

ADRIANNA GRABIZNA, RENATA MACIEJEWSKA,
MARTA MOCZULSKA
Uniwersytet Zielonogórski

TEORIA UMYSŁU A AWERSJA DO NIERÓWNOŚCI – PROJEKT BADANIA Z WYKORZYSTANIEM GRY ULTIMATUM Z UDZIAŁEM OSÓB ZE SPEKTRUM ZABURZEŃ AUTYSTYCZNYCH

WSTĘP

Istnienie teorii umysłu, dzięki której możliwe jest myślenie w kategoriach intencjonalnych o swoich i cudzych przekonaniach, pragnieniach, chęciach, oczekiwaniach, decyzjach, zamiarach etc., było postulowane najpierw przez filozofa i kognitywistę Daniela Dennetta (1971, cf. Seyfarth, Cheney, 2002). Autor mówił w tym kontekście o psychologii naiwnej (*folk psychology*), którą posługujemy się, gdy interpretujemy, wyjaśniamy i przewidujemy zachowanie swoje własne, innych ludzi, zwierząt (ale także robotów i komputerów). Badanie teorii umysłu było jednym z elementów, które wpisało się w rozwój kognitywistyki, która jest ze swej natury interdyscyplinarna (łączy badania z zakresu językoznawstwa, sztucznej inteligencji, neuronauk, etologii czy filozofii umysłu). W tym kontekście teoria umysłu szybko stała się jednym z najbardziej owocnych narzędzi generowania wielu różnych protokołów eksperymentów behawioralnych, zaczynając od etologicznych, badających teorię umysłu u szympan-sów (Premack, Woodruff, 1978), u dzieci (Wimmer, Perner, 1983), a niedługo potem także do badania deficytów teorii umysłu u osób ze spektrum zaburzeń autystycznych (Autism Spectrum Disorder, w skrócie ASD¹) (Baron-Cohen i in., 1985). Po badaniach

¹ Zgodnie z przyjętą w literaturze przedmiotu terminologią (Pisula, 2010; Pietras, Witusik, Gałecki, 2010) będziemy się tu posługiwać terminem „spektrum zaburzeń autystycznych” (*Autism*

behawioralnych nastąpił wysyp badań z zakresu neuronauk (Baron-Cohen i in., 2000; Frith, Frith, 2003²).

Zanim rozpoczęto badania nad teorią umysłu, założono jej istnienie u prymatów ze względu na fakt, że są to gatunki oparte na współpracy. Niedawne badania pokazują też, że wszystkie gatunki, u których badano i stwierdzono awersję do nierówności (tj. reakcję na nierówną dystrybucję nagrody za wykonanie tego samego zadania), także są gatunkami opartymi na współpracy. Mimo to, wciąż nie jest jasne, czy i jak obecność awersji do nierówności wiąże się z obecnością teorii umysłu. W tym artykule chcemy zwrócić uwagę na możliwy związek między teorią umysłu a awersją do nierówności. Przedstawiamy propozycję wykorzystania eksperymentu behawioralnego – o ile nam wiadomo dotąd nieeksplorowanego – w kontekście teorii umysłu i awersji do nierówności, a mianowicie grę przetargu ultimatywnego (*ultimatum game*, którą będziemy tu nazywać grą ultimatum) z udziałem osób ze spektrum zaburzeń autystycznych. Jak pokażemy, angażuje ona zarówno teorię umysłu, jak i awersję do nierówności, ponieważ gra ultimatum oparta na założeniach teorii gier, gdzie (w odróżnieniu od teorii decyzji) decyzje jednego gracza mają wpływ na wartość decyzji drugiego gracza, a gracze, podejmując decyzje, biorą te zależności pod uwagę, wykorzystując do tego teorię umysłu. Ponadto, specyfika tej gry pozwala na wyraźne zaobserwowanie awersji do nierówności. Przy udziale osób ze spektrum autyzmu gra przetargu ultimatywnego to doskonałe narzędzie, żeby na poziomie behawioralnym zbadać związek teorii umysłu i awersji do nierówności. Eksperyment ten ma być punktem wyjścia do dalszych badań z zakresu neuroobrazowania, z zakresu psychologii ekonomicznej i neuroekonomii, a także zarządzania niepełnosprawnością w środowisku pracy. Zaczniemy od wyjaśnienia, czym jest teoria umysłu i awersja do nierówności.

TEORIA UMYŚLU

Teoria umysłu (Theory of Mind, w skrócie ToM) to zdolność do adekwatnego odczytywania intencji i stanów wewnętrznych (w tym stanów emocjonalnych) swoich i innych osób (inna angielska nazwa teorii umysłu to *mindreading*, tj. czytanie umysłu).

Spectrum Disorder), tj. zgodnie z piątym (i najnowszym) wydaniem podręcznika z klasyfikacją zaburzeń psychicznych (Diagnostic and Statistical Manual of Mental Disorders) wydanej w 2013 r. przez Amerykańskie Towarzystwo Psychiatryczne.

² Ponieważ badań tych są ich dziesiątki, a nadrzędnym celem artykułu nie jest dokonywanie ich przeglądu czy syntezy, wymieniamy tylko te względnie wczesne artykuły, które uznaje się za kluczowe, a w dalszej części artykułu odwołujemy się do wybranych badań istotnych z punktu widzenia stawianej hipotezy.

Najkrócej mówiąc, to zdolność do mentalizowania (Frith, 1989/2016). Przedstawiając to obrazowo (za Dennett, 1983, 1987), teoria umysłu jest aktywna, gdy:

- ktoś (X) myśli (czuje, pragnie etc.), że ktoś inny (Y) myśli (czuje, pragnie etc.), że p (badane przez test fałszywych przekonań Sally, Ann, Baron-Cohen i in., 1985);
- gdy X myśli (czuje, pragnie etc.), że X myśli (czuje, pragnie etc.), że p (badane przez zadanie z opakowaniem cukierków smarties, Perner i in., 1989),

wtedy mówi się o teorii umysłu pierwszego rzędu (first-order Theory of Mind), a także gdy:

- X myśli (czuje, pragnie etc.), że Y myśli (czuje, pragnie etc.), że Z myśli (czuje, pragnie etc.), że p,

wtedy mówi się o teorii umysłu drugiego rzędu (second-order Theory of Mind).

