiem przyjmuje się dzień 11 września

1956 r., kiedy to miała miejsce interdyscyplinarna konferencja w Massachusetts Institute of Technology, która zgromadziła przedstawicieli psychologii, lingwistyki, antropologii, filozofii i informatyki, jako moment powstania nowej dziedziny nauki – kognitywistyki. Choć rozmowy o stworzeniu nauk kognitywnych i potrzebie powstania nowego paradygmatu, tj. paradygmatu poznawczego, w ramach psychologii toczyły się już na początku lat 50. ubiegłego wieku (Miller, 2003). Koncepcja neuroetyki jest podobna do innej dyscypliny rozwijanej także w paradygmacie neuronaukowym, tj. neuroestetyki, w której to poszukuje się neurobiologicznych podstaw ocen i emocji estetycznych oraz ich ewolucyjnego wyjaśnienia (Podlipniak, Przybysz, 2013, s. 13-15). Bowiem analogicznie można powiedzieć, iż neuronauka etyki stara się odpowiedzieć na pytanie czy (i jak) sądy i emocje moralne różnią się od innych sądów wartościujących i emocji podstawowych. Dyscyplina ta poszukuje również ich neuronalnego i umysłowego podłoża oraz powodów powstania moralności u ludzi. Natomiast druga tradycja neuroetyki, tj. etyka neuronauk, postrzegana jest jako jedna z etyk stosowanych (zob. Sójka, 2007) czy podkategoria bioetyki. Nie da się ukryć, iż nauki te rzeczywiście dzielą ze sobą część poruszanych w ich ramach problemów. Etyka neuronauk stanowi też odpowiedź na podobne potrzeby „kontroli” nad przebiegiem postępu neuronaukowego jak miało to miejsce kilkadziesiąt lat temu z postępem biomedycyny czy genetyki.

Neuronauka, w Polsce oznaczana częściej terminem neurobiologia, zajmuje się funkcjonowaniem mózgu i układu nerwowego, jego anatomią, biochemią oraz filogenezą i ontogenezą. Z kolei głównym przedmiotem badań kognitywistyki jest umysł, jego sposób działania i istnienia oraz jego związek z mózgiem, znany także jako problem umysł-ciało. Wspólnie, te dwie dyscypliny można określić mianem neuronauk kognitywnych i przyjąć, że poszukują odpowiedzi, *w jaki sposób mózg tworzy umysł?*. Mówiąc bardziej szczegółowo, badają neuronalne korelaty funkcji behawioralnych i kognitywnych człowieka, tj. pamięci, uwagi, zdolności językowych czy sposobów myślenia, rozumowania. Tym samym, niegdyś wyraźny podział przedmiotu badań tych dyscyplin, tj. mózgu dla neurobiologii i umysłu dla kognitywistyki, ulega coraz większemu zatarciu i łączeniu obu tych dziedzin pod wspólną nazwą neuronauk kognitywnych czy neurokognitywistyki (zob. Przybysz, 2012). Dodatkowo jeśli chodzi o neuroetykę, można rozpatrywać ją jako część neurofilozofii, w której „[korzysta się z] osiągnięć neuronauk w refleksji nad tradycyjnymi problemami filozoficznymi” (Kurek, 2011, s. 1). Istotnie wpływ wyników badań neurofizjologicznych na rozumienie natury człowieka i pojęć z nią związanych, jak jaźń czy umysł, zawiera się w problematyce neuroetyki. Tym samym pytania dotyczące tych pojęć sytuują się na

pograniczu filozofii i neuronauk. Ich przykładem jest studiowanie zagadnienia świadomości przez Davida Chalmersa (tamże, s. 7), tj. poszukiwanie odpowiedzi na to, o czym mówimy, kiedy pytamy o świadomość oraz czy świadomość można zredukować do aktywności neuronalnej. Mówiąc bardzo ogólnie, neurofilozofia bada na ile zagadnienia dotyczące problemu umysł-ciało można wyjaśnić w terminach neuronauk poznawczych. Analogicznie, problematyka redukcjonizmu dotyczy neuroetyki (zob. tamże, s. 14-17). Na koniec chciałabym nadmienić, iż w dalszej części pracy będę się posługiwała zamiennie następującymi terminami: 1) neuronauka i kognitywistyka, 2) neurokognitywistyka 3) nauki o mózgu i umyśle oraz 4) neuronauki. Będą się one odnosiły do dziedziny nauki, tj. neurokognitywistyki, z perspektywy której zostanie przedstawiony i omówiony temat najpierw etyki i moralności z perspektywy nauk empirycznych – psychologii i neurokognitywistyki oraz filozofii (rozdział drugi), a następnie etycznych kontrowersji i implikacji związanych z rozwojem wiedzy nauk o mózgu i umyśle (rozdział trzeci).

2. Neuronauka etyki

Za pierwszą próbę przeniesienia klasycznych problemów filozoficznych na grunt nauk empirycznych uznaje się stanowisko Willarda van Ormana Quine'a w sprawie naturalizacji epistemologii w drugiej połowie XX wieku (Klar, 2020). Rewolucje kognitywistyczna, a następnie neuronaukowa pod koniec XX w. również przyczyniły się do powstawania nowych dyscyplin naukowych i wzrostu popularności badań o mózgu i umyśle. Warto wspomnieć, że Edward Osborne Wilson w swojej pracy także sugerował znaturalizowanie etyki i włączenie jej w szerszy krąg badań biologicznych (Wilson, 2000, s. 302-303). Z kolei Patricia Churchland była pierwszą, która usystematyzowała i jednoznacznie przedstawiła program i cele neurofilozofii (zob. Churchland, 1989). Filozofka wymieniła trzy główne postulaty tej dyscypliny, tj. 1) procesy umysłowe są procesami neuronalnymi, 2) teoria wyjaśniająca naturę umysłu powinna odwoływać się i wyjaśniać także strukturę i funkcje mózgu oraz 3) teoria neurofilozoficzna ze względu na swoje globalne podejście przewyższy mocą eksplanacyjną (wyjaśniającą) inne podejścia (Przybysz, 2012). Ponadto, jak sama twierdzi, filozoficzne eksperymenty myślowe niewiele mają wspólnego z rzeczywistością i jeśli rzeczywiście chcemy rozwiązać zagadkę umysłu, w tym trudny problem świadomości i wolnej woli, to należy prowadzić badania empiryczne i testować istniejące hipotezy, a nie jedynie idealizować i „domyślać się”, co jest możliwe (Churchland, 1996). W rzeczy samej zmysł moralny znajduje się w jej programie badawczym jako jedno z klasycznych zagadnień filozoficznych, które powinno zostać znaturalizowane na podobieństwo koncepcji Quine'a (Klar, 2020). Tym samym Churchland postuluje testowanie koncepcji filozoficznych i konfrontowanie ich z ustaleniami neurokognitywistyki, gdyż niepodobna utrzymywać hipotez niespójnych z faktami empirycznymi. W tym względzie neurofilozofia i neuronauka etyki mają podobne czy nawet zbieżne cele. Bowiem za pomocą badań nauki o mózgu i umyśle chcą znaleźć mechanizmy i struktury zaangażowane w myślenie moralne. Jak zachodzi podejmowanie decyzji w dylematach moralnych? Czy i jeśli tak, to czym sądy etyczne różnią się od innych sądów wartościujących? Czy emocje są częścią poznania (umysłu) moralnego? Neuroetyka opiera się w swych badaniach na różnych technikach neuroobrazowania, metodach psychologii oraz przypadkach neuropsychologicznych. Nie planuje przy tym, jak materializm eliminacyjny, zdeklasować pozycję filozofii w etyce, a jedynie wprowadzić nową perspektywę, „rzucić nowe światło” na naturę pojęć etycznych i moralność człowieka. W swych postulatach nie wyraża również

skrajnego redukcjonizmu, zastąpienia np. pojęcia sprawiedliwości jego opisem w terminach aktywności nerwowej, jak ma to miejsce w przypadku materializmu eliminacyjnego Churchland. Pomimo, iż stanowisko to jest przez wielu kwestionowane i krytykowane, niemniej teorie budowane w ramach neurofilozofii mają większą moc eksplanacyjną, ponieważ muszą nie tylko spełniać warunki empirycznej falsyfikacji i empirycznej spójności, lecz także być spójne logicznie oraz podlegać logicznej falsyfikacji (Bremer, 2010, s. 280). Przykładem jest próba wyjaśnienia czym jest Arystotelesowska eudajmonia (szczęście) z perspektywy filozoficznej, psychologicznej oraz neuronaukowej. Wychodząc od teorii filozofii, psychologowie budują model szczęścia (dobrostanu) i operacjonalizują terminy z nim związane, a metody neurokognitywistyki poszukują neuronalnych korelatów (aktywności układu nerwowego) zjawisk z nim związanych (Naor, Ben-Ze'ev, Okon-Singer, 2014). Jeśli chodzi o testowanie teorii, w świetle nowych ustaleń nauk neurokognitywnych systemy etyczne, takie jak deontologizm, postulujące udział jedynie rozumu w ludzkiej moralności mogą okazać się niespójne z faktami empirycznymi. Bowiem dotychczasowe badania prowadzone w ramach neuroetyki wskazują zarówno na rolę racjonalnego myślenia, jak i emocji w poznaniu moralnym. Przez to postulaty deontologizmu mogą okazać się trudne do spełnienia, jeśli niepodobna oddzielić emocji od myślenia o sprawach etycznych. Ponadto zaangażowane w rozumowanie moralne procesy mózgowe mają charakter rozproszonych sieci nerwowych i wszystko wskazuje na to, że nie istnieje specjalny moduł poświęcony wyłącznie moralności. Podobne wnioski wyciągane są na gruncie neuroestetyki i badań nad umysłem estetycznym. Aczkolwiek niektóre struktury mózgu, jak kora orbito-frontalna, przednia kora zakrętu obręczy czy ciało migdałowe, wymieniane są w większości badań, a więc uważa się, że mają kluczowe znaczenie w wyjaśnianiu jej neurokognitywnych podstaw (Racine, Aspler, 2017, s. 24). Zakłada się, że wyniki te mogą mieć znaczenie dla metaetyki i etyki normatywnej. Problemem jest jednak przełożenie faktów empirycznych (deskryptywnych) na terminy norm moralnych – zasad powinnościowych (tamże, s. 94-96). Atutem interdyscyplinarnego podejścia jest szersza perspektywa patrzenia na dane zjawisko, a tym samym lepsza jego operacjonalizacja i większe prawdopodobieństwo uchwycenia wszystkich jego szczegółów i aspektów, co ma istotne znaczenie w przypadku takich złożonych pojęć społecznych jak moralność. Z uwagi na tę złożoność wyróżnia się obecnie wiele specjalizacji czy subdziedzin wśród nauk o mózgu i umyśle, m.in. neuronaukę społeczną (zob. Decety, Keenan, 2006), łączącą perspektywę teorii filozoficznych, modeli psychologii oraz danych neurofizjologicznych w badaniach nad zachowaniami społecznymi człowieka (Cacioppo i in., 2007). Analizując znaczenie pojęć oryginalnie przypisywanych

filozofii, coraz częściej uwzględnia się wyniki badań neuronauk, szczególnie w niedawno powstałej filozofii eksperymentalnej (Knobe, Nichols, 2013), nawet jeśli termin „neurofilozofia” nie jest *explicite* wymieniany w tekście. Na przykład Poczobut stwierdza, że aby zbudować holistyczną teorię jaźni, niepodobna nie uwzględnić w niej dwóch perspektyw, tj. podstaw neuronalnych i analiz fenomenologicznych oraz konkluduje: „Filozof poszukujący wszechstronnego zrozumienia jaźni zmuszony jest zwrócić się także w kierunku neuronauk” (Poczobut, 2013, s. 386). Zwrócenie się filozofów w stronę badań empirycznych umożliwia więc powstawanie nowych koncepcji i wyjaśnień „starych” problemów filozoficznych. Pozwala to jednocześnie na zbudowanie teorii syntetycznych, ujmujących złożoność i dogłębność zjawisk dotyczących istoty człowieczeństwa, w tym moralności. Jednakże same wyniki badań neurokognitywistyki nad biologicznym podłożem takich pojęć, jak tożsamość osobowa czy wolna wola niewiele mówią o znaczeniu tych pojęć. Ich proste sprowadzanie i wyjaśnianie jedynie w kategoriach zjawisk neuronalnych wydaje się również mało prawdopodobne. Postulat materializmu eliminacyjnego Churchlandów o wyeliminowaniu pojęć psychologii potocznej i zastąpieniu ich terminami neurobiologicznymi nie znajduje zatem (na razie) potwierdzenia empirycznego (O'Connor, Joffe, 2013, s. 259). Choć niewątpliwie trwają dyskusje akademickie nad naturą pojęć świadomości, osoby czy odpowiedzialności moralnej. Istotnie, badania nauk o mózgu i umyśle pogłębiają i poszerzają definicje tych pojęć oraz wnoszą nowy punkt widzenia w dyskusji na ich temat. Niemniej bez ich szerszego osadzenia i refleksji w ramach filozofii i nauk społecznych neuronalne korelaty tych zjawisk nie wyjaśniają natury czy istoty człowieczeństwa. Ponadto ścisła współpraca filozofii i neuronauk wydaje się być korzystna dla obydwu dyscyplin. Tym samym neurofilozofia, ergo neuroetyka, jako interdyscyplinarna dyscyplina, łącząca różne metodologie i perspektywy badawcze może okazać się szczególnie ciekawym i owocnym projektem studiowania natury człowieczeństwa i moralności. Do kręgu jej zainteresowań włącza się także badania neuropsychologii, a więc przypadki osób po wszelkich uszkodzeniach mózgu oraz z deficytami i zaburzeniami psychicznymi (szczególnie w sferze społecznej), np. demencją, psychopatią, schizofrenią czy zaburzeniami ze spektrum autyzmu (Racine i in., 2017, s. 329-330). Tym samym próbuje się ustalić czy określone obszary mózgu i deficyty poznawcze lub emocjonalne są istotne (kluczowe) dla rozumowania i postępowania moralnego. Z pomocą metod neuroobrazowania próbuje się również poszukiwać mechanizmów neuronalnych i struktur mózgu zaangażowanych w rozumowanie moralne w trakcie dokonywania wyborów w dylematach etycznych oraz badać udział emocji w poznaniu moralnym i ich związki z racjonalnym myśleniem. Cele neuronauki etyki dzieli się

na te umiarkowane i te bardziej ambitne. Do pierwszej kategorii zaliczyć można wzbogacenie dotychczasowej wiedzy na temat pojęć związanych z moralnością czy znalezienie neuronalnych podstaw moralności (zmysłu moralnego). Z kolei do celów ambitnych i uważanych niekiedy za kontrowersyjne należy stworzenie całkiem nowej teorii etycznej (na miarę nawet etyki uniwersalnej) opartej na faktach empirycznych neurokognitywistyki i psychologii. Zdarzają się jednak też głosy mocno krytyczne, kwestionujące postulaty neuroetyki i dotychczasowe wnioski z badań oraz jej wartość dla refleksji etycznej (tamże, s. 330-332; zob. także Berker, 2009). W obliczu ostatniego kryzysu replikacyjnego w naukach społecznych uzasadnione wydaje się krytyczne podejście do nowych i zaskakujących wyników badań, tym bardziej jeśli wnioski z nich płynące formułowane są na podstawie pojedynczych badań. Nie należy jednak odrzucać programu neuroetyki, zwłaszcza biorąc pod uwagę fakt, iż jest ona dopiero w początkowych etapach rozwoju, jako że neurokognitywistyka udowodniła już wartość swych badań dla nauki, zajmując się zjawiskiem świadomości. Tym samym potwierdziła udział mózgu w powstawaniu umysłu i stanów świadomości i zakwestionowała klasyczny pogląd dualizmu substancjalnego Kartezjusza. Dzisiejsi uczeni, jak wskazuje Churchland (2007), szczególnie ci dopiero rozpoczynający swoją pracę naukową przychylnie podchodzą do programu naturalizacji filozofii i skłonni są brać pod uwagę badania empiryczne w swej refleksji filozoficznej. Zasadna wydaje się zatem postawa optymistyczna wobec mocy eksplanacyjnej nauk neurokognitywnych, także w poszukiwaniu neuronalnych korelatów zmysłu moralnego. Konkludując, pomysł neuroetyki nie jest nowy i wpisuje się w szerszy program naturalizacji zagadnień tradycyjnie przypisywanych filozofii. Nauki empiryczne starają się tu odpowiedzieć na pytanie, czy istnieje osobny moduł (zmysł) moralności w mózgu i towarzyszące mu swoiste procesy psychiczne oraz jakie mechanizmy są podstawą jego funkcjonowania.

2.1 Moralność z perspektywy filozofii

Sekcja ta posłuży za wprowadzenie do tematu moralności i etyki oraz zdefiniowaniu pojęć z nimi związanych. Bardzo ogólnie można powiedzieć, że etyka to nauka o moralności, choć ich szczegółowe definicje nie są jednoznaczne i ścisłe. Przykładowo, Środa (2010, s. 4) przyjmuje, że moralność to „zachowania, postawy, sposoby postępowania, poglądy zgodne z przyjętymi przez daną jednostkę czy społeczność kryteriami dobra i zła”, etyką zaś nazywa „dziedzinę filozofii, której przedstawiciele tworzą, opisują i uzasadniają jakąś koncepcję

dobra (lub powinności, cnoty itd.) i próbują ustalić, jakie dyrektywy powinny kierować ludzkim działaniem. Etyka to zbiór określonych teorii, systemów, koncepcji, poglądów(...)”. Z kolei Sójka (2007, s. 118) pisze, iż etyka jest „bezosobowym projektem ludzkich powinności, w ramach którego wprost dokonuje się rozróżnienia na dobro i zło”, a moralność definiuje jako „zasady przejmowane przez jednostkę od grupy i w efekcie uznawane przez nią za własne. Etyka to pewien wzorzec idealny, zaś moralność to konkretne zasady konkretnych ludzi i grup”. I choć w języku powszechnym terminy „etyczny” i „moralny” są uznawane za synonimiczne, to można je uznać co najwyżej za wyrazy bliskoznaczne. Termin etyka pochodzi od greckiego wyrazu *ethos*, co oznacza charakter, a moralność od łacińskiego *mos, moris* co oznacza zwyczaj. Za twórcę etyki, sięgającej V w p.n.e., uważa się Sokratesa, jednego z największych myślicieli starożytnych. W ramach filozofii zagadnienie etyki dzieli się na trzy działy: etykę normatywną, etykę opisową i metaetykę. Pierwsza formułuje reguły postępowania i kryteria ich oceny jako moralnie dobrych, słuszych lub złych, niewłaściwych. Druga, jak sama nazwa wskazuje, opisuje systemy etyczne, bada moralność (jej rozwój, kształtowanie się) z perspektywy naukowej. A metaetyka jako nauka o nauce bada skąd pochodzą i co znaczą pojęcia etyczne, np. czym jest dobro i czy sądom moralnym można przypisywać prawdziwość lub fałszywość. Poszczególne teorie etyczne starają się odpowiedzieć na pytania, czym jest dobro i zło; co odróżnia czyn moralnie właściwy od niewłaściwego; jakimi zasadami i wartościami powinni się kierować ludzie w życiu codziennym; co oznacza dobre i szczęśliwe życie; do czego ludzie powinni dążyć. W filozofii moralnej wyróżnia się trzy główne teorie etyczne (normatywne), a więc i sposoby odpowiedzi na te pytania. W etyce cnót Arystotelesa doskonałość moralną osiąga się przez stosowanie zasady złotego środka, tj. dzięki znalezieniu umiaru między skrajnościami swoich zachowań. Opanowanie tej zasady prowadzi do trwałych dyspozycji do działania właściwego, dobrego, które Arystoteles nazywa cnotami. Jednocześnie podkreśla, iż sama wiedza odnośnie tego co jest dobre, a co złe nie wystarcza, bowiem należy jeszcze ćwiczyć się w cnotach, tak jak muzyk ćwiczy grę na swym instrumencie (Fingas, 2011, s. 15). Dalej, według Arystotelesa i jego systemu etycznego człowiek cnotliwy to taki, który czerpie przyjemność z właściwego postępowania. Ponadto życie cnotliwe to życie dobre, które pozwala osiągnąć ostateczny cel, czyli szczęście (eudajmonię). Deontologizm, którego najsłynniejszym przedstawicielem był Immanuel Kant, jest teorią obowiązku moralnego, która głosi, iż moralność polega na stosowaniu „takiej maksymy, dzięki której możesz zarazem chcieć, żeby stała się powszechnym prawem” (Środa, 2010, s. 112). Zatem dobre postępowanie jest dobre samo w sobie, właśnie dlatego, że zastosowaliśmy daną regułę. Z

kolei w konsekwencjalizmie (użytylizmie), przypisywanym najczęściej postaciom Jeremy'ego Bentham'a i Johna Stuarta Mill'a, ocenie moralnej podlegają skutki działań. To znaczy, moralność polega na maksymalizacji szczęścia jak największej liczby osób. Intencje sprawcy działań nie są tu brane pod uwagę, jako że właściwe działanie jest działaniem użytecznym, tj. takim, które przynosi więcej szczęścia niż inny czyn w danej sytuacji. Konsekwencjalizm i deontologizm starają się zatem sformułować uniwersalną regułę postępowania moralnego, tj. odpowiedź na pytanie „co powinienem [powinnam] zrobić?”. Z kolei przedmiotem etyki cnót jest człowiek i jego charakter oraz dążenie do znalezienia odpowiedzi na pytanie „jakiego rodzaju człowiekiem powinienem [powinnam] się stać?” (Fingas, 2011, s. 16-17). Systemy te podkreślają wyższość rozumu, racjonalnego myślenia w postępowaniu moralnym. Niemniej byli i tacy filozofowie, którzy w swych koncepcjach zwracali uwagę na rolę emocji w sferze ludzkiej moralności, jak przedstawiciele etyki troski czy sentymentalizmu. Jednym z nich był David Hume, który twierdził, iż rozum pomaga w ocenie danego czynu, lecz sam nie wystarcza do jego potępienia lub aprobaty, sam nie jest w stanie pobudzić nas do jakiegokolwiek działania. To uczucia tworzą u Hume'a podstawę moralności, jako że „Rozum pomaga w wydawaniu ocen, bo trzeba znać należycie fakty, by móc je ocenić, ale to uczucia skłaniają nas do uznania słuszności jakiegoś czynu lub potraktowania go jako występku” (Środa, 2010, s. 118). Dodatkowo szkocki filozof uważał, iż człowiek jest z natury dobry, w którym uczucie sympatii jest podstawowym odruchem reagowania na świat społeczny (Urbaniak, 2014, s. 60), a społeczeństwo istnieje dzięki tym intuicyjnym zachowaniom podyktowanym przez zmysł moralny. Podobne stanowisko zajmował także Arystoteles, uważając człowieka z natury za empatycznego i współpracującego z innymi. Odmiennego zdania był Thomas Hobbes, dla którego ludzie są z natury źli, skrajnie indywidualistyczni i egoistyczni, relacje interpersonalne polegają na wrogości i rywalizacji, a moralność została wymyślona i następnie narzucona przez społeczeństwo, żeby pochamować tę wzajemną wrogość ludzi wobec siebie. Innym filozofom brytyjskim, najpierw Shaftesbury'emu, a następnie Hutcheson'owi na przełomie XVII i XVIII wieku, przypisywana jest idea zmysłu moralnego, definiowanego jako „Zdolność rozpoznawania i tworzenia tego, co dobre lub złe” (...) który to zmysł wyraża się w formie wrodzonego poczucia sympatii wobec innych jednostek” (tamże, s. 59-60). Hume także odwoływał się do idei wrodzonego zmysłu, naturalnej władzy odróżniania dobra i zła. Zatem wedle sentymentalistów racjonalne myślenie jest potrzebne do zrozumienia danej sytuacji i wnioskowania o ewentualnych konsekwencjach określonego działania, lecz ostateczna kategoryzacja tych działań jako dobrych lub złych jest wynikiem naszych

emocjonalnych reakcji na nie. Stwierdzenie to będzie istotne w dalszej części pracy, tj. w części poświęconej psychologii (*sekcja 2.2*) oraz neurokognitywnym podstawom moralności (*sekcja 2.3*). Intuicje szkockiego filozofa odnośnie oceniania czynu jako moralnie dobrego lub naganego w zależności od intencji i motywów jego sprawcy zostały potwierdzone w badaniach eksperymentalnych. Mianowicie działania agenta powodujące nawet pozytywne skutki są oceniane przez większość ludzi jako moralnie złe, jeśli zamiar ich popełnienia był karygodny. Współczesnym filozofem sentymentalistą jest Jesse Prinz porównujący zmysł moralny do gustu kulinarnego czy estetycznego, a więc używając słów „dobry” lub „zły” wobec danego zdarzenia, wyrażamy jedynie nasz stosunek do niego. Innymi słowy są to stwierdzenia wartościujące, oddające nasze reakcje emocjonalne (przyjemność lub nieprzyjemność) i wobec tego nie sposób przypisywać im fałszywości lub prawdziwości, jak zakładają zwolennicy kognitywizmu etycznego. Dodatkowo Prinz odnotowuje aspekt motywacyjny emocji do unikania działań ocenianych negatywnie oraz podtrzymywania czy promowania działań wartościowanych jako pozytywne. Ponadto Prinz jest zaangażowany także w badania empiryczne nad moralnością. Warto także wspomnieć, iż Hume jako jeden z pierwszych zwrócił uwagę na odrębność zdań opisowych i normatywnych, która ma istotne konsekwencje dla etyki (Środa, 2010, s. 119). Problem ten nazywany „gilotyną Hume’a” głosi, że nie można przejść do zdań stwierdzających jak „powinno być” na podstawie prostych faktów opisowych, jakie „coś jest”. Z drugiej strony przyjmuje się, że powinność zakłada możliwość (spełnienia tej powinności). Zatem badania empiryczne mogą dostarczać dowodów i opisów, czy w ogóle człowiek jest w stanie spełnić dane wymagania. Bowiem, przykładowo, przypisywanie odpowiedzialności moralnej zakłada wolność wyboru człowieka, tj. istnienie wolnej woli, a więc jest sprzeczne z założeniami determinizmu neuronalnego i w związku z tym pojawiają się głosy, że należy porzucić ideę odpowiedzialności za czyny, jeśli badania empiryczne dowiodą brak istnienia wolnej woli. Obraz natury etyki i moralności człowieka jaki wyłania się z rozważań filozoficznych nie jest jednoznaczny, a często nawet sprzeczny – jedni zakładają uniwersalizm norm moralnych, inni relatywizm kulturowy, dla jednych sądy etyczne posiadają wartość logiczną, dla innych wyrażają one jedynie nasz subiektywny stosunek czy sympatię lub repulsję wobec określonych czynów, dla wyznawców systemów deontologicznych tylko rozum może odróżnić „dobro” od „zła”, z kolei dla sentymentalistów i przedstawicieli etyki troski to emocje i uczucia są podstawą ocen moralnych. Być może poszczególni filozofowie wskazują także na różne aspekty moralności i dopiero rozważanie ich łącznie jako uzupełniających się, a nie zaprzeczających sobie tez, jest w stanie oddać pełny obraz (biologiczno-społecznej)

natury człowieka, a badania empiryczne dodatkowo mogą sprawdzać intuicje myślicieli i wprowadzić nową perspektywę naturalizującą zagadnienie etyki. Na koniec tej sekcji wspomnę o ciekawej analogii eksperymentu myślowego z Mary-neurobiologką, jakiej dokonał Artur Szutta (2017). Oryginalnie dotyczy on natury qualiów i trudnego problemu świadomości, który mówi, iż sama wiedza odnośnie aspektów jakościowych naszych stanów i przeżyć nie wystarczy, by wytłumaczyć jak to jest być w tym stanie, czym jest to doświadczenie. Analogicznie sama wiedza odnośnie dobrego i złego postępowania oraz zasad etycznych być może nie wystarcza, by wytłumaczyć czym jest moralność i co odróżnia czyn moralny od niemoralnego.

