

## **Michał Białek**

Akademia Leona Koźmińskiego  
Jagiellońska 59, Warszawa  
mbialek@kozminski.edu.pl

### **Autoreferat**

#### **1. Imię i Nazwisko.**

Michał Białek

#### **2. Posiadane dyplomy, stopnie naukowe/ artystyczne.**

2013 *Doktor nauk humanistycznych w dziedzinie psychologii*, Instytut Psychologii Polskiej Akademii Nauk, tytuł pracy doktorskiej: „Wiem więc nie myślę. Znaczenie entymematów w wyjaśnianiu tendencyjności rozumowania”, promotor: dr hab. Maciej Haman, prof. UW

2006 *Magister psychologii*, Instytut Psychologii Uniwersytetu Łódzkiego, tytuł pracy magisterskiej: „Wpływ kompetencji językowych na rozumowanie warunkowe w ujęciu teorii modeli umysłowych”, promotor: dr hab. Hanna Bednarek, prof. SWPS

#### **3. Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych/ artystycznych.**

2017 – dziś, Postdoctoral Research Fellow, University of Waterloo, Waterloo, Ontario, Kanada

2013 – dziś, adiunkt, Akademia Leona Koźmińskiego w Warszawie

2009 – 2013, asystent, Akademia Leona Koźmińskiego w Warszawie

2006 – 2010, wykładowca, Szkoła Wyższa Psychologii Społecznej w Warszawie

#### **4. Wskazanie osiągnięcia wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. 2016 r. poz. 882 ze zm. w Dz. U. z 2016 r. poz. 1311.):**

##### **a) tytuł osiągnięcia naukowego/artystycznego,**

„Procesy poznawcze zaangażowane w sądy moralne. Perspektywa dwusystemowa.”

##### **b) Publikacje wchodzące w skład osiągnięcia naukowego:**

Białek, M., Terbeck, S., & Handley, S. (2014). Cognitive psychological support for the ADC model of moral judgment. *AJOB Neuroscience*, 5(4), 21-23. [Recenzowana publikacja w języku innym niż polski, zamieszczona w innym zagranicznym czasopiśmie naukowym niewymienionym w wykazie czasopism naukowych, Liczba punktów wg aktualnego wykazu MNiSW: 5]

Białek, M., (2015) Przegląd badań współczesnej kognitywistyki nad efektem przekonania. *Przegląd Filozoficzny. Nowa seria*, 95, 91-107. [Lista B, Liczba punktów wg aktualnego wykazu MNiSW: 11]



- Bialek, M., & Terbeck, S. (2016). Can cognitive psychological research on reasoning enhance the discussion around moral judgments? *Cognitive Processing*, 17(3), 329-335. [IF(2016) = 0.974, pięcioletni IF = 1.593, Liczba punktów wg aktualnego wykazu MNiSW: 20]
- Bialek, M., & De Neys, W. (2016). Conflict detection during moral decision-making: Evidence for deontic reasoners' utilitarian sensitivity. *Journal of Cognitive Psychology*, 28(5), 631-639. [IF(2016) = 1.503, pięcioletni IF = 1.766, Liczba punktów wg aktualnego wykazu MNiSW: 20]
- Bialek, M., & De Neys, W. (2017). Dual processes and moral conflict: Evidence for deontological reasoners' intuitive utilitarian sensitivity. *Judgment and Decision Making*, 12(2), 148-167. [IF(2017) = 2.525, pięcioletni IF = 2.868, Liczba punktów wg aktualnego wykazu MNiSW: 35]
- Bialek, M. (2017). Mechanika moralności. dylematy moralne i intuicyjne rozumienie dobra i zła. *Czasopismo Psychologiczne*, 23(1), 9-19. [Lista B, Liczba punktów wg aktualnego wykazu MNiSW: 11]
- De Neys, W., & Bialek, M. (2017). Dual processes and conflict during moral and logical reasoning: A case for utilitarian intuitions? W: B. Trémolière & J. F. Bonnefon (Red.), *Moral Inferences* (pp. 123–136). Hove, UK: Psychology Press [Rozdział w monografii naukowej wieloautorskiej]
- Muda, R., Niszczota, P., Bialek, M., & Conway, P. (2018). Reading dilemmas in a foreign language reduces both deontological and utilitarian response tendencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(2), 321-326. [IF(2017) = 2.319, pięcioletni IF = 3.019, Liczba punktów wg aktualnego wykazu MNiSW: 35]
- Bialek, M., Fugelsang, J., & Friedman, O. (2018). Choosing victims: Human fungibility in moral decision-making. *Judgment and Decision Making*, 13(5), 451-458. [IF(2017) = 2.525, pięcioletni IF = 2.868, Liczba punktów wg aktualnego wykazu MNiSW: 35]
- Bialek, M., & Fugelsang, J. (2019) No evidence for decreased foreign language effect in highly proficient and acculturated bilinguals. A commentary on Čavar and Tytus (2018). *Journal of Multilingual and Multicultural Development*, advance online publication [IF(2017) = 0.942, pięcioletni IF = 1.273, Liczba punktów wg aktualnego wykazu MNiSW: 25]
- Bialek, M., Turpin, M., & Fugelsang, J. (w druku). What is the right question for moral psychology to answer? *Psychological Science*. [IF(2017) = 6.128, pięcioletni IF = 7.370, Liczba punktów wg aktualnego wykazu MNiSW: 45]
- SUMA IF = 16,916, punkty MNiSW = 246.**

**c) omówienie celu naukowego/artystycznego ww. pracy/prac i osiągniętych wyników wraz z omówieniem ich ewentualnego wykorzystania.**

**4.1. Wprowadzenie**

Niniejszy autoreferat przedstawia wyniki moich badań nad procesami poznawczymi w sądach moralnych. W skład prac stanowiących przedstawiane tu osiągnięcie naukowe wchodzi 11 tekstów naukowych, w tym siedem artykułów z listy JCR, dwie publikacje krajowe oraz po jednym rozdziale w anglojęzycznej monografii wieloautorskiej wydanej przez Psychology Press oraz jeden anglojęzyczny artykuł w czasopiśmie nieuwzględnionym przez MNiSW w katalogu czasopism punktowanych. W dalszej części autoreferatu omawiam szczegółowo wspomniane prace, wskazując na obszary prowadzonych przeze mnie badań oraz jak one się łączą w spójny dorobek.

Jednym z głównych efektów prowadzonych przeze mnie badań podoktorskich (po 2013r.) jest krytyka założeń teorii dwusystemowej sądów moralnych (Bialek, Handley i Terbeck, 2014, Bialek, 2015, 2017),

 2

która doprowadziła do stworzenia i walidacji autorskiego, hybrydowego modelu sądów moralnych (Białek i De Neys, 2016; Białek i De Neys, 2017; De Neys i Białek, 2017). Ponadto, w swoich badaniach nad procesami poznawczymi w sądach moralnych pokazałem jak zmieniają się one podczas podejmowania decyzji w języku obcym (Muda, Niszczoła, Białek, i Conway, 2017, Białek i Fugelsang, 2019), oraz jako jeden z pierwszych na świecie badałem jak ludzie dokonują wyboru kogo poświęcić dla większego dobra (Białek, Fugelsang i Friedman, 2018). Ostatni z elementów składających się na osiągnięcia naukowe to analiza problemów metodologicznych tych badań, zarówno w zakresie używanych materiałów (Białek, Turpin i Fugelsang, 2019) jak i metody analizy zebranych danych (Białek i Terbeck, 2016).

Model hybrydowy sądów moralnych osadzony jest w nurcie teorii dwusystemowej i dokonuje w niej znaczących rewizji. W tym modelu sądy i decyzje podmiotu wynikają z dwóch niezależnych od siebie intuicji: chęci postępowania zgodnie z abstrakcyjnymi regułami moralnymi (deontologia) i analizy konsekwencji (użyteczność). Owe intuicje wchodzi w interakcje: współpracując lub będąc ze sobą w konflikcie. W przypadku, gdy obie intuicje są zgodne ze sobą, lub gdy konflikt między nimi jest subiektywnie oceniany jako niewielki, osoba podejmuje szybką decyzję moralną ze stosunkowo dużą pewnością. W przypadku, gdy konflikt jest subiektywnie oceniany jako silny, dochodzi do uruchomienia procesów poznawczych wyższego rzędu, mających na celu albo racjonalizację najsilniejszej intuicji, albo refleksyjny wybór subiektywnie najlepszej opcji. W zależności od kontekstu i cech problemu, poszczególne intuicje są bardziej lub mniej wyraziste i w efekcie być częściej wcielane w życie. Ponieważ każda z intuicji jest niezależna, to poszczególne czynniki mogą wpływać na: (a) zaledwie jedną z nich, (b) na obie jednakowo lub (c) na obie odmiennie.

Dzięki moim rozległym kontaktom międzynarodowym oraz umiejętności pracy w zespole, znaczną część ze wspomnianych publikacji powstała we współpracy z naukowcami z zagranicznych instytucji naukowych (USA, Kanada, Francja, Anglia) oraz z naukowcami z innych krajowych instytucji (UMCS, UEP). Należy zauważyć, że jestem praktycznie zawsze ich pomysłodawcą i pierwszym autorem. Stanowi to niepodważalny dowód na moją samodzielność naukową.

Mimo że omawiane tu prace są w większości świeżo opublikowane, to zyskały już znaczący oddźwięk w pracach innych, z ponad 100 cytowaniami w okresie 2014-2018 wg. Google Scholar. Średnia liczba cytowań znacząco przekracza trzy cytowania rocznie na tekst, co odpowiadałoby wysokiemu jak na czasopiśmie psychologiczne Impact Factorowi równemu 6. Liczba cytowań moich prac w Journal Citation Report jest oczywiście odpowiednio niższa, gdyż część zarówno moich jak i cytujących moje prace tekstów jest klasyfikowana jako "online first" lub "advance online publication", a zatem nieindeksowana przez wspomnianą bazę. Do bazy JCR trafią dopiero w momencie ukazania się w formie drukowanej.

Innym dowodem uznania dla opublikowanych prac jest też fakt, że jestem regularnym recenzentem artykułów na temat sądów moralnych dla czasopism z listy JCR (prawie 20 recenzji w latach 2017-2018, w tym na przykład trzykrotnie dla Cognition), oraz dla organizacji propagujących Otwartą Naukę takich jak Psychological Science Accelerator czy Pipeline Project.

Duża część poniższego autoreferatu powstała w oparciu o artykuł przeglądowy opublikowany w Czasopiśmie Psychologicznym (Białek, 2017). Podobnie jak w tym autoreferacie, omawiam w nim znaczną część swoich publikacji dotyczących sądów moralnych. Zasadniczą różnicą między autoreferatem a wspomnianą publikacją jest to, że artykuł zawiera szersze odniesienie do prac innych badaczy oraz nie omawia prac opublikowanych po 2017 roku.

Niniejszy autoreferat podzieliłem na trzy części. Pierwsza dotyczy metodologii badań nad moralnością, podkreślając użyteczność hipotetycznych dylematów, ale w jasno określonych celach jakimi jest rozumienie intuicji moralnych, a nie przewidywanie realnych wyborów moralnych. Część pierwsza przedstawia ponadto argumenty za rozszerzeniem obecnie stosowanej metodologii o dylematy niekonfliktowe, czyli takie, w których złamanie zasad moralnych nie prowadzi do pozytywnych korzyści oraz przedstawia badanie z jej użyciem na temat moralnego efektu języka obcego.

 3

Część druga autoreferatu przedstawia nurt teoretyczny prowadzonych badań jakim jest teoria dwusystemowa sądów moralnych. W tej części przedstawiam jej założenia, przeprowadzam jej krytykę, przedstawiam argumenty empiryczne podważające założenia teorii dwusystemowej. Ostatecznie przedstawiam własny model hybrydowy sądów moralnych oraz dowody na jego zasadność. Trzecia część autoreferatu opisuje moje pozostałe publikacje związane z sądami moralnymi.

## **4.2. Metodologia badań nad psychologią moralności**

### **4.2.1. Podstawowe pojęcia i definicje**

Podczas II Wojny Światowej trwał pontyfikat Piusa XII. Związana z nim jest poważna kontrowersja: papieżowi zarzucano, że nie potępił Hitlera przez cały okres jego władzy. obrońcy papieża twierdzili, że sam akt potępienia nie dałby nic, może poza silniejszym prześladowaniem Kościoła w Rzeszy. W zamian papież nakazał nuncjaturom by starały się chronić żydów przed hitlerowcami, dzięki czemu ocalało wiele ludzkich istnień. Ocena moralna opisanego postępowania ma dwa wymiary: (a) jego zgodność z pewnymi uniwersalnymi i abstrakcyjnymi zasadami oraz (b) jego konsekwencje praktyczne. Z jednej strony przykazanie „nie zabijaj” jasno sugeruje, należy przeciwstawiać się zabijaniu niewinnych, a jak pisała Zofia Kossak-Szczucka: „Kto milczy w obliczu mordu – staje się współnikiem mordercy. Kto nie potępia – ten przyzwala” (Żukowski, 2015). Z drugiej strony, liczba żyć ludzkich, które ocalono poprzez zachowanie neutralnej postawy Kościoła jest poważnym argumentem za linią postępowania Piusa XII. Osoba, która dokonując oceny moralnej kieruje się kategorycznymi zasadami takimi jak „nie zabijaj”, „nie krzywdź innych” nazywana jest deontologiem; osoba taka uzna postępowanie Piusa za naganne. Inne osoby – tak zwani utylitaryści - uznają za etyczne takie działania, które przynoszą jak najwięcej korzyści jak największej liczbie osób (nawet kosztem naruszenia pewnych innych zasad). Utylitarysta doceni korzyść, jaką przyniosły działania papieża i uzna jego postępowanie za jak najbardziej moralne.