Obecnie obraz, jaki wyłania się z wyników badań, pozwala opisać teorię umysłu, posługując się kategorią modułu *sensu largo* (Leslie, 1991)³. Tak jak istnieje obszar odpowiedzialny za rozpoznawanie twarzy (i jego uszkodzenie prowadzi do prozopagnozji, tj. ślepoty twarzy), tak teoria umysłu jest odpowiedzialna za przetwarzanie informacji na temat stanów umysłowych innych ludzi. U osób z ASD mamy do czynienia z dysfunkcją teorii umysłu, tj. z niezdolnością do mentalizowania, zwaną też ślepotą umysłową (*mindblindness*) (Frith, Frith, 2003).

Teoria umysłu może być opisana jako moduł *sensu largo* po pierwsze dlatego, że przejawia regularność w rozwoju ontogenetycznym (*characteristic ontogeny*): teoria umysłu rozwija się zazwyczaj w wieku około 4 lat i wtedy może być zweryfikowana za pomocą testów werbalnych (choć oczywiście nie pojawia się ona nagle, a badania z wykorzystaniem niewerbalnych zadań pokazują, że już piętnastomiesięczne dzieci wnioskuje na temat przekonań innych osób, Onishi, Baillargeon 2005). Dzieci typowo rozwijające się rozwiązują jeden z podstawowych testów dotyczących teorii umysłu (test fałszywych przekonań) najpóźniej w wieku 5 lat, dzieci z zaburzeniami uczenia się rozwiązują go, gdy osiągną wiek umysłowy pięciu lat, a dzieci z ASD nie przechodzą testu przed osiągnięciem wieku umysłowego 10 lat (i to bez względu na ogólny poziom inteligencji) (Happé, 1995).

Ponadto potrafimy także określić neuronalną lokalizację teorii umysłu (*fixed neuronal architecture*): dzięki technikom obrazowania (głównie tomografii emisyjnej pozytonowej i funkcjonalnemu rezonansowi magnetycznemu) mamy podstawy mówić o osobnym mechanizmie neuronowym, w którym bierze udział skrzyżowanie skroniowo-ciemieniowe (*temporo-parietal junction*, w skrócie TPJ, Saxe, Kanwisher, 2003), a w szczególności prawa bruzda skroniowa górna (*superior temporal sulcus*, w skrócie STS) i płacik ciemieniowy dolny (*inferior parietal lobule*, w skrócie IPL), a także zakręt obręczy (*posterior cingulate*), przedklinek (*medial precuneus*) i środkowa

³ Pojęciem modułu posłużymy się tu instrumentalnie, nawiązując do niektórych kryteriów podanych przez Fodora (1983) i nie wchodząc w spór między lokalizacjonistami a antylokalizacjonistami.

kora przedczołowa (Singer i in., 2006). Badanie fMRI pokazało, że w myślenie o innym (gdy X myśli, że Y myśli, że p, tzw. *perspective taking of other*) jest zaangażowany przedni zakręt obręczy, który jest częścią układu limbicznego (*the anterior cingulate gyrus*), a myślenie o sobie samym (gdy X myśli, że X myśli, że p, tzw. *perspective taking of self*) aktywizuje zakręt obręczy i obszary z prawego TPJ (Keenan, Wheelerb, 2002; Vogeley i in., 2001). Wiadomo również, że lezja prawej półkuli powoduje większą dysfunkcję teorii umysłu niż lezja lewej półkuli (Happé i in., 1999; Stuss i in., 2001; Decety, Sommerville, 2003) oraz że obustronna lezja w obszarze przedniej części lewej półkuli prowadzi do dysfunkcji teorii umysłu, a jednostronna lezja w tym obszarze już nie (Saxe, 2010). Badania wskazują na aktywność prawej półkuli w rozpoznawaniu oszustwa, które to rozpoznawanie umożliwia teoria umysłu (Stuss i in., 2001). Teoria umysłu to moduł w szerokim sensie także dlatego, że wyżej wymienione obszary są wyspecjalizowane dziedzinowo (*domain specific*), tj. są aktywne, gdy myślimy, że ludzie mają stany umysłowe (odczucia, pragnienia, przekonania, myśli, zamiary etc.), ale nie są aktywne wtedy, gdy myślimy o stanach fizycznych, tj. nie są aktywne w teście fałszywych przekonań kontrolnych dotyczących scenariuszy niespołecznych. Została potwierdzona także kolejna charakterystyka teorii umysłu jako modułu, tj. działa ona obligatoryjnie (*obligatory firing*): teoria umysłu ma charakter automatyczny, niezależny od woli i świadomości. Jest tak na przykład w przypadku wykrywania kierunku wzroku i uwspólnionej uwagi (*joint attention* albo *shared attention*), tj. kierowania uwagi na obiekty, na które patrzą inne osoby. Jeśli jakieś obiekty są rozpoznane jako oczy, to przesunięcie uwagi pod ich wpływem nie jest pod kontrolą wolicjonalną i jest wtedy aktywny obszar STS (Ristic, Kingstone, 2005).