2.2 Moralność z perspektywy psychologii

Współcześnie zmysł moralny definiowany jest jako „w sferze kognitywnej – naturalna zdolność wartościowania zdarzeń, obiektów czy zjawisk, ale również – w sferze konatywnej – umiejętność preferowania interesów wspólnoty ponad interesami indywidualnymi. Nie jest to jedynie umiejętność rozpoznawania wartości, ale także hierarchizowania ich ze względu na kontekst sytuacji” (Urbaniak, 2014, s. 62). Z kolei Wojciszke i Baryła (2000, s. 397) zakładają, że przedmiotem moralności jest „ocena działań człowieka z uwagi na ich konsekwencje dla dobra-krzywdy innych ludzi”, a ponadto istnieje u ludzi tendencja do opatrywania sankcją oraz uniwersalizacji sądów o charakterze moralnościowym. Zatem ocenie moralnej podlegają działania, które powodują cudzą krzywdę (fizyczną, psychiczną lub materialną) i które wtedy uznawane są za niewłaściwe oraz takie postępowania, które przyczyniają się do dobra innych i z tego powodu uważane są za właściwe. Choć zachowania mogą być określane jako właściwe lub niewłaściwe także z innych powodów, np. z uwagi na samą ich użyteczność czy efektywność, lecz wtedy nie zalicza się tych ocen do domeny moralności (Wojciszke, Baryła, 2000, s. 397). Niemniej wciąż nierozstrzygnięta pozostaje debata odnośnie wrodzoności moralności, a obecnie można w niej znaleźć trzy główne stanowiska: skrajnie natywistyczne, słabo natywistyczne oraz anty-natywistyczne (Prinz, 2014, s. 105). Pierwsze z nich jest porównywane do poglądu w sprawie wrodzoności języka i które to stanowisko postuluje istnienie „gramatyki moralnej” analogicznej do gramatyki uniwersalnej Chomskiego (tamże, s. 106). Zalicza się tutaj również domenową teorię Turiela, który zakłada, że posiadamy wrodzone zasady odnośnie troski/krzywdy i wobec tego tylko takie zasady można uznać za moralne – pozostałe reguły, odwołujące się do innych wartości niż troska/krzywda, są zwykłymi konwencjami

społecznymi. Choć Wojciszke i Baryła, powołując się na Haidta, uważają to za zbyt wąskie określenie moralności i wnoszą o istnieniu także innych wartości moralnych, które nie dotyczą krzywdy/troski. Jednakże stanowisko sprowadzające wszystkie zasady moralne do tej jednej naczelnej wartości jest możliwe do obronienia. Wojciszke i Baryła (2000, s. 397) podają przykład zniszczenia i zbezczeszczenia flagi narodowej jako naruszenia fundamentu świętości czy lojalności wobec grupy (w koncepcji fundamentów moralnych przedstawionej dalej w tej sekcji), lecz nie troski/krzywdy, aczkolwiek 38% osób w badaniach Haidta (2014) uważało, że ktoś ucierpiał z powodu tego lub podobnych czynów – sam psycholog określa takie sytuacje jako „niekrzywdzące naruszenie tabu”. Krzywda nie musi być jednak bezpośrednio obserwowana i może być jedynie potencjalnie zakładana (przewidywana) – osoby badane uważają bowiem na podstawie własnych doświadczeń i wiedzy o świecie społecznym, że istnieje duże prawdopodobieństwo wywołania u kogoś cierpienia takimi zachowaniami. Ponadto określone postępowanie może w ocenie niektórych nie krzywdzić nikogo, a więc i nie łamać żadnych wartości moralnych, podczas gdy dla pozostałych osób wywołuje ono cierpienie u innych i tym samym narusza jakąś zasadę moralną. Zatem wiarygodne wydaje się twierdzenie o istnieniu naczelnej zasady (wartości) troski/krzywdy i sprowadzaniu do niej pozostałych, które to „na pierwszy rzut oka” maskują ją odwołaniem do innych zasad. Za naturalnością („biologicznym zaprogramowaniem”) zachowań społecznych i zmysłu moralności przemawia również fakt, iż często tego typu zachowania są trudne do wyjaśnienia przez świadome odwołanie się do rozumu. Ponadto badania etologiczne ujawniły powszechność podobnych zachowań nastawionych na współpracę i wzajemną pomoc wśród innych gatunków zwierząt, szczególnie społecznych małych naczelników innych niż człowiek (Prinz, 2014, s. 107). Jednakże Prinz jest sceptyczny, co do podawania dowodów wrodzoności moralności na podstawie badań porównawczych. Bowiem, jak pisze, same zachowania niektórych zwierząt, które z naszej perspektywy wydają się dotyczyć wartości moralnych, takich jak sprawiedliwość i troska, niekoniecznie pociągają za sobą zdolność (władzę umysłową) do wydawania sądów czy oceniania czynów jako dobrych i złych – mogą to być przejawy jedynie podświadomych instynktów. Bowiem zdolność ta zdaje się być podstawą zmysłu moralnego, ewentualnie występującą obok wrodzonych intuicji. Dodatkowo dla moralności wymagana jest samoświadomość, by móc ocenić własne postawy i czyny, a także przyjąć perspektywę innych i zrozumieć ich doświadczenia. Przedmiotem zainteresowań moralności człowieka są również losy osób trzecich i ich posłuszeństwo wobec obowiązujących norm zachowań – według Prinza dotychczasowe badania nie potwierdzają spełnienia tych wymagań przez inne gatunki niż człowiek lub istnieje za mało

danych, aby to stwierdzić. Podsumowując, stanowisko Prinza zakłada, że zwierzęta nie posiadają (abstrakcyjnej) wiedzy odnośnie pojęć dobra i zła, a więc nie potrafią stwierdzić, które zachowania są właściwe, a które karygodne, ergo nie mają zmysłu moralnego. Z drugiej strony można powiedzieć, że niektóre zachowania społeczne wysoce przypominają te ludzkie, ponieważ są na tyle naturalne i wrodzone, że nie potrzeba mieć umiejętności ich (świadomej) oceny, by postępować właściwie – wtedy ta naturalna tendencja (intuicja) do prospołecznych zachowań jest wrodzona także u ludzi. Bowiem do naczelnych należą gatunki zwierząt nazywane wysoce społecznymi, a wzrost inteligencji i posiadanie wyższych funkcji poznawczych u przedstawicieli *Homo sapiens* tylko wzmacnia te naturalne tendencje. Takiego poglądu jest Frans de Waal, który twierdzi, iż moralność obserwowana u ludzi jest daleko rozwiniętą formą zachowań społecznych, występujących także u innych zwierząt. Podobnego zdania jest Patricia Churchland, która moralność oraz umiejętność poruszania się w świecie społecznym przypisuje naturalnym (neuronalnym) mechanizmom motywacji, popędów, skłonności oraz umiejętności kontroli tych zachowań i innych zdolności kognitywnych (Churchland, 2013). W tym najbardziej podstawową wartością jest troska o siebie i krewnych oraz potrzeba przynależności u gatunków społecznych – instynkt samozachowawczy (dbanie o własny organizm) jest obecny u wszystkich istot żywych. Zatem „krąg opieki” u najstarszych przodków obejmował samego siebie i własne potomstwo, a w miarę jak warunki życia się polepszały (obfitość zasobów), do kręgu tego jednostki włączyły także dalszych krewnych i przyjaciół czy nawet obce osobniki. Churchland wskazuje przy tym na jeden neurohormon – oksytocynę jako kluczowy czynnik emergencji zachowań społecznych, związanych z opieką. Ponadto to u małp jako pierwszych wykryto neurony lustrzane, które prawdopodobnie odpowiadają za zdolność do empatii (przyjmowania czyjejś perspektywy, „wchodzenia w czyjeś buty”) oraz naśladowania sposobów postępowania innych. Odnotowano też u nich zachowania świadczące o poczuciu sprawiedliwości i empatii oraz zdolności do rozwiązywania konfliktów społecznych i altruizmu odwzajemnionego (De Waal, 2013). Chociaż niektórzy zarzucają „antropomorfizujący” język wobec porównywania zachowań ludzi i małp, De Waal stanowczo sprzeciwia się tym zarzutom. Bowiem obserwowane zjawiska są na tyle podobne, że trudno nie używać tych samych pojęć do ich opisu – nawet jeśli są to zachowania analogiczne, a nie homologiczne (choć zdaniem De Waala jest bardziej prawdopodobne, że zachowania te pochodzą od wspólnego przodka). Na tej podstawie wnioskuje się, że człowiek i inne zwierzęta społeczne mają naturalną tendencję do życia w grupie. Dodatkowo uważa się to za efektywną strategię przetrwania dla gatunku, przynoszącą bezpośrednie korzyści także

dla pojedynczych osobników. Ostatnimi czasy mówi się też o doborze grupowym, w którym to całe grupy brane są pod uwagę jako podlegające tym samym regułom, co klasyczny dobór naturalny faworyzujący jednostki najlepiej dostosowane (*fitness*). Bowiem życie w grupie posiada wiele korzyści, m.in. zapewnia większe bezpieczeństwo, większą szansę na zdobycie pokarmu czy pomoc w wychowywaniu potomstwa. Chociaż istnieją również pewne ograniczenia nakładane na pojedyncze osobniki „rachunek zysków i strat”, zwłaszcza w perspektywie długoterminowej, wydaje się korzystniejszy dla strategii grupowej – wzajemnej współpracy osobników niż strategii indywidualistycznej. Zatem także popularna hipoteza „samolubnego genu”, która mówi, że ewolucyjnie jesteśmy „zaprogramowani” do dbania o własne interesy, da się pogodzić ze współwystępowaniem i rozwojem tendencji altruistycznych (tamże). W takim wypadku szeroko rozumiane zachowania społeczne, w tym moralność, istnieją dlatego, że sprzyjają utrzymaniu spójności grupy, co ostatecznie przekłada się na sukces ewolucyjny gatunku. Rozwinięcie tej klasy zachowań widocznych u małych i innych zwierząt do postaci zmysłu moralności człowieka zawdzięczamy przede wszystkim wzrostowi inteligencji (Ayala, 2009). To przede wszystkim dzięki umiejętności abstrakcyjnego myślenia i wnioskowania jesteśmy w stanie przewidywać skutki określonych działań i oceniać, czy byłyby one pożądane, czy też nie oraz wymyślać bardziej korzystne alternatywy tych zachowań. Dzięki rozwiniętym zdolnościom intelektualnym i kognitywnym człowiek jest również w stanie zrozumieć złożone sytuacje społeczne, a z pomocą umiejętności „czytania umysłów” wnioskować o intencjach i motywacjach innych. W pewnym sensie moralność jest więc „narzędziem” społecznym, stworzonym na podstawie naturalnej, wspólnej wszystkim organizmom, tendencji do wartościowania bodźców i zdarzeń (wewnętrznego systemu preferencji) oraz specyficznie ludzkich zdolności intelektualnych. Naczelnym zadaniem (przeznaczeniem) tak rozumianej moralności jest więc regulacja stosunków interpersonalnych oraz utrzymywanie społecznego ładu. Zatem przedstawione dotychczas obserwacje naukowe zdają się wspierać stanowisko słabego (czy umiarkowanego) natywizmu, który zakłada, że wrodzone mogą być pewne tendencje do zachowań, lecz konieczny jest również udział czynnika kulturowego do wykształcenia pełnej formy moralności, jak u człowieka. Niemniej procesy psychiczne i mózgowie wymagane dla poznania moralnego niekoniecznie muszą być specjalistyczne i oznaczać istnienie biologicznego modułu moralności – mogą być tymi samymi ogólnymi procesami umysłowymi, angażowanymi do rozwiązywania problemów wszelkiego typu. Tym samym zmysł moralności może być też w pewnym sensie produktem ubocznym (*byproduct*) czy zjawiskiem emergentnym, powstałym poprzez interakcje biologicznych predyspozycji oraz

wpływów kultury – podobnie jak w przypadku zmysłu estetycznego. Ponadto, mimo że pewne zdolności i tendencje są wrodzone, konkretne (dające się zwerbalizować) zasady moralne są dopiero „nadbudowane” na tych naturalnych sprawnościach. Zatem zdolność do wydawania sądów wartościujących jest naturalna, jednak szczegółowe kryteria według których jednostki oceniają dane zdarzenia jako dobre lub złe są wytworem kulturowym. Tym samym obserwowana uniwersalność w zachowaniach moralnych człowieka na całej kuli ziemskiej bierze się z pewnych uwarunkowań biologicznych i napotykania podobnych problemów społecznych, a mniejsze lub większe różnice w tych wzorcach mają swoje źródła w warunkach środowiskowych i indywidualnej historii każdej kultury. Moralność można więc też uznać za formę adaptacji do środowiska (zob. Ayala, 2009).

Choć ostatnimi czasy dokonuje się „zwrot ku uczuciom” w nauce o moralności (Dziarnowska, Przybysz, 2010, s. 57; Żylicz, 2012, s. 78), a ponadto emocje są istotnie związane z niektórymi działaniami, to uczeni na przestrzeni lat badań je pomijali lub traktowali jako czynniki zakłócające racjonalne zachowanie człowieka (Dąbrowski, 2014, s. 142). W drugiej połowie XX wieku powstał i rozwijał się nurt poznawczo-rozwojowy, utożsamiany z Piagetem i Kohlbergiem. Badali oni bowiem sposoby rozumowania dzieci w przedstawianych im dylematach moralnych. Piaget sądził, iż dzieci pojmują zasady moralne analogicznie do sposobu, w jaki opanowują zasadę zachowania ilości – na podstawie własnych doświadczeń i interakcji (zabawie) z innymi dziećmi w określonym etapie rozwoju. Z kolei Kohlberg, na podstawie założeń Piageta, wyróżnił sześć etapowy przebieg rozwoju moralnego. Dwa pierwsze etapy jego modelu nazywa się poziomem przedkonwencjonalnym, zorientowanym na kary i posłuszeństwo, w którym dzieci zachowują się tak, by uniknąć tych kar. Dwa kolejne etapy zwane stadium konwencjonalnym opierają się na rozumieniu i przestrzeganiu norm moralnych jako pewnych obowiązków i przepisów prawa regulujących relacje społeczne. W ostatnich dwóch etapach – poziomie pokonwencjonalnym – jednostka dzięki wykształconej zdolności abstrakcyjnego myślenia próbuje wypracować indywidualny system etyczny, odwołując się do poznanych zasad moralnych i własnych doświadczeń w świecie społecznym (Haidt, 2014; Żylicz, 2012). Dodatkowo dzieci zdają się po osiągnięciu odpowiedniego wieku rozróżniać reguły konwencjonalne, np. nakaz chodzenia do szkoły w mundurku, od reguł moralnych, np. zakaz wyrządzania krzywdy innym, i po otrzymaniu zgody od dorosłych na złamanie tych reguł akceptują naruszenie tych pierwszych, jednak sprzeciwiają się naruszeniu tych drugich. Teorie nurtu poznawczo-rozwojowego zakładają więc, że na moralność człowieka ma wpływ przede wszystkim rozwój struktur poznawczych, w tym zdolność logicznego i abstrakcyjnego myślenia, a także środowisko społeczne

jednostki. Jednakże istnieją uwagi krytyczne wobec tego modelu Kohlberga, jako że w swoich badaniach uwzględniał tylko sposób rozumowania chłopców. Nowsze ustalenia doniosły bowiem o wyraźnych różnicach w rozwoju moralnym chłopców i dziewczynek, co pozwala kwestionować trafność modelu amerykańskiego psychologa (zob. Środa, 2010, s. 167-168). Jednym z czołowych badaczy w nowym nurcie „afektywnym” jest Jonathan Haidt, który w swojej koncepcji fundamentów moralnych (*Moral Foundations Theory*) pierwszoplanowe miejsce przyznał emocjom i intuicjom, zakładając, że racjonalne myślenie służy ludziom jedynie do uzasadniania swych wyborów i przyznając mu rolę subtelного moderatora – metaforycznego jeźdźcy na słoniu (Haidt, 2014). Ponadto uznaje on moralność za specyficzną, wrodzoną adaptacją naszego gatunku, w której różne kultury różnią się jedynie pod względem hierarchii wyróżnianych przez jego teorię kodów (fundamentów). Sformułowana przez Haidta pierwsza zasada psychologii moralności trafnie oddaje jego intuicjonistyczne podejście i brzmi następująco: „Intuicje moralne pojawiają się automatycznie i niemal natychmiast, na długo przed rozpoczęciem rozumowania moralnego. Te pierwsze intuicje nadają kierunek naszemu późniejszemu rozumowaniu” (Haidt, 2014). Tym samym Haidt nawiązuje do i zgadza się z poglądem Hume’a, iż rozum jest jedynie sługą namiętności. Jego teoria zakłada, że moralność człowieka złożona jest z pięciu fundamentów (kodów): 1) troski/krzywdy, 2) sprawiedliwości/oszustwa, 3) lojalności/zdrady, 4) autorytetu/buntu i 5) czystości/upodlenia. W kolejnych latach badań psycholog dodał jeszcze szósty fundament – wolność/ucisk, jednakże większość naukowców w swych analizach odwołuje się do wersji z pięcioma kodami moralnymi, więc i ja w swojej pracy będę odwoływała się do tej wersji teorii. Pierwszy z fundamentów dotyczy reagowania na krzywdę i cierpienie innych, udzielaniu pomocy słabszym i potrzebującym. Drugi zobowiązuje do przestrzegania zasady wzajemności, odwziewania się za przysługę. Trzecia para dotyczy zobowiązań wobec własnej grupy, gotowości do poświęcania się dla jej dobra i potępienia jej zdrajców. Czwarty fundament nawiązuje do przestrzegania hierarchii społecznych oraz poszanowania dla tradycji i osób będących autorytetami dla danej grupy. A fundament piąty odwołuje się do poczucia czystości i nieczystości ciała i duszy. Dodatkowo Haidt przypisuje poszczególnym fundamentom odpowiednie (swoiste) emocje, na których mają one bazować. Mianowicie dla troski/krzywdy jest to uczucie empatii i współczucia, dla sprawiedliwości/oszukiwania są to poczucie wdzięczności za udzieloną pomoc oraz poczucie winy, jeśli nie mamy sposobności odwziewania się za przysługę, para lojalność/zdrada odwołuje się do poczucia dumy i patriotyzmu oraz złości i potępienia nielojalnych członków grupy, autorytet/bunt bazuje na emocjach szacunku i strachu, a fundamentowi

świętości/upodlenia odpowiadają emocje obrzydzenia i wstrętu (Haidt, 2014; Żylicz, 2012. s. 80). Bowiem emocje są procesami „sygnalizowania, że dzieje się coś istotnego z punktu widzenia dobrostanu jednostki” (Dąbrowski, 2012, s. 318) czy też realizacji jej celów i potrzeb. Dodatkowo dla pojawienia się emocji moralnych zdarzenie musi mieć znaczenie dla dobra grupy, do której jednostka należy i z którą się identyfikuje lub przynajmniej dla innej osoby niż sama jednostka. Zadaniem procesów poznawczych jest wtedy nadawanie/rozpoznanie znaczenia bodźców, a zadaniem emocji ocena tego znaczenia („nadanie etykiety” wartościującej). Jednocześnie emocje posiadają charakter motywujący – wiążą się z nimi określone dyspozycje do działania, tj. odpowiednio „zachęcające” jednostkę do zbliżenia się do przedmiotu lub skłaniające ją do wycofania się z otoczenia danego obiektu (Dąbrowski, 2014, s. 131-142). Niektórzy jednakże kwestionują jakoby różnym treściom moralnym, reprezentowanym przez pięć fundamentów, odpowiadały oddzielne i specyficzne emocje (zob. Gray, Schein, Cameron, 2017). Ponadto choć psychologowie i inni badacze odwołują się do komponentu afektywnego moralności, to nie zawsze precyzują, co dokładnie rozumieją pod pojęciem tych reakcji afektywnych i jak przebiegają one w organizmie czy też jakie zmiany wywołują. Bowiem współcześnie w psychologii termin „emocja” jest definiowany co najmniej na trzy różne sposoby w zależności od przyjętej teorii, a sposób jego rozumienia może mieć istotny wpływ na odpowiedź na pytanie, jak emocje wiążą się z procesami poznawczymi i racjonalnym myśleniem oraz jaka jest ich rola w sferze moralności. Mianowicie są to teorie: emocji podstawowych, emocji ocenowych i konstruktywistyczna. W pierwszej z nich wyróżnia się skończoną liczbę, najczęściej 6 emocji podstawowych – strach, złość, radość, smutek, wstręt, zaskoczenie – i uznaje się, że są to unikalne, biologicznie zdeterminowane, automatyczne mechanizmy ewaluacji bodźców. Ich uruchomienie wywołuje specyficzny dla każdej emocji zestaw reakcji afektywnych i zmian fizjologicznych (cielesnych) – subiektywne doznanie, program behawioralny i ekspresję mimiczną (Jasielska, 2013, s. 209). Dodatkowo twórcy teorii dopuszczają możliwość indywidualnej nauki autoewaluacji określonych bodźców i ich kulturowego przekazu. W teoriach ocenowych emocja powstaje dopiero w wyniku złożonej oceny (poznawczej), w której dodatkowo wyróżnić można dwa komponenty – poznawczy odpowiadający za pozyskiwanie informacji o zdarzeniach i identyfikowanie ich znaczenia oraz wartościujący, oceniający to zdarzenie w kategoriach korzyści lub szkód/strat dla organizmu. Oba te elementy oceny (poznawczej) stanowią podstawę dla pojawienia się odpowiedniej emocji pozytywnej (subiektywnie odczuwanej jako przyjemna dla organizmu) lub negatywnej (odczuwanej jako nieprzyjemna). Kluczowy jest tu zatem sposób interpretacji danego

zdarzenia, jako że związek między daną oceną, a powstaniem określonej emocji jest automatyczny i wrodzony. W tych teoriach emocje również składają się z procesów fizjologicznych, zdarzeń mózgowych, subiektywnych odczuć oraz dyspozycji behawioralnych (Dąbrowski, 2012, s. 320). Tym samym emocje pozwalają na czy przygotowują do odpowiedniej reakcji na daną sytuację, a więc i do skutecznego radzenia sobie z nią. Inaczej emocja jest rozumiana w teorii konstruktywistycznej, gdzie emocją nazywa się jedynie subiektywny aspekt odczuwania zmian zachodzących w ciele. Gray i współpracownicy proponują ideę emocji i ocen moralnych jako zjawisk emergentnych, powstałych przez interakcję kilku podstawowych mechanizmów (procesów) psychicznych, które można przyrównać do nakładających się na siebie okręgów w diagramach Venna (Gray i in., 2017, s. 41). Tym samym konstruowana emocja jest wynikiem synergicznego scalenia: 1) rdzennego pobudzenia afektywnego o danym stopniu intensywności (wysoki vs niski) i znaku pozytywnym lub negatywnym, 2) osobistych doświadczeń jednostki i 3) obowiązujących norm kulturowych (Jasielska, 2013, s. 215). Analogicznie ich teoria diadycznej moralności głosi, że u podstaw oceny działania jako moralnie złego leżą trzy elementy – negatywny stan rdzennego pobudzenia afektywnego, naruszenie obowiązujących w danej społeczności norm/standardów oraz postrzeganie czynu jako (potencjalnie) powodującego u kogoś krzywdę. W tym rozumieniu kluczowa dla moralności jest zdolność do postrzegania cierpienia u drugiego człowieka, wyrządzonego przez intencjonalnego agenta wobec (bezbronnego) pacjensa. Zatem każdy czyn postrzegany jako wywołujący u kogoś cierpienie będzie oceniany jako moralnie zły i odwrotnie, jeśli dane działanie określane jest mianem niemoralnego, to znaczy, że zostało w nim dostrzeżone rzeczywiste lub potencjalnie możliwe spowodowanie u kogoś krzywdy. Teoria ta zakłada więc zdolność ludzi do czytania innych umysłów (mentalizacji i symulacji stanów innych), którą również inni autorzy postrzegają jako kluczowy składnik moralności (Gray, Young, Waytz, 2012). Dodatkowo badania wskazują, że ludzie oceniają intencjonalne wykroczenia jako bardziej karygodne niż przypadkowe, nieumyślne uczynki – tym samym intencje sprawców mają znaczenie dla oceny moralnej. Różnice kulturowe teoria diadycznej moralności wyjaśnia poprzez odmienne postrzeganie krzywdy, tzn. dla jednej społeczności określony czyn będzie rozpoznawany jako przyczyna czyjegoś cierpienia, a dla innych już nie. Koncepcja ta poddaje zatem w wątpliwość oddzielność (odrębność) wyróżnianych przez Haidta modułów moralnych, uznając, że poszczególne fundamenty są jedynie wariacjami czy modyfikacjami leżącej u ich podstaw, naczelnej wartości troski/krzywdy. Tym samym w zmysł moralny zaangażowane są te same ogólne mechanizmy i zdolności poznawcze, co w innych zadaniach wymagających

rozumowania. Teza o rekrutacji ogólnych, a nie oddzielnych, specyficznych procesów w rozwiązywaniu dylematów moralnych znajduje potwierdzenie w badaniach z użyciem technik obrazowania mózgu (przedstawionych w sekcji 2.3). Ponadto na poparcie swojego stanowiska Schein i Gray podają przykłady badań, w których większość ludzi oceniała dany czyn jako jednocześnie dotyczący Haidtowskiego fundamentu czystości i troski, a w analizach statystycznych korelacja tych dwóch modułów jest zazwyczaj na tyle silna, iż zaciera się granica między nimi. Podobnie wysokie korelacje uzyskiwane są także między pozostałymi fundamentami (zob. Schein, Gray, 2018, s. 53-55). Ich zdaniem fundamenty moralne można rozpatrywać co najwyżej jako praktyczne klasyfikacje „moralnych gustów”, lecz nie jako oddzielne moduły czy mechanizmy psychiczne i mózgowie. Teoria Haidta niewiele mówi też o procesach leżących u podstaw tych modułów, których opisanie i wyjaśnienie powinno być istotnym zadaniem empirycznej nauki o moralności. W literaturze przedmiotu można znaleźć również rozróżnienie między emocjami a innymi stanami afektywnymi. Przykładowo sentymenty są bardziej trwałymi dyspozycjami do emocjonalnego i poznawczego reagowania oraz oceniania zdarzeń w określony sposób – „odnosi się je do trwałych sympatii bądź antypatii” (Frijda, 1998, s. 61). Ponadto nabywane są w wyniku doświadczeń i społecznego uczenia jednostek, a niektóre z nich mogą być też wrodzone. Podobnie sentymenty moralne definiuje Prinz (2014, s. 101). Zatem można sądzić, iż istnieje różnica definicyjna zarówno między emocjami a sentymentami (i innymi stanami afektywnymi), jak i pomiędzy samymi emocjami, którą nie zawsze precyzują badacze piszący o moralności. Warto także podkreślić, iż emocje nie wskazują na obiektywne cechy zdarzeń, lecz na ich znaczenie i implikacje dla dobrostanu oraz posiadanych celów i wartości danej jednostki, a w przypadku emocji moralnych również grup, do których jednostka należy. Można teraz zapytać, w jaki sposób emocje wiążą się z procesami poznawczymi i racjonalnego myślenia? Głos w tej sprawie oddam Haidtowi. Według amerykańskiego psychologa moralne rozumowanie polega na świadomym, tj. kontrolowanym i intencjonalnym przetwarzaniu i manipulowaniu informacji o charakterze społecznym. Z kolei intuicja jest w ujęciu Haidta natychmiastową, automatyczną oceną danego zdarzenia, gdzie uświadamiany jest co najwyżej wynik końcowy tego procesu oceny – akceptacja lub potępienie. Niemniej Haidt podkreśla, iż przeciwstawienie intuicji i rozumu (racjonalności) nie oznacza jednocześnie przeciwstawiania sobie procesów emocjonalnych i poznawczych (Haidt, 2008, s. 1029). Bowiem emocje zawierają zarówno elementy intuicji, afektywnego wartościowania, jak i rozumowania oraz ocen poznawczych – mają wartość informacyjną, a więc można je uznać za pewną formę poznania (*cognition*). Dodatkowo, choć nie pisze o tym