Badania pokazują, że wybory większości ludzi raz zgodne są z deontologią, a raz z utylitaryzmem. Niespójność ludzi w dokonywanych wyborach moralnych świetnie obrazuje problem między sytuacją a jej interpretacją (Tomaszewski, 1975). Badani, w takich samych obiektywnie warunkach (na przykład ocena dopuszczalności poświęcenia jednej osoby dla ocalenia piętki innych osób) zdają się dokonywać innych interpretacji sytuacji w zależności od jej drugorzędnych elementów, takich jak rodzaj czynności którą należy wykonać (pośrednio czy bezpośrednio, Royzman i Baron, 2002), albo tego, czy osoba jest celowo poświęcana, czy też ginie „przy okazji”, jako skutek uboczny próby ratowania pięciu osób (Waldmann i Dieterich, 2007). Wpływ na oceny moralne ma na przykład także fakt czy decydent jest u władzy, czy nie (Lammers, Galinsky, Dubois, i Rucker, 2015; Wojciszke, 2011) oraz język, w jakim przedstawiono problem (ojczysty vs. obcy; (Geipel, Hadjichristidis, i Surian, 2015). Wszystkie te elementy nie stanowią osi problemu moralnego, ale zdają się mieć istotne znaczenie dla ostatecznych decyzji. Sami decydenci rzadko zdają sobie sprawę z zewnętrznych czynników wpływających na ich wybory, mając przekonanie o własnej racjonalności (Farsides, Sparks, i Jessop, 2018). Kluczowe dla zrozumienia moralności wydają się zatem cechy problemu oraz to, w jaki sposób są one przetwarzane przez decydentów.

Badania sądów moralnych prowadzą między innymi filozofowie eksperymentalni, neuropsychologowie, psychologowie rozwojowi, społeczni, osobowości, kulturowi oraz poznawczy. Poza tym, że każda z tych grup stosuje nieco różną metodologię, to dodatkowo także każda z nich stara się odpowiedzieć na nieco inne pytania. Pomimo różnic, elementem spajającym wiele z tych badań jest osadzenie ich w nurcie teorii dwusystemowej. Przyjmuje się w niej, że decyzje, nie tylko moralne, składają się z dwóch komponentów: (a) automatycznego, szybkiego i intuicyjnego, oraz (b) refleksyjnego, powolnego i wymagającego zasobów poznawczych (np. zasobów pamięci roboczej). Prowadzone badania starają się wyjaśnić jakie wybory moralne wywoływane są przez procesy automatyczne, jakie przez procesy refleksyjne, oraz jak decydenci rozstrzygają ewentualny konflikt między tymi procesami. Większość z moich prac stanowiących omawiane

tu osiągnięcie naukowe rozwija i modyfikuje tą teorię by doprowadzić ostatecznie do stworzenia własnego hybrydowego modelu sądów moralnych.

#### 4.2.2. Hipotetyczne dylematy moralne

Poza szerokim kontekstem teoretycznym jakim jest teoria dwusystemowa, inną cechą wspólną różnych podejść do badań nad moralnością jest stosowanie hipotetycznych scenariuszy, w których zadaniem badanego jest określenie dopuszczalności przekroczenia norm moralnych, by w efekcie doprowadzić do „większego dobra”. Klasycznym przykładem jest tu dylemat wagonika. Należy w nim podjąć decyzję czy przełączyć zwrotnicę i przekierować pędzący wagonik tak, by zamiast pięciu osób zabił jedynie jedną osobę. Innymi słowy, ocenie podlega dopuszczalność złamania jednej normy moralnej (zabicie człowieka), aby w ten sposób zaspokoić inną zasadę moralną (ocalenie życia innym ludziom).

Metoda hipotetycznych dylematów moralnych spotkała się ostatnio z krytyką, gdyż okazuje się, że hipotetyczne wybory różnią się od tych dokonywanych w wirtualnej rzeczywistości (Patil i inni, 2014), oraz nie są dobrym predyktorem późniejszych rzeczywistych zachowań (Bostyn, Sevenhant, i Roets, 2018). Rozwiązując hipotetyczne dylematy badani zdają się przeceniać swoje inklinacje deontologiczne, podczas gdy w rzeczywistości ich wybory są przeważająco utylitarystyczne. W związku z tym niektórzy badacze postulują odejście od stosowania hipotetycznych dylematów moralnych (tzw. „wagonikologii” czy trolleyology).

Podjąłem dyskusję na temat tego zagadnienia w swojej publikacji w *Psychological Science* (Białek, Turpin, i Fugelsang, 2019). Tekst ten ma on charakter teoretyczno-polemiczny w stosunku do wspomnianych doniesień. Argumentuję, że nietrafność prognostyczna wyborów hipotetycznych tylko pozornie zaskakuje, a hipotetyczne wybory w dylematach moralnych mają nadal swoje miejsce w przyszłych badaniach. Po pierwsze, większość badań nad deklarowanymi preferencjami jest słabym predyktorem rzeczywistych wyborów, zatem problem niskiej predyktywności nie jest szczególną przypadłością sądów moralnych. Jest to wielka bolączka praktyków, choćby w marketingu (Kuhberger, 2002). Po drugie, ponieważ wybory takie jak te w dylematach moralnych są niezwykle rzadkie, to praktycznie nigdy nie odwołują się do realnego doświadczenia badanego. W związku z brakiem doświadczenia, jedyny sposób, by podjąć hipotetyczną decyzję w dylemacie moralnym, to odwołać się do swoich intuicji na temat tego co wydaje się właściwe. Intuicje te, z kolei, są abstrakcyjne i nie uwzględniają efektów kontekstualnych decyzji oraz ignorują okoliczności, które nie są kluczowe dla problemu, ale mają realny wpływ na to jakich wyborów decydent dokona. Na przykład intuicje na temat tego, co jest moralnie akceptowalne nie uwzględniają dostatecznie incydentalnego wpływu nastroju, emocji czy dostępności zasobów poznawczych.

Moim zdaniem taki stan rzeczy pozwala przededefiniować cele badania nad moralnością – odpowiedzi udzielane w dylematach nie powinny służyć przewidywaniu zachowania. Nawet jeśli ludzie zachowują się inaczej niż deklarują, nie ma to wielkiego znaczenia praktycznego, bowiem większość z nas nigdy nie znajdzie się w podobnej sytuacji. W zamian, hipotetyczne dylematy moralne pozwalają zrozumieć intuicje jakimi kierują się ludzie oceniając innych. Jest to niezwykle ważne, gdyż każdy z nas praktycznie codziennie ocenia moralność innych i to właśnie te sądy przekładają się na kluczowe działania takie jak wybory znajomych, partnerów społecznych, głosy w wyborach, wyroki sądowe, czy nawet inwestycje. Zatem to przekonania i intuicje moralne a nie realne zachowania stanowią najbardziej praktyczną informację na temat ludzkiej moralności. Podsumowując, w omawianym tu tekście argumentuję za dalszym używaniem hipotetycznych dylematów moralnych, podkreślając ich wyjątkową użyteczność w testowaniu intuicji moralnych i kryteriów, którymi ludzie posługują się oceniając innych.

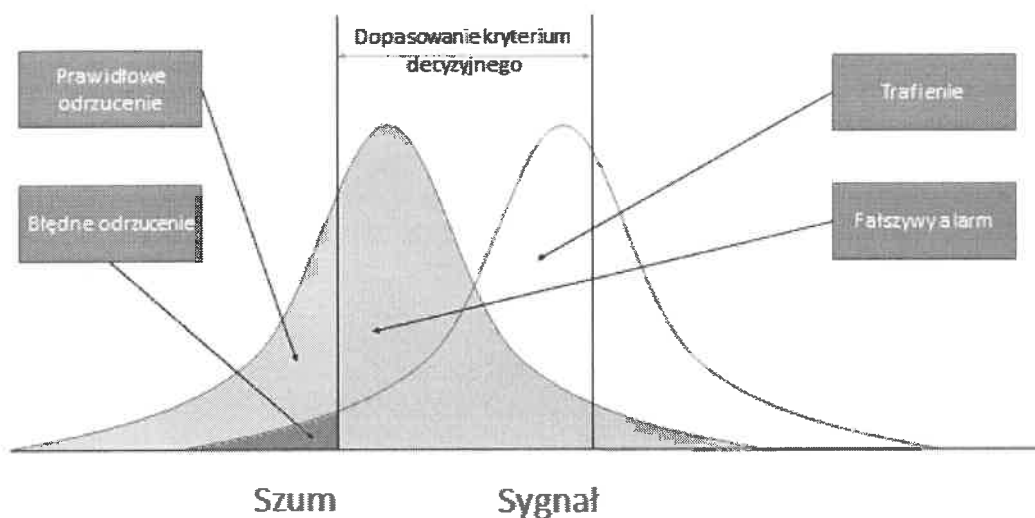
Uwzględniając powyższe podobieństwa i różnice w badaniu decyzji moralnych przez osoby wywodzące się z różnych subdyscyplin psychologii i filozofii, można uznać, że badania nad teorią dwusystemową stanowią pewien rodzaj badań podstawowych dla tego nurtu, a stosowana metodologia hipotetycznych dylematów

*Patil* 5

moralnych stanowi dobre narzędzie do testowania przypuszczeń tej teorii. O ile sama idea stosowania hipotetycznych scenariuszy ma racje bytu, to jest z nią jednak pewien problem. Popularnie stosowane dylematy testują jedynie pewien specyficzny warunek, w którym wybór działania zawsze prowadzi do pozytywnych skutków, a niepodjęcie działania do negatywnych skutków. Zatem działanie i pozytywne skutki są w pewien sposób wymieszane. Wracam do tego problemu w kolejnych omawianych tu pracach.

#### 4.2.3. Propozycja udoskonalenia stosowanej metodologii

W swoim artykule w Cognitive Processing (Białek i Terbeck, 2016) zaproponowałem użycie teorii detekcji sygnału, by uzyskać większy wgląd w procesy poznawcze zaangażowane w sądy moralne. Podstawy teorii detekcji sygnałów w kontekście wydawania sądów opisywałem w Przeglądzie Filozoficznym (Białek, 2014). Teoria detekcji sygnału to model opisujący jak decydent dokonuje wyboru między dwoma bodźcami: sygnałem i szumem. Istnieją dwie poprawne decyzje klasyfikacyjne. Zaklasyfikowanie sygnału jako sygnał nazywamy trafieniem, a szumu jako szum prawidłowym odrzuceniem. Istnieją także dwa rodzaje błędów klasyfikacyjnych – uznanie szumu za sygnał, nazywane fałszywym alarmem oraz zaklasyfikowanie sygnału jako szum, nazywane ominięciem (Rys. 1). Na fałszywy alarm przyjęło się mówić także Błąd I Rodzaju, a na ominięcie Błąd II Rodzaju.



**Rys 1.** Adaptacja teorii detekcji sygnału. Pionowe linie przedstawiają liberalne (lewo) i konserwatywne (pravo) kryterium decyzyjne. Przykładowe rodzaje błędu, oznaczone na kolorowo, odnoszą się do liberalnego kryterium decyzyjnego.

Najłatwiej zilustrować teorię detekcji sygnału przy pomocy zadania percepcyjnego. Przyjmijmy, że sygnał i szum to bodźce świetlne o różnej jasności, sygnał jest jaśniejszy, podczas gdy szum jest mniej jasny. W związku z niedoskonałością systemu percepcyjnego, postrzegany bodziec nie jest idealnie odwzorowywany: raz wydaje się nieco jaśniejszy, a raz nieco ciemniejszy niż jest w rzeczywistości. Niestety, w związku z tą niedoskonałością percepcji sygnał i szum mogą zostać ze sobą pomyłone. Im precyzyjniejsze odwzorowanie siły bodźca przez system percepcyjny (większa czułość systemu percepcyjnego), tym wahania w spostrzeganej jasności bodźca są mniejsze, a liczba pomyłek mniejsza. Zapisując to matematycznie, przy większej czułości rozkład rejestrowanej subiektywnie siły sygnału i szumu nachodzą

*Redy 6*

na siebie w mniejszym stopniu i tym mniejsza szansa na to, że zostaną ze sobą pomyłone. Zatem czułość zmniejsza liczbę błędów klasyfikacyjnych, ale oczywiście nie eliminuje ich zupełnie. Podczas każdej decyzji nadal istnieje ryzyko popełnienia błędu, a nie zawsze każdy z tych błędów ma jednakową wagę. Na przykład w prawie preferuje się błąd II rodzaju nad błąd I rodzaju, gdyż skazanie niewinnej osoby podważa zaufanie do całego systemu prawnego bardziej, niż nieskazanie winnego. Przeciwnie do prawa, w medycynie często preferuje się błąd I rodzaju, gdyż lepiej leczyć nieistniejącą niż przeoczyć realną chorobę. W zależności od problemu, różne błędy mają zatem różne znaczenie. Za to, który rodzaj błędu jest preferowany przez decydenta odpowiada jego kryterium decyzyjne. Prawo przyjmuje konserwatywne, a medycyna liberalne kryterium decyzyjne.

Rysunek 1 przedstawia liberalne i konserwatywne kryterium decyzyjne. Mówi ono o progu, po przekroczeniu którego decydent będzie uznawał bodziec za sygnał. Wracając do przykładu jaśniejszych i ciemniejszych świateł, decydent może podwyższyć swoje kryterium decyzyjne tak, by tylko naprawdę jasne bodźce były klasyfikowane jako sygnał. Doprowadzi to do spadku liczby fałszywych alarmów, ale kosztem wzrostu liczby ominięć. Mówimy w takim przypadku o konserwatywnym kryterium decyzyjnym. Decydent może też podjąć decyzję by obniżyć kryterium decyzyjne i w efekcie klasyfikować nieznacznie jaśniejsze obiekty jako sygnał. Doprowadzi to do spadku ominięć, ale kosztem wzrostu fałszywych alarmów. Mówimy w takim przypadku o liberalnym kryterium decyzyjnym.