AWERSJA DO NIERÓWNOŚCI

Awersję do nierówności zbadano w znanym eksperymencie Sary Brosnan i Fransa de Waala (2003) z udziałem małpek kapucynek (*Cebus apella*). Eksperyment odbywał się w dwóch boksach ze szkła akrylowego tak, że małpki widziały nawzajem siebie i eksperymentatora, a dodatkowo w ścianie między nimi a eksperymentatorem znajdowała się duża liczba otworów. Wpierw nauczono każdą z nich, że jeśli poda eksperymentatorowi jeden z kamyków, których stosik znajdował się w każdym boksie, w zamian otrzymają kawałek ogórka. Gdy kapucynki dobrze się tej zasady nauczyły, wprowadzono modyfikację, która polegała na tym, że kapucynka z pierwszego boksu za to samo zadanie (podać kamyczek eksperymentatorowi) nie dostawała już ogórka, które lubi, ale winogrono, które uwielbia. W tym czasie kapucynka z boksu obok za podanie kamyka wciąż otrzymywała kawałek ogórka. Reakcje kapucynek z drugiego

boksu, które przecież widziały, jakie zadanie wykonała małpka z pierwszego boksu i że dostała za nie atrakcyjniejszy pokarm, były gwałtowne. Nie tylko zaprzestawały wykonywania zadania, ale też nie chciały jeść otrzymanego ogórka – ciskały nim w eksperymentatora i potrząsały ścianką z pleksiglasu, chwytając ją mocno za otwory. Podobne wyniki przyniosły badania z udziałem szympanów (Brosnan, Schiff, de Waal, 2005) i ludzi, zarówno dorosłych, jak i dzieci (Fehr, Bernhard, Rockenbach, 2008).

Wyjaśnienie w kategoriach ultymatywnych, funkcjonalnych awersji do nierówności jest takie, że osobniki, które reagują negatywnie na nierówność, przestają wchodzić w interakcje z osobnikami, które nie zachowują się wobec nich fair, szukają lepszego partnera. Hipotezę tę wspiera to, że kapucynki reagują na nierówność tylko w kontekście zadania. Jeśli eksperymentator daje jednej małpie ogórka, a drugiej winogrono, ale nie jest to warunkowane, druga małpa przyjmie ogórka i nie będzie reagowała negatywnie (Dindo, de Waal, 2006). Podobne reakcje zaobserwowano u szympanów (*Pan troglodytes*) i tamaryn (*Saguinus oedipus*) (Brosnan i in., 2010; Neiworth i in., 2009). Jak można się było spodziewać, prawdopodobieństwo, że osobnik zareaguje negatywnie w sytuacji, gdy inny osobnik dostaje lepszą nagrodę za zadanie, jest większe ze strony osobnika dominującego (Brosnan et al., 2010).

Obszar mózgu, który bierze udział w awersji do nierówności i altruistycznym karaniu, to prawa brzuszno-boczna kora przedczołowa. Zbadano to, stosując powtarzalną przezczaszkową stymulację magnetyczną o niskiej częstotliwości (*low-frequency repetitive transcranial magnetic stimulation*, w skrócie rTMS). Polega ona na tym, że aplikuje się miejscowo powtarzaną serię impulsów magnetycznych, które powodują długotrwałe wzmocnienie lub osłabienie synaptyczne i tym samym dłużej utrzymującą się zmianę w czynności danego obszaru mózgu. Osłabienie aktywności prawej (ale nie lewej) brzuszno-bocznej kory przedczołowej (*right dorsolateral prefrontal cortex*) znacząco redukuje chęć respondentów do odrzucenia oferty egoistycznej (intencjonalnej, tj. niewygenerowanej przez komputer), zwiększa prawdopodobieństwo przyjęcia takiej oferty. Co ciekawe, respondenci dalej oceniają oferty jako niebędące w porządku (*unfair*) (Knoch i in., 2006), co być może wskazuje na zachowanie aspektu kognytywnego, poznawczego, a utratę aspektu odczuciowego i awersyjnego w sytuacji nierówności.

Wyniki te rezonują z podziałem empatii na kognitywną i afektywną. Teoria umysłu odpowiada za aspekt kognitywny empatii, ale już nie za jej aspekt afektywny. Osoby z ASD mają zachowane obwody odpowiedzialne za instynktowne współczucie, na przykład na widok twarzy wyrażających smutek i strach dzieci dotknięte autyzmem reagują podobnie jak dzieci typowo rozwijające się (Blair, 1999). Empatii afektywnej nie przejawiają natomiast dzieci ze skłonnościami psychopatycznymi i dorośli psychopaci. Osoby z ASD mają deficyty w teorii umysłu, która z kolei u psychopatów działa bardzo dobrze, a często lepiej niż u średniej populacji (Blair, 1996).

METODA BADANIA: GRA PRZETARGU ULTYMATYWNEGO

Gra ultimatum należy do teorii gier, która została sformalizowana matematycznie na potrzeby analiz ekonomicznych (von Neuman, Morgenstern, 1944). Jak wiadomo, narzędzie, jakim jest teoria gier, było od tamtej pory wykorzystywane do modelowania biologicznych procesów ewolucyjnych (od ewolucji płci do ewolucji współpracy, na czele z koncepcjami altruizmu odwzajemnionego – *reciprocal altruism*, doboru krewniczego – *kin selection* i dostosowania łącznego – *inclusive fitness*), do opisu konfliktów społecznych i politycznych, do modelowania procesów ekonomicznych etc. Gra ultimatum (pierwszy raz opisana przez Güth i in., 1982) modeluje sytuację, w której wolontariusze przydzieleni są losowo do ról odpowiednio oferenta i respondenta. Respondent i oferent pozostają dla siebie anonimowi zarówno w trakcie gry, jak i po jej zakończeniu. Oferent otrzymuje określoną sumę pieniędzy (np. 100 zł). Następnie oferent zostaje poproszony o podzielenie tej sumy między siebie a respondenta, zostając przy tym poinformowany, że możliwy jest każdy podział. Oferent może zatem złożyć propozycję podziału egoistycznego (chcąc zatrzymać dla siebie większość kwoty wyjściowej, np. 80 zł/20 zł), przez podział oscylujący wokół 50% kwoty wyjściowej, aż do propozycji podziału altruistycznego (proponując respondentowi większość kwoty wyjściowej, np. 10 zł/90 zł). Po tym jak oferent złoży propozycję podziału, respondent może albo zaakceptować jego ofertę i wtedy obydwie strony dostają sumę pieniędzy ustaloną przez oferenta, albo może odrzucić ofertę i wtedy żadna ze stron nie dostanie nic (0 zł/0 zł). Respondent ma zatem w tej grze postąpić zgodnie z zasadą ultimatum „either take it or leave it”, od której gra bierze swoją nazwę. Gra jest jednorazowa, a nie iterowana, tj. po decyzji respondenta nie odbywają się kolejne rundy.