Haidt, emocje można uznać za (w pewnym sensie) racjonalne – bowiem pomagają ludziom zachowywać się zgodnie z posiadanymi standardami moralnymi/jakiegoś typu wiedzą, a więc zakładają możliwość posługiwania się przy tym regułami wnioskowania i racjonalnego rozumowania (zob. Scherer, 2011). Ponadto amerykański psycholog przyrównuje intuicje moralne do heurystyk myślowych, jako że ludzie posługują się nimi z tych samych powodów, tj. ograniczonych zasobów poznawczych i wydajności pracy umysłu. Bowiem są one formą „skrótów myślowych”, opierają się na zasadzie angażowania jak najmniejszego wysiłku poznawczego, a ponadto sprawdzają się w większości sytuacji życiowych i dopiero nietypowe czy wysoce konfliktowe sytuacje wymagają zaangażowania złożonych procesów świadomego rozumowania (tamże, s. 1032). Haidt wspomina również o badaniach związku intuicji i rozumowania z zachowaniem moralnym. Bowiem doniesienia Kohlberga o dodatniej korelacji między IQ, a wynikami w jego teście rozwiązywania dylematów moralnych okazują się mieć ograniczone potwierdzenia w nowszych badaniach. Przypadki psychopatii również zaprzeczają temu związkowi. Tym samym większość badań wskazuje na silną zależność emocji moralnych, a nie zdolności intelektualnych, i zachowania moralnego. Osoby ze zdiagnozowaną psychopatią posiadają bowiem wiedzę odnośnie zasad moralnych, potrafią odróżnić dobro i zło czy przewidywać krzywdzące konsekwencje danych zdarzeń, jednakże nie potrafią oszacować, czy ktoś będzie cierpiał z powodu danych działań, czy też nie (tamże, s. 1038; Cima, Tonnaer, Hauser, 2010). Wykazują oni deficyty niektórych reakcji emocjonalnych, np. wyrzutów sumienia, poczucia winy i smutku oraz w obrębie mechanizmu empatii. Badania nad osobami z psychopatią zdają się także wskazywać na istotność związku emocji, a nie racjonalnego rozumu, z motywacją i posiadaniem dyspozycji do określonych zachowań moralnych. Poglądom Haidta najbliższe zdaje się być do teorii emocji ocenowych lub połączenia założeń teorii ocenowych i teorii emocji podstawowych. Wobec takiego przedstawienia problemu dominująca odpowiedź na postawione wyżej pytanie głosi, iż emocje (czy też inne prostsze reakcje afektywne) pojawiają się już na początku wraz z odbiorem i gromadzeniem informacji o zdarzeniach. Tym samym mają one wpływ na występujące w dalszej kolejności procesy decyzyjne (rozumowania), nadawania „wag”/ustalania hierarchii ważności różnym informacjom oraz ostateczną odpowiedź i działanie jednostki. Zatem stany afektywne nie są tylko dodatkiem, lecz współwystępują i ściśle współpracują z „chłodnym” rozumowaniem w poznaniu moralnym. Warto również przywołać Haidtowski schemat powstawania ocen moralnych, który przedstawia się następująco: 1) bodziec wywołuje automatyczną reakcję cielesno-emocjonalną, 2) dochodzi do reprezentacji sytuacji i jej możliwych rozwiązań, 3) następuje ocena tych możliwości wraz

z pojawieniem się odpowiedniej emocji moralnej, 4) podjęcie decyzji, 5) interpretacja i uzasadnienie własnej reakcji (Urbaniak, 2014, s. 62-63). Wrodzone są tu zarówno same reakcje cielesno-emocjonalne „akceptacji i zbliżenia” oraz „odrzućenia i unikania” pewnych bodźców, jak i związek między tą pierwszą reakcją i pojawieniem się emocji pozytywnej oraz tą drugą reakcją i emocją negatywną. Są to odpowiedzi wyselekcjonowane przez dobór naturalny i dodatkowo wzmocniane przez społeczne praktyki i funkcjonowanie tabu (tamże). W prezentowanym podejściu intuicjonistycznym pierwsze (pierwotne) reakcje cielesno-emocjonalne mają postać prostych afektów (charakteryzujących się jedynie znakiem pozytywnym lub negatywnym czy też przyjemnością lub przykrością) i które to zdają się ukierunkowywać dalsze procesy poznawcze oceny i rozumowania. Niemniej istnieje możliwość pojawienia się wtórnych emocji moralnych – tj. reakcji afektywnych bardziej konkretnych i intencjonalnych, dotyczących danego aspektu sytuacji/bodźca, będących już rezultatem świadomych procesów myślenia. Możliwa jest również zmiana pierwszych intuicyjnych decyzji. Inspiracją dla koncepcji Haidta była, jak sam psycholog wspomina, trójwymiarowa teoria Shwedera dzieląca wartości moralne na 3 grupy, tj. etykę autonomii, etykę wspólnoty i etykę boskości oraz teoria modeli relacyjnych Alana Fiske, klasyfikująca moralność na podstawie czterech motywów: jedności, hierarchii, równości i proporcjonalności (Żylicz, 2012, s. 79-80). Każda z nich zakłada pewien stopień uniwersalności zasad moralnych i jednocześnie pewien relatywizm kulturowy w obrębie wyróżnionych wartości. Na podstawie teorii fundamentów powstał również kwestionariusz postaw moralnych – *Moral Foundation Questionnaire* (Graham i in., 2011), w polskiej adaptacji znany jako *Kwestionariusz kodów moralnych* (Jarmakowski-Kostrzanowski, Jarmakowska-Kostrzanowska, 2016). Dodatkowo Haidt wiąże tu uzyskanie określonych wyników przez badanych z ich politycznymi poglądami. Mianowicie uzyskanie wysokich wyników na skali troska i sprawiedliwość oraz niskich w pozostałych trzech skalach przypisywane jest liberałom/badanym z poglądami lewicowymi. Z kolei uzyskanie wielu punktów we wszystkich pięciu wymiarach oznacza, iż badany opowiada się po stronie konserwatystów/poglądów prawicowych. Choć koncepcja Haidta należy do najpopularniejszych współczesnych teorii moralności na gruncie psychologii, istnieją również głosy krytyczne zarówno wobec niej, jak i kwestionariusza (zob. Graham i in., 2013, s. 98-107; Schein, Gray, 2018). Niektórzy wskazują, że teoria ta nie odnotowuje istotnych różnic między altruizmem krewniaczym i odwzajemnionym, zakładając że tak samo ważna będzie dla człowieka pomoc jego rodzinie, co nieznajomym – choć niektóre badania wskazują inaczej (zob. Curry, Whitehouse, Mullins, 2019, s. 49-50). Ponadto kwestionowana

jest wrodzoność pięciu fundamentów moralnych oraz założenie co do ich modułowości i odrębności względem siebie (Suhler, Churchland, 2011). Jeśli chodzi o związki emocji i racjonalnego myślenia proponowane stanowisko teorii fundamentów moralnych (rozum jako sługa emocji) jest konfrontowane z twierdzeniem o współwystępowaniu trybu automatycznych reakcji i „chłodnego” rozumowania w neurokognitywnej teorii dwóch procesów Greene’a (Graham i in., 2013, s. 101). Z tego względu Oliver Curry i współpracownicy (Curry i in., 2019) zaproponowali teorię moralności jako współpracy (*Morality-as-Cooperation*), którą można uznać za nieco zmodyfikowaną i rozbudowaną wersję koncepcji Haidta. Bowiem wyróżnia się w niej siedem głównych wartości moralnych, którymi kierują się ludzie, tj. pomaganie członkom rodziny, lojalność grupowa, wzajemność (odwzajemnianie przysług), dzielność, szacunek, sprawiedliwość i poszanowanie własności. Ich głównym założeniem jest traktowanie moralności jako biologiczno-kulturowej odpowiedzi na potrzebę współpracy społecznej, a więc i jako strategię przetrwania gatunku. Tym samym na mechanizm moralności składają się wrodzone instynkty (sposoby reagowania) i intuicje oraz inwencje kulturowe i instytucje społeczne (tamże, s. 48). Curry i współpracownicy zadają tym samym pytanie, czy reguły moralne mają charakter uniwersalny, czy też są wysoce zróżnicowane kulturowo? Na podstawie uzyskanych wyników odwołują się do poglądów Hume’a na naturę moralności (podobnie jak Haidt) i konkludują, iż nie znajdują dowodów na skrajny relatywizm kulturowy, lecz międzykulturową zgodność co do przyjmowania określonych wartości i praktyk jako moralnie dobrych lub złych (tamże, s. 55). Ponadto z ich artykułu można też wyprowadzić tezę, że nie istnieją zasady, które można by spisać i uznać za obowiązujące we wszystkich społecznościach i kulturach, a raczej są to pewne wartości, które można uważać za uniwersalne i naturalne tendencje człowieka i dopiero na podstawie których społeczeństwa formułują odpowiednie reguły moralnego postępowania – część w formie pisemnych zakazów i nakazów, pozostałe w formie podświadomej wiedzy, nabywanej w wyniku interakcji społecznych – dlatego tak trudno też jest znaleźć i wyartykułować konkretne uniwersalne zasady moralne, a łatwiej odkryć pewne wartości jako naturalne dyspozycje przejawiane przez ludzi międzykulturowo. Z tego względu konkretne zasady mogą się różnić w zależności od kultury, ale wartości jakie leżą u ich podstaw i na których one bazują są uniwersalne. Zatem dowody psychologii zdają się wskazywać na założenia słabego czy umiarkowanego natywizmu jako najbardziej trafnego stanowiska w sprawie istnienia i rozwoju zmysłu moralności.

2.3 Moralność z perspektywy neurokognitywistyki

Współcześnie istnieją dwa główne paradygmaty badań nad udziałem emocji i racjonalnego rozumowania w nauce o moralności. Pierwszy przeciwstawia sobie oba systemy, zakładając ich rywalizację w procesach rozwiązywania dylematów i wydawania ocen moralnych. Jednakże drugi nurt badań poszukuje dowodów na współpracę systemu emocjonalnego i racjonalnego zmysłu moralności (De Oliveira-Souza, Moll, Grafman, 2011). Ponadto w badaniach tych widoczny jest coraz większy udział zespołów interdyscyplinarnych, łączących perspektywy psychologii, filozofii, neuronauk, antropologii, biologii czy nawet ekonomii. Bowiem klasyczne dziedzinowe podejście, wąsko patrzące na zagadnienie moralności w obliczu stałego wzrostu publikacji i wiedzy naukowej może nie zdołać uchwycić wszystkich kluczowych elementów i szczegółów, skupiając się wyłącznie na skali mikro, podczas gdy do opisu i wyjaśnienia złożonych zjawisk społecznych, takich jak moralność, potrzebna jest również skala makro. Tym samym po wprowadzeniu filozoficznym i przeglądzie badań psychologii (która sama w znacznym stopniu uwzględnia odkrycia antropologii) na temat moralności przejdę teraz do perspektywy neurokognitywistyki, której badania dopiero ostatnimi laty zostały włączone do dyskursu nad tym zagadnieniem, a szczególnie rolę emocji moralnych i ich związku z rozumowaniem. Dziarnowska i Przybysz definiują emocje moralne jako „pojawiające się w szczególnych sytuacjach społecznych, tj. jako stany afektywne towarzyszące ocenie lub poprzedzające ocenę jakiegoś zachowania w kategoriach „dobra”/„zła”, „słuszności”/„niesłuszności”, „powinności”/„niepowinności”” (Dziarnowska, Przybysz, 2010, s. 58) – a więc nie są tylko reakcjami na dokonany wybór, jak twierdzą niektórzy. Ponadto przywołują klasyfikację emocji społecznych Antoniego Damasio, do których zalicza on zakłopotanie, wstyd, pogardę, oburzenie, współczucie, litość, respekt, uniesienie, wdzięczność i dumę – wśród nich wyróżnia dodatkowo węższą klasę emocji moralnych. „Troska o innych” jest właściwością pozwalającą na odróżnienie tego rodzaju emocji od innych, np. emocji podstawowych, oraz określanie ich mianem społecznych. Dodatkowa właściwość – ocenianie zdarzeń w kategoriach zgodności z obowiązującymi standardami moralnymi – czyni ponadto część z nich emocjami moralnymi. Podobnej klasyfikacji dokonują Kroll i Egan (2004), wskazując emocje wstydu, poczucia winy, skruchy (wyrzutów sumienia) i żalu (ubolewania). Dziarnowska i Przybysz (2010, s. 58) dokonują jeszcze dodatkowego podziału emocji moralnych na epistemiczne i nieepistemiczne. Te pierwsze posiadają bogatą treść poznawczą – wiemy kogo lub czego dotyczą oraz w wyniku jakiego bodźca się pojawiły, z kolei te

drugie są jedynie prostym odruchem akceptacji lub odrzucenia wobec pewnych zachowań – ludziom trudno przychodzi wytłumaczenie tych odczuć (wąska treść poznawcza). Moll i współpracownicy (Moll i in., 2007) także proponują autorski podział emocji moralnych na trzy grupy: 1) promujące zachowania i postawy prospołeczne, np. wzajemną pomocą i współpracę, 2) emocje afiliacyjne/empatyzowania z innymi (przywiązania społecznego), 3) emocje krytyczne/strategiczne, w tym karzące za nieprzestrzeganie zasad czy zdradę własnej społeczności oraz utrzymujące spójność grupy. Zakładają też, że każdej z nich odpowiada inny wzór aktywności neuronalnej. Dalej wskazują obszary kory orbito-frontalnej i skroniowej, skrzyżowanie skroniowo-ciemieniowe oraz podkorowe struktury limbiczne jako centra „moralnej wrażliwości” oraz istotnie związanych z nią pojęć agencyjności, intencjonalności oraz zdolności wnioskowania o stanach umysłu innych osób i przewidywania skutków zdarzeń. Przy okazji warto wspomnieć o jeszcze innym podziale na negatywne i pozytywne emocje moralne, jako że przez długi czas wyróżniano zasadniczo tylko tę pierwszą grupę, czyli wstyd, pogardę, złość/oburzenie, poczucie winy i zażenowanie/zakłopotanie, a grupę pozytywnych emocji moralnych – uniesienie, współczucie, wdzięczność i moralną dumę uwydatniono dopiero ostatnimi czasy (Dziarnowska, Przybysz, 2010; Tangney, Stuewig, Mashek, 2007). Zdaniem Prinza sentymentalizm zawdzięcza naukom kognitywnym swój powrót do dyskusji o naturze moralności. Ostatnie dwadzieścia lat badań postaw i sposobów rozwiązywania dylematów moralnych z wykorzystaniem technik neuroobrazowania mózgu przychyliła się do obserwacji niektórych filozofów o zaangażowaniu czy głównej roli emocji w wydawaniu sądów i podejmowaniu decyzji moralnych (Prinz, 2014, s. 101). Tym samym większość neurokognitywistów poddała ostrej krytyce stanowisko racjonalizmu etycznego Kanta. Badania eksperymentalne wprowadzające ludzi w dane stany emocjonalne wykazały ich wpływ na końcową odpowiedź osób badanych. Mianowicie wywoływanie emocji wstrętu/obrzydzenia przez przykre zapachy lub przebywanie w brudnym pomieszczeniu skutkowało surowszym ocenianiem czynów jako moralnie złych oraz częstszym wyborem utylitarystycznych rozwiązań. Przeciwny efekt uzyskiwano po wprowadzeniu badanych w pozytywne stany emocjonalne. Ponadto, gdy poproszono ich o podanie powodów swoich ocen i sądów, badani polegli raczej na własnych odczuciach niż faktach i racjonalnych dowodach. Na tej podstawie Prinz wnioskuje, że emocje nie tylko są konieczne dla rozumowania moralnego, lecz same wystarczą dla oceniania zdarzeń jako moralnie dobrych lub złych (tamże). Chociaż chodzi tu o szczególną kategorię – emocje moralne, wśród których wyróżnić można te nakierowane na innych (*other-directed*), np. złość/oburzenie,

pogarda oraz te skierowane na własną osobę (*self-directed*), np. poczucie winy, wstyd i duma/uniesienie. Dodatkowo określa się je mianem emocji refleksyjnych, też samoświadomościowych (Haidt, 2003, s. 859; Tangney i in. 2007; zob. także Imbir, 2012, s. 45-47). Konkretnie wartości moralne są wtedy dyspozycjami do emocjonalnego reagowania (Prinz nazywa je także sentymentami) na zdarzenia dotyczące własnego „ja” i/lub innych. Są one „kompasem” zgodności określonych zachowań i postaw z posiadanymi przez daną osobę i jej grupę standardami (wartościami, zasadami) moralnymi. Tym samym Prinz zajmuje stanowisko identyczne do Hume’a, twierdząc, że rozum służy do zbierania informacji i ogólnej reprezentacji sytuacji, lecz to emocje skłaniają do oceny słuszności lub niesłuszności danego zdarzenia – rozum podległy emocjom. Inni badacze opowiadają się za hipotezami o współwystępowaniu i rywalizacji procesów afektywnych i racjonalnych (zob. Greene, Nystrom, Engell, Darley, Cohen, 2004; Greene, 2014), czy też ich wzajemnej współpracy (zob. Damasio, 1999), a ponadto niektórzy podkreślają, że choć emocje mogą być ważnym komponentem zmysłu moralnego, to ich wpływ na podejmowanie decyzji w dylematach moralnych może być mniejszy, niż sądzono (Horne, Powell, 2016; zob. także Huebner, Dwyer, Hauser, 2009) – czyżby zatem udział emocji był warunkiem koniecznym, lecz niewystarczającym dla formułowania sądów o charakterze moralnym? Aby sprawdzić te przypuszczenia, badane są sposoby rozumowania i zachowania moralne zarówno u osób zdrowych, jak i z różnymi zaburzeniami psychopatologicznymi. Szczególnie często brane pod uwagę przypadki osób z psychopatią. Bowiem zdają się one mieć zachowaną zdolność do racjonalnego myślenia, lecz wykazują głębokie deficyty w sferze emocjonalnej, szczególnie jeśli chodzi o odczuwanie emocji negatywnych i samoświadomościowych (autorefleksyjnych), co przekłada się także na zaburzenia w sferze moralności. Odnotowano u nich zmniejszoną aktywność w ciele migdałowatym oraz brzuszno-przyśrodkowej korze przedczołowej (VMPFC), czyli struktur zaangażowanych w przetwarzanie bodźców emocjonalnych oraz w trakcie rozwiązywania dylematów moralnych (Blair, 2007; May, Workman, Han, Haas, 2020). Ponadto wskazuje się na obniżoną umiejętność uczenia się instrumentalnego opartego o kary oraz wspomniane w poprzedniej sekcji deficyty empatii, czyli zdolności przyjmowania perspektywy innych, rozpoznawania i odpowiedniego reagowania na stany emocjonalne (zwłaszcza cierpienie) innych osób. Efekty te są prawdopodobnie wynikiem nieprawidłowości funkcjonowania wymienionych wyżej struktur. Inni wskazują dodatkowo na deficyty funkcjonowania w obrębie zakrętu obręczy i zakrętu kąтового (Glenn, Raine, Schug, 2009). Choć w badaniach tych przeważają doniesienia o nieprawidłowościach (głównie obniżonej aktywności) okolic ciała migdałowatego i VMPFC

oraz połączonych z nimi struktur limbicznych, nowa metaanaliza (Deming, Koenigs, 2020) wskazuje na powiązanie psychopatii ze zredukowaną aktywnością przedniej części kory obręczy (ACC) i jednocześnie wzmożoną aktywnością tylnej części kory obręczy (PCC), przedklinka, grzbietowo-przyśrodkowej kory przedczołowej oraz kory orbito-frontalnej, środkowej części płata skroniowego, ciała migdałowatego i okolic hipokampa w prawej półkuli. Tym samym Deming i Koenigs wskazują na dysfunkcję dwóch dużych sieci nerwowych, tj. sieci aktywności bazowej (*default mode network*) i sieci istotności (*salience network*) w psychopatii (czasem także trzeciej sieci czołowo-ciemieniowej). Ponadto, jak wynika z badań nad osobami z uszkodzoną VMPFC, zdaje się, że istnieje okres krytyczny w rozwoju dziecka, w którym uczy się ono norm i zachowań moralnych oraz społecznych, tj. nabywa wiedzę moralną. Bowiem jeśli uszkodzenie mózgu w obrębie okolic VMPFC miało miejsce w dorosłym życiu pacjenta, to nadal posiada on wiedzę odnośnie zasad moralnych i właściwych zachowań społecznych, jednak częściowo traci zdolność postępowania według tej wiedzy – stąd często nazywa się ich przypadki nabytą socjopatią (Damasio, 2007). Chociaż dzięki warunkowaniu emocjonalnemu, które jest istotne dla rozwoju moralnego, mogą oni nadal reagować adekwatnie do obowiązujących standardów społecznych wobec znanych im sytuacji. Niemniej jeżeli do urazu doszło w okresie dzieciństwa, osoby te mają wyraźne trudności z opanowaniem reguł dotyczących życia społecznego i moralności (zob. Damasio, 1999) – takie przypadki określa się z kolei terminem rozwojowej socjopatii. Istotnie istnieją badania świadczące, że osoby z psychopatią posiadają wiedzę odnośnie dobra i zła oraz potrafią odróżniać czyny moralnie właściwe od niewłaściwych, jednakże nie umieją postępować zgodnie z tą wiedzą (Cima, Tonnaer, Hauser, 2010). Prawdopodobnie nie potrafią oni w pełni zrozumieć odczuwanego przez drugiego człowieka cierpienia (McAuliffe, 2019). Osoby te częściej wybierają także rozwiązania utylitarystyczne (Bartels, Pizarro, 2011.; zob. także Koenigs, Kruepke, Zeier, Newman, 2012), a ponadto traktują normy moralne jak zasady konwencjonalne czy formy zwyczajów.