Prawdopodobieństwo poprawnego zaklasyfikowania sygnału jako sygnał określa proporcja powierzchni jego rozkładu, która przekracza kryterium decyzyjne do powierzchni która go nie przekracza. Analogicznie, prawdopodobieństwa błędnego uznania szumu za sygnał określa proporcja jego rozkładu, która przekracza kryterium decyzyjne. Jak widać z Rysunku 1, zmieniając kryterium decyzyjne wywołujemy nieliniowo zmiany między ominięciami a fałszywymi alarmami. Na przykład niewielki wzrost prawdopodobieństwa trafienia może współtowarzyszyć lawinowemu wzrostowi prawdopodobieństwa fałszywego alarmu.

Teoria detekcji sygnału stanowi dobrą podstawę teoretyczną do analizowania sądów moralnych. Jeśli przyjmiemy użyteczność za sygnał, to wybór działania, które maksymalizuje użyteczność będzie trafieniem, a odmowa dokonania działania, które obniża użyteczność będzie prawidłowym odrzuceniem<sup>1</sup>. Jednocześnie działanie, które ignoruje możliwość maksymalizacji użyteczności należy uznać za ominięcie, a podjęcie działania, które zmniejsza ogólną użyteczność, za fałszywy alarm. Przekładając tę terminologię na problemy moralne, działanie w dylemacie wagonika należy uznać za trafienie, a odmowę tego działania za ominięcie. Decyzje podejmowane przez badanych w eksperymentach psychologicznych dają nam zatem informację o proporcji trafień do ominięć. Jest to jednak informacja niekompletna, gdyż ignoruje informację na temat prawidłowych odrzuceń i fałszywych alarmów.

Wróćmy do przykładu świateł jasnych i mniej jasnych jako sygnału i szumu. Wyobraźmy sobie detektor świetlny, który reaguje na każdy bodziec uznając go za sygnał. Z punktu widzenia trafień byłby to idealny detektor, gdyż reagowałby zawsze, gdy mamy do czynienia z sygnałem. Niestety, detektor ten uznawałby za sygnał każdy bodziec, także szum. Detektor ten generowałby zatem także maksymalną liczbę fałszywych alarmów. A skoro dowolny bodziec klasyfikowany jest jako sygnał, to dzięki użyciu takiego detektora nie wiemy ani odrobinę więcej, niż wiedzielibyśmy po prostu zgadując. Pomimo idealnego wyniku w zakresie trafień taki detektor jest bezużyteczny, to i w sądach moralnych analiza, która patrzy jedynie na trafienia jest suboptymalna. Także w tym przypadku warto przyjrzeć się liczbie fałszywych alarmów.

Funkcję „szumu” dla sądów moralnych stanowią dylematy, w których działanie obniża ogólną użyteczność. Ponieważ zarówno deontologia jak i utilitaryzm sugerują w takim przypadku brak działania, nazywam

<sup>1</sup> Nie należy utożsamiać jednak utilitaryzmu z czymś dobrym, a deontologii z czymś negatywnym. Jest to jedynie konwencja nazewnicza.

*Pał 7*

takie dylematy niekonfliktowymi, i odróżniam je od klasycznych dylematów, które nazywam konfliktowymi. Jeśli decydent decyduje się działać w dylematach konfliktowych i niekonfliktowych to w takim przypadku nie mówimy o utylitarystyce, ale raczej o psychopacie, który niezależnie od tego czy skutki tego są pozytywne czy negatywne dla innych osób dąży do wyrządzenia krzywdy innym. Utylitarysta to zatem osoba, która jest skłonna poświęcić jedną osobę, ale tylko wtedy, gdy przyczyni się to do większych pozytywnych skutków.

Użycie zwykłego zestawu konfliktowych dylematów moralnych nie pozwala na odróżnienie psychopaty od utylitarysty. W konsekwencji, istnieje kilka badań empirycznych, które faktycznie wskazują, że utylitarystyka koreluje dodatnio z psychopatią (Koenigs, Kruepke, Zeier i Newman, 2011) i ujemnie z empatią (Patil i Silani, 2014). Inną konsekwencją używania jedynie dylematów konfliktowych jest problem z interpretacją czynników, które eksperymentalnie zwiększają lub obniżają chęć wyborów utylitarystycznych. Na przykład skłonność do refleksji sprzyja częstszym wyborom utylitarystycznym (Patil i Tremoliere, 2018). Nie jest jednak jasne czy ten wzrost trafień nie jest jednocześnie powiązany ze wzrostem fałszywych alarmów.

Jak wspominałem, teoria detekcji sygnału do opisu procesu decyzyjnego posługuje się dwoma parametrami: czułością oraz kryterium decyzyjnym. Terminom tym nadałem znaczenie psychologiczne w kontekście sądów moralnych (Białek i Terbeck, 2016). Zaproponowałem, aby inklinację utylitarystyczną utożsamić z czułością decydenta, a inklinację deontologiczną utożsamić z kryterium decyzyjnym. W takim przypadku, osoba o silnej inklinacji utylitarystycznej (czyli wysokiej czułości) zwraca szczególną uwagę na konsekwencje i doskonale różnicuje problemy moralne pod względem ich konsekwencji. Deontolog nie musi analizować konsekwencji, gdyż ocenia on jedynie samo działanie i odrzuca je, jeśli uzna je za łamiące pewne zasady moralne. Zatem osoba o silnej inklinacji deontologicznej stosuje konserwatywne kryterium decyzyjne i niezależnie od swojej czułości uznaje większość działań (lub nawet wszystkie) w dylematach za niemoralne. Zgodnie z teorią detekcji sygnału, kryterium decyzyjne jest niezależne od analizy konsekwencji. Zatem deontologia i utylitarystyka stanowią niezależne od siebie komponenty sądów moralnych.

O ile czułość wydaje się być stosunkowo stałą cechą decydenta, to w zaproponowanym przeze mnie modelu, kryterium decyzyjne byłoby wrażliwe na rodzaj działania, które należy podjąć, by ocalić większą liczbę osób. Bardziej odpychające działanie (np. konieczność popchnięcia osoby na tory, by zatrzymać pędzący wagonik vs. użycie w tym celu zwrotnicy) powoduje przyjęcie bardziej konserwatywnego kryterium decyzyjnego. W rezultacie czego, nawet osoba o dość silnej inklinacji utylitarystycznej będzie wrażliwa na sam rodzaj wymaganego działania i mniej skłonna będzie działać w dylemacie który wymaga bezpośredniego wyrządzenia krzywdy innej osobie, niż w dylemacie, który wymaga jedynie działania pośredniego.

Zaproponowany przeze mnie model oparty o teorii detekcji sygnałów dopuszcza deliberatywną regulację kryterium decyzyjnego. Osoba, która ma przed sobą problem moralny może dowolnie regulować przyjęte kryterium decyzyjne: podwyższyć lub obniżyć je. To banalne założenie ma daleko idące implikacje, bo skoro decydent może podwyższyć kryterium decyzyjne, to model ten przewiduje refleksyjną deontologię. Zjawisko to nieprzewidywane było dotychczas przez żaden inny model teorii dwusystemowej sądów moralnych, a doczekało się późniejszych dowodów dostarczonych przez niezależne zespoły badawcze. Tekst opublikowany w *Journal of Experimental Psychology: General* wykazał, że teoria detekcji sygnału faktycznie lepiej niż dotychczasowe modele przewiduje wybory moralne (Cohen i Ahn, 2016). Inny tekst, opublikowany w *Journal of Experimental Social Psychology*, wykazuje, że osoby religijne podejmują wybory deontologiczne w wyniku refleksji, a nie impulsu (McPhetres, Conway, Hughes, i Zuckerman, 2018), dowodząc w ten sposób, że refleksja może także skłaniać osoby do podejmowania decyzji przypisywanych dotychczas procesom Typu 1.

P. A. H. 8



### 4.3. Efekt języka obcego

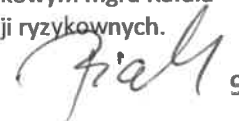
Opisanej w poprzednim podrozdziale metody hipotetycznych dylematów moralnych, rozszerzając klasycznie stosowane dylematy konfliktowe o dylematy niekonfliktowe, użyłem z sukcesem w szeregu badań. Przykładem jest badanie moralnego efektu języka obcego. Efekt ten polega na zmianie decyzji moralnych na bardziej utylitarystyczne w przypadku, gdy problem przedstawiony jest badanemu w jego języku obcym. Mój artykuł na ten temat opublikowany został w *Journal of Experimental Psychology: Learning, Memory, and Cognition* (Muda, Niszczoła, Białek, i Conway, 2018), a w recenzji znajdują się dwa kolejne<sup>2</sup>.

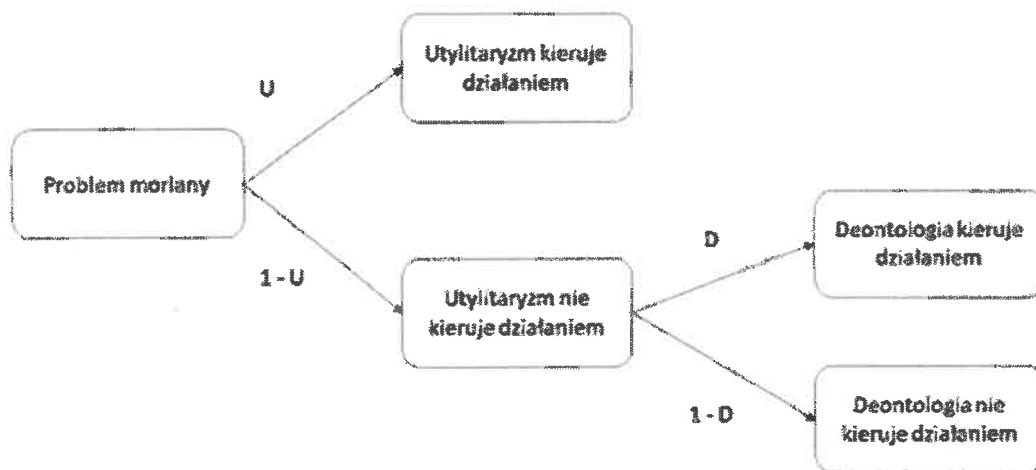
Wcześniejsze badania sugerują, że decydenci są bardziej skłonni dokonywać wyborów utylitarystycznych gdy podejmują decyzję w języku obcym (Geipel i inni., 2015). Proponowanym wyjaśnieniem efektu języka obcego w sądach moralnych jest albo spadek reakcji emocjonalnych związanych z wyrządzeniem krzywdy innym (zatem spadek inklinacji deontologicznych), albo wzrost zaangażowania poznawczego w rozwiązywanie problemu moralnego (zatem wzrost inklinacji utylitarystycznych) (Hayakawa, Costa, Foucart, i Keysar, 2016; Hayakawa, Tannenbaum, Costa, Corey, i Keysar, 2017). Wcześniejsze badania nie były w stanie rozstrzygnąć pomiędzy tymi wyjaśnieniami, gdyż w stosowanym powszechnie paradygmacie dylematów moralnych wybór deontologiczny jest z konieczności anty-utyliitarystyczny.

Aby zweryfikować te dwa przypuszczenia przeprowadziłem badania z zastosowaniem metody dysocjacji procesu (process dissociation, Conway i Gawronski, 2013). Metoda ta pozwala na odseparowanie dwóch potencjalnie niezależnych procesów, które generują decyzję (deontologię i utylitarystykę) i wymaga zastosowania klasycznych dylematów moralnych wraz z ich komplementarną wersją niekonfliktową. Dla przypomnienia, dylemat konfliktowy wymaga naruszenia zasad moralnych w celu uzyskania większych korzyści dla większej liczby osób, na przykład pozwala na torturowanie podejrzanego o planowanie zamachu terrorystycznego, by mu zapobiec i ocalić w ten sposób wiele osób. Przeciwstawia sobie zatem inklinację deontologiczną i utylitarystyczną. Dylemat niekonfliktowy, z kolei, pyta o naruszenie zasad moralnych, które nie prowadzi do uzyskania większych korzyści dla większej liczby osób. Na przykład oceniana jest dopuszczalność tortur, by zapobiec aktowi wandalizmu. Ponieważ w tym przypadku koszty nie przeważają korzyści, to w zarówno inklinacja deontologiczna jak i inklinacja utylitarystyczna skłaniają decydenta do odmowy działania. Rysunek 2 przedstawia drzewo decyzyjne sądów moralnych, z uwzględnieniem motywów ich podjęcia.

---

<sup>2</sup> Aktualnie realizuję jako kierownik projektu grant SONATA poświęcony efektowi języka obcego w procesach poznawczych, z szczególnym uwzględnieniem sądów moralnych, oraz jestem opiekunem naukowym mgra Rafała Mudy w PRELUDIUM poświęconemu tematowi efektu języka obcego w podejmowaniu decyzji ryzykownych.