Wzajemna sprawiedliwość (*reciprocal fairness*) wymaga:

- 1) powściągnięcia działania na rzecz własnego interesu (*self-interest*),
- 2) możliwości karania zachowań niesprawiedliwych (*unfair*), co łączy się zarówno z awersją do nierówności, jak i tego, co się nazywa w literaturze przedmiotu karaniem altruistycznym (*altruistic punishment*, Fehr, Gächter, 2002).

Jeśli chodzi o powściągnięcie działania na rzecz własnego interesu, to być może działanie altruistyczne wynika z samej tylko zdolności do empatii, rozumianej tutaj w sensie aspektu odczuciowego, współodczuwania i nie wymaga teorii umysłu? Jeśli tak, to nie można na tej podstawie wnioskować o teorii umysłu u zwierząt. W klasycznym już eksperymencie Massermana i jego współpracowników (1964) najpierw małpy były nauczone, że jeśli pociągną za łańcuch przymocowany do sufitu, a w klatce były dwa takie łańcuchy, to dostaną jedzenie i wodę. Po tym, jak je tego nauczono, wprowadzono modyfikację, która polegała na tym, że pociągnięcie jednego z łańcuchów (ale nie pociągnięcie drugiego) oprócz jedzenia dla ciągnącej go małpy powodowało też rażenie

prądem rezusa z sąsiedniego boksu, którego widać było przez ściankę z pleksiglasu. Po tej modyfikacji większość badanych małp konsekwentnie wybierała bezpieczny łańcuch, niekrzywdzący sąsiada, a niektóre z nich na wiele dni (na 5, a nawet 12) nie ciągnęły za żaden łańcuch, co oznaczało, że przez te dni nie przyjmowały żadnych pokarmów. Autorzy eksperymentu wskazali na czynniki, które mogły sprzyjać takim altruistycznym zachowaniom, jak na przykład to, czy osobniki się wcześniej znały i czy wcześniej w eksperymencie same znalazły się na miejscu osobnika rażonego prądem. Natomiast nie zauważyli zależności między częstością zachowań altruistycznych a wiekiem, płcią, wzrostem czy miejscem w hierarchii społecznej rezusa, który znalazł się za ścianką z pleksiglasu. Niewątpliwie było to działanie altruistyczne, gdzie poniesiony koszt to głodzenie się, a rezusy ponosiły go dla dobra innego rezusa (tym dobrem jest uniknięcie rażenia prądem), ale być może nie wymagało ono teorii umysłu.

Jeśli zaś chodzi o drugi warunek wzajemnej sprawiedliwości (*reciprocal fairness*), jakim jest możliwość karania zachowań niesprawiedliwych, którą to rolę odgrywa awersja do nierówności i karanie altruistyczne (*altruistic punishment*), to potwierdzają go zarówno badania behawioralne, jak i neuroobrazujące. Ciekawe wyniki przyniosły badania behawioralne nad psami (Range et al., 2009). Najpierw uczy się psy, że za podniesienie łapy otrzymują nagrodę w postaci jedzenia. Okazuje się, że odmawiają one współpracy wtedy, gdy pies obok dostaje nagrodę, a one nie dostają żadnej. Nie reagują odmową współpracy w dwóch sytuacjach. Pierwsza to taka, gdzie podczas wykonywania zadania w obecności trenera nie jest obecny drugi pies. Wynika z tego, że psu nie przeszkadza to, że podnosi łapę i nie dostaje nagrody, ale to, że podnosi łapę i nie dostaje nagrody, podczas gdy drugi pies podnosi łapę i nagrodę dostaje (cf. Horowitz, 2012). Druga sytuacja to taka, że (w przeciwieństwie do małp, psy nie reagują odmową współpracy, jeśli dostają inny rodzaj pokarmu bądź inną ilość pokarmu za wykonane zadanie niż pies z pary, dalsze badania pokazały, że mimo braku reakcji behawioralnej w sytuacji eksperymentu – przez działanie takie jak zaprzestanie współpracy), w późniejszym czasie psy potraktowane nierówno nie dzieliły się jedzeniem z psami z pary i unikały go w czasie wolnym, co wskazuje na to, że w czasie eksperymentu nastąpiła jednak reakcja emocjonalna (Brucks i in., 2016)⁴.

Podobnie istnienie mechanizmów karania altruistycznego wspierają badania neuroobrazujące. W jednym z nich (Seymour i in., 2007) uczestnicy wpierw zostali poproszeni o zagranie w grę ekonomiczną (w tym badaniu wybrano dylemat więźnia), a następnie zobrazowano metodą fMRI aktywność ich mózgow w czasie, gdy zadawa-

⁴ Pojawia się teraz pytanie, czy małpy i psy mają poczucie sprawiedliwości (*sense of fairness*), które polega na tym, że chcą równego traktowania nie tylko w stosunku do siebie, ale także w stosunku do innych, czy po prostu nie chcą dostawać mniej niż współbratymcy? Być może nie ze względu na awersję do nierówności psy i małpy odrzucają gorsze nagrody, ale dlatego, że dostępne są lepsze (cf. Wynne 2004).

no ból im i ich współgraczom. W tej części eksperymentu uczestnik podłączony do skanera fMRI widział zarówno swoje ręce, jak i ręce dwóch graczy, z których jeden zagrał fair w dylemacie więźnia, drugi zaś zagrał nie fair. Poprzez elektrody stymulowano ręce wszystkich trzech uczestników, a skanowany uczestnik widział na mierniku dla kogo stymulacja, której ich kolejno poddawano, jest bolesna czy też niebolesna. Niezależnie od płci respondenci wykazali się empatią w stosunku do graczy, którzy byli boleśnie stymulowani, ale zagrali fair (aktywność w wyspie i zakręcie obręczy). Jednak u mężczyzn aktywność tych obszarów była znacznie mniejsza, gdy obserwowali oni ból zadawany graczom, którzy zagrali nie fair, natomiast wzrosła aktywność obszarów nagrody. Wzmacnia to zdaniem autorów dowody na istnienie obwodów biorących udział w karaniu altruistycznym.