Najsłynniejszą koncepcją zmysłu moralnego na gruncie neurokognitywistyki jest teoria dwóch procesów (*The Dual-Process Theory*), znana też pod nazwą moralnych trybów (*Moral Tribes*), Joshuy Greene’a. Zakłada się w niej, że mamy silną tendencję do automatycznego i emocjonalnego reagowania na krzywdę innych – a więc decyzji o charakterze deontologicznym, jednakże możemy polegać także na procesach świadomego, racjonalnego rozumowania i kontroli poznawczej – i tym samym dokonywać wyborów utylitarystycznych („większego dobra”). Jej założenia są podobne do tych w koncepcji moralnych fundamentów Haidta na gruncie psychologii, choć istnieją też istotne różnice

między nimi. W formułowaniu swojej teorii Greene oparł się na wynikach korelacyjnych odpowiedzi i aktywności mózgu osób badanych podczas rozwiązywania różnych problemów moralnych, a najczęściej tzw. dylematu wagonika (Greene, 2014). W wersji tego dylematu ze zwrotnicą (*switch dilemma*) rozpędzony wagonik zmierza w kierunku pięciu osób znajdujących się na torach bez możliwości ucieczki. Jednak można je uratować, przestawiając zwrotnicę i tym samym skierować wagonik na tor, gdzie znajduje się tylko jedna osoba. Zadaniem osoby badanej jest dokonanie wyboru o przestawieniu zwrotnicy i uratowaniu pięciu osób lub zaniechaniu tego działania i ocaleniu jednej osoby. Natomiast w dylemacie wagonika w wersji z kładką (*footbridge dilemma*) rozpędzony wagonik także zmierza w kierunku pięciu osób stojących na torach, jednak badany może ich ocalić, jeśli zdecyduje się własnoręcznie zepchnąć osobę stojącą na kładce kolejowej. Bowiem wtedy masa ciała takiej osoby zatrzyma wagonik, jednakże jednocześnie doprowadzi to do jej śmierci. Warto również wspomnieć, iż tym co odróżnia dylematy moralne od innych dylematów jest fakt, że w tych pierwszych osoba badana musi wybierać pomiędzy dwoma alternatywami działania, z których każda przyczynia się w jakimś stopniu do cierpienia innych lub też narusza jakąś normę moralną, a w tych drugich osoba wybiera pomiędzy dwoma opcjami o neutralnej wartości moralnej, nie powodując krzywdy u innych (Przybysz, Dziarnowska, 2012, s. 48). Dodatkowo wyróżnia się dylematy moralne osobiste i nieosobiste. W tych pierwszych badany wyrządza krzywdę osobiście, przez bezpośredni fizyczny kontakt, a w tych drugich zadaje cierpienie „na odległość”, za pomocą jakiegoś narzędzia (pośrednika). Tym samym dylemat wagonika w wersji ze zwrotnicą zalicza się do dylematów nieosobistych, natomiast wersja z kładką ma charakter osobisty. Co ciekawe w wersji ze zwrotnicą większość badanych decyduje się na jej przestawienie, a więc dokonanie wyboru utylitarystycznego, podczas gdy w wersji z kładką większość osób deklaruje, że nie zepchnęliby człowieka, tym samym wybierając opcję deontologiczną. Ponadto tym rozbieżnościom w odpowiedziach odpowiadają różnice w aktywności neurofizjologicznej osób badanych, mierzonej za pomocą funkcjonalnego obrazowania metodą rezonansu magnetycznego (fMRI). Podczas rozwiązywania dylematów personalnych zanotowano zwiększoną aktywność mózgu w okolicach związanych z emocjami i przetwarzaniem bodźców społecznych, tj. przyśrodkowej kory przedczołowej – szczególnie VMPFC, tylnej części zakrętu obręczy, przedklinka, górnej bruzdy skroniowej, zakrętu kąтового, skrzyżowania skroniowo-ciemieniowego oraz ciała migdałowatego. Z kolei w dylematach niepersonalnych obserwuje się większą aktywność mózgu w obszarach związanych z „chłodnym” rozumowaniem (pamięcią roboczą, abstrakcyjnym myśleniem, procesami

kalkulowania zysków i strat) w porównaniu z „okolicami afektywnymi”, czyli w grzbietowo-bocznej korze przedczołowej (DLPFC) i dolnej części płatów ciemieniowych. Dodatkowo obserwuje się różnice czasów reakcji między tymi dwoma rodzajami dylematów. Mianowicie w dylematach o charakterze osobistym badanym decydującym się na zepchnięcie człowieka z kładki dłużej zajęło dokonanie wyboru w porównaniu z czasami reakcji osób sprzeciwiających się temu czynowi. Ponadto średni czas podejmowania decyzji w dylematach nieosobistych był krótszy od średniego czasu obu typów decyzji w dylematach personalnych. Zatem trudniej przychodzi nam osobiste skrzywdzenie drugiego człowieka – prawdopodobnie potrzeba więcej czasu, by stłumić silną reakcję emocjonalną (Przybysz, Dziarnowska, 2012, s. 50-51). Ponadto zakłada się, że badani w podejmowaniu takich decyzji świadomie lub nieświadomie stosują czy odwołują się do trzech zasad (czynników): 1) intencjonalności czynu – celowe działanie prowadzące do czyjejś krzywdy jest oceniane gorzej, niż w przypadku krzywdy będącej niezamierzonym skutkiem ubocznym działania, 2) działania – krzywda spowodowana działaniem jest oceniana jako gorsza, niż w przypadku jej wyrządzenia przez zaniechanie, 3) kontaktu – bezpośrednie wyrządzenie krzywdy (w fizycznym kontakcie z drugim człowiekiem) jest gorsze niż cierpienie zadane pośrednio, „na odległość” (Dziarnowska, Przybysz, 2007, s. 59; Greene i in., 2009). Bowiem wyrządzenie krzywdy przez bezpośredni kontakt fizyczny z drugim człowiekiem jest traktowane jako „osobista przemoc”, którą dodatkowo Greene określa mianem ewolucyjnie pierwotnej. Odwołuje się przy tym do wyjaśnienia ewolucyjnego, że w historii gatunku dominowała bezpośrednia przemoc fizyczna, nie było „narzędzi” do zadawania jej pośrednio. Z tego względu z osobistymi dylematami wiąże się też zwiększony udział procesów emocjonalnych. Jest to „ewolucyjnie stary” mechanizm radzenia sobie z tego typu problemami społecznymi, rodzi się już z pewnym stopniem gotowości do emocjonalnego reagowania, a badania obserwacyjne wskazują, że również inne naczelne przejawiają szereg emocji społecznych. Jednakże naczelne posiadają mniejsze zdolności intelektualne w porównaniu z człowiekiem, przez co emocjonalne reagowanie staje się ich jedyną strategią radzenia sobie w sytuacjach konfliktów interindywidualnych. Tym samym ludzie także posiadają te wrodzone, intuicyjne mechanizmy emocjonalnego reagowania, które w połączeniu ze znacznie większymi zdolnościami do abstrakcyjnego myślenia i wnioskowania doprowadziły do powstania drugiego sposobu radzenia sobie z tego typu problemami – racjonalnym rozumowaniem i kalkulowaniem zysków i strat wynikających z danego zachowania. Różnice w zapisie fMRI odnotowano również pomiędzy badanymi dokonującymi wyborów deontologicznych i utylitarystycznych. Wyniki badań wskazują na zwiększoną aktywność obszarów związanych

z procesami emocjonalnymi i przetwarzaniem bodźców społecznych, jeśli badany zdecydował się na odpowiedź deontologiczną. Z kolei w wypadku wyborów utylitarystycznych zanotowano zwiększoną aktywność okolic związanych z procesami *stricte* kognitywnymi i kontrolnymi, tj. DLPFC, przedniego zakrętu kory obręczy, dolnej części płatów ciemieniowych oraz w częściach lewego (zakręcie dolnym) i prawego (zakręcie górnym i środkowym) płata skroniowego (Greene, Sommerville, Nystrom, Darley, Cohen, 2001; Greene i in., 2004). W tych nowszych badaniach Greene i współpracownicy zwracają szczególną uwagę na obszar ACC, jako że prawdopodobnie jest on związany z występowaniem konfliktu poznawczego, choć niektórzy wskazują także na jego powiązanie z emocjami. Grupa badawcza Greene’a stara się to wyjaśniać obecnością silnego konfliktu wewnętrznego pomiędzy reakcją emocjonalną na cierpienie innych i racjonalną kalkulacją korzyści i strat w sytuacji dylematu moralnego. Tym samym dowodzą na rzecz moralności jako złożonej zarówno z komponentu afektywnego, jak i racjonalnego – co oznacza, że sprzeciwiają się czysto racjonalistycznemu podejściu oraz skrajnemu emotywizmowi. Wyjaśniają, że istnienie tych dwóch trybów (systemów) moralności jest pewnego typu kompromisem między wydajnością (efektywnością) i elastycznością pracy mózgu, wypracowanym prawdopodobnie przez dobór naturalny. Niemniej Greene w swoich nowszych pracach twierdzi, że choć intuicje są przydatne w radzeniu sobie z łatwymi problemami, to nie wystarczą do rozwiązywania złożonych dylematów moralnych i tym samym postuluje „wyższość” racjonalnej strony moralności (zob. Greene, 2014). Amerykański naukowiec formułuje jednocześnie „odważne” wnioski o charakterze normatywnym, tj. odrzucenie deontologizmu jako zbyt automatycznego, nieelastycznego sposobu rozwiązywania złożonych dylematów moralnych i słuszność postępowania według założeń utylitarystyki – który nazywa też głębokim pragmatyzmem. Można oceniać tę propozycję jako „zbyt odważną”, biorąc pod uwagę niekonkluzywność i niejednoznaczność dotychczasowych wyników badań. Niemniej należy docenić próbę Greene’a, jako że właśnie wypracowanie sposobu przejścia od czysto deskryptywnych wniosków do formułowania tych o charakterze normatywnym jest jednym z głównych zadań neuroetyki. Podsumowując wyniki badań Greene’a i współpracowników, decyzje deontologiczne i dylematy osobiste angażują w większym stopniu procesy emocjonalne – przede wszystkim VMPFC, zaś w trakcie podejmowania decyzji utylitarystycznych i dylematach nieosobistych, odnotowuje się znacznie bardziej aktywne okolice DLPFC związane z „chłodnym myśleniem” i procesami racjonalnego kalkulowania. Jest to główny wniosek płynący z badań Greene’a i współpracowników, potwierdzany także w badaniach innych grup naukowców zarówno w

odniesieniu do zdrowych uczestników badań, jak i przypadków neuropsychologicznych. Bowiem w zadaniach rozwiązywania personalnych dylematów moralnych istniało większe prawdopodobieństwo, że osoby z uszkodzeniami VMPFC wybiorą opcję utylitarystyczną w porównaniu z grupami kontrolnymi (Ciaramelli, Muccioli, Làdavas, Di Pellegrino, 2007; Koenigs i in., 2007). Mendez, Anderson i Shapira (2005) wskazują na uzyskiwanie podobnych rezultatów w grupie osób z czołowo-skroniową demencją, w której najpierw dochodzi do uszkodzeń części brzusznej tych płatów. Badacze wiążą więc obszar VMPFC z kategorią emocji moralnych, a także samoświadomościowych (Ciaramelli i in., 2007). Zdaje się zatem, że aktywność i procesy związane z VMPFC stanowią „wskazówkę” co do słuszności danego zachowania czy też jego zgodności z obowiązującymi standardami etycznymi. Antonio Damasio podkreśla tutaj wybiórczość (specyficzność) uszkodzeń, które mają znaczenie dla postępowania moralnego. To znaczy, uszkodzenie okolic kory orbito-frontalnej, szczególnie VMPFC, wywołuje objawy nabytej socjopatii, a defekty DLPFC nie powodują podobnych zaburzeń w sferze afektywnej, lecz *stricte* deficyty poznawcze (Damasio, 2007). Do takich samych wniosków dochodzą również inni naukowcy i grupy badawcze (zob. Young, Koenigs, 2007). Stąd Damasio wnosi, iż jedną z funkcji VMPFC jest wyzwalanie adekwatnych do sytuacji programów postępowania społecznego (w tym emocjonalnego reagowania), nabytych w drodze osobistych doświadczeń za pomocą mechanizmu warunkowania emocjonalnego. Region ten posiada ponadto połączenia z DLPFC oraz innymi obszarami korowymi i podkorowymi, co sugeruje powiązanie czy nawet współpracę dwóch systemów (emocjonalnego i racjonalnego) w podejmowaniu trafnych decyzji. Jednakże uszkodzenie którejś ze współzależnych okolic, np. kory somatosensorycznej w prawej półkuli, kluczowej dla zdolności przyjmowania czyjejś perspektywy, także może skutkować określonymi „deficytami moralności”. Według Damasia wyniki te wskazują, że emocje nieodzownie łączą się z rozumem, wspomagając procesy nabywania wiedzy o wartościach moralnych i zasadach życia społecznego oraz podejmowania decyzji zgodnych z tymi standardami. Dodatkowo proponuje własną hipotezę markerów somatycznych jako szeregu reakcji psychofizjologicznych i ich odczuć emocjonalnych. Jednocześnie wnosi, iż wrodzone są tu pewne dyspozycje (system preferencji unikania bólu/przykrości i szukania przyjemności), a mechanizmy ich wywoływania, czyli kojarzenie somatyczne (warunkowanie emocjonalne) określonych typów bodźców z odpowiednią reakcją zostają uformowane w drodze socjalizacji i uczenia (Damasio, 1999, s. 199-208). Jednakże Damasio w przeciwieństwie do sentymentalistycznego stanowiska Prinza uważa, że samo reagowanie emocjonalne nie wystarcza do podjęcia trafnych decyzji czy

wydania sądu moralnego i konieczne jest zaangażowanie również „chłodnych” procesów analizy i wnioskowania. „Markery somatyczne zwiększają prawdopodobnie trafność i efektywność procesów decyzyjnych. Ich brak natomiast ową trafność i efektywność redukuje” (Damasio, 1999, s. 200). Dalej konkluduje, że są to mechanizmy ukierunkowujące, uwypuklające niektóre opcje, co w rezultacie umożliwia wypracowanie szybkiej i adekwatnej do sytuacji odpowiedzi. Ponadto zdaniem Damasia wybieranie opcji utylitarystycznych zarówno przez zdrowych uczestników badań, jak i pacjentów z uszkodzeniami VMPFC można określić jako nieroztropność czy brak „zdrowego rozsądku” – zdają się być wynikiem „wyłączenia” wiedzy odnośnie wypracowanych w przeszłości sposobów rozwiązywania dylematów moralnych i społecznych. Jednocześnie neurobiolog zwraca uwagę, iż procesy umysłowe wywoływane aktywnością tych obszarów przedczołowych są uniwersalne, tzn. biorą udział we wszelkiej aktywności poznawczej człowieka – argumentując przeciwko koncepcjom istnienia specjalistycznego modułu moralności w mózgu. Bowiem tym co odróżnia wiedzę moralną i sposób rozwiązywania dylematów moralnych jest jedynie ich wyraźny i istotny czynnik afektywny. Niemniej sfera moralności jest tylko jedną z kategorii wiedzy oraz sposobów funkcjonowania w świecie społecznym, która to sfera społeczna należy z kolei do ogólnego zbioru wiedzy człowieka. Przeto, choć w badaniach neurokognitywistyki nad moralnością systematycznie odnotowuje się aktywność pewnych obszarów mózgu, to nie ma dowodów na istnienie specjalnego modułu (zmysłu) moralnego w mózgu i wszelkie próby jego znalezienia wydają się być skazane na niepowodzenie (tamże, s. 5). Do podobnych wniosków dochodzą Young i Dungan (2012). Niemniej, pomimo przeważającej liczby artykułów głoszących o udziale elementu afektywnego w konstruowaniu sądów moralnych, Huebner, Dwyer i Hauser (2009) zauważają, że nie ma jeszcze „twardych” dowodów, potwierdzających słuszość takich koncepcji. Bowiem zapis fMRI nie ma na tyle odpowiedniej rozdzielczości czasowej, by stwierdzić, czy obserwowana aktywność mózgu badanych w zadaniach rozwiązywania dylematów moralnych pojawia się przed, w trakcie czy dopiero po podjęciu decyzji czy wydaniu danej oceny. Ponadto mimo popularności teorii dwóch procesów Greene’a i współpracowników oraz wiarygodności jej założeń, istnieją wobec niej także uwagi krytyczne i wyniki badań sprzeczne z jej twierdzeniami i hipotezami. Przykładowo Bago i De Neys (2019) poprosili osoby badane najpierw o intuicyjne odpowiedzi w dylematach moralnych, a następnie dali im czas na ich ponowne przemyślenie. Okazało się, że większość badanych, którzy po zastanowieniu wybierali opcję utylitarystyczną, już w pierwszych odpowiedziach przyjmowali to stanowisko – co podważa wyjaśnienia o konieczności angażowania wymagających czasu i

wysiłku poznawczego procesów racjonalnego rozumowania. Inne badania eksperymentalne donoszą również o większej tendencji do wyboru tego typu odpowiedzi wśród osób o zmniejszonej awersji do krzywdzenia innych, uzależnionych od alkoholu/narkotyków, o obniżonym poziomie empatii i serotoniny oraz z osobowością psychopatyczną i makiaweliczną. Takie same rezultaty uzyskano także u osób z deficytami w przetwarzaniu informacji społecznych, m.in. w zaburzeniach ze spektrum autyzmu (Duke, Begue, 2015, s. 122). Niemniej notuje się mniej przypadków wyboru opcji utylitarystycznych w próbach z presją czasową i przeciążeniem poznawczym osób badanych – co potwierdza założenia teorii dwóch procesów Greene’a, że duże znaczenie w wyborze takich odpowiedzi mają świadome procesy rozumowania. Nawiązując do innych jeszcze założeń tej koncepcji, zwiększona reaktywność emocjonalna i ograniczenia w funkcjonowaniu wyższych czynności poznawczych powinny mieć wpływ na odpowiedzi w dylemacie wagonika (szczególnie w wersji z kładką), tj. zwiększać frekwencję odpowiedzi deontologicznych. Jednakże wyniki badań Duke i Begue okazały się przeciwne do tych założeń i stężenie alkoholu we krwi (wywołujące wspomniane wyżej skutki emocjonalne i poznawcze) pozytywnie korelowało ze skłonnością do wyboru odpowiedzi utylitarystycznych (tamże, s. 123-124). Ponadto badaczki twierdzą, że wyjaśnianie wyborów deontologicznych przez odwołanie się do zmniejszonego poziomu empatii badanych wydaje się tak samo prawdopodobne, jak oryginalne tłumaczenia tych wyborów zwiększonym stopniem zaangażowania świadomych procesów racjonalnych. Wyniki te nie potwierdzają zatem do końca założeń modelu Greene’a i zachęcają do dalszych badań weryfikujących teorię dwóch procesów. Ponadto w niektórych badaniach osoby z uszkodzoną VMPFC przejawiały silną tendencję do emocjonalnego (i jednocześnie mniej racjonalnego) reagowania w grze w ultimatum – co zdaniem niektórych naukowców świadczy o bardziej deontologicznym charakterze tych decyzji (zob. Moll, De Oliveira-Souza, 2007). Podobne wyniki dają eksperymenty, w których osobom badanym aplikowano stymulację magnetyczną w rejonie DLPFC, wywołującą tymczasowe lezje tego obszaru. Niemniej badania neuroobrazowe Molla i współpracowników (2011) potwierdzają obserwacje grupy Greene’a, co do neuronalnych korelatów społecznych reakcji emocjonalnych, empatyzowania i wnioskowania o stanach innych – wskazują na brzuszną i grzbietową korę przedczołową, obszary płatów skroniowych, ciało migdałowe, przednią część wyspy i szereg obszarów podkorowych. Szczególnie wyróżniają okolice przyśrodkowej kory przedczołowej, podkolanowej części zakrętu obręczy, górnej bruzdy skroniowej, części podstawnej kresomózgowia i podwzgórze jako struktury powiązane z zachowaniami i uczuciami prospołecznymi. Z kolei z awersją wobec innych, negatywnymi emocjami

moralnymi, jak oburzenie/złość, pogarda i wstręt, a także chęcią do ukarania innych za karygodne czyny wiąże się aktywacją obszarów grzbietowo-przyśrodkowej i bocznej kory przedczołowej, wyspy i ciała migdałowatego. Ponadto uszkodzenia kory frontalnej i obszaru przegrody, a także obniżony poziom metabolizmu (odpowiedź hemodynamiczna) w rejonach podkorowych (m.in. ciele migdałowatym) korelowały z deficytami w sferze zachowań prospołecznych (moralnych sentymentów), takich jak współczucie i poczucie winy (Moll i in., 2011). Założenie Greene’a co do istnienia dwóch „trybów” działania zmysłu moralnego znajduje pewne potwierdzenie w badaniach Cassa Sunsteina (2005) nad heurystykami, czyli mentalnymi skrótami myślowymi wypracowanymi na podstawie ówczesnych doświadczeń. Amerykański badacz wnosi, iż heurystyki są wykorzystywane przez ludzi także w sferze moralności, a przez to, że są to bardzo ogólne zasady, pozbawione kontekstu sytuacji mogą prowadzić też do nietrafnych czy nieracjonalnych decyzji. Podobnie jak Greene twierdzi, iż metoda odruchowych intuicji sprawdza się w większości prostych przypadków, jednak gdy człowiek znajdzie się w wysoce konfliktowej sytuacji (jak w dylemacie wagonika z kładką), odpowiednim zdaje się zaangażowanie dodatkowych procesów świadomego rozumowania – Sunstein nie zgłębia jednak tematu emocji w rozwiązywaniu problemów o charakterze moralnym. Dodatkowo w literaturze przedmiotu istnieją dwie wersje teorii dwóch procesów – szeregową i równoległą (Białek, 2017, s. 11-12). Pierwsza zakłada, że proces intuicyjnej oceny (nazywany też systemem typu 1) jest każdorazowo, automatycznie uruchamiany w odpowiedzi na relewantny bodziec, a zaangażowanie procesu świadomej refleksji (systemu typu 2) jest jedynie opcjonalne – jeśli intuicyjna odpowiedź jest dla nas odpowiednia i wystarczająca to ją przyjmujemy, jednak gdy procesy racjonalizacji tej decyzji podpowiadają nam, że jest ona niepełna czy niewłaściwa, zostają uruchomione procesy świadomej refleksji nad sytuacją poszukujące lepszego rozwiązania. Druga wersja teorii mówi, że oba typy procesów zostają w umyśle uruchomione jednocześnie, jednak odpowiedź intuicyjna pojawia się szybciej i można w tym momencie przyjąć ją za właściwą wraz z zaniechaniem trwających równolegle procesów wolicjonalnej analizy – chociaż można też zdecydować się na kontynuowanie świadomej refleksji nad sytuacją i przyjąć odpowiedź wypracowaną na jej podstawie, zastępując tym samym tę pierwszą, intuicyjną decyzję. Ponadto Greene twierdzi, iż decyzje utylitarystyczne wymagają więcej czasu, dlatego że angażują procesy świadomej refleksji, jednak takie wyjaśnienie jest konfrontowane z innym, odwołującym się do zależności czasu odpowiedzi badanych i siły ich wewnętrznego konfliktu pomiędzy wyborem opcji utylitarystycznej i deontologicznej (Białek, 2017, s. 13). Białek wskazuje dodatkowo, iż takie alternatywne wyjaśnienie tego efektu świadczy na korzyść wersji równoległej teorii

dwóch procesów, lecz nie szeregowej. Aktywność mózgu i sposoby reakcji badanych testuje się także w eksperymentach z udziałem gry w ultimatum. Polega ona na tym, że jeden z dwóch graczy dostaje pewną kwotę pieniędzy, np. 100 zł, i jego zadanie polega na rozdzieleniu owej kwoty między siebie i drugiego gracza. Jeżeli ten drugi gracz zaakceptuje ofertę podziału, 100 zł zostanie odpowiednio rozdzielone między obydwoma graczami. Jednakże jeśli drugi gracz odrzuci ofertę pierwszego, żaden z nich nie dostanie pieniędzy. Teoretycznie zakłada się, że każda kwota, nawet 1 zł, powinna zostać zaakceptowana przez drugiego gracza, lecz w badaniach systematycznie obserwuje się odrzucanie ofert przez drugiego gracza, jeżeli pierwszy gracz zaoferuje mu mniej niż 20% sumy pieniędzy (McAuliffe, 2019). Ponadto notuje się wtedy u drugich graczy wzmożoną aktywność neuronalną obszarów związanych z negatywnymi emocjami i przetwarzaniem bodźców społecznych – DLPFC, VMPFC, przedniej części wyspy i przedniej części zakrętu obręczy (Fehr, Camerer, 2007). Badania z udziałem stymulacji magnetycznej DLPFC powodowały wzrost akceptacji niesprawiedliwych ofert – choć, co ciekawe, badani ci nadal oceniali te oferty jako nieuczciwe. Stąd badacze wnoszą o roli jego obszaru w kontrolowaniu impulsywnego reagowania, jak w sytuacji doświadczenia niesprawiedliwości oraz w monitorowaniu zgodności zachowań z obowiązującymi normami społecznymi (McAuliffe, Blake, Steinbeis, Warneken, 2017). McAuliffe (2019) wskazuje również na istnienie dodatkowego czynnika w produkowaniu ocen moralnych – mianowicie ludzie reagują emocjonalnie przede wszystkim na sytuacje mające znaczenie dla ich własnego dobrostanu lub też ich bliskich i przyjaciół. To znaczy, choć ludzie oceniają czyny wyrządzające krzywdę nieznanym jako moralnie złe, to nie obserwuje się u nich silnej emocji oburzenia/złości oraz motywacji do ukarania sprawców tych czynów, jak w sytuacji gdyby poszkodowanymi byli ich bliscy. Do podobnych wniosków doszła grupa badawcza Molly Crockett (2020), która wykazała, że w wydawaniu sądów o charakterze moralnym, ludzie biorą pod uwagę nie tylko samo zachowanie, ale również osobę wobec której dokonano tego czynu. Bowiem rodzaj relacji łączącej wykonawcę (agensa) i odbiorcę (pacjensa) danego zachowania okazał się istotnym czynnikiem odpowiedzi badanych oceniających określone zachowania jako moralnie złe – np. zachowania, które były oceniane jako nieodpowiednie między przyjaciółmi, nie były oceniane tak samo negatywnie, jeśli zachodziły między sąsiadami. Zatem człowiek przejawia prawdopodobnie różny (zmienny) poziom wrażliwości moralnej i „surowości” wydawanych sądów w zależności od tego, kogo dotyczy dane zachowanie. Wniosek ten przemawia także na korzyść diadycznej teorii moralności, opisaney w poprzedniej sekcji. Neurokognitywne podstawy podobieństw i różnic indywidualnych w

moralności podsumowują Young i Saxe (2011). Badaczki wskazują m.in. na istnienie większego prawdopodobieństwa preferowania decyzji utylitarystycznych przez osoby przejawiające większą potrzebę poznania, tj. „tendencję do angażowania się w wymagające poznawczo aktywności i czerpania z tego przyjemności” (Matusz, Traczyk, Gąsiorowska, 2011) oraz z większą pojemnością pamięci roboczej. Na różnice w ocenach moralnych mają wpływ także różnice we wnioskowaniu o intencjach sprawcy danego czynu i przypisywaniu moralnej sprawczości. Strukturą zaangażowaną w tego typu wnioskowaniach (przypisywania określonego stanu mentalnego innym) jest skrzyżowanie skroniowo-ciemieniowe w prawej półkuli (RTPJ). Badani mający za zadanie osądzić zachowanie człowieka w sytuacjach nieumyślnego spowodowania krzywdy oceniają taką osobę jako mniej winną za powstałe cierpienie innych, jeśli bardziej aktywny był u nich obszar RTPJ. Ponadto osoby przejawiające wzmożoną skłonność do reagowania emocjonalnego, np. osoby częściej doświadczające negatywnej emocji wstrętu mają również większą tendencję do wydawania surowszych ocen moralnych – częściej wskazują, że jakieś zachowanie narusza normy moralne. Część tych czynników jest dodatkowo kształtowana kulturowo.