**Rys. 2.** Drzewo decyzyjne sądów moralnych i zaangażowane w nie inklinacje moralne.

Powyższe ścieżki oraz odpowiadające im motywy decydenta można zoperacjonalizować matematycznie. Prawdopodobieństwo tego, że utylitaryzm będzie sterować zachowaniem oznaczmy  $U$ , a tego, że deontologia będzie sterować zachowaniem oznaczmy  $D$ . Z tego wynika, że  $1-U$  opisywać będzie przypadki, w których utylitaryzm nie kierował decyzją, a  $1-D$  przypadki, w których deontologia nie kierowała decyzją. Wyrządzenie krzywdy innej osobie bez uzasadnienia (jak w dylemacie niekonfliktowym) jest zatem możliwe jedynie gdy ani deontologia ( $1-D$ ) ani utylitaryzm ( $1-U$ ) nie zaktywują działania. Wyrządzenie krzywdy innej osobie w celu osiągnięcia korzyści dla innych osób (jak w dylemacie konfliktowym) uznane będzie za niedopuszczalne, gdy utylitaryzm nie kieruje działaniem ( $1-U$ ), oraz gdy jednocześnie deontologia kieruje działaniem ( $D$ ). Działanie takie zostanie jednak uznane za dopuszczalne w przypadku, gdy utylitaryzm kieruje działaniem ( $U$ ), albo gdy ani deontologia ( $1-D$ ) ani utylitaryzm ( $1-U$ ) nie kierują działaniem. Możemy powyższe rozważania zapisać wzorami:

$$p(\text{niedopuszczalne / niekonfliktowy}) = U + [(1-U) \times D] \quad (1)$$

$$p(\text{dopuszczalne / niekonfliktowy}) = (1-U) \times (1-D) \quad (2)$$

$$p(\text{dopuszczalne / konfliktowy}) = U + [(1-U) \times (1-D)] \quad (3)$$

Z powyższych wzorów można wyprowadzić wzór na inklinację utylitarystyczną  $U$

$$U = p(\text{niedopuszczalne / niekonfliktowy}) - p(\text{niedopuszczalne / konfliktowy}) \quad (4)$$

W kolejnym etapie, mając wartość  $U$ , możemy wyliczyć wartość inklinacji deontologicznej  $D$

$$D = p(\text{niedopuszczalne / konfliktowy}) / (1 - U) \quad (5)$$

Podsumowując, znając odpowiedzi badanego na konfliktowe i niekonfliktowe dylematy moralne możemy wyliczyć dwa, niezależne od siebie parametry, które wyrażają liczbowo siłę inklinacji utylitarystycznej oraz deontologicznej. Jest to zdecydowany postęp w stosunku do klasycznych badań prowadzonych przy użyciu jedynie dylematów konfliktowych.

*Pał*

W opublikowanym przede mnie eksperymencie, dwujęzyczni polscy badani byli przydzielani do grupy kontrolnej, w której czytali polskojęzyczny zestaw 20 dylematów (10 konfliktowych oraz 10 niekonfliktowych) albo do grupy eksperymentalnej, w której czytali ten sam zestaw 20 problemów, tyle że w języku angielskim. Po zapoznaniu się z dylematem badani oceniali dopuszczalność podjęcia opisanego działania np. dokonania aborcji ze względów medycznych (dylemat konfliktowy) lub ze względów ekonomicznych (dylemat niekonfliktowy). Aby uniknąć problemów z interpretacją potencjalnych różnic wynikających z niezrozumienia tekstu, badani byli także oceniani pod kątem zrozumienia problemów które czytali. Ponadto, badani oceniali swoje zrozumienie tekstu oraz deklarowali stopień znajomości języka angielskiego. Ci, którzy wypadli źle na którymś z tych elementów byli usuwani z badania przed dokonaniem analizy danych, ale dotyczyło to jedynie niewielkiej liczby uczestników (17% z n = 167 badanych).

Omówiona powyżej analiza z wykorzystaniem dysocjacji procesu pozwoliła mi prześledzić zachodzące niezależnie od siebie zmiany w inklinacji deontologicznej oraz utylitarystycznej spowodowane przez język, w którym prezentowano dany problem. Po pierwsze, wyniki potwierdziły wcześniejsze przypuszczenia, że deontologia i utylitarystyka są niezależnymi od siebie procesami. Po drugie, nowatorskim wynikiem jest zaobserwowany spadek znaczenia obu inklinacji moralnych, gdy problem przedstawiano w języku obcym. Oznacza to, że problemy moralne w języku obcym tracą nieco ze swojego unikatowego statusu, a stają się „zwykłymi” problemami decyzyjnymi, znanymi choćby z ekonomii. Zmiany w sądach moralnych wynikają zatem ze zmiany statusu problemu z moralnego na amoralny, a nie ze zmienioną identyfikacją z konkretną filozofią moralną.

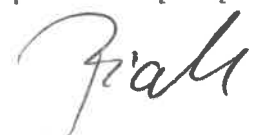
Równocześnie z moimi badaniami, identyczny projekt badawczy, także używając dysocjacji procesu prowadzili naukowcy z University of Chicago, którzy opublikowali swoje wyniki w *Psychological Science* (Hayakawa i inni, 2017). Badacze ci uzyskali identyczne z moimi wyniki: także w ich badaniach zaobserwowano spadek inklinacji deontologicznych oraz utylitarystycznych w sądy moralne. W świetle tych danych oba proponowane dotychczas wyjaśnienia moralnego efektu języka obcego wydają się nieadekwatne. Aktualnie, we współpracy z Bertramem Gawronskim, prowadzę dalsze badania analizujące opisywany tu efekt (tekst w recenzji w *Journal of Experimental Social Psychology*).

#### **4.4. Modele teoretyczne procesów poznawczych zaangażowanych w wydawanie sądów moralnych**

##### **4.4.1. Teoria dwusystemowa (dual-process theory)**

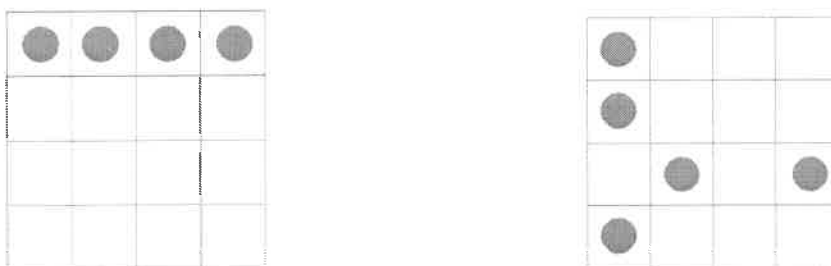
Dobłą ilustracją różnicy między deontologią i utylitarystyką jest seria historii stworzonych przez Haidta (Haidt, Koller, i Dias, 1993). Jedną z takich historii opisuje pewnego mężczyznę, który co weekend kupuje mrożonego kurczaka, na osobności z nim kopuluje, a następnie gotuje go i zjada. Nikt tego nie widzi ani o tym nie wie. Opisane zachowanie jest odrażające. Nic zatem dziwnego, że duża część z nas w oparciu o to uczucie uznaje zachowanie opisywanego mężczyzny za niemoralne (Wheatley i Haidt, 2005). Są jednak ludzie, którzy zaczynają się zastanawiać komu szkodzi to, że mężczyzna robi to, co robi (przecież nikt nie widzi jego zachowania, nie grozi ono żadną chorobą, a nawet jedzenie się nie marnuje). Analiza konsekwencji opisanego działania prowadzi do wniosku, że zachowanie mężczyzny – mimo że odrażające – nie jest szkodliwe, zatem nie jest niemoralne.

W oparciu o powyższy przykład widzimy wyraźnie jak namysł i wynikająca z niego szersza analiza konsekwencji pozwalają przewyciężyć odruch oceniania moralności w oparciu o emocje. Haidt sugeruje, że emocje i intuicja prowadzą do podejmowania częstszych wyborów deontologicznych, podczas gdy namysł pełniej analizuje konsekwencje czynu i skłania decydentów do wyborów utylitarystycznych. I właśnie rozróżnienie na jakościowo różne procesy intuicyjne i refleksyjne jest podstawową tezą teorii dwusystemowej (Greene, 2007).



Teoria dwusystemowa rozwijana jest w ramach psychologii rozumowania od lat 80—tych, ale swoją popularność zawdzięcza przeniesieniu jej do obszarów mniej abstrakcyjnych niż rozumowanie, m.in. do sądów moralnych. Joshua Greene uchodzi za pioniera w tym zakresie i to właśnie z jego nazwiskiem utożsamiana jest teoria dwusystemowa sądów moralnych. Pokrótce, teoria ta zakłada, że decyzje moralne realizowane są w umyśle na dwóch poziomach: (a) intuicji i emocji, które koncentrują się na samym działaniu i jego bezpośrednich skutkach, oraz (b) chłodnej, nieemocjonalnej refleksji, która analizuje szerszy kontekst decyzji, oraz która rozważa zarówno niebezpośrednie skutki działania, jak i skutki, które nie wystąpią, gdy decydent zaniecha podjęcia danego działania. Ilustrując to na przykładzie dylematu wagonika: intuicja koncentruje się głównie na samym akcie, czyli na przełączeniu zwrotnicy i spowodowaniu w ten sposób śmierci jednej, niewinnej osoby. Refleksja z kolei uwzględnia także skutki zaniechania działania, czyli śmierć pięciu osób. Zatem intuicja sugeruje odmowę przełączenia zwrotnicy, podczas gdy refleksja sugeruje jej przełączenie. Nas poziomie filozofii moralnej mówimy o analogii tych wyborów do deontologii (która nie pozwala łamać zasad moralnych pod żadnym pozorem) i utylitaryzmu (który uznaje za moralne takie działania, które maksymalizują szczęście). Łącząc oba te podejścia, możemy uprościć teorię dwusystemową, mówiąc, że intuicja faworyzuje deontologię, a refleksja utylitaryzm.

Teorię dwusystemową sądów moralnych wspiera szereg badań, prowadzonych głównie przy zastosowaniu dylematów moralnych będących wariacją wspomnianego już dylematu wagonika. Badania wskazują, że osoby, które zmuszone są do odpowiadania pod presją czasu dokonują znacząco mniej utylitarystycznych wyborów (Suter i Hertwig, 2011) niż osoby mające czas do namysłu; podobny efekt daje zajęcie zasobów poznawczych innym zadaniem (Conway i Gawronski, 2013; Trémolière i Bonnefon, 2014). W efekcie ludziom najwidoczniej brakuje zasobów poznawczych, by skupić się na konsekwencjach wyboru, a w zamian bardziej polegają na swoich emocjach. Wynik ten potwierdziłem we własnych badaniach, opublikowanych w *Judgment and Decision Making* (Białek i De Neys, 2017). W trzech eksperymentach badani czytali dwa konfliktowe i dwa niekonfliktowe problemy moralne, każdy z nich rozbitý na dwie części. Pierwsza część każdego z dylematów przedstawiała opis problemu (np. „nadjeżdża wagonik, który zabije pięć nieświadomych niczego osób”), a druga część opisywała jego potencjalne rozwiązanie („możesz ocalić je przełączając zwrotnicę i kierując wagonik na zapasowy tor. Niestety, pracuje na nim jedna osoba, a przekierowanie wagonika z pewnością ją zabije”). Badani proszeni byli o ocenę, na ile skłonni byłiby podjąć opisane działanie (Eksperyment 1) lub o ocenę dopuszczalności danego działania (Eksperyment 2). Pomiędzy częściami dylematu badany obserwował przez kilka sekund macierz 4x4, z umiejscowionymi w niej kilkoma kropkami (Rysunek 3). Po zapamiętaniu macierzy i zapoznaniu się z dylematem badan, podejmował decyzji, i na końcu próbował rozpoznać macierz, którą widział wcześniej spośród czterech dostępnych opcji.



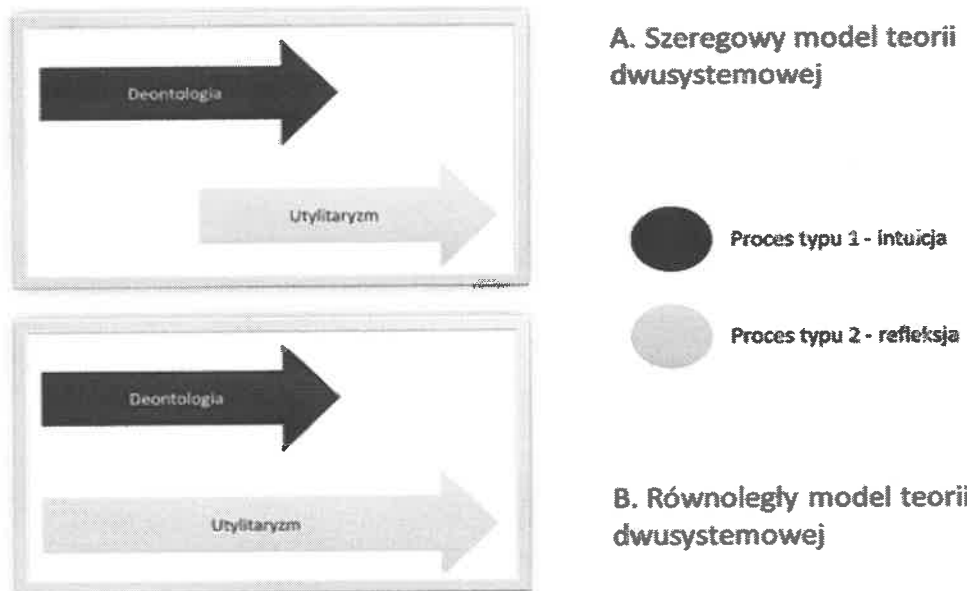
**Rys.3.** Macierze wykorzystane w badaniach opublikowanych Białek i De Neys, 2017. Lewa macierz stanowiła niskie obciążenie poznawcze, a prawa wysokie obciążenie poznawcze.