PROBLEMATYZACJA

W celu zbadania relacji między awersją do nierówności i teorią umysłu proponujemy przeprowadzić grę ultimatum z udziałem osób z ASD, na próbie celowej, zróżnicowanej względem płci, wieku, wykształcenia, a także doświadczenia zawodowego (tj. osoby, które już pracowały i które jeszcze nie pracowały). Dlaczego chcemy zastosować akurat grę przetargu ultimatywnego? Skłaniają nas do tego wyniki dotychczas przeprowadzonych eksperymentów z wykorzystaniem gry ultimatum. Eksperymenty były przeprowadzone w wielu kulturach, od dużych społeczeństw zindustrializowanych (w krajach Europy Zachodniej, Ameryki Północnej, w Japonii i Izraelu) (Roth i in., 1991), do małych społeczeństw, które nie opierają się na nowoczesnej technologii (Henrich i in., 2004). Na podstawie tych badań zaobserwowano pewne zależności i pewne stałe. Za interesującą zależność można uznać tą, że odsetek altruistycznych propozycji podziału oraz odsetek odrzuceń propozycji podziału altruistycznego są wprost proporcjonalne do stopnia wzajemnej zależności (np. w zdobywaniu pożywienia) i współpracy poza kręgiem rodzinnym (Hauser, 2006). Natomiast ważną stałą dotyczy oferentów i wskazuje, że ich propozycje podziału kwoty wyjściowej rzadko są skrajnie egoistyczne (a zawsze są wyższe niż te, które wynikałyby z równowagi Nasha⁵). Oferenci najczęściej chcą zachować dla siebie więcej kwoty wyjściowej niż respondenci, a drugim najczęstszym sposobem podziału nagrody jest podzielenie się nią równo (Fehr, Schmidt, 1999). Stałą dotyczącą respondentów jest zaś to, że najczęściej odrzucają oni oferty egoistyczne. Jeśli propozycja podziału zostawia 20% i mniej dla respondenta, to jest 50% szans, że zostanie ona przez niego odrzucona.

⁵ Równowaga Nasha to w teorii gier sytuacja, w której żaden z graczy nie zyskałby na zmianie strategii (rezygnując ze strategii kooperacyjnej, w nomenklaturze teorii gier zwanej strategią gołębia, bądź rezygnując ze strategii łamiącej zasady współpracy, zwanej strategią jastrzębia).

Skłania to do przypuszczenia, że gra przetargu ultymatywnego to bardzo dobre narzędzie do behawioralnego badania związku teorii umysłu i awersji do nierówności. Wyniki pokazują, że rzadko propozycje podziału są skrajnie egoistyczne, gdyż najczęściej takie są przez respondentów odrzucane, co można w uzasadniony sposób zinterpretować w kategoriach teorii umysłu i awersji do nierówności, zarówno w przypadku oferenta, jak i respondenta. Oferent nie składa propozycji podziału, co do których spodziewa się, że będą odrzucone przez respondenta – co znaczy, że u oferenta działa teoria umysłu, gdyż oferent myśli, że respondent pomyśli, że p. Działa też awersja do nierówności, ponieważ oferent spodziewa się, że decyzja respondenta będzie powodowana awersją do nierówności. Natomiast respondent, odrzucając egoistyczną propozycję podziału kwoty wyjściowej, znosi stan nierówności między graczami, gdyż obydwie strony mają wtedy zysk o zł – co oznacza, że w uzasadniony sposób można sądzić, że respondent kieruje się awersją do nierówności. Przypisuje też intencje oferentowi.

W naszym badaniu chcemy przeprowadzić gry z osobami z ASD w roli respondentów, raz mówiąc im, że propozycję podziału kwoty wyjściowej przedstawił człowiek, a innym razem mówiąc im, że propozycję wygenerował komputer. Skłaniają nas do tego wyniki badań neuroobrazujących, które pokazały, że w grze ultimatum oferty egoistyczne aktywowały u respondentów lewy i prawy obszar przedni wyspy (*anterior insula*), grzbietowo-boczną korę przedczołową (*dorsolateral prefrontal cortex*) i zakręt obręczy (*cingulate cortex*) (Sanfey i in., 2003). Istotne w naszym kontekście jest jednak to, że aktywność tych obszarów u respondenta była dużo większa, gdy był przekonany, że egoistyczna oferta pochodzi od osoby, niż gdy respondent myślał, że została ona wygenerowana przez komputer (Sanfey i in., 2003). Co by było, gdyby respondentem była osoba z ASD i wiedziała, że oferentem jest komputer? Czy byłaby jakaś różnica? Można przypuszczać, że nie, ponieważ nie mając zdolności do mentalizowania, osoby z ASD powinny zareagować podobnie do osób o typowym rozwoju, tj. nie przypisywać oferentowi intencji, nawet przy skrajnie egoistycznej ofercie. Nie powinny zatem wykazać się awersją do nierówności.

Ostatnim aspektem, który chcemy zbadać w celu analizy związku między awersją do nierówności a teorią umysłu, jest reakcja osób z ASD w sytuacji nierówności. W przeciwieństwie do dzieci ze skłonnościami psychopatycznymi, dzieci z ASD potrafią odróżnić zasady moralne (np. nie wolno krzywdzić innych) od zasad konwencjonalnych (np. nie wolno rozmawiać podczas lekcji). W badaniu, opierając się na grze przetargu ultymatywnego, warto zatem przeanalizować także wybory osób z ASD w roli oferentów i zapytać ich o uzasadnienie ich decyzji. Powstaje pytanie, jak osoby z ASD reagują na sytuację nierówności? Być może jednak reagują awersyjnie, gdyż *de facto* reagują na łamanie ustalonych zasad (np. za wykonanie zadania należy się nagroda).