3. Etyka neuronauk

Druga część definicji neuroetyki odnosi się do kontrowersji etycznych związanych z wdrażaniem rozwiązań opartych na wiedzy o mózgu i umyśle, a także do oceny wpływu tej nowej wiedzy na pojęcie tego kim jesteśmy. Etyka neuronauk jest więc czasem uważana za jedną z subdyscyplin etyki stosowanej, przyjmując za Sójkę jej określenie jako „refleksji, w której chodzi o stosowanie dorobku filozoficznego do rozsądzania kwestii praktycznych, w bliskim kontakcie z praktyką (...). Nacisk położony jest tutaj nie na odwoływanie się do koncepcji filozoficznych, lecz na efekt praktyczny.” (Sójka, 2007, s. 122). Porównuje się ją również z bioetyką, a także wskazuje się na podobieństwo poruszanych w obu dziedzinach problemów (Roskies, 2002; por. Bołoz, 2003). Mimo, iż bioetyka jest stosunkowo młodą dziedziną, której początki sięgają lat siedemdziesiątych ubiegłego wieku, zdołała już wypracować spójny zestaw reguł postępowania – własny system norm etycznych. Motywację stojącą za powstaniem najpierw bioetyki, a następnie neuroetyki trafnie oddają następujące słowa Bołoz:

Bioetyka wzbudza u większości ludzi nadzieję, że mimo nowych biotechnologii pozwalających na (...) transplantację ludzkich tkanek i organów oraz inne formy manipulacji ludzkim życiem, nie zostaną naruszone prawa osoby i wypracowane przez stulecia sposoby odnoszenia się do pacjenta wyrastające z przekonania, że każdy człowiek posiada nienaruszalną godność. (Bołoz, 2003, s. 161-162)

W dalszej części pracy autor podkreśla również autonomiczność bioetyki jako nauki i jej nieredukowalność do etyki filozoficznej czy etyki medycznej (Bołoz, 2003, s. 166). Ponadto, nauka ta nie kieruje się jednym, konkretnym systemem etycznym, lecz stara się wypracować na ich podstawie własny system norm. System ten, definiowany jako pryncypializm, zawiera praktyczne wskazówki, którymi powinni kierować się badacze i klinicyści w obszarze biomedycyny. Odnoszą się one do czterech zasad wypracowanych przez filozofów Beauchampa i Childressa, tj. przynoszenia korzyści, nieszkodliwości, autonomii i sprawiedliwości (Hain, Saad, 2016). Tworzą one minimalny zbiór, fundament, do którego można dodawać kolejne zasady, w miarę trudności pojawiających się problemów etycznych w konkretnych przypadkach. Tym samym są punktem wyjścia debat i dyskusji etycznych, lecz nie skończonym zbiorem nakazów. Choć pryncypializm pierwotnie został wypracowany

przez bioetykę jako standard dla etyki biomedycznej, można przyjąć go za system trafny także w etyce badań neurokognitywnych z udziałem ludzi. W połączeniu z ośmioma zasadami sformułowanymi przez Greely i in. (2018) w ramach *The Neuroethics Working Group of the National Institutes of Health BRAIN Initiative* tworzą one szeroki, a zarazem syntetyczny zbiór reguł etycznych, które należy brać pod uwagę podczas planowania i prowadzenia projektów neuronaukowych. Zasady wymieniane przez Greely i in. (2018) dotyczą następujących kwestii: bezpieczeństwa i świadomej zgody osób badanych, ochrony danych, skutków tzw. podwójnego (tj. właściwego i niewłaściwego) zastosowania narzędzi neurokognitywistyki i ich wdrażania w sferze publicznej, informowania i edukacji opinii publicznej o działaniach neurokognitywistyki oraz sprawiedliwego dostępu do najnowszych narzędzi i metod neurokognitywistyki. Zatem mimo podobieństwa do bioetyki neuroetyka porusza nieco inne, także dodatkowe i bardziej specyficzne problemy – zagadnienia, takie jak autonomia jednostki, poczucie własnego „ja” czy świadomość są szczególnie ważne. Nie można więc sprowadzić ich do wspólnej definicji. Bowiem mózg nie jest podobny do innych organów, takich jak wątroba czy serce, ponieważ to on stanowi o wyjątkowości, osobliwości każdego człowieka, jego zachowaniach i osobowości. Wszelkie uszkodzenia mózgu skutkują zazwyczaj ogromnymi zmianami w funkcjonowaniu człowieka, zaburzając tym samym jego poczucie tożsamości i ciągłości doświadczeń. Nie wspominając nawet o ograniczonych metodach leczenia chorób i urazów mózgu czy jego przeszczepie. Zatem również środki i metody wykorzystywane do ingerowania w układ nerwowy wymagają stawiania innych, swoistych dla neuroetyki pytań. Do ich omawiania zapraszani są nie tylko przedstawiciele zawodów medycznych i filozofowie, ale też specjaliści z nauk społecznych, humanistycznych i prawnych. Wskazuje to na złożoność problematyki oraz interdyscyplinarny charakter neuroetyki. Rommelfanger i in. (2018, s. 24) przedstawiają listę pytań wypracowanych przez grono naukowców nie tylko z różnych dziedzin, lecz i z różnych kultur. Bowiem odpowiedzi na dylematy etyczne, choć często są podobne, to mogą też znacznie się różnić w zależności od wyznawanych i przyjmowanych w danym kręgu kulturowym wartości czy definicji, np. świadomości lub koncepcji osoby. Zatem sformułowane przez Rommelfanger i in. (tamże) pytania są pytaniami otwartymi, na które powinny odpowiedzieć całe zespoły badawcze oraz każdy z naukowców z osobna przed, w trakcie oraz po przeprowadzeniu eksperymentu. Przykładowo, jedno z nich brzmi następująco: *czy i jak wyniki badań neurokognitywnych wpłyną na posiadane przez jednostki oraz całe społeczeństwa koncepcje i wyobrażenia na temat zaburzeń mentalnych czy natury ludzkiej? Czy nowe wnioski nie doprowadzą do stygmatyzacji dotkniętych nimi osób, czy obarczania ich winą za powstanie choroby?* Jeśli

chodzi o rozumienie pojęć klasycznie zarezerwowanych dla i definiowanych na gruncie filozofii, neuronauki wprowadzają nową perspektywę patrzenia na nie, a tym samym „zmuszają” filozofów do włączania ich do swych rozważań. Stwarza to możliwość, albo wręcz konieczność, rozwoju neurofilozofii, ergo neuroetyki. Przykładowo, w świetle nowych ustaleń neuronauk „Osobowa jaźń, chociaż rzeczywista, istnieje w sposób ontycznie zależny i niesamodzielny, zaś jej względna autonomia i tożsamość są następstwem szczególnej organizacji mózgu (względnie stabilnych wzorców neuronowych)” (Poczobut, 2013, s. 375). Z perspektywy neuronauk rozumienie tożsamości osobowej oraz tego czym/kim jest „osoba” przedstawia również Bremer (2010). Ustalenia neurokognitywistyki nie zrewolucjonizowały obecnego sposobu rozumienia tożsamości osobowej czy wolnej woli (O'Connor, Joffe, 2013), jednakże poszerzyły wiedzę społeczeństwa na ich temat. Ponadto O'Connor i Joffe (2013, s. 261-262) odnotowują też negatywny trend w kwestii stygmatyzacji osób z zaburzeniami psychicznymi w świetle nowych faktów o ich biologicznym podłożu. Badania neuronauk nad istnieniem wolnej woli chociaż pozwalają na przewidywanie decyzji nawet do kilku sekund przed jej świadomym podjęciem, to są to badania wobec stosunkowo prostych decyzji, nie przekraczające 60 – 70% trafności. Ponadto wnioski wyciągane na podstawie tych wyników są w skrajnej opozycji do siebie, tzn. część twierdzi, iż jest to ostateczny dowód na brak istnienia wolnej woli, inni konkludują, że aktywność neuronalna w tych badaniach jedynie pokazuje pracę mózgu podczas podejmowania decyzji, przyjmując ją za korelat wolnej woli. Ponadto badani zgadzają się, że nawet jeśli to „mózg podejmuje decyzje”, to jest on częścią ich osoby, więc odpowiedzialność za własne działania nadal istnieje (Nahmias, 2015).

Przewijające się w literaturze przedmiotu pytanie i jego rozstrzygnięcie, tj. czy neuroetyka jest odrębną i samodzielną dyscypliną nauki, czy (tylko) interdyscyplinarnym projektem badawczym, nie umniejsza ważności stawianych przez nią pytań i podejmowanych w jej ramach zagadnień. Nauka ta bowiem jest odpowiedzią na powstałą potrzebę debaty nad rozwojem nauk o mózgu i umyśle oraz prowadzonych w ich ramach badań, podobnie jak miało to miejsce w przypadku biologii molekularnej i badań DNA w latach dziewięćdziesiątych ubiegłego wieku (zob. Collins, 1998). Historyczny kontekst powstania etyki neuronauk opisują Illes i Bird (2006) oraz wymieniają badania kliniczne, naukowe i komercyjne oraz potrzebę publicznego dialogu nad neuronaukami jako główne obszary zainteresowania neuroetyki. Podkreślają, iż współczesny nacisk na rozważanie nie tylko korzyści, ale i zagrożeń związanych z rozwojem nauki bierze się z historii badań i praktyk naukowych, szczególnie biologicznych. Odwołują się przy tym do wątpliwych i haniebných

zabiegów dokonywanych w III Rzeszy, które miały miejsce jeszcze kilkadziesiąt lat temu. Zarazem zwracają uwagę, iż ośrodki akademickie powinny zadbać o edukację swoich studentów w zakresie krytycznego myślenia oraz kodeksu etyki prowadzenia badań i pracy naukowca. Martin, Becker, Darragh i Giordano (2016) także podają listę tematów poruszanych obecnie przez neuroetyków, opierając się na przeprowadzonym przez nich przeglądzie literatury przedmiotu. A są to kwestie dotyczące: 1) neuroobrazowania, 2) neurogenetyki, 3) neurofarmakologii, 4) stymulacji mózgu, 5) komórek macierzystych i mózgowych organoidów, 6) pacjentów pediatrycznych w badaniach neuronaukowych oraz 7) podwójnego zastosowania (*dual-use*) neurokognitywistyki. Dla niektórych może wydawać się nieintuicyjne debatowanie o kwestiach z podpunktu 5), czyli sztucznych „proto-mózgach w probówkach”, co zauważa i o czym pisze Kozhevnikova (2019, s. 132), skoro rokrocznie przeprowadzanych jest tysiące inwazyjnych eksperymentów na żywych, a także cierpiących zwierzętach. Jednakże „należy wspomnieć, że postęp technologiczny odbywa się w zawrotnym tempie i często wyprzedza refleksję etyczną. Toteż, nawet jeśli w danym momencie problem ten ma marginalny i czysto hipotetyczny charakter, to nie oznacza, że takim pozostanie w przyszłości.” (tamże). Ostrożność należy również zachować w wyciąganiu wniosków z badań neuroobrazowych, jako że podają one jedynie korelaty danych zachowań i funkcji poznawczych, jednak nie ma jeszcze pewności czy można je rozpatrywać jako związki przyczynowo-skutkowe, tj. aktywność neuronalna w danym obszarze (np. w ciele migdałowatym) – określone zachowanie (np. reakcja lękowa). Zatem komisje neuroetyczne, na podobieństwo bioetycznych, powinny również dbać o poprawność wniosków wyciąganych z wyników badań nad mózgiem i umysłem. Bowiem różni badacze, a także osoby spoza kręgu naukowego mogą inaczej tłumaczyć konkluzje płynące z tych badań i w rezultacie głosić nie do końca poprawne czy wręcz błędne teorie, np. na temat chorób psychicznych czy istnienia wolnej woli. Wydaje się, że neuroetyka powinna także w pewnym stopniu obserwować i czuwać nad projektami komercyjnymi, informując konsumentów danych neurotechnologii czy korzystanie z nich rzeczywiście może pozytywnie wpływać na układ nerwowy i/lub poznawczy człowieka, czy jednak obietnice składane przez producentów tych technologii nie mają dowodów empirycznych i wręcz mogą tworzyć zagrożenie i doprowadzić do uszczerbku na zdrowiu. Naukowcy zaangażowani w sferze neuroetyki powinni również zadbać o merytoryczny dialog z opinią publiczną, tłumacząc w zrozumiały sposób cele, wyniki, potencjalne korzyści, a także i kontrowersje płynące z badań mózgiem i umysłem. Słowem zakończenia tego wprowadzenia warto podkreślić, że tak rozumiana neuroetyka nie chce uniemożliwić czy zatrzymać rozwój

neuronauk. Pragnie jedynie, by rozwijały się one zgodnie ze standardami XXI wieku, a nie dążyły do rewolucji technologicznych i naukowych za wszelką cenę, nie bacząc na metody jakimi to osiągnie.

Wzrost popularności nauk neurokognitywnych uwidacznia się w rosnącej liczbie powstających programów badań nad mózgiem i umysłem. Do największych tego typu inicjatyw zaliczają się: powołany przez Unię Europejską The Human Brain Project, japoński Brain/MINDS, kanadyjski Brain Canada, amerykańskie programy Brain Research through Advancing Innovative Neurotechnologies (BRAIN), Human Connectome Project i Decade of the Mind Project czy chiński China Brain Project (Rommelfanger i in., 2018, s. 20-21). Głównym celem tych inicjatyw jest pogłębianie wiedzy na temat budowy i funkcjonowania układu nerwowego i systemu poznawczego człowieka. Tym samym, w sześciu poniższych podrozdziałach zostaną opisane sposoby wykorzystywania wyników badań neurokognitywistyki, kolejno w medycynie, wojsku, ekonomii i marketingu, prawie, sztucznej inteligencji oraz edukacji. W każdej z tych sekcji poruszę także problemy etyczne zgłaszane przez środowisko etyków i innych badaczy związane ze stosowaniem danych metod i praktyk wypracowywanych przez nauki neurokognitywne.

3.1 Medycyna

Wypracowana przez neurokognitywistykę wiedza dotycząca powiązań między selektywnymi uszkodzeniami czy nieprawidłowościami funkcjonowania mózgu i określonymi deficytami poznawczymi stanowi pomoc dla lekarzy w diagnozie i leczeniu zaburzeń psychicznych i neurologicznych (Cacioppo i in., 2007; Cooper, Korb, Akil, 2019; Hall, Carter, Morley, 2003; Reynolds, Lewis, Detre, Schatzberg, Kupfer, 2009). Szczególnie jeśli chodzi o trafną identyfikację danego zaburzenia, gdyż od tego zależy dobór i w rezultacie skuteczność dalszej terapii, a więc i dobrostan, i sprawności funkcjonowania jednostki. Techniki, takie jak głęboka stymulacja mózgu (DBS) czy przezczaszkowa stymulacja magnetyczna (TMS) znajdują już zastosowanie w leczeniu parkinsonizmu, depresji, zaburzeniach obsesyjno-kompulsywnych, zaburzeniach lękowych, schizofrenii, zaburzeniach afektywnych dwubiegunowych i anoreksji (Lefaucheur i in., 2020; Park, Singh, Pike, Tan, 2017). Niemniej stymulacja mózgu nadal jest zaliczana do metod eksperymentalnych w leczeniu zaburzeń psychiatrycznych i neurologicznych i nie jest jeszcze ustabilizowaną czy szeroko akceptowaną techniką leczenia (Hariz, Amadio, 2015). Przeważnie stosuje się ją w przypadkach chorób opornych na leczenie innymi metodami.

Rozwój tych nowych technik motywowany jest po części wynikami badań nauk neurokognitywnych. Bowiem odkrycie neuronalnych podstaw (strukturalnych i funkcjonalnych) chorób psychicznych i neurologicznych znacząco przyczyniło by się do wyjaśnienia przyczyn tych zaburzeń oraz lepszego zrozumienia ich etiologii i wypracowania skutecznych sposobów leczenia (Arbuckle, Travis, Ross, 2017; Kupfer, Regier, 2011). Pomogłoby to również w informowaniu pacjentów o naturze ich choroby i przyczynach oraz w „zdjęciu ciężaru odpowiedzialności” za pojawienie się i rozwój choroby psychicznej u pacjentów, którzy mogą niesłusznie się za to obwiniać. Pomogłoby to również zmniejszyć ewentualną stygmatyzację chorych na nie osób (Hall i in., 2003, s. 867). Chociaż pojawiają się też głosy ostrzegające o nadmiernym upraszczaniu modeli powstawania zaburzeń psychicznych i sprowadzania ich jedynie do czynników biologicznych (Stupak, 2020). W rozwoju takich zaburzeń należy brać pod uwagę również szerszy kontekst życia człowieka, tj. jego środowisko postnatalne i społeczne, którym zajmuje się subdziedzina neurokognitywistyki społecznej. Z rozwojem nowych metod leczenia oraz prowadzeniem wszelkich badań biomedycznych wiążą się również trudne i często niejednoznaczne kontrowersje etyczne. Zalecenia postępowania przed zastosowaniem wysoce ryzykownych metod leczenia czy badania dotyczą następujących kwestii: rozważenia wszelkich korzyści i zagrożeń związanych z daną metodą oraz poinformowania o tym pacjenta lub osoby decydującej o jego postępowaniu medycznym, rozważenia i zastosowania w pierwszej kolejności inwazyjnych metod leczenia oraz ostrożności i powściągliwości w stosowaniu tych metod, jeśli ich skuteczność nie została potwierdzona w danych jednostkach chorobowych (Park i in., 2017). Nie tylko bowiem spowodowanie niepożądanych skutków ubocznych, ale również dawanie zbyt wysokiej nadziei na powrót do zdrowia może spowodować utratę wiary u pacjentów na wyleczenie, jeśli, mimo obietnic, nie widzą oni poprawy swojego stanu zdrowia. Nauki neurokognitywne mogą również pomóc oszacować poziom funkcjonowania poznawczego pacjentów, czy są oni zdolni do decydowania o swoim zdrowiu i życiu, czy potrafią dokonywać świadomych wyborów. Na przykład osoby chore na anoreksję często określają swoją chorobę jako część ich tożsamości i odmawiają leczenia, jednak czy tak samo twierdziłyby, gdyby były zdrowe? Czy są one w stanie obiektywnie ocenić swoją sytuację i konsekwencje swojego wyboru? Czy ich procesy kognitywne nie zostały w trakcie choroby zaburzone? Czy osoba chora jest sobą, swoim prawdziwym „ja”, czy choroba zmieniła tożsamość tej osoby, a przez leczenie lekarze mogą przywrócić pacjentowi jego prawdziwe „ja”? Czy jest wręcz przeciwnie i to leczenie np. poprzez stymulację mózgu czy leki zmienia tożsamość człowieka? (por. Bremer, 2010, s. 283-286). Podobnie w przypadku przyjmowania

środków farmakologicznych poprawiających nastrój – czy przez ich stosowanie ludzie w pewien sposób modyfikują swoją osobowość? Bowiem ktoś może być bardziej neurotyczny czy introwertyczny niż przeciętna osoba, lecz nie oznacza to jeszcze, iż jest chory czy cierpi na jakieś zaburzenia. Co w przypadku dzieci, u których istnieje ryzyko wystąpienia takich zaburzeń? Bowiem z jednej strony wczesne leczenie może zapobiec rozwojowi choroby, jednak z drugiej na ile dokładnie można przewidzieć rozwój choroby czy stwierdzić, że przejawiane przez dziecko zachowanie jest wynikiem danego zaburzenia psychicznego a na ile „testowaniem granic” i objawem dynamicznych przemian, jakich doświadcza dziecko czy nastolatek w trakcie swojego rozwoju.

Współtworzone przez neurokognitywistykę i medycynę są również programy badań nad lekami oraz technikami, takimi jak interfejsy mózg-komputer (*brain-computer interfaces*, BCIs) czy treningi kognitywne oparte na sprzężeniu zwrotnym (*neurofeedback*) poprawiającymi funkcjonowanie poznawcze, np. polepszające pamięć u osób starszych czy koncentrację uwagi u cierpiących na zespół nadpobudliwości psychoruchowej z deficytem uwagi. BCIs są wykorzystywane także w nowoczesnych neuro-protezach i rehabilitacji psycho-ruchowej pacjentów z uszkodzeniami układu nerwowego (Lebedev, Nicolelis, 2017). Trwają również badania z ich udziałem jako potencjalnych metod komunikacji z osobami z zaburzeniami świadomości (Giacino i in., 2014). Ponownie pojawiające się tu pytania o kwestie etyczne nie są łatwe, co potwierdza brak jednoznacznych odpowiedzi na nie (Racine, Aspler, 2017, s. 19-39). Niemniej wśród najczęściej wymienianych jest problem świadomej zgody – czy pacjenci są w stanie zrozumieć zarówno korzyści, jak i ryzyko związane z zastosowaniem danej metody? Czy będą wtedy zdolni podjąć najlepszą możliwą dla siebie decyzję? Czy wszczepienie urządzenia do stymulacji mózgu u chorych, np. na chorobę Parkinsona, nie zaburzy ich poczucia autonomiczności i jedności? Co w przypadku usterki wszczepionego neuroimplantu i ewentualnego uszczerbku na zdrowiu chorego? Kto ponosi odpowiedzialność za naprawę urządzenia i/lub koszty leczenia – producent, lekarz czy pacjent, który był świadomy takiego ryzyka przy decydowaniu o wszczepieniu implantu? Ponadto nie są znane jeszcze wszystkie następstwa i skutki uboczne działania środków farmakologicznych i nowych technologii na układ nerwowy, zwłaszcza w perspektywie długookresowej. Niemniej dotychczas odnotowano przypadki zmiany osobowości i zaburzenia nastroju u niektórych pacjentów z wszczepionymi neurostymulatorami (Jotterand, Giordano, 2011) czy też pojawienie się uzależnienia od hazardu (Smeding i in., 2007). Co w takich przypadkach z poczuciem „ja”, tożsamością takich osób? Czy pacjenci z chorobami neurodegeneracyjnymi, modyfikującymi ich pamięć i zachowanie są tymi samymi osobami,

co sprzed choroby? Poszukiwane w naukach neurokognitywnych neuronalne korelaty świadomości mogą okazać się niezmiernie ważne w ocenie stopnia świadomości u pacjentów z uszkodzeniami układu nerwowego i w stwierdzaniu śmierci mózgu (Giacino i in., 2014). Już dzisiaj różnorodne metody neuroobrazowania rejestrujące aktywność metaboliczną, hemodynamiczną i elektryczną mózgu, okazują się być pomocnym i trafnym narzędziem przy ocenie poziomu świadomości oraz śmierci mózgu (Walter, Fernández-Torre, Kirschstein, Laureys, 2018). Stworzony na podstawie teorii zintegrowanej informacji Perturbacyjny Indeks Złożoności (*Perturbational Complexity Index*), czyli narzędzie do pomiaru poziomu świadomości, także może okazać się prostym i szybkim sposobem jej oceny w warunkach klinicznych i rehabilitacyjnych (Koculak, Kałwak, 2014). Niemniej nie są to na razie metody niezawodne, chociażby dlatego, że nie zawsze badają one aktywność neuronalną w sposób bezpośredni (jak w przypadku fMRI). Nie ma jeszcze także jednoznacznej (wystandaryzowanej) odpowiedzi na pytanie jaki jest „próg świadomości”, jaki stopień aktywności mózgu występuje u pacjentów w minimalnym stanie świadomości, a jaki w stanie wegetatywnym? Poszczególne metody neuroobrazowania mają także różne rozdzielczości czasowe i przestrzenne, przez co rozsądne wydaje się łączenie różnych technik przy ocenie funkcjonowania układu nerwowego pacjentów. Zdają się to potwierdzać Gosseries, Laureys i Boly (2014), którzy dowiedli, że opieranie się jedynie na testach behawioralnych przy ocenie świadomości okazuje się być błędne aż w 40% przypadków. Zlokalizowanie w mózgu ośrodków bólu pozwala ocenić jego nasilenie i w konsekwencji podawanie odpowiednich środków przeciwbólowych pacjentom, którzy sami nie są w stanie zakomunikować lekarzom swojego cierpienia. Rozwój implantów najnowszej technologii pozwala już pacjentom z barwnikowym zapaleniem siatkówki i zwyrodnieniem plamki żółtej odzyskać jakość widzenia (Merabet, 2011; Waltz, 2017). Jednak jak wskazują autorzy w wyżej wymienionych artykułach, okazuje się, że problemem w rozwoju tego typu narzędzi nie jest stopień zaawansowania technologicznego, lecz niekompletność wiedzy odnośnie szczegółowej budowy mózgu i jego konektomu (precyzyjnej mapy sieci połączeń neuronalnych). Dlatego tak ważny jest rozwój badań w naukach o mózgu i umyśle oraz wspieranie powstających w ich ramach inicjatyw. Niemniej jednak w projektach komercyjnych, takich jak Neurolink Elona Muska (2019), należy także zapytać o zasadność tego typu eksperymentów. Czy nie są one podejmowane tylko „na pokaz” jako chwyt marketingowy? Czy nie narażają wykorzystywanych w nich zwierząt na nieuzasadnione cierpienie, nie wnosząc tym samym żadnej wiedzy naukowej? Jaki naprawdę motyw kryje się za tego typu eksperymentami?