Manipulacja eksperymentalna polegała na tym, że część badanych znajdowała się w warunkach niskiego obciążenia poznawczego (4 kropki ułożone w linię), a inna część badanych znajdowała się w warunkach

wysokiego obciążenia poznawczego (5 kropek rozmieszczonych losowo). Zaobserwowałem, że gdy obciążenie poznawcze było większe badani rzadziej deklarowali chęć ocalenia pięciu osób kosztem jednej niewinnej osoby, ale ich wybory w dylematach niekonfliktowych nie ulegały zmianie. Wynik ten zdaje się być zgodny z przypuszczeniami teorii dwusystemowej – utylitarystyczna decyzja o poświęceniu jednej osoby, by ocalić pięć wydaje się być uzależniona od dostępności zasobów poznawczych, a nie wynika jedynie ze zmiany kryterium decyzyjnego na bardziej liberalne.

Kolejnym źródłem dowodów wspierających teorię dwusystemową jest fakt, że sądy utylitarystyczne pojawiają się częściej u osób, które preferują refleksyjny styl poznawczy (Paxton, Ungar, i Greene, 2012) oraz u których reakcje emocjonalne są osłabione w wyniku aleksytymii (Patil i Silani, 2014). Więcej utylitarystycznych wyborów obserwujemy także, gdy prosimy osoby by rozważyły problem moralny posługując się językiem obcym zamiast ojczystym (Cipolletti, McFarlane, i Weissglass, 2016; Geipel i inni, 2015). Ten ostatni efekt autorzy tłumaczą niższą aktywnością emocji, gdyż język obcy nie ma tak silnych afektywnych konotacji, jakie ma język ojczysty. W efekcie, w języku obcym słabiej odczuwamy emocje takie jak oburzenie, złość czy odrazę. Gdy rozważamy problemy moralne w języku obcym mniejsze znaczenie mają też intencje (Geipel, Hadjichristidis, i Surian, 2016) oraz normy społeczne (Gawinkowska, Paradowski, i Bilewicz, 2013). Wszystkie te elementy odpowiadają za podejmowanie wyborów deontologicznych i stanowią procesy poznawcze klasyfikowane jako procesy Typu 1. Istnieje też cała seria danych z neuroobrazowania, które łączą utylitarystykę z ośrodkami odpowiedzialnymi za funkcje wykonawcze czy myślenie abstrakcyjne (Greene, Nystrom, Engell, Darley, i Cohen, 2004), a deontologię z ośrodkami odpowiedzialnymi za emocje (Moll, Zahn, de Oliveira-Souza, Krueger, i Grafman, 2005).

Podsumowując tą część, można przyjąć, że teoria dwusystemowa sądów moralnych zakłada odmienne źródło wyborów moralnych spójnych z deontologią i utylitarystyką (odpowiednio intuicyjne procesy Typu 1 vs refleksyjne procesy Typu 2). O ile na tym poziomie ogólności opis teorii dwusystemowej sądów moralnych wydaje się trafny, problemem staje się to w jaki sposób owe procesy współpracują. Istnieją mianowicie dwa konkurencyjne wyjaśnienia: model szeregowy i model równoległy (Stanovich i Evans, 2013, De Neys i Białek, 2017). Przedstawiam je schematycznie na Rysunku 4.



**Rys. 4.** Szeregowy i równoległe ujęcie teorii dwusystemowej. Procesy Typu 1 są szybkie, automatyczne, niekontrolowalne wolicjonalnie i nie wymagają zasobów pamięci roboczej. Procesy Typu 2 są wolicjonalnie uruchamiane, powolne i uzależnione od pojemności pamięci roboczej. Długość strzałek i miejsce gdzie się zaczynają i kończą odpowiada charakterystyce czasowej procesów.

Pierwszy z proponowanych modeli postuluje że decydent zawsze stara się uruchamiać oba typy procesów, a decyzję opierać na tym, który generuje subiektywnie lepsze rozwiązanie. Model ten nazywa się modelem równoległym. Drugi z proponowanych modeli postuluje że procesy intuicyjne uruchamiają się automatycznie, a procesy refleksyjne uruchamiane są wtedy, gdy decydent uzna to za stosowne (np. gdy problem jest abstrakcyjny, gdy decyzja ma duże znaczenie, lub gdy poziom motywacji decydenta jest wysoki). Model ten nazywa się modelem szeregowym.

Równoległy model teorii dwusystemowej zakłada, że oba typy procesów startują równocześnie, gdy w otoczeniu pojawi się stosowny bodziec, ale procesy Typu 1 są szybsze i z reguły przejmują kontrolę nad działaniem. Gdy decydent pozna odpowiedź intuicyjną, to w zależności od jej subiektywnej jakości, procesy typu 2 są wygaszane albo kontynuowane. Jedynie w tym drugim przypadku mają one szansę na przejęcie kontroli nad zachowaniem (Stanovich, 2009). Model równoległy narusza jednak zasadę skąpstwa poznawczego, w świetle której umysł skonstruowany jest tak by pracować jak najefektywniej przy jak najmniejszym użyciu zasobów poznawczych (Fiske i Taylor, 2013).

Szeregowy model teorii dwusystemowej zakłada, że ludzie uruchamiają procesy Typu 1 zawsze, gdy natrafią na odpowiedni bodziec ze środowiska. Nie mają nad tym żadnej kontroli. Od decyzji człowieka zależy jedynie to czy uruchomi dodatkowo procesy Typu 2. Taka wersja (model default-interventionist) jest popularna np. w obszarze rozumowania, gdzie zakłada się, że człowiek dokonuje subiektywnej analizy swojej intuicyjnej odpowiedzi i gdy czuje, że odpowiedź ta nie jest satysfakcjonująca, to stara się skorygować ją wykorzystując do tego procesy Typu 2 (Evans, 2008; Evans i Stanovich, 2013).

Greene nie dookreśla którą wersją teorii dwusystemowej jest jego model sądów moralnych, szczęśliwie unikając konieczności przyjęcia wszystkich jej założeń (Białek i Terbeck, 2016; De Neys i Białek, 2017; Gürçay i Baron, 2017). Nie wiemy zatem czy utylityzm pojawia się wtedy, gdy intuicja deontologiczna daje niesatysfakcjonującą odpowiedź, czy też równolegle prowadzona jest refleksja utylitystyczna. Na to pytanie szukałem odpowiedzi we własnych badaniach. To pytanie jest o tyle ważne, że jeśli modelu Greena nie daje się osadzić spójnie w ramach żadnego z modeli – szeregowego bądź równoległego – to oznacza to, że teoria ta jest nadmiernie uproszczona, albo nawet całkiem błędna. Oczekujemy przecież w nauce pewnej spójności w zakresie tworzonych modeli i nieracjonalne wydaje się tworzenie modelu wydawania sądów, a odmiennego modelu dla wydawania sądów moralnych (Białek i Terbeck, 2016; Osman i Wiegmann, 2017).

Przedstawiona powyżej teoria dwusystemowa stanowiła obiekt moich badań. Kilka artykułów teoretycznych analizowały założenia owego modelu, a kilka kolejnych testowało empirycznie przypuszczenia modeli szeregowego oraz równoległego. W kolejnych podrozdziałach przedstawiam szczegółowo te artykuły.

#### **4.4.2. Krytyka założeń teorii dwusystemowej sądów moralnych**

Jak wspomniałem wcześniej, jednym z podstawowych założeń teorii dwusystemowej jest refleksyjny charakter deontologii. Pierwszą z publikacji, w której podważam to założenie, jest tekst teoretyczny w *American Journal of Bioethics – Neuroscience* (Białek, Terbeck, i Handley, 2014). Tekst ten jest rozwinięciem wątków z mojej dysertacji, w której zajmowałem się intuicyjnym rozumieniem logiki. W typowych badaniach nad rozumowaniem, badani analizują syllogizmy, które są prawomocne (lub nieprawomocne) i których wniosek jest wiarygodny (lub niewiarygodny). Zadaniem badanych jest ocena prawomocności wniosku ignorując jego wiarygodność. Na przykład rozumowanie:

*Wszystkie rośliny potrzebują wody*

*Róże potrzebują wody*

*Zatem, wszystkie róże to rośliny*



jest nieprawomocne logicznie (wniosek nie wynika koniecznie z przesłanek), ale wniosek jest wiarygodny (faktycznie różne są roślinami). Mimo to, większość badanych uznaje powyższy wniosek za prawomocny. Podobnie zatem do rozumowania moralnego, badani częściej opierają swoje sądy na intuicyjnie dostępnych informacjach, a tylko niektórzy analizują problem od strony formalnej.

Zdarza się jednak, że ludzie udzielający błędnej odpowiedzi w zadaniach konfliktowych jak to powyżej (takich w których prawomocność i wiarygodność sugerują różne oceny wniosku) są mniej pewni siebie, potrzebują więcej czasu na udzielenie odpowiedzi oraz pamiętają więcej szczegółów problemu, niż gdy udzielają takiej samej odpowiedzi w zadaniach niekonfliktowych (gdy wniosek jest zarówno prawomocny jak i wiarygodny). Zjawisko to nazwano detekcją konfliktu (De Neys, 2014). Detekcja konfliktu jest o tyle zaskakująca, że pomimo iż ocena wniosku oparta została o intuicję, to w jakiś sposób uwzględnia ona także strukturę logiczną wniosku. Zatem, nawet gdy wniosek jest intuicyjny, to w tle przeprowadzana jest analiza refleksyjna struktury logicznej sylogizmu. Dane zdają się zatem sugerować, że model równoległy sądów moralnych jest bardziej adekwatnym opisem przebiegu procesu rozumowania.

Sprawa jest jednak bardziej złożona. Detekcja konfliktu pojawia się nawet wtedy, gdy badani poddani są obciążeniu poznawczemu. W swoim zamyśle obciążenie poznawcze powinno utrudnić lub nawet uniemożliwić realizację procesów Typu 2 odpowiedzialnych za analizę struktury logicznej sylogizmu. Badania pokazują, że w obciążeniu poznawczym badani popełniają więcej błędów w ocenie rozumowania, jeszcze częściej niż zwykle opierając ocenę prawomocności wniosku na jego wiarygodności. Mimo tego pogorszenia poprawności odpowiedzi, detekcja konfliktu nie zostaje jednak zaburzona. Wniosek, jaki z tego wyciągnięto jest taki, że analiza struktury logicznej posiada charakterystykę procesów Typu 1, czyli że ludzie posiadają intuicję logiczną. Podobne wyniki jak w zakresie rozumowania odkryto także w przypadku ignorowania proporcji podstawowej (base rate neglect, Pennycook, Trippas, Handley, i Thompson, 2014), gdzie wykazano intuicyjne rozumienie i przetwarzanie prawdopodobieństwa warunkowego.

Powyższe badania sugerują zatem, że sądy i decyzje, które dotychczas uznawano za procesy Typu 2, mogą być realizowane przez procesy Typu 1. Analogicznie do oceny prawdopodobieństwa oraz oceny prawomocności wniosku, argumentuję w swoim artykule, także sądy użyteczne mogą być realizowane przez procesy Typu 1 (Białek, Handley i Terbeck, 2014), a założenie o refleksyjnym charakterze użyteczności może zatem być nietrafne.

Innych argumentów przeciw refleksyjnemu charakterowi użyteczności dostarczam w artykule teoretyczno-przeglądowym opublikowanym w *Cognitive Processing* (Białek i Terbeck, 2016). Przytaczamy tam dane, które pokazują, że pozytywna zależność między refleksyjnym stylem poznawczym a wyborami użytecznymi zachodzi tylko, gdy problem rozpatrywany jest na poziomie konkretnym (concrete construal level). Gdy problem moralny rozpatrywany jest na poziomie abstrakcyjnym (abstract construal level), to refleksyjny styl poznawczy jest odwrotnie skorelowany z sądami użytecznymi (Körner i Volk, 2014). Takie wyniki badań są niespójne z przypuszczeniami teorii dwusystemowej.

Innym przypuszczeniem teorii dwusystemowej sądów moralnych jest to, że sądy deontologiczne mają swoje źródło w procesach Typu 1 (m. in. w emocjach). Terbeck i inni (2012) pokazali jednak, że farmakologiczna blokada reakcji emocjonalnych prowadzi w efekcie do wzrostu sądów użytecznych. Także ten wynik jest niespójny z przypuszczeniami teorii dwusystemowej.

Ostatni omawiany przeze mnie argument przeciw teorii dwusystemowej pochodzi z prac Guya Kahane z Oksfordu, który pokazał, że w niektórych dylematach pierwszą, intuicyjnie dostępną odpowiedzią jest ta użyteczna. Na przykład, gdy wiemy, że nasz znajomy przechowuje zakazane książki na strychu, odruchowo wolimy skłamać pytającym nas o to policjantom, aby uchronić go przed karą (Kahane i in., 2014). W tym kontekście Kahane mówi, że dylematy „wagonikopodobne” są po prostu

*Fialy* 15

przedstawicielami grupy problemów, w których utylitaryzm jest intuicyjny, podkreślając jednocześnie, że nie jest to grupa reprezentatywna dla wszystkich problemów moralnych (Kahane, 2015).