PERSPEKTYWY DALSZYCH BADAŃ

Perspektywy dalszych badań są szerokie i mają charakter interdyscyplinarny. Jako pierwszą perspektywę wymienimy badania nieobrazujące. Jak już wcześniej wyjaśniono, przyjmujemy, że teoria umysłu odpowiada za aspekt kognitywny empatii, ale nie za jej aspekt afektywny. W dalszym etapie badań warto zatem zbadać u osób z ASD aktywność prawej brzuszo-bocznej kory przedczołowej (*right dorsolateral prefrontal cortex*), tj. obszaru mózgu, który u zdrowych osób bierze udział w awersji do nierówności i w altruistycznym karaniu (odrzuconiu oferty egoistycznej). Zdrowe osoby w roli respondentów przy zaburzonej pracy tego obszaru mózgu wciąż oceniają oferty egoistyczne jako nie w porządku (*unfair*) (Knoch i in., 2006), co wskazywałoby na zachowanie aspektu kognitywnego, poznawczego, a utratę aspektu odczuciowego i awersyjnego w sytuacji nierówności. Badanie tego obszaru u osób z ASD miałoby szczególnie sens w świetle różnic bądź braku różnic w wersji gry ultimatum z ofertą intencjonalną (tj. niewygenerowaną przez komputer) i ofertą wygenerowaną przez komputer.

Kolejną perspektywą, którą chcemy tu nakreślić, jest perspektywa etologiczna. Badania porównawcze nad awersją do nierówności u różnych gatunków zwierząt społecznych pokazały, że wszystkie gatunki, u których zbadano awersję do nierówności, są gatunkami opartymi na współpracy (np. niektóre gatunki z rzędu ssaków naczelnych, takich jak wcześniej wymienione szympansy, kapucynki, tamaryny czy psy). W interpretacji badań nad teorią umysłu u zwierząt innych niż ludzie wciąż wisi widmo behawioryzmu. Kwestia zależności między teorią umysłu a awersją do nierówności, którą tu analizujemy, jest potencjalnie pomocna w badaniach nad teorią umysłu u innych zwierząt.

Awersja do nierówności i teoria umysłu składają się na behavior określany mianem współpracy (*cooperation*). Ich badanie ma sens nie tylko w przypadku szeroko rozumianych zachowań społecznych (Brosnan, Freeman, de Waal, 2006), ale także w kontekście zachowań ekonomicznych. (Fehr, Schmidt, 1999). Według Ernsta Fehra i Klaus Schmidta (1999) jednym z najsilniejszych motywów decydujących o zachowaniach ekonomicznych (tu w kontekście pracy) jest właśnie dążenie do sprawiedliwego podziału (*fairness*). Chodzi tu na przykład o działania związane z zarabianiem pieniędzy⁶, o kwestie współpracy i rywalizacji w środowisku pracy⁷, kwestie podziału wynagrodzenia, mając na uwadze często dokonywane przez pracowników porównywanie

⁶ Można je rozpatrywać jako przedmiot wymiany (zarówno o charakterze ekonomicznym, jak i społecznym) między organizacją a pracownikiem.

⁷ Wyniki badań ukazują, że w działaniach kolektywnych istotne jest m.in. zaufanie, dotychczasowe działania uczestników, przewidywany podział korzyści, komunikowanie się, w tym niewerbalne (Ostrom, 2010)

proporcji osiągniętych przez siebie i przez kolegów wyników finansowych w stosunku do wkładu pracy⁸. Jak to wygląda u osób ze spektrum autyzmu? Wpływ (nie)zdolności do symbolicznego postrzegania pieniądza na kształtowanie relacji interpersonalnych, postawy wobec pieniędzy, wobec podziału wynagrodzeń, motywacji etc. z punktu widzenia osób z ASD, ale także pracodawców zatrudniających osoby z ASD.

Przedsiębiorstwa, które zdecydują się na zatrudnienie pracowników z ASD, powinny wpisać w politykę personalną firmy i jej kulturę organizacyjną tematykę zarządzania niepełnosprawności w miejscu pracy⁹. Znalezienie miejsca pracy dopasowanego do oczekiwań, możliwości i ofert rynku pracy dla osoby z ASD jest zadaniem wielowymiarowym, uwzględniającym nie tylko posiadany przez nią potencjał i zainteresowania, ale także wiele czynników psychologicznych, motywacyjnych oraz ograniczeń wynikających z jej niepełnosprawności, a także specyfiki przedsiębiorstwa. Zarządzanie zasobami ludzkimi nie przynosi jasnych odpowiedzi na pytania stojące przed pracodawcą chcącym zatrudniać osoby z ASD. Jak przeprowadzić proces rekrutacji i selekcji pracowników, nie narażając się na łamanie przepisów o dyskryminacji? Jak konstruować systemy motywacyjne oraz przeprowadzać ocenę pracowniczą? Jak wspierać relacje interpersonalne między pracownikami? Autorki są zdania, że konieczna jest indywidualizacja programów przygotowania do zatrudnienia. Postulowane badania relacji między awersją do nierówności i teorią umysłu z udziałem osób z ASD byłyby pierwszym krokiem przy ich tworzeniu.

LITERATURA

- Baron-Cohen, S., Leslie, A., Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21, 37–46.
- Baron-Cohen, S., Tager-Flusberg, H., Cohen, J.D. (2000). *Understanding Other Minds: Perspectives From Developmental Cognitive Neuroscience*. New York: Oxford University Press.
- Blair, R. (1996). Brief report: Morality in the autistic child. *Journal of Autism and Developmental Disorders*, 26, 571–579.
- Blair, R. (1999). Psychophysiological responsiveness to the distress of others in children with autism. *Personality and Individual Differences*, 26, 477–485.
- Brosnan, S., de Waal, F. (2003). Monkeys reject unequal pay. *Nature*, 425, 297–299.