3.2 Wojsko

Dla armii postęp badań neuronaukowych ma zastosowanie przede wszystkim w sferze polepszania działania poznawczego, rozwijając i ulepszając sposoby nabywania wiedzy i nowych umiejętności czy podejmowania decyzji przez żołnierzy w ciężkich i stresujących warunkach. Nowoczesne techniki neuroobrazowania, interfejsy mózg-komputer oraz neuromodulacja jako przykłady metod wypracowanych przez badania neurokognitywistyki znajdują już swoje zastosowania w wojsku (National Research Council, 2009). W cytowanym raporcie autorzy wymieniają „inteligentne” okulary połączone z urządzeniem EEG wbudowanym w hełm, które rejestruje aktywność mózgu. Technologia ta wykorzystywana jest do szybkiego wykrywania potencjalnego niebezpieczeństwa w otoczeniu żołnierza. Bazuje ona na nieświadomej aktywności mózgu i bezpośrednio odwołuje się do rozwijanych w ramach neurokognitywistyki teorii świadomości. Mianowicie dowiedziono, że umysł jest w stanie „rejestrować” cały obszar pola wzrokowego, jednakże tylko niewielka część tego pola staje się świadoma. Reszta obszaru pozostaje w podświadomości, przynajmniej do czasu, aż zwrócimy uwagę na nowy fragment pola wzrokowego. Niemniej neurony przejawiają aktywność także na bodźce nieuświadomione i właśnie ta właściwość układu nerwowego została wykorzystywana w technologii „inteligentnych” okularów. W ramach „rozszerzonego umysłu” testowany jest również wpływ stymulacji mózgu i niektórych środków farmakologicznych na układ nerwowy w celu polepszenia jego funkcjonowania. Wymienia się tu substancje psychoaktywne takie jak modafinil, amfetamina, nikotyna i kofeina, a także beta-blokery jak propranolol. Dotychczas została potwierdzona skuteczność ich stosowania w sferze poprawy wydajności funkcjonowania poznawczego (uwagi i pamięci) oraz w redukowaniu poczucia zmęczenia (National Research Council, 2008). Z kolei metody stymulacji mózgu, jak przezczaszkowa stymulacja stałoprądowa (tDCS), TMS i DBS okazały się mieć pozytywny wpływ zarówno na kodowanie i odtwarzanie nowego materiału (werbalnego, wizualnego, liczbowego), jak i konsolidację pamięci semantycznej i proceduralnej (umiejętności motoryczne). Ponadto odnotowano polepszenie funkcjonowania poznawczego w przypadku zmęczenia oraz zmniejszenie częstości przywoływania fałszywych wspomnień (Dresler i in., 2013, s. 536). Podejmowane są również próby identyfikowania stanów umysłu na podstawie zapisu aktywności mózgu. Badania neurokognitywistów w wojsku skupiają się także na poszukiwaniu markerów somatycznych poziomu zmęczenia/wycieńczenia układu nerwowego. Wszelkie metody neuroobrazowania są doskonałe i ulepszone, by w

przyszłości służyły jako precyzyjne narzędzia do oceny funkcjonowania psychicznego, rozszerzania możliwości pamięciowych i zdolności do nabywania nowego materiału szkoleniowego czy w nowoczesnych „wykrywaczach kłamstw” i interfejsach mózg-komputer do zdalnego sterowania urządzeniami na polu walki (National Research Council, 2008). Niemniej jeśli chodzi o wszelkiego typu ingerencje w układ nerwowy należy brać pod uwagę bezpieczeństwo danych metod, ich krótko- i długotrwałe (także potencjalnie negatywne) skutki neurologiczne. Nauki o mózgu i umyśle dostarczają również wiedzy na temat zespół stresu pourazowego, szeroko rozpowszechnionego zaburzenia wśród żołnierzy i weteranów. Dzięki temu współtworzą metody jego diagnozowania i skutecznego leczenia, np. poprzez terapię behawioralno-kognitywną czy stosowanie niektórych beta-blokerów modyfikujących pamięć, np. traumatycznych obrazów wojny (Tanielian, Tanielian, Jaycox, 2008). Jednakże czy żołnierze będą mogli odmówić stosowania niektórych metod, czy też będą zmuszeni do ich bezwzględnej zaakceptacji? Czy wraz z wyjazdem na misję wojskową tracą oni możliwość do indywidualnego decydowania o własnym organizmie, bezwzględnie podporządkowując się wszelkim rozkazom przełożonych? Kto kontroluje przestrzeganie podstawowych praw żołnierzy jako autonomicznych jednostek w „oddalonych od cywilizacji” bazach wojskowych? Wiadomo, iż wojsko stanowi trzon postępu technologicznego i rewolucji badawczych, ale czy żołnierze nie stają się tym samym obiektem testowym tych nowych wynalazków? (zob. Greely i in., 2008, s. 703) Stwarza to możliwość zarówno bezpośredniego, jak i pośredniego wywierania nacisku do stosowania środków psychofarmakologicznych i technologicznych. Bowiem żołnierze mogą czuć się winni ewentualnych niepowodzeń misji wskutek własnego zmęczenia i obniżenia zdolności psychomotorycznych, którym teoretycznie mogliby zapobiec gdyby zastosowali dane metody „wzmocnienia”. Metody stymulacji mózgu często dają pozytywne rezultaty, niemniej ich skuteczność nie została ostatecznie potwierdzona. Ponadto fizjologiczne mechanizmy ich działania nie są jeszcze do końca znane, a skutki ich zastosowania różnią się między osobami i zależą od wielu czynników (Dayan, Censor, Buch, Sandrini, Cohen, 2013). Znane są również przypadki zwiększenia sprawności jednych funkcji kognitywnych z jednoczesnym pogorszeniem innych funkcji (Voarino, Dubljević, Racine, 2017). Wśród skutków ubocznych przeczaszkowej stymulacji mózgu wymienia się m.in. bóle głowy i napady padaczkowe (Dresler i in., 2013, s. 536), co mogłoby być szczególnie niebezpieczne dla żołnierzy, jeśli napady te wystąpiłyby w trakcie walki. Wreszcie, czy niektóre metody nie staną się pewnego typu bronią neurobiologiczną stosowaną przez siły zbrojne wobec przeciwnika, czy nie staną się one nowymi technikami tortur. Czy taka „cyborgizacja” żołnierzy nie wykracza daleko

poza naturalne możliwości człowieka i nie podważa filozoficznej idei sprawiedliwej wojny? Obecnie pytania te mogą wydawać się nader wybiegające w przyszłość, jednak tempo rozwoju badań nad mózgiem i umysłem każe z wyprzedzeniem zadbać o kwestie etyczne postępu neuro-technologicznego.

3.3 Ekonomia i marketing

Nauki o mózgu i umyśle mogą wspomóc nauki ekonomiczne w dokładniejszym zrozumieniu teorii podejmowania decyzji przez ludzi. Tezę tę zdaje się potwierdzać Kahneman (2011), który w swej książce opisał zbiór heurystyk (tj. skrótów myślowych) i błędów poznawczych, których używamy na co dzień. Wnioski płynące z pracy Kahnemana każą poważnie się zastanowić nad pojęciem *homo oeconomicus*. Termin ten przyjmowany jest w ekonomii na określenie konsumenta jako jednostki racjonalnej, świadomej wszelkich niezbędnych informacji i posiadającej nieograniczone zdolności poznawcze w procesie podejmowania decyzji. Niemniej Gigerenzer (1999) ocenia heurystyki jako pomocne narzędzia funkcjonowania poznawczego człowieka, szczególnie jeśli weźmiemy pod uwagę nadmiar docierających do nas informacji w dzisiejszych czasach. Wtedy pewne „automatyczne” (nawykowe) zachowania wydają się wystarczająco dobre (*satisficing*) do wykonania danego zadania, czego dowodzą Klein, Calderwood i Clinton-Cirocco (2010) w swoich badaniach podejmowania decyzji pod presją czasu. Badali oni procesy decyzyjne u zawodowych strażaków w eksperymentalnej sytuacji pożaru. Okazało się, że sposób ich rozumowania i podejmowania decyzji polega na identyfikacji kluczowych aspektów sytuacji i następnie ich dopasowaniu do schematów „akcji strażackich” znanych i wypracowanych przez lata praktyki. Chociaż dokonany wybór nie zawsze musi być najlepszym z możliwych, to jest on wystarczająco trafny i szybki by „być tym właściwym”. Z kolei Knutson i Huettel (2015) piszą, iż na podstawie wzoru aktywności mózgu można przewidzieć podejmowanie ryzykownych zachowań finansowych u osób badanych. Pytanie czy wiedza odnośnie sposobów podejmowania decyzji czy skłonności niektórych osób do ryzyka, nie zostanie w przyszłości wykorzystana przeciwko nim., np. przez kasyna i firmy pożyczkowe budujące swój kapitał na tych „słabościach” ludzi.

Ponadto w marketingu nauki neurokognitywne stwarzają możliwość dokładniejszego zrozumienia preferencji i emocji przejawianych przez konsumentów wobec pewnych produktów. Bowiem ludzie mają często trudności z określaniem i rozpoznawaniem własnych emocji i innych odczuć oraz wnioskowaniem o ich przyczynach. Wobec tego badania

neurokognitywistyki pomogłyby dowiedzieć się markom, co naprawdę czują ich odbiorcy. Najpopularniejsze są tu badania neuroobrazowania mózgu, które pomagają działom marketingu w testowaniu produktów pod względem wizualnym czy smakowym, a także w tworzeniu produktów mających wywołać dane wrażenia u jego użytkowników, jak np. w produkcji zwiastunów filmowych (Ariely, Berns, 2010) i spotów reklamowych. Farah (2012, s. 577) podaje nawet przykłady korzystania z tych metod w badaniu preferencji ludzi wobec określonych kandydatów w kampaniach politycznych. Z kolei Astolfi i in. (2008) na podstawie samej aktywności mózgu z zapisu EEG byli w stanie przewidzieć, która reklama zostanie lepiej zapamiętana przez osoby badane. McClure i in. (2014) badali za pomocą neuroobrazowania preferencje ludzi względem dwóch popularnych napojów – Coca Coli i Pepsi. Wnioski okazały się dość zaskakujące, potwierdzając jednocześnie ważność przekazu reklamowego w budowaniu marki produktu. Mianowicie w tzw. ślepych teście, gdy badani nie wiedzieli, który z dwóch napojów jest przez nich próbowany, preferencje względem jednego lub drugiego rozkładały się równo po połowie (50/50). Kiedy natomiast badani byli wcześniej poinformowani, czy piją Coca Colę, czy Pepsi inne obszary mózgu były u nich aktywne, a ponadto uwidoczniła się wyraźna preferencja wyboru jednego lub drugiego produktu. Z kolei Knutson, Rick, Wimmer, Prelec i Loewenstein (2007) na podstawie samego zapisu z urządzenia fMRI byli w stanie przewidzieć decyzje zakupowe badanych. Popularną metodą w neuromarketingu jest również okulografia (*eye-tracking*), szczególnie w badaniach *user experience*, tj. doświadczeń użytkowników (Marecka, 2018). Chociaż metody badań nauk o mózgu i umyśle są efektywne i coraz popularniejsze, to nie zastąpią całkowicie tradycyjnych narzędzi, np. badań ankietowych, lecz będą stanowić ich uzupełnienie. Budzącą kontrowersje etyczne jest sprawa prywatności danych o funkcjonowaniu psychicznym konsumentów. Bowiem w badaniach neuroobrazowych na jaw mogą wyjść pewne fakty, których świadomie (tzn. w badaniach tradycyjnych, np. ankietowych) badani nie chcieli o sobie ujawniać. Badani powinni także być informowani o celach badania i ewentualnych wnioskach, wyciąganych na podstawie ich aktywności układu nerwowego, gdyż ludzie mogą nie być świadomi, do czego zostaną wykorzystane ich dane. Poruszana jest także kwestia przypadkowych odkryć nieprawidłowości w budowie czy działaniu mózgu (Illes i in., 2006). Czy osoby badane powinny być informowane o wykrytych zaburzeniach ich układu nerwowego? Co w przypadku, gdy okażą się one tylko „fałszywym alarmem”, lecz spowodują niepotrzebny stres u tych osób? Wreszcie, czy osoby przeprowadzające te badania są wystarczająco kompetentne do stwierdzania, co jest normą, a co zaburzeniem? Problemem może być również sposób wykorzystywania wyników badań nauk o mózgu i umyśle.

Przykładowo znajomość mechanizmu tworzenia nawyków czy nawet uzależnień od pewnych produktów i usług, może zostać wykorzystana przez ich producentów do „manipulowania” klientów (Campos i in., 2020; Ariely, Berns, 2010). Dowiedziono na przykład, iż zwiększona zawartość cukru w produktach spożywczych jest silniej „odczuwana” w pewnych ośrodkach mózgowych, m.in. w układzie nagrody i jądrze półleżącym, co w konsekwencji może prowadzić do nadmiernej konsumpcji takich wyrobów i związanych z tym problemów zdrowotnych, np. otyłości czy cukrzycy (Gordon, Ariel-Donges, Bauman, Merlo, 2018; Smith, Robbins, 2013). Wiedza z działy językoznawstwa kognitywnego również dostarcza wskazówek odnośnie sposobów „manipulowania słowami”. To znaczy, poprzez określone konstrukcje językowe można kierować uwagę odbiorców na pewne aspekty informacji, a „ukrywać” inne, np. przez tzw. profilowanie językowe (Evans, Green, 2006, s. 41-43). Wiedza ta może zostać wykorzystana przez działy marketingu, np. w tworzeniu haseł reklamowych czy w dyskusjach politycznych. Niemniej informacje przekazywane odbiorcom powinny być nie tylko prawdziwe, lecz także jasne i zrozumiałe, więc czy taka „gra słów” celowo nie wprowadza odbiorców tych przekazów w błąd? Skandal z udziałem Cambridge Analytica z 2016 r. wywołany prezydencką kampanią marketingową, profilującą osoby podatne na pewnego typu przekazy każe kwestionować etyczność tego typu praktyk oraz brak ich akceptacji ze strony społeczeństwa.

3.4 Prawo

Jeśli chodzi o korzystanie z wiedzy neurokognitywistyki w dziedzinie prawa, zdania są podzielone. Część zaangażowanych w debatę uważa, że nauka ta nie wniesie nic nowego, z kolei druga część zwiastuje istną rewolucję w prawie (Goodenough, Tucker, 2010, s. 65-66; Greely, 2009, s. 688). Niewątpliwie badania neurokognitywistów nad świadomością i innymi stanami umysłu mogą zrewidować stosowane w prawie karnym definicje poczytalności, wolnej woli czy odpowiedzialności. Powstaje jednak pytanie czy pojęcia te zostaną jedynie sprecyzowane, czy całkowicie zmieniony zostanie sposób myślenia o nich. Jak pisze Fischborn (2016) na razie nowe doniesienia neuronaukowe nie wpłynęły na powszechny sposób rozumienia koncepcji wolnej woli, choć nadal trwają dyskusje czy mózg jest z góry zdeterminowany, czy też nie i niewykluczone, iż w przyszłości może nastąpić rewizja definicji tego pojęcia. Jednakże podważenie czy zmniejszenie wiarygodności zeznań naocznych świadków można uznać za rewolucję w prawie, dokonaną dzięki badaniom psychologii kognitywnej pod koniec XX wieku. Czołową badaczką, która zwróciła uwagę na

ten problem była Elizabeth Loftus. Począwszy od lat siedemdziesiątych ubiegłego wieku opublikowała ona serię badań dotyczących funkcjonowania pamięci (Loftus, 2005). Wskazała m.in. możliwe sposoby jej manipulacji, liczne ograniczenia systemu pamięciowego, podatność ludzi na dezinformację oraz implementację fałszywych wspomnień (Schacter, Loftus, 2013). Schacter i Loftus w swoim artykule (2013) wskazali również sposoby badania pamięci i wiarygodności zeznań, m.in. za pomocą fMRI. Z kolei Edersheim, Brendel i Price (2012) piszą o stosowaniu badań neurokognitywnych w ocenie sprawności funkcjonowania poznawczego oskarżonych. Za pomocą neuroobrazowania mózgu udało się wykazać pewne nieprawidłowości w funkcjonowaniu psychicznym i budowie anatomicznej układu nerwowego. W kilku przypadkach doprowadziło to do złagodzenia wymierzonej kary czy nawet uniewinnienia oskarżonych i skierowania ich, w zamian, do specjalistycznych ośrodków leczniczych. Prowadzone są również badania nad wspomnianymi już wykrywaczami kłamstw, bazującymi na aktywności neurofizjologicznej układu nerwowego (Greely, 2009, s. 698-699). Obecnie przeszkodą w ich wykorzystywaniu jest stosunkowo niska trafność, a w konsekwencji wiarygodność takich metod „czytania umysłu”. Ponadto o kłamstwie badanego wnioskuje się przede wszystkim na zwiększonej aktywności części mózgu odpowiedzialnej za kontrolę poznawczą, która może być zaangażowana także do ogólnej kontroli zachowania w sytuacji stresowej, a nie tylko do maskowania prawdziwych odpowiedzi przez badanego. Może to dodatkowo zaburzać trafność takiego urządzenia, a więc i jego użyteczność. Jeśli udałoby się odczytywać treści, o których myślą badani, to dodatkowym pytaniem jest, czy samo myślenie o zbrodni stanowi dowód na jej popełnienie. Na ile też możemy wnioskować o stanie umysłu oskarżonego w chwili popełnienia przestępstwa na podstawie obecnej aktywności neurofizjologicznej, co mówi obecny stan mózgu? Wyniki badań nauk o mózgu i umyśle doprowadziły również do stosowania nowych metod w resocjalizacji i leczeniu skazanych, m.in. środków psychofarmakologicznych czy głębokiej stymulacji mózgu w redukcji agresji i hiperseksualności (Greely, 2009, s. 16-17). Problemem etycznym jest tutaj kwestia świadomej zgody skazanych na tego typu zabiegi, czy nie są to metody naruszające autonomię i poczucie spójności obrazu własnej osoby (Farah, 2012, s. 585). W dziedzinie kryminologii wiedzę o mózgu i umyśle wykorzystuje się w przewidywaniu podatności do zachowań przestępczych u małoletnich i dorosłych. Tym sposobem próbuje się zredukować liczbę przestępstw i recydywistów na podstawie wzorów aktywności neuronalnej czy markerów somatycznych (Greely, 2009, s. 692-693). Zatem można uznać to za próby odpowiedzi na pytanie czy mózgi przestępców różnią się czymś od mózgów osób prawych, przestrzegających prawa. Etyczne kontrowersje może tu budzić

pytanie, czym jest norma w kwestii budowy mózgu i funkcjonowania psychicznego. Istnieje bowiem spora zmienność międzyosobnicza i wewnątrzosobnicza pod względem funkcjonowania poznawczego i budowy anatomicznej układu nerwowego u ludzi (Jones, Buckholtz, Schall, Marois, 2009). Przegląd możliwości wdrażania wiedzy nauk neurokognitywnych w polskim prawie analizuje Zyzik (2013). Oprócz kwestii korzystania z nauk neurokognitywnych na salach sądowych, neuroetycy mogą zostać również zaangażowani w procesy legislacyjne dotyczące nowych neurotechnologii, oferując swoją wiedzę odnośnie funkcjonowania układu nerwowego oraz wyników badań w dziedzinie nauk o mózgu i umyśle oraz ewentualnych problemów etycznych z nimi związanych.

3.5 Sztuczna inteligencja

Rozwój robotyki i sztucznej inteligencji (AI) zależy po części od rozwoju nauk o mózgu i umyśle. Jak twierdzi Brooks (2003) w swojej książce, prawdopodobnie istnieje pewna luka w naszej wiedzy o systemach biologicznych, która uniemożliwia stworzenie prawdziwie inteligentnych, a może nawet świadomych robotów. Jego zdaniem nie odkryliśmy jeszcze właściwego „algorytmu biologicznego” (lub też algorytmów), który odpowiada za nasze funkcjonowanie poznawcze, np. zdolność do rozpoznawania obiektów czy uczenia się i działania pamięci. Z kolei projekt standardowego modelu umysłu i próby jego implementacji w systemach komputerowych (robotach kognitywnych) przedstawiają w swojej pracy Laird, Lebiere i Rosenbloom (2017). Autorzy piszą o ważności współpracy badaczy z różnych dziedzin, m.in. kognitywistyki i robotyki, w celu wypracowania zgodnego modelu umysłu i inteligencji, dającego się jednocześnie zaprogramować w systemach AI. Wymieniają także modele wypracowanych dotychczas architektur kognitywnych, wzorowanych na wiedzy nauk o mózgu i umyśle (zob. także Goertzel i in., 2010). Tym samym neuronauki podają pewne wzory i wskazówki dla twórców sztucznej inteligencji, w jaki sposób można projektować ludzkie zachowania w robotach. Floreano, Ijspeert i Schaal (2014) również piszą o współpracy nauk neurokognitywnych i robotyki, w tym o wzajemnym testowaniu swoich biologicznych i komputerowych modeli (teorii funkcjonowania) układu nerwowego i umysłu. Podają również przykłady „robotycznych zwierząt”, np. salamandry czy szczurów, w których modelowane były sposoby lokomocji i nawigacji w środowisku tych zwierząt. Mechanizmy działania takich neurorobotów są *stricte* wzorowane na sposobach działania układów nerwowego i poznawczego (Armstrong, Drapeau, 2010, s. 169). Istnieją także próby implementacji systemu neuronów lustrzanych, dzięki któremu robot mógłby

nabywać nowe umiejętności poprzez naśladowanie (imitację) sekwencji ruchów człowieka (tamże, s. 171) Z kolei w tzw. robotach humanoidalnych, wzorowanych na wyglądzie i sposobie zachowania człowieka, projektowane i testowane są na razie proste zachowania ludzkie, jak kontrola równowagi czy sięganie i chwytanie przedmiotów (Floreano, Ijspeert, Schaal, 2014). Niemniej trwają także prace nad zaimplementowaniem dwunożnego chodu czy bardziej złożonych zachowań, opartych na wyższych funkcjach poznawczych, takich jak planowanie sekwencji złożonych ruchów czy uczenie się robotów na „własnych błędach”. Znane obecnie idee sztucznych sieci neuronowych, metody uczenia maszynowego, np. uczenie ze wzmocnieniem (*reinforcement learning*), czy koncepcja hierarchicznej budowy systemu wzrokowego implementowane w robotach zbudowane zostały na wiedzy o sposobie funkcjonowania układu nerwowego zwierząt (Arel, Rose, Karnowski, 2010; Cox, Dean, 2014). Hassabis, Kumaran, Summerfield i Botvinick (2017) również piszą o wiedzy na temat budowy i działania mózgu jako o źródle inspiracji do projektowania i walidacji modeli układu poznawczego, wypracowanych w ramach sztucznej inteligencji. Przykładem są tu algorytmy typu *Long Short-Term Memory* czy implementowane w modelach głębokiego uczenia maszynowego mechanizmy uwagowe, wzorowane na ludzkiej uwadze wzrokowej. Zrozumienie neurokognitywnych podstaw inteligencji i jej poznawczych komponentów pomoże wyznaczyć kierunek badań nad inteligentnymi maszynami (Wang i in., 2020), a tym samym stworzyć możliwość zaprojektowania w nich zdolności dorównujących ludzkim (Lake, Ullman, Tenenbaum, Gershman, 2017, s. 3). Podobnie Voss (2007) pisze, że ogólną sztuczną inteligencję (*Artificial General Intelligence*, AGI) osiągniemy dopiero, gdy uda się zamodelować odpowiedni poziom złożoności zachowań i stanów maszyny, dorównujący ludzkiemu umysłowi i jego funkcjom poznawczym. AGI musi bowiem posiadać ogólne zdolności kognitywne, a nie tylko pojedyncze umiejętności (choćby nawet przewyższające ludzi) w wąskich dziedzinach, jak gra w szachy czy rozpoznawanie twarzy. Tym samym Voss (2007, s. 132) wymienia trzy podstawowe cechy AGI, tj. autonomiczność (system samodzielnie, bez bezpośredniej ingerencji człowieka musi poznawać środowisko, zbierać informacje o nim), zachowanie nakierowane na cel (umiejętność rozwiązywania wskazanych problemów) i adaptacyjność (musi umieć dopasowywać swoje zachowanie do warunków sytuacji, w której się znajduje, a także posiadać w pewnym sensie zdolność do samodoskonalenia). Jakkolwiek fascynujący postęp dokonał się w dziedzinie robotyki i informatyki, nie istnieje jeszcze egzemplarz tak rozumianej ogólnej sztucznej inteligencji. Obecnie systemy AI posiadają przewagę jedynie w pojedynczych, wąskich dziedzinach, lecz ludzką inteligencję cechuje jej uniwersalność czy wieloaspektowość, co trafnie ilustruje np.