#### 4.4.3. Krytyka modeli szeregowego i równoległego teorii dwusystemowej

Jak wspomniałem, w psychologii poznawczej rozpatruje się dwa warianty teorii dwusystemowej – szeregowy i równoległy (Stanovich i Evans, 2013). Oba te modele podlegają krytyce, a szczegółowo omawiam te argumenty w swoich pracach w Przeglądzie Filozoficznym – Nowa Seria (Białek, 2015) oraz w Cognitive Processing (Białek i Terbeck, 2016). Argumentuję tam, że model równoległy teorii dwusystemowej ma szereg wad koncepcyjnych. Na przykład wydaje się on być nieekonomiczny – jeśli przyjąć, że nadrzędną zasadą funkcjonowania umysłu jest wydatkowanie jak najmniejszej ilości energii by osiągnąć zamierzony cel (skąpstwo poznawcze, Fiske i Taylor, 1991), to nie ma sensu każdorazowo uruchamiać procesów refleksyjnych mających duże zapotrzebowanie na zasoby poznawcze. Efektywniej byłoby opierać umysł o działanie intuicyjne, a posiłkować się refleksją tylko w razie konieczności. Faktycznie, w pozamoralnych decyzjach wydaje się, że większość z naszych decyzji i wniosków oszczędza energię korzystając z prostych schematów myślenia – heurystyk (Kahneman, 2011). Model równoległy teorii dwusystemowej ma ponadto wadę w postaci potrzeby powołania jakiegoś systemu arbitrażowego. Musi bowiem istnieć jakiś trzeci system, który mógłby ocenić rozwiązania proponowane przez intuicję i refleksję. Aby optymalnie dokonać takiej oceny, system ten powinien znać najlepszą odpowiedź na rozwiązywany problem lub mieć specjalny system wytycznych do oceny potencjalnych rozwiązań. System ten de facto podejmowałby decyzje porównując zaproponowane rozwiązania do optymalnego. Otwartymi pytaniami są: na jakiej zasadzie działa ten mechanizm, jakich używa kryteriów do podjęcia decyzji i czy on sam jest intuicyjny czy refleksyjny. Adaptując ten model do sądów moralnych wymagałby on, by decydent intuicyjnie określał, że dane działanie jest moralnie niedopuszczalne, ale jednocześnie refleksyjnie analizował konsekwencje podjęcia i zaniechania działania. Decyzję decydent podejmuje poprzez porównanie wagi argumentów wspierających działanie i niez działanie, posługując się w tym celu pewnym, bliżej nieokreślonym procesem kontrolnym, lub używając do tego refleksji. Evans (2008) porównuje taki model do wyścigu koni - szybkiego i wolnego. Na metę pierwszy dobiega szybki koń, ale musi on poczekać na wolnego konia, który i tak ostatecznie decyduje, kto wygrał wyścig.

Krytyka rozwiązania równoległego nie oznacza jednak, że model szeregowy jest lepszy i należy go przyjąć. Podstawowym problemem modelu szeregowego jest określenie kiedy decydent powinien uruchomić procesy Typu 2. Decydent musi bowiem skądś wiedzieć, że intuicyjne rozwiązanie jest potencjalnie błędne i że warto skorygować je poprzez procesy Typu 2. Teoria dwusystemowa nie jest jasna, co do tego jak przebiegać miałby ów monitoring procesów Typu 1, ponad postulat metapoznawczy. Odwołuje się on do pewnego dość niejasnego procesu, który subiektywnie ocenia „jakość” intuicji.

Przykładem modelu szeregowego w przypadku sądów moralnych jest model SIM Haidta (2007). Zakłada on, że jeśli intuicyjnemu sądowi towarzyszą negatywne emocje to dana intuicja jest oceniana jako niezadowolająca, i decydent może uruchomić dodatkowe procesy mające za zadanie dogłębną analizę danego problemu (czyli właśnie procesy Typu 2). Przekładając powyższe rozważania bezpośrednio na terminologię teorii dwusystemowej, model szeregowy postuluje, że decydent najpierw generuje intuicję deontologiczną (proces typu 1), a następnie ocenia ją pod względem jej subiektywnej jakości. Gdy ocena ta jest pozytywna to intuicja zostaje uznana za poprawną, gdy ocena jest niezadowolająca to decydent próbuje dogłębnie analizować dany problem moralny przy pomocy procesów Typu 2. Proces oceny intuicji nie może jednak być intuicją, która sama się ocenia, nie może też być procesem Typu 2, gdyż nie ma możliwości by znać jego wyniki przed jego rozpoczęciem. Charakterystyka procesu kontrolnego jest zatem najłagodniejszą stroną modelu szeregowego teorii dwusystemowej. Nie argumentuję w swoim tekście, że problemu tego nie da się przewyciężyć, ale jedynie, że jego konceptualizacja jest niepełna i wymaga doprecyzowania w oparciu o dane empiryczne.





#### 4.4.4. Weryfikacja empiryczna zarzutów stawianych modelom szeregowemu i równoległemu dwusystemowemu

Założenie o refleksyjnym charakterze utylitaryzmu stanowiło oś trzech moich publikacji napisanych wspólnie z Wimem De Neyssem (Sorbona). W pierwszej z nich, opublikowanym w *Journal of Cognitive Psychology* (Białek i De Neys, 2016) wykazaliśmy, że analiza utylitarystyczna odbywa się zawsze, nawet w przypadku podjęcia teoretycznie intuicyjnej decyzji deontologicznej.

Badani w omawianym eksperymencie proszeni byli o podjęcie decyzji w czterech dylematach moralnych. Dwa z dylematów były konfliktowe: przedstawiały wybór, w którym należy poświęcić jedną osobę, by ocalić pięć innych osób. Dla przypomnienia, w tych wyborach deontologia sugerować powinna odmowę poświęcenia, gdyż niedopuszczalne jest aktywne wyrządzenie krzywdy innym, niezależnie od pozytywnych konsekwencji jakie ten akt za sobą pociąga. Z kolei utylitaryzm sugerowałby, że należy dokonać poświęcenia, gdyż celem wyborów moralnych jest maksymalizacja szczęścia dla jak największej liczby osób. W tych problemach deontologia i utylitaryzm są ze sobą w konflikcie. Według teorii dwusystemowej sądów moralnych odmowa dokonania poświęcenia opiera się w tych zadaniach na procesach Typu 1, podczas gdy decyzja o poświęceniu jednej osoby wynika z procesów Typu 2.

Kolejne dwa dylematy moralne były niekonfliktowe, przedstawiały podobny scenariusz, z kluczową różnicą w postaci liczby osób, które należało poświęcić oraz które mogą zostać ocalone. W tych dylematach należało bowiem poświęcić pięć osób, by ocalić zaledwie jedną osobę. W tych dylematach deontologia i utylitaryzm wspierają odmowę poświęcenia, gdyż wymaga to aktywnego skrzywdzenia innych (deontologia) oraz nie prowadzi do wzrostu szczęścia jak największej liczby osób (utyliaryzm).

Zasadnicza różnica pomiędzy decyzjami deontologicznymi decydentów w tych dylematach polega na tym, że inklinacja utylitarystyczna wspiera intuicję deontologiczną w dylematach niekonfliktowych a jest z nią w konflikcie w dylematach konfliktowych. W świetle omawianej wcześniej wersji szeregowej teorii dwusystemowej sądów moralnych wybór deontologiczny opiera się jedynie na intuicji, zatem dla deontologa pozytywne czy negatywne konsekwencje dokonania poświęcenia nie powinny mieć dużego znaczenia, gdyż nie zostały nawet przeanalizowane przez decydenta. Intuicja deontologiczna była bowiem zadowolająca i nie było potrzeby uruchamiać procesów Typu 2, które mogłyby wygenerować argumenty na rzecz wyboru utylitarystycznego.

Wyniki tego eksperymentu wykazały jednak, że badani, którzy podjęli decyzje deontologiczne byli mniej pewni swojego wyboru oraz potrzebowali więcej czasu by go dokonać. Ten efekt wynika z detekcji konfliktu między inklinacją deontologiczną i utylitarystyczną w dylematach konfliktowych. Decydenci zdawali sobie sprawę z atrakcyjności istniejącej, aczkolwiek niewybranej, opcji utylitarystycznej. Oznacza to, że mimo iż intuicyjnie podjęta została decyzja moralna, to jednak zawsze uwzględniała konsekwencje uznawane dotychczas za produkt procesów refleksyjnych. Dane te są zatem niespójne z modelem szeregowym sądów moralnych. Model równoległy przyjmuje, że deontologia i utylitaryzm są przetwarzane równoległe, zatem nie jest zaskakujące byłoby zaobserwowanie naszych wyników. Być może refleksyjna analiza konsekwencji w tak prostych problemach jak dylematy moralne, które de facto sprowadza się do porównania dwóch liczb, może zostać dokonana stosunkowo szybko i interferować z wyborami intuicyjnymi. Nawet jednak taka zgrubna analiza, w świetle modelu równoległego jest procesem Typu 2 i z definicji wymaga dostępu do zasobów poznawczych pamięci roboczej.

Przypuszczenia modelu równoległego zweryfikowałem w serii trzech eksperymentów opublikowanych w *Judgment and Decision Making* (Białek i De Neys, 2017). W tych badaniach, podobnie jak w omawianym wcześniej badaniu z *Journal of Cognitive Psychology* (Białek i De Neys, 2016) użyliśmy zestawu dylematów konfliktowych oraz ich niekonfliktowych odpowiedników. Tym razem decydenci dokonywali wyborów w warunku silnego lub niskiego obciążenia poznawczego. Szczegóły procedury obciążenia przedstawiłem już

*Pidk* 17

w części autoreferatu omawiającej teorię dwusystemową. Zgodnie z modelem równoległym, analiza utylitarystyczna wymaga dostępu do zasobów poznawczych. Ich redukcja powinna zatem efektywnie utrudnić lub nawet uniemożliwić realizację procesów Typu 2 i w efekcie osłabić inklinację utylitarystyczną. W takim przypadku osłabiona analiza utylitarystyczna nie powinna wywoływać konfliktu z intuicją deontologiczną (lub wywoływać ten konflikt słabszy). Z powyższego wywodu wynika, że jeśli model równoległy teorii dwusystemowej sądów moralnych jest prawdziwy, to w wysokim obciążeniu poznawczym detekcja konfliktu (czyli obserwowany przez nas spadek pewności siebie decydentów deontologicznych w dylematach konfliktowych) zaniknie lub przynajmniej zostanie osłabiona. Jeśli jednak detekcja konfliktu nie zostanie zachwiana, to najprawdopodobniej zachodzi ona pomiędzy dwoma procesami Typu 1, a nie między procesem Typu 1 a Typu 2.

Trzy eksperymenty miały podobną do siebie strukturę. W pierwszym, badani oceniali swoją chęć do podjęcia działania w dylematach konfliktowych oraz niekonfliktowych. W drugim, oceniali dopuszczalność moralną dokonania poświęcenia. W obu eksperymentach zaobserwowałem niższą pewność siebie deontologów w problemach konfliktowych niż w problemach niekonfliktowych. Efekt ten był niezależny od tego jak silne wzbudzone obciążenie poznawcze. Aby nie polegać jedynie na niebezpośredniej mierze detekcji konfliktu (czasie reakcji czy deklarowanej pewności siebie), w badaniu trzecim pytaliśmy jego uczestników wprost o to: (a) na ile dany problem wydawał im się trudny i (b) na ile czuli się wewnętrznie rozdarci podejmując daną decyzję. Także w tym eksperymencie wyniki były spójne z eksperymentami 1 i 2: w przypadku zadań konfliktowych badani częściej deklarowali bycie rozdartymi oraz uznawali rozwiązywane dylematy za trudniejsze, a obciążenie poznawcze nie zmniejszało tego efektu.

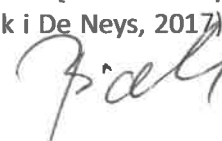
Sam brak różnic w sile detekcji konfliktu pomiędzy warunkami eksperymentalnymi nie dowodzi jednak, że należy przyjąć model zakładający brak różnic jako właściwy. Aby to potwierdzić użyliśmy statystyk bayesowskich (Domurat i Białek, 2016), które pozwalają porównać dowody wspierające model zerowy oraz kierunkowy i ocenić który z nich jest bardziej prawdopodobny. Analiza bayesowska wskazała, że model zerowy jest 32-rzy bardziej prawdopodobny niż model kierunkowy (czyli zakładający osłabienie detekcji konfliktu wraz ze wzrostem obciążenia poznawczego).

Wyniki tej serii eksperymentów sugerują, że model równoległy nie jest właściwym w wyjaśnieniu obserwowanych danych. Jeśli utylitaryzm miałby swoje źródło w procesach Typu 2, to zwiększone obciążenie poznawcze powinno skutkować osłabieniem inklinacji utylitarystycznej i w efekcie zredukować lub nawet wyeliminować zjawisko detekcji konfliktu. Eksperyment z wcześniej omawianej publikacji podważa jednak również model szeregowy teorii dwusystemowej. Podsumowując, moje badania pokazują, że ani model szeregowy, ani model równoległy nie są adekwatnym opisem przebiegu procesów poznawczych leżących u podstaw sądów moralnych. W miejsce obu tych modeli zaproponowałem hybrydowy model teorii dwusystemowej sądów moralnych. Omawiam go poniżej.

#### **4.5. Model hybrydowy sądów moralnych**

Z omawianych powyżej artykułów wynika jednoznacznie, że zarówno model szeregowy jak i model równoległy dwusystemowej teorii sądów moralnych jest niedoskonałym opisem tego, jak w rzeczywistości decydent dokonuje wyborów i wydaje sądy moralne. W przeciwieństwie do założeń modeli szeregowych i równoległych, utylitaryzm należy uznać za proces intuicyjny. Taki wniosek wspiera zarówno przedstawiona przeze mnie analiza teoretyczna (Białek i Terbeck, 2016; Białek, Handley i Terbeck, 2014) jak i zebrane przeze mnie dane empirycznie (Białek i De Neys, 2016, 2017). Mimo, że utylitaryzm jest intuicyjny, to liczne dane wskazują jednak, że brak dostępu do odpowiedniej ilości zasobów poznawczych ogranicza liczbę wyborów utylitarystycznych.