⁸ Zgodnie z teorią sprawiedliwości Adamsa (Sikorski, 2004, 43), głównym czynnikiem motywacji do pracy jest stopień sprawiedliwości, jaki ludzie przypisują swojej sytuacji pracy. Oceniając ją, pracownicy dokonują porównania wielkość wkładu własnego i innych w stosunku do osiągniętych korzyści (swoich i innych).

⁹ Nomenklaturą „zarządzanie niepełnosprawnością” posługujemy się tu za Kodeksem Postępowania – Zarządzanie Niepełnosprawnością w Miejscu Pracy wydanym w 2003 r. w Polsce przez Ministerstwo Gospodarki, Pracy i Polityki Społecznej (dostępnym na stronie Biura Pełnomocnika Rządu ds. Osób Niepełnosprawnych). Kodeks ten jest dokumentem Międzynarodowej Organizacji Pracy z 2002 r. zatytułowanym *Managing disability in the workplace*.

- Brosnan, S., Freeman, C., de Waal, F. (2006). Partner's behavior, not reward distribution, determines success in an unequal cooperative task in capuchin monkeys. *American Journal of Primatology*, 68, 713–724.
- Brosnan, S., Schiff, H., de Waal, F. (2005). Tolerance for inequity may increase with social closeness in chimpanzees. *Proceedings of the Royal Society B*, 272, 253–258.
- Brosnan, S., Talbot, C., Ahlgren, M., Lambeth, S., Schapiro, S. (2010). Mechanisms underlying responses to inequitable outcomes in chimpanzees, Pan troglodytes. *Animal Behavior*, 79, 1229–1327.
- Brucks, D., Essler, J., Marshall-Pescini, S., Range, F. (2016). Inequity Aversion Negatively Affects Tolerance and Contact-Seeking Behaviours towards Partner and Experimenter. *PLoS ONE*, 11(4), e0153799.
- Decety, J., Sommerville, J.A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends in Cognitive Science*, 7, 527–533.
- Dennett, D. (1971). Intentional systems. *Journal of Philosophy*, 68, 87–106.
- Dennett, D. (1983). Intentional systems in cognitive ethology: the 'Panglossian Paradigm' defended. *Behavioral and Brain Sciences*, 6, 343–390.
- Dennett, D. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.
- Dindo, M., de Waal, F. (2006). Partner effects on food consumption in brown capuchin monkeys. *American Journal of Primatology*, 69(4), 448–456.
- Fehr, E., Bernhard, H., Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, 454, 1079–1084.
- Fehr, E., Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137–140.
- Fehr, E., Schmidt, K.M. (1999). A theory of fairness, competition, and cooperation. *Quarterly journal of Economics*, [brak numeru], 817–868.
- Fodor, J. (1983). *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge: Mass., MIT Press.
- Frith, U. (1989/2016). *Autyzm. Wyjaśnienie tajemnicy*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Frith, U., Frith, C. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 358, 459–473.
- Güth, W., Schmitterberger, R., Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, 3, 367–388.
- Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66, 843–855.
- Hauser, M. (2006). *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*. New York: Ecco/Harper Collins.
- Heinrich, J., Bowles, S., Boyd, R., Camerer, C., Fehr, E., Gintis, H., McElreath, R. (2001). In Search of Homo Economicus: Behavioral Experiments in 15 Small-scale Societies. *American Economic Review*, 91, 73–78.
- Horowitz, A. (2012). Fair is Fine, but More is Better: Limits to Inequity Aversion in the Domestic Dog. *Social Justice Research*, 25(2), 195–212.
- Keenan, J., Wheeler, M. (2002). Elucidation of the brain correlates of cognitive empathy and self-awareness. *Behavioral and brain sciences*, 25, 40–41.
- Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, 3, 314(5800), 829–832.
- Leslie, A. (1991). The theory of mind impairment in autism: Evidence for a modular mechanism of development? In: A. Whiten (ed.), *Natural Theories of Mind*. Oxford: Blackwell, 63–78.

- Masserman, J., Wechkin, S., Terris, W. (1964). Altruistic behavior in Rhesus Monkeys. *The American Journal of Psychiatry*, 121, 584–585.
- Neiworth, J., Johnson, E., Whillock, K., Greenberg, J., Brown, V. (2009). Is a sense of inequity an ancestral primate trait? Testing social inequity in cotton top tamarins (*Saguinus oedipus*). *Journal of Comparative Psychology*, 123(1), 10–17.
- Onishi, K., Baillargeon, R. (2005). Do 15-month old infants understand false beliefs? *Science*, 308, 255–258.
- Ostrom, E. (2010). Analyzing collective action. *Agricultural Economics*, 41, 155–166.
- Perner, J., Frith, U., Leslie, A., Leekam, S. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief and communication. *Child Development*, 60, 688–700.
- Pietrasa, T., Witusik, A., Gałęcki, P. (2010). *Autyzm – epidemiologia, diagnoza i terapia*. Wrocław: Wydawnictwo Continuo.
- Pisula, E. (2010). *Autyzm. Przyczyny, symptomy, terapia*. Gdańsk: Wydawnictwo Harmonia.
- Premack, D., Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Science*, 1, 515–526.
- Range, F., Horn, L., Viranyi, Z., Huber, L. (2009). The absence of reward induces inequity aversion in dogs. *PNAS*, 106(1), 340–345.
- Ristic, J., Kingstone, A. (2005). Taking control of reflexive social attention. *Cognition*, 94(3), B55–65.
- Roth, A., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S. (1991). Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An Experimental Study. *American Economic Review*, 81, 1068–1095.
- Sanfey, A., Rilling, J., Aronson, J., Nystrom, L., Cohen, J. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, 300(5626), 1755–1758.
- Saxe, R. (2010). The right temporo-parietal junction: a specific brain region for thinking about thoughts. W: A. Leslie, T. German (ed.), *Handbook of Theory of Mind*.
- Saxe, R., Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in “theory of mind”. *NeuroImage*, 19(4), 1835–1842.
- Seyfarth, R., Cheney, D. (2002). Dennett's contribution to research on the animal mind. W: A. Brook, D. Ross (eds.), *Daniel Dennett*. Cambridge: Cambridge University Press, 117–139.
- Seymour, B., Singer, T., Dolan, R. (2007). The neurobiology of punishment. *Nature Reviews Neuroscience*, 8, 300–311.
- Sikorski, C. (2004). *Motywacja jako wymiana. Modele relacji między pracownikiem a organizacją*. Warszawa: Wydawnictwo Difin.
- Singer, T., Seymour, B., O'Doherty, J., Stephan, K., Dolan, R., Frith, Ch. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439(26), 466–469.
- Stuss, D., Gallup, G., Alexander, M. (2001). The frontal lobes are necessary for 'theory of mind'. *Brain*, 124(2), 279–286.
- Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happe, F., Falkai, P., Maier, W., Shah, N., Fink, G., Zilles, K. (2001). Mind reading: Neural mechanisms of theory of mind and self-perspective. *Neuro-Image*, 14, 170–181.
- von Neumann, J., Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton: University Press.
- Wimmer, H., Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128.
- Wynne, C. (2004). Animal behaviour: Fair refusal by capuchin monkeys. *Nature*, 428, 140.