teoria Inteligencji Wielorakich Gardnera (Strelau, Doliński, 2008, s. 734). Zatem dopiero pomyślnie zintegrowanie tych umiejętności w jednym systemie pozwoli na powstanie AGI, która mogłaby dorównywać inteligencji człowieka (Voss, 2007, s. 147). W tym względzie wiedza neurokognitywna i psychologiczna odnośnie systemu poznawczego człowieka to jedynie pewny koncept do naśladowania, lecz bynajmniej nie chodzi o dokładne odzwierciedlenie w AI, jeden do jednego, umysłu człowieka. Voss (2007, s. 145) podkreśla zarazem, iż niektóre idee rozwijane w ramach kognitywistyki miały duży wpływ na obecny sposób projektowania systemów sztucznej inteligencji, jak np. koncepcja umysłu ucieleśnionego. Poruszane obecnie kontrowersje etyczne dotyczące połączenia wiedzy neurokognitywistyki i robotyki odnoszą się przeważnie do przyszłych rozwiązań. To znaczy, rysowane są możliwe scenariusze i eksperymenty myślowe dotyczące świadomych, autonomicznych maszyn, zdolnych do rozumowania – jednakże stosownym pytaniem jest tu, w jaki sposób mielibyśmy mierzyć czy sprawdzać obecność i/lub poziom świadomości u maszyn i na którym etapie rozwoju technologicznego „realnie” pojawi się taka potrzeba. O ile minie jeszcze sporo czasu, zanim będzie można mówić (jeśli w ogóle) o świadomej sztucznej inteligencji, o tyle problem odpowiedzialności za szkody spowodowane przez autonomiczne maszyny jest już codziennością (Hevelke, Nida-Rümelin, 2015). Z tego względu pojawiają się głosy, iż należałoby zaimplementować jakiś system wartości, odpowiednik ludzkiej moralności, w systemach sztucznej inteligencji. W rzeczy samej istnieją już badania w tym temacie i pierwsze próby programowania takiego zachowania. Tolmeijer, Kneer, Sarasua, Christen, Bernstein (2020) podają trzy sposoby podejścia do implementowania etycznego zachowania w maszynach. Podejście top-down wychodzi od danej teorii etycznej, której zasady są programowane w systemie komputerowym. Jednakże mogą pojawić się sytuacje konfliktowe (dylematy etyczne), których dana teoria nie obejmie i w których AI nie będzie wiedziała, jak postąpić. Z kolei w strategii bottom-up maszyna uczy się dobrego postępowania na konkretnych, pojedynczych przykładach zachowania własnego i ludzi obecnych w jej otoczeniu. Pojawia się jednak prawdopodobieństwo, że AI nauczy się nie tylko dobrych, ale i złych praktyk od ludzi. Istnieją również sposoby hybrydowe, łączące oba typy podejść. Rozwiązania te mają zapobiec „buntowi maszyn”, którego niektórzy interesariusze i opinia publiczna się obawiają. Są zdania, że pojawi się egzystencjalne ryzyko związane ze sztuczną inteligencją i tym samym domagają się regulacji i kontroli nad rozwojem AI czy tzw. superinteligencji. Jeśli chodzi o pojazdy autonomiczne, Awad i in. (2018) przeprowadzili badanie postaw, opinii ludzi wobec tego, jak powinien się zachować taki pojazd w sytuacjach analogicznych do dylematu wagonika. Otóż na podstawie kilku

milionów odpowiedzi, autorzy sformułowali trzy główne wnioski, tj. tendencję badanych do przedkładania życia ludzi nad życiem zwierząt, do ratowania w pierwszej kolejności osób młodszych (niż starszych) oraz ocalania jak największej liczby osób (tzn. badani wybierali sytuacje, w których można było uratować trzy osoby względem dwóch). Ponadto widoczna była wyraźna zmienność międzykulturowa odpowiedzi badanych pochodzących z ponad dwustu państw, co może prowadzić do kwestii spornych, jeśli interesariusze planują sformułować uniwersalne kodeksy działania sztucznej inteligencji. De facto już dzisiaj istnieją pierwsze próby regulacji oraz implementacji zasad etyki w robotach (Vanderelst, Winfield, 2018), co ma stanowić ekwiwalent ludzkiej moralności czy samokontroli. Autorzy powyższego artykułu postulują konsekwencjalizm jako najbardziej odpowiednią teorię dla robotów. Jednakże z systemem tym wiążą się pewne paradoksy, np. uznając, że ludzkie życie czy szczęście można w prosty sposób sumować i traktować w sposób instrumentalny, jedynie jako środki do osiągnięcia celu. Bowiem zgodnie z utylitaryzmem, kierując się „rachunkiem szczęścia (przyjemności)” można dojść do kuriozalnego wniosku, iż „trzeba skrzywdzić jedną osobę, by uszczęśliwić setki” (Środa, 2010, s. 123). Jednakże człowiek posiada godność i wartość samą w sobie, a wnioski, jak te powyżej są nieakceptowalne. Czy maszyna wiedziałaby (nie wspominając już o rozumieniu) czym są prawa człowieka? Dobrym rozwiązaniem mogłoby okazać się zaimplementowanie dodatkowych reguł czy wartości, na podobieństwo imperatywu Kanta czy praw Asimova (zob. Campa, 2011, s. 67). Dopiero po ich spełnieniu maszyna mogłaby wykonać utylitarystyczny „rachunek szczęścia”. Różne instytucje, takie jak Unia Europejska podjęły już pierwsze kroki w opracowaniu prawnych i etycznych kodeksów rozwoju i użytkowania inteligentnych robotów (zob. Rezolucja Parlamentu Europejskiego, 2017). Parlament Europejski wskazuje, by działanie takich systemów było zgodne z czterema zasadami pryncypializmu i tym samym wzywa projektantów i producentów sztucznej inteligencji do przestrzegania najwyższych standardów etycznych. Ponadto szczególna uwaga poświęcona została kwestii odpowiedzialności za czyny takich autonomicznych systemów, gdzie jednym z rozwiązań mogłoby być przypisanie maszynom „opiekunów prawnych”, którzy byliby odpowiedzialni za wszelkie działania maszyn, podobnie jak rodzice są odpowiedzialni za swoje niepełnoletnie dzieci. W cytowanej rezolucji sformułowane zostały też zalecenia dla użytkowników sztucznej inteligencji. Nauki neurokognitywne wraz z psychologią badają również interakcję człowiek-maszyna oraz zjawisko doliny niesamowitości (Wang, Lilienfeld, Rochat, 2015). Wyniki badań tych nauk pozwolą interesariuszom zaprojektować jak najbardziej odpowiednie, bezpieczne modele sztucznej inteligencji, a tym samym ich akceptację ze strony społeczeństwa oraz przyjazną

interakcję ludzi z maszynami. Badania prowadzone przez Darling (2016; zob. także Coeckelbergh, 2018) pokazują, że już dzisiaj niektórym ludziom dość łatwo przychodzi antropomorfizowanie i empatyzowanie ze zwykłymi robotami, a nawet emocjonalne przywiązanie czy zaufanie do nich na podobieństwo relacji ze zwierzęcymi pupilami. Uzasadnione wydaje się zatem pytanie, czy ludzie będą domagali się przyznania jakichś praw swoim robotom. Czy po awarii takiego robota będą czuli smutek, czy ból, jak po stracie kogoś bliskiego? Czy systemy te nie staną się środkami do manipulowania ludzkimi zachowaniami? Czy w związku z tymi faktami odnośnie interakcji człowiek-maszyna należy projektować roboty, które jak najbliżej przypominają ludzi (czy też zwierzęta) i to nie tylko z zachowania, ale i wyglądu? Czy może lepszym, właściwszym rozwiązaniem byłoby tak je zaprojektować, by zachować wyraźną odrębność, granicę między człowiekiem a maszyną, by ludzie w interakcji z nimi nie mieli wątpliwości, czy rozmawiają z drugim człowiekiem (względnie zwierzęciem), czy z systemem komputerowym.

3.6 Edukacja

Wiedza jak działają mózg i umysł może przynosić korzyści również w systemie nauczania, zarówno dzieci, jak i dorosłych. Mayer i Moreno (2003) wymieniają sposoby redukcji przeciążenia poznawczego w nauce nowego materiału, opierając się na teoriach wypracowanych w ramach kognitywistyki. Piszą m.in., iż przekaz multimodalny, tj. jednoczesne podążanie wzrokiem za tekstem i słuchanie czytanego materiału, jest lepszym sposobem nauki niż przekaz tylko w jednej modalności, tj. albo wzrokowej, albo słuchowej. Niemniej nadmiar bodźców otaczających zarówno dzieci, jak i dorosłych może przynieść odmienne skutki. Bowiem wyniki badań nauk neurokognitywnych nad wielozadaniowością (*multitasking*) pokazują szkodliwość tej praktyki dla przyswajania nowego materiału i wyników w nauce (Junco, Cotten, 2012; Kirschner, Bruyckere, 2017). Z kolei Haier (2016) przedstawia w swojej książce zagadnienie inteligencji z perspektywy nauk o mózgu. Okazuje się, że szybkość przesyłania impulsów nerwowych, grubość kory mózgowej czy ilość połączeń asocjacyjnych (w obrębie jednej półkuli mózgu) i międzypółkulowych koreluje ze wskaźnikami inteligencji. Na podstawie tej wiedzy można by tworzyć specjalne programy nauczania dostosowane do poziomu sprawności układu nerwowego danej osoby, tym samym wspomagać jej rozwój i funkcjonowanie poznawcze. Metody neurokognitywistyki mogłoby być, obok standardowych testów wiedzy i testów behawioralnych, dodatkowym narzędziem analizy i oceny rozwoju neurokognitywnego (Chai i in., 2017) oraz efektów nauczania u

dzieci i dorosłych (Bathelt, Scerif, Nobre, Astle, 2019). Kwestią kontrowersyjną jest tu ewentualny podział ludzi na tych z bardziej i tych z mniej sprawnym układem nerwowym. Bowiem mózg jest do pewnego stopnia plastyczny i przez to może się zmieniać, niwelując np. jego wolniejsze tempo przetwarzania informacji czy przesyłania impulsów nerwowych. Niemniej są to w dużym stopniu zmienne genetyczne i nie każdy deficyt czy niższy poziom funkcjonowania mózgu można zmienić poprzez edukację. Sama definicja inteligencji i sposoby jej pomiaru również budzą kontrowersje i to nie tylko wśród uczonych. Z historii badań nad inteligencją wiadomo, iż osoby o niższym ilorazie inteligencji były stygmatyzowane i marginalizowane. Zatem pojawia się pytanie, czy poszukiwanie jej neurokognitywnych podstaw nie wznowi „starych” kontrowersji i krzywdzących podziałów (zob. Meisenberg, 2019). Dzięki wiedzy neuronauk o sposobie przyswajania wiedzy przez mózg nauczyciele mogą wykorzystać fakt iż, nauka rozłożona w czasie („porcjowanie” materiału) jest bardziej efektywną metodą niż nauka całości materiału za jednym razem. Próbę połączenia wiedzy o mózgu i umyśle z praktyką nauczania przedstawia też Schunk (2012), podając gotowe przykłady i wskazówki. Podkreśla on m.in. rolę emocji i motywacji uczniów w procesie nauczania oraz przewagę stosowania metod nauki opartych na praktycznym rozwiązywaniu problemów i aktywnych dyskusjach tematów w przeciwieństwie do zwykłego nauczania faktów na pamięć. Zatem znajomość zagadnień dotyczących funkcjonowania poznawczego, takich jak mechanizmy uwagi czy konsolidacja pamięci, może pomóc nauczycielom i innym specjalistom w systemie edukacyjnym w wypracowywaniu jak najefektywniejszych metod nauki. Pytaniem, jakie może się tu jednak pojawić jest, czy wybór danej metody nauczania nie będzie faworyzował jednych uczniów, pomijając innych, np. o specjalnych potrzebach, czy po prostu stosujących inne strategie poznawcze w uczeniu niż większość rówieśników. Farah włącza badania neuronauk do dyskusji na temat wpływu statusu socjoekonomicznego na rozwój dzieci i wykazuje, że rzeczywiście dzieci z rodzin o niskim statusie rozwijają się wolniej, nie osiągając prawdopodobnie całości swojego potencjału (Farah, 2018, s. 33-34). Wiedząc zatem jak kluczową rolę odgrywają warunki w jakich rozwija się dziecko, czy rządy państw powinny zadbać o stworzenie odpowiedniego środowiska dla wszystkich dzieci, szczególnie biorąc pod uwagę sytuację dzieci z mniej zamożnych rodzin? Dodatkowo nierówności społeczne w zakresie dostępu do edukacji i wykształcenia mogą się jeszcze powiększyć za sprawą tzw. dopingu czy wzmocnienia kognitywnego (*cognitive enhancement*). Termin ten odnosi się do korzystania ze środków farmaceutycznych i technik stymulacji mózgu w celu poszerzania normalnych (naturalnych) możliwości człowieka w sferze poznawczej. I choć

badania często potwierdzają ich skuteczność, nie u wszystkich badanych notuje się poprawę funkcjonowania. Efekty zależą również od dawek i intensywności stosowania danych metod czy też miejsca stymulacji mózgu (Voarino i in., 2017). Regularne stosowanie psychostymulantów może powodować też uzależnienie od tych substancji. Istotnie, w literaturze przedmiotu wymienia się trzy główne zastrzeżenia odnośnie stosowania dopingu kognitywnego, tj. bezpieczeństwo, autonomię wyboru odnośnie jego stosowania oraz uczciwość (Greely i in., 2008, s. 703). Ponadto nie są znane jeszcze wszystkie krótko- i długotrwałe skutki ich aplikowania, zatem stosunek ryzyka i zysków czerpanych z ich korzystania nie jest do końca jednoznaczny i jasny (Dresler i in., 2012, s. 536-537). Szczególną uwagę należy zachować w przypadku stosowania wzmocnienia poznawczego u dzieci, jako że ich układ nerwowy jest jeszcze bardzo plastyczny i podatny na wszelkie bodźce środowiskowe. Nie wiadomo więc, jakie zmiany neurobiologiczne mogą wywołać u nich w dłuższej perspektywie czasowej pewne środki farmaceutyczne czy stymulacja mózgu. Nie wiadomo również, czy ewentualne zmiany byłyby odwracalne, czy wprowadzałyby trwałą (potencjalnie także negatywną) modyfikację układu nerwowego. Stosowanie tych metod każe również kwestionować ideę uczciwości i sprawiedliwości, jako że pisanie egzaminu czy wykonywanie pracy i zyskiwanie awansu pod wpływem środków zwiększających pamięć, czy uwagę może być uważane za oszukiwanie, podobnie jak ma to miejsce w przypadku dopingu sportowego. Jednak z drugiej strony, rozszerzanie naturalnych możliwości człowieka może być kolejnym stopniem w osiąganiu doskonałości w jakiejś dziedzinie funkcjonowania, do czego ludzie dążą od wieków. Ponadto osoby zatrudnione w niektórych zawodach, np. lekarze, którzy podczas wielogodzinnych operacji decydują o zdrowiu i życiu pacjentów, mogłyby poprzez stosowanie takich środków zwiększyć swoją uwagę czy koncentrację, a więc i efektywność czy bezpieczeństwo wykonywanych przez nich zadań. Czy zatem w takich sytuacjach dopuszczalne, czy nawet wymagane byłoby korzystanie z odpowiednich środków farmaceutycznych czy zabiegów stymulacji mózgu zwiększających możliwości poznawcze? Czy wtedy wręcz nieetyczne byłoby ich niestosowanie skoro sprzyjałyby funkcjonowaniu poznawczemu, np. procesom podejmowania decyzji u medyków? Ponadto badania podkreślają większy pozytywny wpływ na funkcjonowanie poznawcze metod konwencjonalnych, takich jak odpowiednia dieta, odpowiednia długość snu czy regularne ćwiczenia fizyczne niż nootropików („smart drugs”), czy stymulacji mózgu, i to te pierwsze są propagowane w różnych kampaniach społecznych i szkołach. Niemniej osoby cierpiące na zaburzenia i deficyty poznawcze mogą skorzystać z tych nowoczesnych metod i to właśnie ze względu na właściwości terapeutyczne są

prowadzone badania z udziałem nowych substancji i technik stymulacji mózgu, a u osób zdrowych ich wpływ jest najczęściej niewielki. Wydają się one atrakcyjnym wyborem przede wszystkim ze względu na łatwość i szybkość ich stosowania. Z drugiej strony są one wzmocnieniem, a nie zastąpieniem edukacji, nauki czy ogólnie wykonywania pracy – nadal potrzebne jest włożenie jakiegoś wysiłku w pracę. Zatem jeśli zostanie potwierdzone bezpieczeństwo ich stosowania, nie sądzę by istniały poważne przesłanki przemawiające za zakazem ich produkcji i dystrybucji.

4. Podsumowanie

Prawdopodobne podejmowanie decyzji i wydawanie ocen moralnych, będących naczelnymi zadaniami zmysłu moralności, zachodzą według schematu zaproponowanego przez Haidta (*sekcja 2.2*). W mechanizmie tym biorą udział zarówno pierwotne (automatyczne) reakcje afektywne, wtórne emocje moralne, jak i świadome procesy poznawcze oraz racjonalnego rozumowania. Ponadto założenia te zdają się mieć potwierdzenie zarówno w badaniach psychologów, jak i neuronaukowców. Bowiem te pierwotne reakcje afektywne prawdopodobnie zachodzą drogą dolną w koncepcji mózgu emocjonalnego LeDoux, tj. przez obszary podkorowe (wzgórze i ciało migdałowate), a emocje moralne, jako kategoria emocji refleksyjnych, wiążą się z górną drogą powstawania emocji, a więc aktywnością kory nowej (Imbir, 2012, s. 34-37). Ponadto przyjmując, że pierwsze reakcje cielesno-emocjonalne są wrodzonymi mechanizmami o charakterze automatycznym, intuicyjnym, podświadomym, to możemy je dzielić także z niektórymi gatunkami zwierząt. Z kolei (wtórne) emocje moralne mają już charakter refleksyjny (też samoświadomościowy), zakładają zatem udział świadomej, wolicjonalnej aktywności poznawczej oraz powiązanie z procesami intelektualnymi, a więc ich występowanie u innych gatunków zwierząt jest prawdopodobnie ograniczone. Przeto moralność człowieka ma inną formę niż zachowania społeczne innych gatunków – chociaż niewykluczone, że one również posiadają jakąś prostszą postać zmysłu moralnego, szczególnie jeśli chodzi o inne naczelne. Choć takie pierwotne reakcje afektywne są jednym z pierwszych etapów całego procesu formułowania ocen (sądów) moralnych, a zatem mogą mieć istotny wpływ na powstanie określonej emocji moralnej i końcowe wydanie sądu moralnego, to prawdopodobnie same intuicje nie wystarczają dla wydania sądu moralnego. Niemniej emocje moralne mają bardziej złożoną postać niż intuicje, ponieważ oprócz tego, że wiążą się z określoną walencją (charakterem pozytywnym/przyjemnością lub negatywnym/nieprzyjemnością), zakładają też odwołanie się do posiadanych przez jednostkę koncepcji dobra i zła oraz standardów moralnych (osobistych – idealnych i społecznych – powinnościowych), a więc istnieje możliwość, że w takiej formie emocje mogą być podstawą decyzji i sądów o charakterze moralnym, jak zakłada sentymentalizm i jeden z jego współczesnych przedstawicieli Jesse Prinz. W tym sensie emocje te mogą posiadać także element racjonalny, co stanowi stosunkowo nowy wniosek wobec przyjmowanego przez dziesięciolecia poglądu o nieracjonalności emocji i ich niższości względem rozumu. Niemniej za najbardziej

powszechne i trafne obecnie uważa się teorie moralności zakładające mniejszą lub większą współpracę procesów emocjonalnych, poznawczych i racjonalnych. Tym samym część zmysłu moralnego jest wrodzona, a część nabywana i wykształcana przez indywidualne doświadczenia życia w danej kulturze. Nauki empiryczne przemawiają zatem zarówno za pewnym stopniem uniwersalności oraz relatywizmu kulturowego w formułowanych i przyjmowanych zasadach moralnych. Bezpośrednie osobiste wyrządzanie krzywdy drugiemu człowiekowi angażuje w większym stopniu ośrodki w mózgu związane z przetwarzaniem informacji społecznych i emocjonalnych, natomiast sytuacje dylematów nieosobistych, nie wymagające bezpośredniej bliskości fizycznej agensa i pacjensa aktywują znacznie bardziej ośrodki mózgowe związane z świadomymi procesami racjonalizacji oraz kalkulowania zysków i strat. Dodatkowo na podstawie wyników tych badań, wnosi się o istnieniu dwóch głównych sposobów odpowiedzi/radzenia sobie z dylematami moralnymi – te o charakterze deontologicznym (związane z „etyką obowiązku”) oraz utylitarystycznym (związane z wyborem „większego dobra”). Na koniec warto dodać, że badacze w swoich analizach różnic aktywności neuronalnej osób rozwiązujących dylematy moralne odwołują się przede wszystkim do systemów utylitarystycznych i deontologicznych. Tym samym niemal całkowicie pomija się trzeci popularny system etyczny, tj. Arystotelesowską etykę cnót – czasem uczeni piszą o podziale na odpowiedzi utylitarystyczne i nie-utylistarystyczne, jednak nie odwołują się *explicite* do koncepcji Arystotelesa. Jako, że teoria ta mówi przede wszystkim o kształtowaniu moralnego charakteru przez jednostki, możliwe że w zapisie aktywności neurofizjologicznej szczególnie dominowałyby obszary związane z samoświadomością czy innymi procesami psychicznymi odnoszącymi się do własnej osoby. Ponadto osoby badane wskazują postawy i zachowania moralne oraz powiązane z nimi własności charakteru jako fundamentalne w obrazie własnego „ja” oraz jako te, które w największym stopniu świadczą o tożsamości czy osobowości drugiego człowieka (Strohminger, Nichols, 2014). Mogłoby to wskazywać na ważną część natury moralności człowieka, tj. jego charakter. Dodatkowo badania donoszą o istotnej różnicy między deklaracjami osób badanych odnośnie własnych postaw i sposobów postępowania w hipotetycznych sytuacjach, a ich rzeczywistymi zachowaniami (zob. Doliński, 2020; Tangney i in., 2007). Podobne wnioski w odniesieniu do sytuacji o charakterze moralnym formułuje McAuliffe (2019) – przykładowo, badani w jednym z eksperymentów deklarowali w przedstawianej im hipotetycznej sytuacji niesprawiedliwego potraktowania jakiejś osoby, że czuliby złość wobec sprawcy tego nieuczciwego czynu, jednakże gdy te osoby badane zostały postawione w rzeczywistej sytuacji obserwowania niesprawiedliwości wobec jakiejś

osoby, to deklarowały, że nie czują złości/oburzenia na tę sytuację. Tym samym także realne zachowanie w sytuacjach analogicznych do zepchnięcie człowieka z kładki i/lub przekierowanie zwrotnicy i w konsekwencji pozbawienie kogoś życia mogłoby okazać się o wiele trudniejsze do dokonania, niż deklarują badani. Różnice te w przewidywaniach i rzeczywistych zachowaniach moralnych ludzi również powinny zostać zweryfikowane przez przyszłe badania neuroetyki.

Jeśli zaś chodzi o etykę neuronauk, to zarówno konkretne przypadki badań oraz wykorzystywania nowych technologii, jak i dziedziny ich przeznaczenia mają wpływ na „stopień ich kontrowersyjności” i chęci regulacji etyczno-prawnych. Bowiem naukowcy i pacjenci w określonych stanach chorobowych są w stanie przyjąć większe ryzyko oraz inny rachunek zysków i strat (zagrożeń) z nimi związanych, niż podczas prowadzenia badań i wykorzystywania neuro-technologii w celach rozrywkowych – jedynie rozszerzających możliwości poznawcze u zdrowych osób, a nie redukujących określone deficyty funkcjonowania psycho-ruchowego. Zatem również odwołanie się do i kierowanie założeniami określonego systemu etycznego w znacznym stopniu zależy od kontekstu sytuacji i osób zaangażowanych (decyzyjnych). Niemniej postaram się podsumować, który system zasad etycznych moim zdaniem, zdaje się być tym najbardziej odpowiednim w przedstawionych wyżej obszarach życia publicznego. Jeśli chodzi o ocenę etyczną wdrażania osiągnięć neuronauk w dziedzinach medycznych, zasady pryncypializmu oraz Arystotelesowskiej etyki cnót wydają się najbardziej odpowiednie, przynajmniej jako założenia wyjściowe dla analiz konkretnych przypadków. Pytanie, który system wartości etycznych jest najbardziej odpowiedni do oceny wykorzystywania postępu neurokognitywistyki w wojsku, sugerują Lin i współpracownicy (2014, s. 147-150). Przedkładają oni postępowanie według etyki cnót dla interesariuszy zaangażowanych w sferze wojskowości nad inne systemy etyczne. Piszą tym samym, że wzmocnienie kognitywne metodami farmakologicznymi czy stymulacji mózgu jest dopuszczalne, o ile pomoże to żołnierzom działać według cnót/chlubnych cech charakteru, tj. dzięki redukcji stresu lub zmęczenia będą oni zdolni do lepszej oceny sytuacji i tym samym bardziej adekwatnego zachowania, zgodnego z ich postępowaniem w sytuacji gdyby to zmęczenie czy stres i lęk nie występowały. Jeżeli chodzi o obszar marketingu i ekonomii zasadnym wydaje się stworzenie reguł wzorowanych na założeniach deontologii, które ustalałyby dopuszczalność wykorzystywania metod neuronauk w tych dziedzinach. Podobnie w prawie powinny zostać ustalone czy przestrzegane obecnie istniejące zasady postępowania ze skazanymi, jednak uwzględniać także założenia konsekwencjonalizmu i kierowanie się

regułą „większego dobra” dla społeczeństwa. Twórcom sztucznej inteligencji niewątpliwie przyświeca idea polepszenia jakości życia ludzi, czyli maksymalizowanie dobra. Niemniej dodatkowa uwaga powinna zostać skierowana na to, jakimi metodami pragnie się to osiągnąć, a więc i wypracowanie zasad rozwoju AI będących rozwinięciem praw Asimova i nawiązujących najsilniej do deontologizmu, wydaje się odpowiednie. Z kolei w obszarze edukacji, tak jak w poprzednich dziedzinach większość zachowań da się uzasadnić odwołując się do zasady „większego dobra” dla ogółu, jednak pewne sposoby postępowania zdaje się, że nie powinny zyskać miana pożądanых. Tym samym w odniesieniu do sposobów wzmocnienia poznawczego w nauce dzieci powinny zostać ustalone pewne nieprzekraczalne granice/zasady, a jeśli chodzi o dorosłych to odwołanie do wartości kształtowania Arystotelesowskiego moralnego charakteru – dodatkowo społeczną aprobatę pewnych cech i zachowań – wydaje się najbardziej stosownym rozwiązaniem.

Bibliografia

- Aicardi, C., Akintoye, S., Fothergill, B. T., Guerrero, M., Klinker, G., Knight, W., ... & Ulnicane, I. (2020). Ethical and social aspects of neurorobotics. *Science and Engineering Ethics*, 26(5), 2533-2546.
- Arbuckle, M. R., Travis, M. J., & Ross, D. A. (2017). Integrating a neuroscience perspective into clinical psychiatry today. *JAMA psychiatry*, 74(4), 313-314.
- Arel, I., Rose, D. C., & Karnowski, T. P. (2010). Deep machine learning-a new frontier in artificial intelligence research [research frontier]. *IEEE computational intelligence magazine*, 5(4), 13-18.
- Ariely, D., & Berns, G. S. (2010). Neuromarketing: the hope and hype of neuroimaging in business. *Nature reviews neuroscience*, 11(4), 284-292.
- Armstrong, R. E., & Drapeau, M. D. (Eds.). (2010). *Bio-inspired innovation and national security*. NDU Press.
- Astolfi, L., Fallani, F. D. V., Cincotti, F., Mattia, D., Bianchi, L., Marciani, M. G., ... & Babiloni, F. (2008). Neural basis for brain responses to TV commercials: a high-resolution EEG study. *IEEE Transactions on Neural Systems and rehabilitation engineering*, 16(6), 522-531.
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ... & Rahwan, I. (2018). The moral machine experiment. *Nature*, 563(7729), 59-64.
- Ayala, F. J. (2009). What the Biological Sciences can and cannot Contribute to Ethics. *Contemporary Debates in Philosophy of Biology*, 316-336.
- Bago, B., & De Neys, W. (2019). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*, 148(10), 1782.
- Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, 121(1), 154-161.
- Bathelt, J., Scerif, G., Nobre, A. C., & Astle, D. E. (2019). Whole-brain white matter organization, intelligence, and educational attainment. *Trends in neuroscience and education*, 15, 38-47.
- Berker, S. (2009). The normative insignificance of neuroscience. *Philosophy & Public Affairs*, 37(4), 293-329.

- Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in cognitive sciences*, 11(9), 387-392.
- Bołoz, W. (2003). Bioetyka. Jej historia i sposoby ujmowania. *Studia Ecologiae et Bioeticae*, 1, 161-175.
- Bremer, J. (2010). „Status ontyczny osoby z perspektywy neuronauk.”. *Analiza i Egzystencja*, 11, 264-290.
- Brooks, R. A. (2003). *Flesh and machines: How robots will change us*. Vintage.
- Cacioppo, J. T., Amaral, D. G., Blanchard, J. J., Cameron, J. L., Carter, C. S., Crews, D., ... & Quinn, K. J. (2007). Social neuroscience: Progress and implications for mental health. *Perspectives on Psychological Science*, 2(2), 99-123.
- Campa, R. (2011). Kodeksy etyczne robotów: zagadnienie kontroli sprawowanej przez człowieka. *Pomiary Automatyka Robotyka*, 15, 66-70.
- Campos, D., Escudero-Marín, M., Snitman, C. M., Torres-Espínola, F. J., Azaryah, H., Catena, A., & Campoy, C. (2020). The Nutritional Profile of Food Advertising for School-Aged Children via Television: A Longitudinal Approach. *Children*, 7(11), 230.
- Chai, L. R., Khambhati, A. N., Ciric, R., Moore, T. M., Gur, R. C., Gur, R. E., ... & Bassett, D. S. (2017). Evolution of brain network dynamics in neurodevelopment. *Network Neuroscience*, 1(1), 14-30.
- Churchland, P. S. (1996). The hornswoggle problem. *Journal of Consciousness Studies*, 3(5-6), 402-408.
- Churchland, P. S. (1989). *Neurophilosophy: Toward a unified science of the mind-brain*. Cambridge, MA: MIT Press.
- Churchland, P. S. (2007). Neurophilosophy: the early years and new directions. *Functional Neurology*, 22(4), 185-195.
- Churchland, P. S. (2013). *Moralność mózgu. Co neuronauka mówi o moralności*. Przeł. Hohol, M., Marek, N. Kraków: Copernicus Center Press.
- Ciaramelli, E., Muccioli, M., Làdavas, E., & Di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social cognitive and affective neuroscience*, 2(2), 84-92.
- Cima, M., Tonnaer, F., & Hauser, M. D. (2010). Psychopaths know right from wrong but don't care. *Social cognitive and affective neuroscience*, 5(1), 59-67.
- Coeckelbergh, M. (2018). Why care about robots? Empathy, moral standing, and the language of suffering. *Kairos. Journal of Philosophy & Science*, 20(1), 141-158.

- Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., & Walters, L. (1998). New goals for the US human genome project: 1998-2003. *Science*, 282(5389), 682-689.
- Cooper, J. J., Korb, A. S., & Akil, M. (2019). Bringing Neuroscience to the Bedside. *Focus (American Psychiatric Publishing)*, 17(1), 2-7.
- Cox, D. D., & Dean, T. (2014). Neural networks and neuroscience-inspired computer vision. *Current Biology*, 24(18), R921-R929.
- Curry, O., Whitehouse, H., & Mullins, D. (2019). Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Current Anthropology*, 60(1).
- Damasio A. (1999). *Błąd Kartezjusza. Emocje, rozum i ludzki mózg*. Tłum. M. Karpiński. Poznań: Dom Wydawniczy Rebis.
- Damasio, A. (2007). Neuroscience and Ethics: Intersections. *The American Journal of Bioethics*, 7(1): 3–7.
- Darling, K. (2016). Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In *Robot law*. Edward Elgar Publishing.
- Dayan, E., Censor, N., Buch, E. R., Sandrini, M., & Cohen, L. G. (2013). Noninvasive brain stimulation: from physiology to network dynamics and back. *Nature neuroscience*, 16(7), 838-844.
- Dąbrowski, A. (2012). Wpływ emocji na poznawanie. *Przegląd Filozoficzny. Nowa Seria*, (3), 315-335.
- Dąbrowski, A. (2014). Czym są emocje?. *Analiza i Egzystencja: czasopismo filozoficzne*, (27), 123-146.
- De Waal, F. (2013). *Małpy i filozofowie: skąd pochodzi moralność?*. Kraków: Copernicus Center Press.
- Decety, J., & Keenan, J. P. (2006). Social neuroscience: A new journal. *Social Neuroscience*, 1(1), 1-4.
- Deming, P., & Koenigs, M. (2020). Functional neural correlates of psychopathy: a meta-analysis of MRI data. *Translational psychiatry*, 10(1), 1-8.
- Dresler, M., Sandberg, A., Ohla, K., Bublit, C., Trenado, C., Mroczko-Wąsowicz, A., ... & Repantis, D. (2013). Non-pharmacological cognitive enhancement. *Neuropharmacology*, 64, 529-543.
- Duke, A. A., & Bègue, L. (2015). The drunk utilitarian: Blood alcohol concentration predicts utilitarian responses in moral dilemmas. *Cognition*, 134, 121-127.

- Dziarnowska, W., & Przybysz, P. (2010). Emocje moralne jako podstawa zmysłu moralnego. Defekty zmysłu moralnego. *Rocznik Kognitywistyczny*, 4.
- Earp, B. D., McLoughlin, K., Monrad, J., Clark, M. S., & Crockett, M. (2020). How social relationships shape moral judgment.
- Edersheim, J. G., Brendel, R. W., & Price, B. H. (2012). 10 Neuroimaging, Diminished Capacity and Mitigation. *Neuroimaging in Forensic Psychiatry: From the Clinic to the Courtroom*.
- Evans, V., & Green, M. (2006). *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Farah, M. J. (2005). Neuroethics: the practical and the philosophical. *Trends in cognitive sciences*, 9(1), 34-40.
- Farah, M. J. (Ed.). (2010). *Neuroethics: an introduction with readings* (Vol. 28). Mit Press.
- Farah, M. J. (2012). Neuroethics: the ethical, legal, and societal impact of neuroscience. *Annual review of psychology*, 63, 571-591.
- Farah, M. J. (2018). Socioeconomic status and the brain: prospects for neuroscience-informed policy. *Nature Reviews Neuroscience*, 19(7), 428-438.
- Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in cognitive sciences*, 11(10), 419-427.
- Fingas, B. (2011). Etyka cnót, jako odradzający się paradygmat filozofii moralnej. *Ogrody Nauk i Sztuk*, (1), 11-21.
- Fischborn, M. (2016). Libet-style experiments, neuroscience, and libertarian free will. *Philosophical Psychology*, 29(4), 494-502.
- Floreano, D., Ijspeert, A. J., & Schaal, S. (2014). Robotics and neuroscience. *Current Biology*, 24(18), R910-R920.
- Frijda, N. (1998). Różnorodność afektu: emocje i zdarzenia, nastroje i sentymenty. W: P. Ekman, R. Davidson (red.), *Natura emocji*, (s. 56-63). Gdańsk: GWP.
- Giacino, J. T., Fins, J. J., Laureys, S., & Schiff, N. D. (2014). Disorders of consciousness after acquired brain injury: the state of the science. *Nature Reviews Neurology*, 10(2), 99.
- Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford University Press, USA.
- Giordano, J., & Shook, J. R. (2018). Neuroethics: What it is, does, and should do. *Health Care Ethics USA*, 1-5.
- Glenn, A. L., Raine, A., & Schug, R. A. (2009). The neural correlates of moral decision-making in psychopathy. *Molecular psychiatry*, 14(1), 5-6.

- Goertzel, B., Lian, R., Arel, I., De Garis, H., & Chen, S. (2010). A world survey of artificial brain projects, Part II: Biologically inspired cognitive architectures. *Neurocomputing*, 74(1-3), 30-49.
- Goodenough, O. R., & Tucker, M. (2010). Law and Cognitive Neuroscience. *Annual Review of Law and Social Science*, 6, 61-92.
- Gordon, E. L., Ariel-Donges, A. H., Bauman, V., & Merlo, L. J. (2018). What is the evidence for “food addiction?” A systematic review. *Nutrients*, 10(4), 477.
- Gosseries, O., Di, H., Laureys, S., & Boly, M. (2014). Measuring consciousness in severely damaged brains. *Annual Review of Neuroscience*, 37, 457-478.
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of personality and social psychology*, 101(2), 366.
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology* (Vol. 47, pp. 55-130). Academic Press.
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological inquiry*, 23(2), 101-124.
- Gray, K., Schein, C., & Cameron, C. D. (2017). How to think about emotion and morality: Circles, not arrows. *Current opinion in psychology*, 17, 41-46.
- Greely, H. T. (2009). Law and the Revolution in Neuroscience: An Early Look at the Field. *Akron Law Review*, 42(3), 687-715.
- Greely, H. T., Grady, C., Ramos, K. M., Chiong, W., Eberwine, J., Farahany, N. A., Johnson, L. S. M., Hyman, B. T., Hyman, S. E., Rommelfanger, K. S., Serrano, E. E. (2018). Neuroethics guiding principles for the NIH BRAIN initiative. *Journal of Neuroscience*, 38(50), 10586-10588.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105-2108.
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44(2), 389-400.
- Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, 111(3), 364-371.
- Greene, J. D. (2014). Beyond point-and-shoot morality: Why cognitive (neuro) science matters for ethics. *Ethics*, 124(4), 695-726.

- Haidt, J. (2003). The moral emotions. In: R. J. Davidson, K. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852–870). Oxford, England: Oxford University Press
- Haidt, J. (2008). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. In: J. E. Adler, L. J. Rips, *Reasoning: Studies of human inference and its foundations (1024-1053)*. Cambridge: Cambridge University Press.
- Haidt J. (2014) *Prawy umysł. Dlaczego dobrych ludzi dzieli religia i polityka?* Sopot: Smak Słowa.
- Haier, R. J. (2016). *The neuroscience of intelligence*. Cambridge University Press.
- Hain, R., & Saad, T. (2016). Foundations of practical ethics. *Medicine*, 44(10), 578-582.
- Hall, W. D., Carter, L., & Morley, K. I. (2003). Addiction, neuroscience and ethics. *Addiction*, 98(7), 867-870.
- Hariz, M., & Amadio, J. P. (2015). The new era of neuromodulation. *AMA Journal of Ethics*, 17(1), 74-81.
- Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95(2), 245-258.
- Hevelke, A., & Nida-Rümelin, J. (2015). Responsibility for crashes of autonomous vehicles: an ethical analysis. *Science and engineering ethics*, 21(3), 619-630.
- Horne, Z., & Powell, D. (2016). How large is the role of emotion in judgments of moral dilemmas?. *PloS one*, 11(7), e0154780.
- Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in cognitive sciences*, 13(1), 1-6.
- Illes, J., & Bird, S. J. (2006). Neuroethics: a modern context for ethics in neuroscience. *Trends in neurosciences*, 29(9), 511-517.
- Illes, J., Kirschen, M. P., Edwards, E., Stanford, L. R., Bandettini, P., Cho, M. K., ... & Michael, D. B. (2006). Incidental findings in brain imaging research: what should happen when a researcher sees a potential health problem in a brain scan from a research subject?. *Science (New York, NY)*, 311(5762), 783.
- Illes, J., & Sahakian, B. J. (Eds.). (2013). *Oxford handbook of neuroethics*. Oxford University Press.
- Imbir, K. (2012). *Odmienność emocji automatycznych i refleksyjnych: poszukiwanie różnicowania neurobiologicznego i psychologicznego*. Rozprawa doktorska. Uniwersytet Warszawski Wydział Psychologii, Warszawa.

- Insel, T. R., Landis, S. C., & Collins, F. S. (2013). The NIH brain initiative. *Science*, 340(6133), 687-688.
- Jasielska, A. (2013). Reprezentacja współczesnych modeli emocji w dziełach sztuki. *Teksty Drugie*, (6), 205-221.
- Johnson, J. (2013). *Designing with the mind in mind: simple guide to understanding user interface design guidelines*. Elsevier.
- Jones, O. D., Buckholtz, J., Schall, J. D., & Marois, R. (2009). Brain Imaging for Legal Thinkers: A Guide for the Perplexed. *Stanford Technology Law Review*, 5, 10-09.
- Jotterand, F., & Giordano, J. (2011). Transcranial magnetic stimulation, deep brain stimulation and personal identity: ethical questions, and neuroethical approaches for medical practice. *International review of psychiatry (Abingdon, England)*, 23(5), 476.
- Junco, R., & Cotten, S. R. (2012). No A 4 U: The relationship between multitasking and academic performance. *Computers & Education*, 59(2), 505-514.
- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
- Kaye, W. H., Wierenga, C. E., Bailer, U. F., Simmons, A. N., & Bischoff-Grethe, A. (2013). Nothing tastes as good as skinny feels: the neurobiology of anorexia nervosa. *Trends in neurosciences*, 36(2), 110-120.
- Kirschner, P. A., & De Bruyckere, P. (2017). The myths of the digital native and the multitasker. *Teaching and Teacher Education*, 67, 135-142.
- Klar, P. (2020). What is neurophilosophy: Do we need a non-reductive form?. *Synthese*, 1-25.
- Klein, G., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid decision making on the fire ground: The original study plus a postscript. *Journal of Cognitive Engineering and Decision Making*, 4(3), 186-209.
- Knobe, J., & Nichols, S. (Eds.). (2013). *Experimental Philosophy: Volume 2* (Vol. 2). Oxford University Press.
- Knutson, B., & Huettel, S. A. (2015). The risk matrix. *Current Opinion in Behavioral Sciences*, 5, 141-146.
- Knutson, B., Rick, S., Wimmer, G. E., Prelec, D., & Loewenstein, G. (2007). Neural predictors of purchases. *Neuron*, 53(1), 147-156.
- Koculak, M., & Kałwak, W. (2014). Opisać świadomość na skali od zera do jeden: perturbacyjny Indeks Złożoności jako naukowa próba pomiaru świadomości na poziomie indywidualnym. *Rocznik Kognitywistyczny*, 7, 9-20.

- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, 446(7138), 908-911.
- Koenigs, M., Kruepke, M., Zeier, J., & Newman, J. P. (2012). Utilitarian moral judgment in psychopathy. *Social cognitive and affective neuroscience*, 7(6), 708-714.
- Kozhevnikova, M. (2019). Żywe–nie żywe: mózgowie organoidy i doświadczenia na zwierzętach. *ZOOPHILOLOGICA. Polish Journal of Animal Studies*, (5).
- Kroll, J., & Egan, E. (2004). Psychiatry, moral worry, and the moral emotions. *Journal of Psychiatric Practice*®, 10(6), 352-360.
- Kupfer, D. J., & Regier, D. A. (2011). Neuroscience, clinical evidence, and the future of psychiatric classification in DSM-5. *American Journal of Psychiatry*, 168(7), 672-674.
- Kurek, Ł. (2011). Neurofilozofia jako filozofia w kontekście nauki. W: B. Brożek, J. Mączka, W. Grygiel, M. Hohol (red.), *Oblicza racjonalności. Wokół myśli Michała Hellera*, 63-82.
- Kushniruk, A. W., & Patel, V. L. (2004). Cognitive and usability engineering methods for the evaluation of clinical information systems. *Journal of biomedical informatics*, 37(1), 56-76.
- Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *Ai Magazine*, 38(4), 13-26.
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and brain sciences*, 40.
- Lebedev, M. A., & Nicolelis, M. A. (2017). Brain-Machine Interfaces: From Basic Science to Neuroprostheses and Neurorehabilitation. *Physiological Reviews*, 97(2), 767-837.
- Lefaucheur, J. P., Aleman, A., Baeken, C., Benninger, D. H., Brunelin, J., Di Lazzaro, V., ... & Jääskeläinen, S. K. (2020). Evidence-based guidelines on the therapeutic use of repetitive transcranial magnetic stimulation (rTMS): an update (2014–2018). *Clinical neurophysiology*, 131(2), 474-528.
- Levy, N. (2007). *Neuroethics: Challenges for the 21st century*. Cambridge University Press.
- Lin, P., Mehlman, M., Abney, K., French, S., Vallor, S., Galliot, J., ... & Schuknecht, S. (2014). Super Soldiers (Part 2): The Ethical, Legal, and Operational Implications. In *Global Issues and ethical considerations in human enhancement technologies* (pp. 139-160). IGI Global.
- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & memory*, 12(4), 361-366.

- Marecka, A. (2018). Istota badań pomiarowych z wykorzystaniem narzędzi Eye-tracking. *Zeszyty Naukowe WSP*, (3), 40-57.
- Markram, H. (2012). The human brain project. *Scientific American*, 306(6), 50-55.
- Martin, A., Becker, K., Darragh, M., & Giordano, J. (2016). A four-part working bibliography of neuroethics: part 3—"second tradition neuroethics"—ethical issues in neuroscience. *Philosophy, Ethics, and Humanities in Medicine*, 11(1), 1-55.
- Matusz, P. J., Traczyk, J., & Gąsiorowska, A. (2011). Kwestionariusz Potrzeby Poznania—konstrukcja i weryfikacja empiryczna narzędzia mierzącego motywację poznawczą. *Psychologia Społeczna*, 62(17), 113-128.
- May, J., Workman, C., Han, H., & Haas, J. (2020). The Neuroscience of Moral Judgment: Empirical and Philosophical Developments. In: *Neuroscience and Philosophy. Bringing the Fields Together*. W. Sinnott-Armstrong & F. De Brigard, Eds.: 1–21. Cambridge, MA: MIT Press.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, 38(1), 43-52.
- McAuliffe, K., Blake, P. R., Steinbeis, N., & Warneken, F. (2017). The developmental foundations of human fairness. *Nature Human Behaviour*, 1(2), 1-9.
- McAuliffe, W. H. (2019). Do emotions play an essential role in moral judgments?. *Thinking & Reasoning*, 25(2), 207-230.
- McClure, S. M., Li, J., Tomlin, D., Cypert, K. S., Montague, L. M., & Montague, P. R. (2004). Neural correlates of behavioral preference for culturally familiar drinks. *Neuron*, 44(2), 379-387.
- Meisenberg, G. (2019). Should cognitive differences research be forbidden?. *Psych*, 1(1), 306-319.
- Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgement in frontotemporal dementia. *Cognitive and behavioral neurology*, 18(4), 193-197.
- Merabet, L. B. (2011). Building the bionic eye: an emerging reality and opportunity. In *Progress in brain research* (Vol. 192, pp. 3-15). Elsevier.
- Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in cognitive sciences*, 7(3), 141-144.
- Moll, J., & de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends in cognitive sciences*, 11(8), 319-321.

- Moll, J., Zahn, R., de Oliveira-Souza, R., Bramati, I. E., Krueger, F., Tura, B., ... & Grafman, J. (2011). Impairment of prosocial sentiments is associated with frontopolar and septal damage in frontotemporal dementia. *Neuroimage*, 54(2), 1735-1742.
- Musk, E. (2019). An integrated brain-machine interface platform with thousands of channels. *Journal of medical Internet research*, 21(10), e16194.
- Nahmias, E. (2015). Why we have free will. *Scientific American*, 312(1), 76-80.
- Naor, N., Ben-Ze'ev, A., & Okon-Singer, H. (2014). The modern search for the Holy Grail: is neuroscience a solution?. *Frontiers in human neuroscience*, 8, 388.
- National Research Council. (2008). *Emerging Cognitive Neuroscience and Related Technologies*
- National Research Council. (2009). *Opportunities in neuroscience for future army applications*. National Academies Press.
- Niebrój, L. T. (2013). Neuroetyka: Nowa Jakość Etyki Medycznej? [Neuroethics: New Quality of Medical Ethics?]. *Roczniki Pomorskiej Akademii Medycznej W Szczecinie*, 59(1):130–136.
- O'Connor, C., & Joffe, H. (2013). How has neuroscience affected lay understandings of personhood? A review of the evidence. *Public Understanding of Science*, 22(3), 254-268.
- Park, R. J., Singh, I., Pike, A. C., & Tan, J. O. (2017). Deep brain stimulation in anorexia nervosa: hope for the hopeless or exploitation of the vulnerable? The oxford neuroethics gold standard framework. *Frontiers in Psychiatry*, 8, 44.
- Pascual, L., Gallardo-Pujol, D., & Rodrigues, P. (2013). How does morality work in the brain? A functional and structural perspective of moral behavior. *Frontiers in integrative neuroscience*, 7, 65.
- Poczobut, R. (2013). Naturalizowanie jaźni. *Przegląd Filozoficzny. Nowa Seria*, (2), 373-388.
- Podlipniak, P., & Przybysz, P. (2013). Sztuka, mózg, muzyka: perspektywy neuroestetyki muzyki.
- Prinz, J. J. (2014). Where do morals come from?—A plea for a cultural approach. In *Empirically informed ethics: Morality between facts and norms* (pp. 99-116). Springer, Cham.
- Przybysz, P. (2012). Neurofilozofia i filozofia neuronauk. W: M. Miłkowski, R. Poczobut (red.), *Przewodnik po filozofii umysłu*. Kraków: Wydawnictwo WAM, 689-734.
- Racine, E., & Aspler, J. (Eds.). (2017). *Debates about neuroethics: perspectives on its development, focus, and future*. Springer.

Racine, Eric; Dubljević, Veljko; Jox, Ralf J.; Baertschi, Bernard; Christensen, Julia F.; Farisco, Michele; Jotterand, Fabrice; Kahane, Guy; Müller, Sabine (2017). *Can Neuroscience Contribute to Practical Ethics? A Critical Review and Discussion of the Methodological and Translational Challenges of the Neuroscience of Ethics*. *Bioethics*, 31(5), 328–337.

Reynolds III, C. F., Lewis, D. A., Detre, T., Schatzberg, A. F., & Kupfer, D. J. (2009). The future of psychiatry as clinical neuroscience. *Academic medicine: journal of the Association of American Medical Colleges*, 84(4), 446.

Rezolucja Parlamentu Europejskiego z dnia 16 lutego 2017 r. zawierające zalecenia dla Komisji w sprawie przepisów prawa cywilnego dotyczących robotyki (2015/2103(INL)).
Pobrano (01.03.2021) z:
https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_PL.html

Rommelfanger, K. S., Jeong, S. J., Ema, A., Fukushima, T., Kasai, K., Ramos, K. M., Salles, A., Singh, I. (2018). Neuroethics Questions to Guide Ethical Research in the International Brain Initiatives. *Neuron*, 100(1), 19-36.

Roskies, A. (2002). Neuroethics for the new millenium. *Neuron*, 35(1), 21-23.

Schacter, D. L., & Loftus, E. F. (2013). Memory and law: what can cognitive neuroscience contribute?. *Nature neuroscience*, 16(2), 119-123.

Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, 22(1), 32-70.

Scherer, K. R. (2011). On the rationality of emotions: or, When are emotions rational?. *Social Science Information*, 50(3-4), 330-350.

Schunk, D. H. (2012). *Learning theories an educational perspective sixth edition*. Pearson.

Shearer, B. (2018). Multiple Intelligences in teaching and education: Lessons learned from neuroscience. *Journal of Intelligence*, 6(3), 38.

Smeding, H. M. M., Goudriaan, A. E., Foncke, E. M. J., Schuurman, P. R., Speelman, J. D., & Schmand, B. (2007). Pathological gambling after bilateral subthalamic nucleus stimulation in Parkinson disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 78(5), 517-519.

Smith, D. G., & Robbins, T. W. (2013). The neurobiological underpinnings of obesity and binge eating: a rationale for adopting the food addiction model. *Biological psychiatry*, 73(9), 804-810.

Sójka, J. (2007). Idea etyki stosowanej. Pomiedzy wyzynami filozofii a wymogami praktyki. *Annales. Ethics in Economic Life*, 10(1), 117-126.

Strelau, J., & Doliński, D. (2008). *Psychologia: podręcznik akademicki*. Gdańskie Wydawnictwo Psychologiczne.

- Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, 131(1), 159-171.
- Stupak, R. (2020). Model biomedyczny w psychopatologii. Krytyczny szkic historyczny, współczesny kontekst i problemy etyczne. *Diametros*, 17(66), 34-51.
- Suhler, C. L., & Churchland, P. (2011). Can innate, modular “foundations” explain morality? Challenges for Haidt's moral foundations theory. *Journal of cognitive neuroscience*, 23(9), 2103-2116.
- Sunstein, C. R. (2005). Moral heuristics. *Behavioral and brain sciences*, 28(4), 531-541.
- Szutta, A. (2017). Artur Szutta: Spór o emocje moralne, *Filozofuj*. Pobrane z (dostęp 26.03.2021): <https://filozofuj.eu/artur-szutta-spor-o-emocje-moralne/>
- Środa, M. (2010). *Etyka dla myślących*. Warszawa: Wydawnictwo Czarna Owca.
- Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annu. Rev. Psychol.*, 58, 345-372.)
- Tanielian, T. L., Tanielian, T., & Jaycox, L. (2008). *Invisible wounds of war: Psychological and cognitive injuries, their consequences, and services to assist recovery* (Vol. 1). Rand Corporation.
- Tolmeijer, S., Kneer, M., Sarasua, C., Christen, M., & Bernstein, A. (2020). Implementations in machine ethics: A survey. *ACM Computing Surveys (CSUR)*, 53(6), 1-38.
- Urbaniak, M. (2014). Rozpoznawanie stanów emocjonalnych jako biologiczne źródło zmysłu moralności. *Rocznik Kognitywistyczny*, (7), 59-86.
- Vanderelst, D., & Winfield, A. (2018). An architecture for ethical robots inspired by the simulation theory of cognition. *Cognitive Systems Research*, 48, 56-66.
- Voarino, N., Dubljević, V., & Racine, E. (2017). tDCS for memory enhancement: analysis of the speculative aspects of ethical issues. *Frontiers in human neuroscience*, 10, 678.
- Voss, P. (2007). Essentials of general intelligence: The direct path to artificial general intelligence. In *Artificial general intelligence* (pp. 131-157). Springer, Berlin, Heidelberg.
- Walter, U., Fernández-Torre, J. L., Kirschstein, T., & Laureys, S. (2018). When is “brainstem death” brain death? The case for ancillary testing in primary infratentorial brain lesion. *Clinical Neurophysiology*, 129(11), 2451-2465.
- Waltz, E. (2017). *French Regulators Approve Human Trial of a Bionic Eye*. IEEE Spectrum. Data dostępu: 11.12.2020. Pobrane z: <https://spectrum.ieee.org/the-human-os/biomedical/bionics/french-regulators-approve-human-trial-of-a-bionic-eye>
- Wang, Y., Kwong, S., Leung, H., Lu, J., Smith, M. H., Trajkovic, L., ... & Kinsner, W. (2020). Brain-inspired systems: A transdisciplinary exploration on cognitive cybernetics, humanity,

- and systems science toward autonomous artificial intelligence. *IEEE Systems, Man, and Cybernetics Magazine*, 6(1), 6-13.
- Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, 19(4), 393-407.
- Wilson, E., O. (2000). *Socjobiologia. Wydanie popularnonaukowe*. Tłum. Siemiński, M. Poznań: Zysk i S-ka.
- Wojciszke, B., & Baryła, W. (2000). Potoczne rozumienie moralności: pięć kodów etycznych i narzędzie ich pomiaru. *Przegląd Psychologiczny*, 43(4), 395-421.
- Young, L., & Koenigs, M. (2007). Investigating emotion in moral cognition: a review of evidence from functional neuroimaging and neuropsychology. *British medical bulletin*, 84(1), 69-79.
- Young, L., & Saxe, R. (2011). Moral universals and individual differences. *Emotion Review*, 3(3), 323-324.
- Zyzik, R. (2013). Dowody Neuronaukowe W Polskim Prawie Dowodowym (Neuroscientific Evidence in the Polish Law of Evidence). In *Forum Prawnicze* (Vol. 2, No. 16, pp. 23-34).
- Żylicz, O. (2012). O trzech orientacjach moralnych. *Studia Psychologiczne (Psychological Studies)*, 50(2).