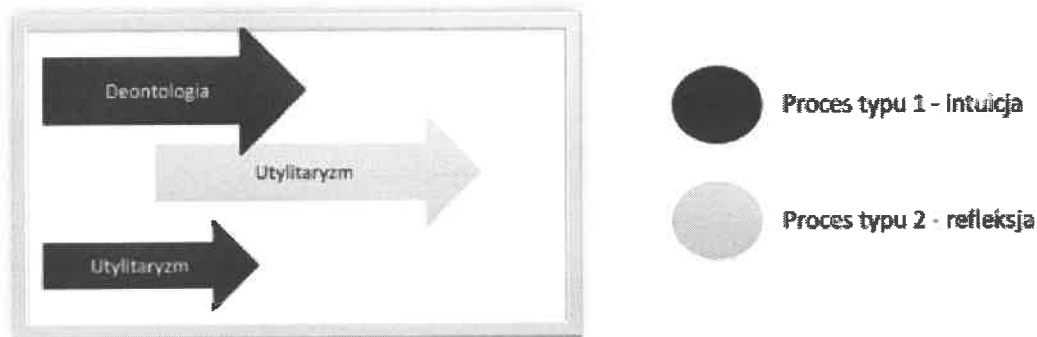
W miejsce modeli szeregowego i równoległego, zaproponowałem hybrydowy model sądów moralnych (Rysunek 5), którego zarys przedstawiłem w Judgment and Decision Making (Białek i De Neys, 2017), a

 18

szczegółowo omawiam w rozdziale monografii wydanej przez prestiżowe wydawnictwo Psychology Press *Moral Inferences* (De Neys i Białek, 2017) i w *Czasopiśmie Psychologicznym* (Białek, 2017). W zaproponowanym przeze mnie modelu zarówno deontologia jak i utilitaryzm stanowią intuicyjną reakcję na składowe dylematu moralnego: reakcja deontologiczna koncentruje się głównie na samym akcie i konieczności aktywnego (vs. pasywnego) wyrządzenia szkody drugiej, niewinnej osobie; reakcja utilitarystyczna z kolei koncentruje się głównie na szkodzie, która wyrządzona zostanie innej osobie w przypadku niepodjęcia działania. Obie te reakcje są procesami Typu 1, zachodzą automatycznie oraz nie wymagają zasobów pamięci roboczej.

Siła intuicji odczuwana jest subiektywnie jako jej poprawność (feeling of rightness, Thompson, Prowse-Turner i Pennycook, 2011). Jeśli jedna z intuicji jest subiektywnie postrzegana jako słuszniejsza, to decydent podejmuje stosowną decyzję z dużą pewnością siebie i pozornie bezrefleksyjnie. Gdy siła obu intuicji jest bardziej wyrównana, to silniejsza z nich przejmuję kontrolę nad zachowaniem, ale zachodzi w międzyczasie detekcja konfliktu, dzięki której decydent zdaje sobie sprawę z tego, że istniała inna, potencjalnie dobra opcja wyboru. W końcu, gdy obie intuicje są porównywalnie silne to detekcja konfliktu osiąga krytyczną siłę, która uruchamia procesy Typu 2. Te, w przeważającej większości dotychczasowych badań, skłaniały decydenta do wyboru utilitarystycznego.

### Hybrydowy model teorii dwusystemowej



**Rys 5.** Propozycje hybrydowego modelu teorii dwusystemowej. Refleksja, podobnie jak w innych modelach szeregowych uruchamia się, gdy procesy intuicyjne nie dają satysfakcjonującego rozwiązania (np. deontologia i utilitaryzm są o podobnej sile i tworzą silny wewnętrzny konflikt) Wielkość strzałek odpowiada sile intuicji, a ich długość i miejsce, gdzie się zaczynają i kończą, charakterystyce czasowej.

Model hybrydowy jest o wiele bardziej plastyczny niż modele szeregowy czy równoległy, gdyż pozwala wyjaśniać jak zmiany na poziomie procesów Typu 1 mogą wpływać na deontologię oraz utilitaryzm. Ponieważ obie inklinacje moralne są procesami Typu 1, to są niewrażliwe na obciążenie poznawcze. Z tego wynika, że model hybrydowy, w przeciwieństwie do modelu szeregowego czy równoległego, jest spójny z przedstawionymi powyżej danymi eksperymentalnymi. Pozwala on też zintegrować niespójne dotychczas wyniki badań nad sądami moralnymi, gdyż przewiduje intuicyjne oraz refleksyjne źródła utilitaryzmu.

*Fieda*

Przypuszczenia te znalazły potwierdzenie w późniejszych badaniach prowadzonych niezależnie na Oksfordzie (Kahane i inni, 2018), w oparciu o które stworzono dwukomponentowy model utylityzmu: pozytywny (troskę o innych) i negatywny (gotowość do skrzywdzenia innych dla osiągnięcia celu). W tym modelu utylityzm nabiera charakteru emotywnego, jako odczuwana empatię w stosunku do osób które można ocalić, oraz refleksyjnego, w którym należy przezwyciężyć awersję do krzywdzenia innych by ocalić inne osoby. Zatem model ten, podobnie jak model hybrydowy, zakłada, że utylityzm ma swoje źródło zarówno w procesach Typu 1, jak i Typu 2.

#### 4.6. Inne, powiązane tematycznie badania

##### 4.6.1. Wymiennność ludzkiego życia

Jeśli pożyczymy od sąsiada 20 złotych by zrobić zakupy na obiad, to możemy mu później oddać dowolny banknot lub kombinację monet o sumarycznej wartości 20 złotych. Gdybyśmy jednak pożyczymy od niego patelnię, to nie możemy oddać mu innej patelni o tej samej wartości czy zestawu nieco podniszczonych patelni o sumarycznej wartości takiej jak oryginalna patelnia. Pieniądze w tym kontekście są wymienne (*fungible*), a patelnia nie. Moja seria trzech eksperymentów dotyczyła problemu, kiedy ludzkie życie traktowane jest jako wymienne, a kiedy jako niewymienne. Problem jest o tyle ważny, gdyż na przykład zamachowcy, gdy chcą się zemścić na żołnierzach USA za nich niewłaściwe postępowanie to nie starają się zrewanżować dokładnie tym osobom, które zawiniły, ale atakują losowych obywateli USA. W tym kontekście Amerykanie są traktowani jako wymienne zasoby, a szkoda wyrządzona jednym jednostkom stanowi karę dla całej grupy za niewłaściwe zachowanie innych jej członków. Artykuł opisujący wyniki omawianych tu eksperymentów opublikowany został w *Judgment and Decision Making* (Białek, Fugelsang, i Friedman, 2018).

Badani w każdym z trzech eksperymentów wyobrażali sobie, że są dowódcami międzynarodowych sił NATO i proszeni byli o wybór osoby, którą należy wysłać na samobójczą misję. Celem tej misji było ocalić trzech innych żołnierzy oraz odzyskać zdobyte przez nich informacje wywiadowcze. Trzech żołnierzy zawsze było do siebie jakoś podobnych, np. wszyscy byli Łotyszami, mieli po 21 lat, bądź urodzili się w marcu. Żołnierz, którego można było wysłać by ich ocalić był albo podobny to tej grupy, albo niepodobny (np. był Estończykiem, urodził się w Kwietniu, bądź miał 23 lata).

Wyniki eksperymentów 1 oraz 2 pokazały, że badani znacznie częściej decydowali się wysłać żołnierza podobnego do pozostałej trójki jedynie wtedy, gdy podobieństwo było definiowane przez narodowość. W pozostałych przypadkach badani nie mieli preferencji co do tego, którego żołnierza należy poświęcić. W artykule argumentuję, że w przypadkach, gdy przynależność do grupy jest definiowana przez istotne cechy, dana grupa postrzegana jest, jako jedność (*entitative*, Lickel i inni, 2000), a ludzie do niej przynależący stanowią jej wymienne elementy. Z kolei, gdy grupa stanowi przypadkowy konglomerat osób, a cecha definiująca jest incydentalna, to jej członkowie zachowują swoją indywidualność. Wydaje się zatem, że w pewnych przypadkach ludzie mogą być traktowani jak wymienne elementy zbioru.

Badanie 3 testowało alternatywne wyjaśnienie obserwowanych efektów, mianowicie związane z księgowaniem mentalnym (Thaler, 1990) oraz teoria perspektywy (Tversky i Kahneman, 1992). W świetle tej drugiej dwukrotna strata 10 złotych jest bardziej bolesna niż jednorazowa strata 20 złotych, pomimo że oba przypadki strat nominalnie są jednakowo duże. Łatwiej zatem zaakceptować straty, gdy uda się je zagregować w jedną większą. Innym twierdzeniem teorii perspektywy jest to, że zyski odczuwane są mniej intensywnie niż straty. Zatem zysk 20 złotych przy jednoczesnej stracie 10 złotych jest mniej satysfakcjonujący niż zysk 10 złotych. Analogicznie do powyższego przykładu finansowego, śmierć jednej osoby prowadząca do ocalenia trzech osób będzie subiektywnie korzystniejsza, gdy uda się potraktować ją jako jedno zdarzenie niż jako dwa niezależne zdarzenia (zaksięgować zdarzenia do jednego zamiast do dwóch kont mentalnych). Aby przeciwstawić wyjaśnienie wynikające z teorii księgowania mentalnego z

*P. Białek* 20

proponowanym przeze mnie podejściem zakładającym wymiennosc życia, badanie 3 opisywało (a) scenariusz wymiany, w którym poświęcenie jednej osoby mogło ocalić trzy inne osoby, oraz (b) scenariusz nie-wymiany, w którym trzech żołnierzy zginęło, a samobójcza misja wymagała odzyskania zdobytych przez nich informacji. W obu przypadkach księgowanie mentalne przewiduje, że badani powinni wysłać podobnego żołnierza, gdyż podobieństwo między żołnierzami pozwala agregować zdarzenia i w ten sposób zwiększyć ich subiektywną użyteczność. Z kolei podejście traktujące członków grupy jako wymiennych zakłada, że uniemożliwienie traktowania zdarzenia jako wymiany zasobów wewnątrz puli zlikwiduje preferencję do poświęcenia podobnego żołnierza. W eksperymencie 3 zaobserwowaliśmy preferencję do poświęcenia żołnierza tej samej narodowości, co trzej inni żołnierze jedynie w warunku wymiany, ale nie w warunku nie-wymiany. Wydaje się zatem, że postulowany model wymiennosci życia ludzkiego trafniej opisuje zebrane dane niż teoria księgowania mentalnego.

Podsumowując, powyższy projekt rzuca światło na to kiedy może dochodzić do dehumanizacji i na możliwość traktowania ludzkiego życia jako wymiennego zasobu, zamiast jako niezależnej i nienaruszalnej wartości. Ponadto, jako jeden z pierwszych na świecie postawiłem pytanie nie o to, gdy poświęcenie jest konieczne, to kogo powinno się wybrać na ofiarę. Aktualnie kontynuuję tą linię badań, by doprecyzować jakie warunki musi spełniać grupa, by jej członków traktować jako wymienne elementy zamiast jako niezależne podmioty.

#### **4.6.2. Akulturacja jako mechanizm likwidujący moralny efekt języka obcego**

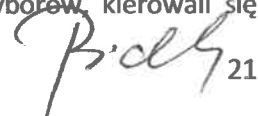
Ostatnia publikacją stanowiącą omawiane tu osiągnięcia naukowe jest mój tekst w Journal of Multilingual and Multicultural Development (Białek i Fugelsang, 2019), w którym krytykuję wcześniej opublikowany tam tekst (Cavar i Tytus, 2018). Krytykowany tekst sugeruje, że badani, którzy przechodzą akulturację niepodatni są na efekt języka obcego, a ich sądy moralne są jednakowe w języku ojczystym i obcym. W oparciu o przedstawione statystyki opisowe i rezultaty testów statystycznych przeprowadziłem re-analizę danych empirycznych zebranych przez autorki i wykazałem, że brak różnic między grupami wynika prawdopodobnie z niedostatecznej mocy (w badaniu uczestniczyło zaledwie około 60 badanych) i nie stanowi dowodu na efekt akulturacji. Ponadto, dokonałem analizy siły dowodu (a raczej braku takiego dowodu) na rzecz braku efektu używając statystyk Bayesowskich (Domurat i Białek, 2017) oraz metody TOSTER (Lakens, Scheel, i Isager, 2018). Ponadto zdiagnozowałem poważny problem metodologiczny – autorki analizują nowy efekt w zmienionych warunkach, nie replikując go jednak w warunkach kontrolnych. To znaczy, autorki nie dostarczają dowodów, że ich metodologia wywołałaby efekt języka obcego u osób, które nie zostały zakulturowane. Nawet zatem jeśli przyjąć, że autorki dostarczają dowodu, że w ich badaniu moralny efekt języka obcego nie występuje, to ten sposób trudno jednoznacznie określić czy badanie wykazuje efekt akulturacji, czy jest po prostu nieudaną replikacją wcześniej raportowanego efektu. W świetle moich własnych, opublikowanych badań (Muda i inni, 2018) oraz badań innych autorów, w których także nie zaobserwowano zmiany wyborów moralnych (Hayakawa i inni, 2017), można przypuszczać, że efekt języka obcego nie zawsze jest obserwowalny i być może zależy od wyboru konkretnych problemów moralnych lub innych, niewyjaśnionych dotychczas czynników.