TEORIA UMYSŁU A AWERSJA DO NIERÓWNOŚCI – PROJEKT BADANIA Z WYKORZYSTANIEM GRY ULTIMATUM Z UDZIAŁEM OSÓB ZE SPEKTRUM ZABURZEŃ AUTYSTYCZNYCH

STRESZCZENIE: Teoria umysłu to zdolność do adekwatnego odczytywania intencjonalnych stanów wewnętrznych swoich i innych osób, jest podstawą do mentalizowania. Jest ważna np. w terapii niektórych zaburzeń osobowości i kluczowa w rozumieniu spektrum zaburzeń autystycznych (Autism Spectrum Disorder, ASD). Jednak zanim teoria umysłu trafiła do psychologii, jej istnienie było postulowane i badane u prymatów ze względu na to, że są to gatunki społeczne, oparte na współpracy. Gatunki, u których zbadano awersję do nierówności (tj. reakcję na nierówną dystrybucję nagrody za wykonanie tego samego zadania), to także gatunki oparte na współpracy. Mimo to, wciąż nie zostało zbadane, czy i jak obecność awersji do nierówności wiąże się z teorią umysłu. W tym artykule chcemy zwrócić uwagę na możliwy związek właśnie między nimi. Przedstawiamy propozycję wykorzystania eksperymentu behawioralnego – o ile nam wiadomo dotąd nieeksplorowanego – tj. gry ultimatum (*ultimatum game*) z udziałem osób z ASD. Jak pokażemy, angażuje ona zarówno teorię umysłu, jak i awersję do nierówności. Badania neuroobrazujące pokazały, że w grze ultimatum oferty egoistyczne aktywowały u neurotypowych respondentów lewy i prawy obszar przedni wyspy, grzbietowo-boczną korę przedczołową i zakręt obręczy, ale aktywność tych obszarów była dużo słabsza, gdy respondent był przekonany, że egoistyczna oferta pochodzi od komputera, a nie od osoby. Chcemy pokazać, że warto zbadać, czy będzie różnica w tym, jak osoby ze spektrum zaburzeń autystycznych (Autism Spectrum Disorder, ASD) przypisują intencje oferentom/osobom oraz oferentom/komputerom, a także czy przejawiają one awersję do nierówności. Badania takie pozwoliłyby ustalić, czy istnieje zależność teorii umysłu i awersji do nierówności. Metoda behawioralna oparta na wykorzystaniu gry ekonomicznej może być punktem wyjścia do wielu dalszych badań: neuroobrazujących, z zakresu psychologii ekonomicznej i neuroekonomii. Zdaniem autorek ta prosta metoda może być wykorzystana także do kształtowania możliwości pracy z osobami z ASD, sposobów ich zatrudniania czy konstruowania systemów motywowania.

SŁOWA KLUCZOWE: spektrum zaburzeń autystycznych, teoria umysłu, awersja do nierówności, gra ultimatum.

THEORY OF MIND AND INEQUITY AVERSION – A RESEARCH PROJECT USING THE ULTIMATUM GAME WITH PEOPLE WITH AUTISM SPECTRUM DISORDER AS PLAYERS

SUMMARY: The Theory of Mind (ToM) is the cognitive capacity to attribute intentional states to self and others, capacity on which mentalization is based. It plays an important role in psychotherapies (e.g. for personality disorders) and it is crucial to understand the Autism Spectrum Disorder (ASD). However, before the ToM got to psychology, its existence was postulated and examined in primates because of their social and cooperative behavior. Species in which the inequity aversion (i.e. negative reaction in response to an unequal reward distribution) was observed are species which rely on cooperation as well. Despite this fact, the relation between the inequity aversion and the ToM has never been examined. In this paper, we want to argue in favour of the possible relation between these two. We present a design for a behavioural experiment – which, as far as we know, has never been explored – namely, the ultimatum game with people with ASD as players. The game involves both ToM and inequity aversion. Importantly, neuroimaging research has showned that in neurotypical players involved in the ultimatum game selfish offers activated anterior insula, dorsolateral prefrontal cortex and cingulate cortex, but the activity of these areas was much weaker if the responder was

convinced that the offer was made by a computer program, and not by a human. We argue that this is worth to be examined whether people with ASD attribute different intentions to offerers/humans than to offerers/computer programs and whether they display the inequity aversion. This will help to establish whether between the ToM and the inequity aversion there is a relation and of which kind. The behavioral experiment would yield results needed for further research: in neuroimaging, in economic psychology and in neuroeconomics. The authors wish to make this simple method profitable for inclusion in employment of people with ASD and ultimately they wish to create employer's tool kits to hiring, retaining and motivating employees with ASD.

KEYWORDS: autism spectrum disorder, theory of mind, inequity aversion, ultimatum game.