### **5. Omówienie pozostałych osiągnięć naukowo – badawczych.**

#### **5.1. Pozostałe publikacje naukowe o zasięgu międzynarodowym**

##### **5.1.1. Redukcja błędów w wyborach międzyokresowych i ryzykownych poprzez przyjęcie perspektywy eksperta**

Ludzie charakteryzują się awersją do ryzyka oraz niechęcią do czekania na większą wypłatę. W efekcie często dokonują nieracjonalnych ekonomicznie decyzji, np. wolą pewne 4 złote niż 50% szans na uzyskanie 10 złotych. Proszeni o ocenę jak takich wyborów dokonałby racjonalny i uczciwy doradca badani zamiast (jak oczekiwaliśmy w oparciu o literaturę) dokonywać bardziej racjonalnych wyborów, kierowali się

 21

stereotypami na temat społecznie akceptowanych wartości – unikania ryzyka oraz cierpliwości. W efekcie, gdy badani przyjmowali perspektywę eksperta, ich impulsywność w wyborach międzyokresowych spadała ale awersja do ryzyka wzrosła. Przyjmowanie perspektywy eksperta nie wydaje się zatem dobrą metodą „prostowania” błędów poznawczych w decyzjach ekonomicznych.

**Białek, M., & Sawicki, P. (2014).** Can taking the perspective of an expert debias human decisions? The case of risky and delayed gains. *Frontiers in Psychology, 5*, 989. [IF(2014) = 2.560, pięcioletni IF = 3.039, Liczba punktów wg aktualnego wykazu MNiSW: 35]

### 5.1.2. Alternatywna metoda pomiaru preferencji w odroczonej loterii

Pomiar preferencji w wyborach z ryzykiem odroczonej w czasie jest niezwykle trudny metodologicznie, a stosowane metody często niezadowolające (Green i Myerson, 2004). Zaproponowaliśmy zastosowanie alternatywnej metody (conjoint) aby wyliczyć bardziej stabilne w czasie i bardziej predyktywne preferencje badanych. Metoda conjoint polega na wyborze najatrakcyjniejszej i najmniej atrakcyjnej z przedstawionych kilku loterii A, B, C i D. Gdy badany wybierze A jako najlepszą i D jako najmniej atrakcyjną, to uzyskany w ten sposób zestaw pozwala określić cały szereg innych zależności między opcjami, np. ( $A > B$ ,  $A > C$ ,  $A > D$ , oraz  $B > D$  i  $C > D$ ). Inną zaletą takich wyborów jest, że ludzie łatwiej określają swoje preferencje na skali nominalnej (co wole bardziej), niż na skali ilorazowej (o ile bardziej lubię dana opcję). Zaproponowana metoda wydaje się spełniać zakładane w niej nadzieje i generować stabilne w czasie i trafne oceny preferencji w odroczonej loterii.

**Białek, M., Markiewicz, Ł., & Sawicki, P. (2015).** Introducing conjoint analysis method into delayed lotteries studies: its validity and time stability are higher than in adjusting. *Frontiers in psychology, 6*, 23. [IF(2014) = 2.463, pięcioletni IF = 2.885, Liczba punktów wg aktualnego wykazu MNiSW: 35]

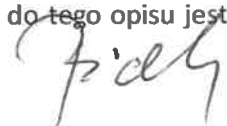
### 5.1.3. Zastosowanie analizy bayesowskiej do dowodzenia hipotezy zerowej

Logika testów statystycznych opierających się na testowaniu hipotezy zerowej pozwala na określenie na ile nieprawdopodobne jest uzyskanie wyników takie jak zebrane przy założeniu, że tak naprawdę hipoteza zerowa zakładająca brak związku między zmiennymi jest prawdziwa. Ten sposób wnioskowania nie pozwala jednak dowodzić prawdziwości hipotezy zerowej, a jedynie ją obalać lub nie. Alternatywnym podejściem statystycznym jest zastosowanie Bayes Factor. Stanowi on numeryczny wskaźnik na ile bardziej prawdopodobna jest jedna z dwóch porównywanych hipotez, zatem może skutecznie dowodzić prawdziwości zarówno hipotezy zerowej, jak i hipotezy alternatywnej. Artykuł przedstawia re-analizę opublikowanych przez innych autorów danych empirycznych, pokazując przewagę stosowania Bayes Factora nad testowaniem hipotezy zerowej.

**Domurat, A., & Białek, M. (2016).** Dowodzenie hipotez za pomocą czynnika bayesowskiego (bayes factor): przykłady użycia w badaniach empirycznych. *Decyzje, 26*, 109-141. [Lista B, Liczba punktów wg aktualnego wykazu MNiSW: 11]

### 5.1.4. Proporcja podstawowa wpływa na percepcję stereotypów

Liczne badania wskazują, że ludzie preferują informacje mniej diagnostyczne (np. stereotypy) nad bardziej diagnostyczne dane statystyczne (Kahneman i Tversky, 1973). Zjawisko to wyjaśniane jest jako poleganie na procesach Typu 1 (stereotypy), kosztem procesów Typu 2 (statystyka). Klasycznie, pyta się badanych o to, na ile prawdopodobne jest że losowa osoba z grupy 995 prawników i 5 inżynierów jest inżynierem, gdy jej opis pasuje do stereotypu inżyniera. Wykazałem, że badani oceniają prawdopodobieństwo starając się integrować oba źródła informacji, i że robią to intuicyjnie. Ponadto, pokazałem że skład grupy (proporcja podstawowa) nie tylko nie jest ignorowana, ale wpływa nawet na niepowiązane sądy jak dopasowanie opisu danej osoby do stereotypu: jest ono tym lepsze im więcej osób pasujących do tego opisu jest w grupie.



**Białek, M. (2017).** Not that neglected! Base rates influence related and unrelated judgments. *Acta Psychologica*, 177, 10-16. [IF(2017) = 1.862, pięcioletni IF = 2.223, Liczba punktów wg aktualnego wykazu MNiSW: 30]

#### **5.1.5. Test myślenia refleksyjnego (CRT) jest trafny nawet wśród badanych którzy już go znają**

Test myślenia refleksyjnego (CRT) jest popularnym testem oceniającym skłonność do refleksji vs. polegania na intuicji. Test składa się z trzech zadań matematycznych, które mają intuicyjnie oczywistą, ale niepoprawną odpowiedź. Badacze sugerują (i znalazło to potwierdzenie w naszej metaanalizie ponad 20 publikacji), że badani, którzy znają to narzędzie mają w nim wyższy wynik (Haigh, 2016). Nasza analiza danych sugeruje jednak, że nawet ten wyższy wynik CRT koreluje identycznie z innymi zmiennymi, zatem trafność CRT nie jest w żaden sposób zaburzona jego poprzednią znajomością przez badanych.

**Białek, M., & Pennycook, G. (2018).** The Cognitive Reflection Test is robust to multiple exposures. *Behavior Research Methods*, 50, 1953-1959. [IF(2017) = 3.597, pięcioletni IF = 4.885, Liczba punktów wg aktualnego wykazu MNiSW: 45]

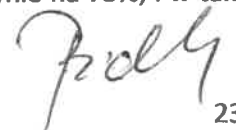
#### **5.1.6. Test myślenia refleksyjnego (CRT) przewiduje wybory międzyokresowe oraz ich stabilność na wskazówki kontekstualne**

Impulsywność w wyborach oraz w sposobie myślenia stanowią dwa odrębne, choć potencjalnie skorelowane konstrukty psychologiczne. W serii dwóch eksperymentów wskazaliśmy, że osoby o wyższym wyniku CRT są mniej impulsywne, oraz mniej podatne na narzucony punkt odniesienia. Punkt odniesienia (*reference point*) to taka opcja wyboru do której porównywana jest inna opcja, np. w problemie „Ile oczekujesz w zamian by poczekać na wypłatę miesiąc, zamiast żeby dostać 200 zł dziś” punktem odniesienia jest opcja natychmiastowej. Badania wskazują, że ludzie mają preferencję do opcji która jest punktem referencyjnym, i dyskontują silniej w scenariuszu odrzucania zysków niż przybliżania zysków (Appelt, Hardisty & Weber, 2011). W świetle naszych badań efekt ten występuje głównie u osób o nieskorefleksyjnym stylu poznawczym.

**Białek, M., & Sawicki, P. (2018).** Cognitive Reflection Effects on Time Discounting. *Journal of Individual Differences*, 39(2), 99-106. [IF(2017) = 1.283, pięcioletni IF = 1.730, Liczba punktów wg aktualnego wykazu MNiSW: 20]

#### **5.1.7. Replikacje mogą powodować powątpiewanie w postęp nauki**

Od 2011 roku trwa tak zwany kryzys replikacyjny, w którym psychologii zarzuca się niską replikowalność publikowanych badań spowodowanych selektywną publikacją eksperymentów które „zadziałały” oraz stosowaniem wątpliwych etycznie metod zbiórki i analizy danych. Propozycja Zwaana i innych (2018) by replikacje stały się „manistreamem” w czasopiśmie niesie ze sobą niestety pewne negatywne konsekwencje. Argumentuję, że każde badanie charakteryzuje się mocą mniejszą niż 1, zatem zawsze istnieje tylko pewne prawdopodobieństwo uzyskania istotnego statystycznie wyniku gdy faktycznie badany efekt jest prawdziwy. Nawet przy braku jakichkolwiek innych błędów w projekcie replikacyjnym w efekcie doprowadzi do pewnej liczby nieudanych replikacji prawdziwych efektów psychologicznych. Jest to duże zagrożenie dla badaczy (skonsternowanych w co wierzyć, a w co nie), oraz przede wszystkim dla praktyków i laików. Ci ostatni oczekiwać mogą od psychologii propozycji skutecznych rozwiązań problemów oraz twardych faktów znanych z innych dyscyplin jak matematyka i fizyka. Komunikacja stopnia (nie) pewności co do publikowanych odkryć psychologicznych może zapobiec w pewnym stopniu spadkowi zaufania do psychologii w świetle nieudanych replikacji. Dla przykładu, wynik testu statystycznego z  $p = 0.05$  najczęściej odpowiada Bayes Factor = 3, czyli temu że hipoteza testowana jest 3 razy bardziej prawdopodobna niż hipoteza zerowa. Odkrycie jest zatem pewne jedynie na 75%, i w taki sposób można komunikować wyniki badań psychologicznych.



**Białek, M.** (2018). Replications can cause distorted belief in scientific progress. *Behavioral and Brain Sciences*, 41, e122. [IF(2017) = 15.071, pięcioletni IF = 21.181, Liczba punktów wg aktualnego wykazu MNiSW: 50]

### **5.1.8. Analityczny styl poznawczy, a nie zdolności poznawcze, przewidują nadmierną pewność siebie**

W tej pracy wykazałem błędy w interpretacji danych empirycznych przez autora, który błędnie wywnioskował że za nadmierną pewność siebie odpowiadają niskie zdolności poznawcze. Re-analiza danych zebranych przez autora pozwoliła wykazać, że tak na prawdę kluczowe w nadmiernej pewności siebie jest poleganie na intuicji i brak refleksji nad własnymi sądami.

**Białek, M., & Domurat, A.** (2018). Cognitive abilities, analytic cognitive style and overconfidence: a commentary on Duttler (2016). *Bulletin of Economic Research*, 70(1), E119-E125. [IF(2017) = 0.264, pięcioletni IF = 0.604, Liczba punktów wg aktualnego wykazu MNiSW: 15]

### **5.1.9. Efekt utopionych kosztów w medycynie**

Lekarze często, mimo negatywnych efektów, kontynuują raz zaordynowaną terapię. Dzieje się to, mimo że gdyby stawiali mu diagnozę na nowo nie zaordynowaliby oni danej terapii pacjentowi. Wilkinson i inni (2018) twierdzą, że jest to przykład nowego błędu poznawczego – awersji do wycofywania się. Argumentuję, że zjawisko to wynika z eskalacji zaangażowania wynikającego z efektu utopionych kosztów. Ponadto, wraz ze współpracownikami dokonuję krytycznej ewaluacji proponowanych przez Wilkinsona mechanizmów minimalizacji omawianego zjawiska, oraz proponuję własne rozwiązania wynikające z badań nad utopionymi kosztami.

**Turpin, M., Meyers, E., Fugelsang, J., Friedman, O., & Białek, M.** (2019). Sunk cost and withdrawal aversion. *American Journal of Bioethics*, 19(2), 8-9. [IF(2017) = 4.847, pięcioletni IF = 5.597, Liczba punktów wg aktualnego wykazu MNiSW: 50]

## **5.2. Opieka nad młodymi badaczami**

Jednym z moich osiągnięć jest efektywna opieka nad młodymi osobami zaczynającymi karierę naukową. Przez rok przed wyjazdem do Kanady opiekowałem się kołem naukowym na ALK. Poza wypromowaniem kilkunastu licencjatantów i magistrantów na ALK, University of Waterloo i Uniwersytecie Warszawskim, byłem też współautorem sześciu pierwszych publikacji osób zaczynających karierę naukową. Dwie z tych publikacji znajdują się na liście JCR i ocenione są na 45 i 50 punktów MNiSW (w druku: *American Journal of Bioethics* oraz *Psychological Science*). Część z tych publikacji nie jest wykazana w osiągnięciach, gdyż nie udało mi się zdobyć od nich oświadczeń o wkładzie współautorów (dwie publikacje w *Psychologii Ekonomicznej*, jedna w *Polish Psychological Bulletin* i w *Przeglądzie Badań Edukacyjnych*). Dodatkowo, jestem opiekunem naukowym grantu Preludium przyznany magistrowi Rafałowi Mudzie (UMCS). Uważam, że jestem w stanie pomóc innym, początkującym naukowcom w pierwszych krokach w akademii za swoje duże osiągnięcie.

## **5.3. Zaangażowanie w Otwartą Naukę**

Otwarta Nauka stanowi dla mnie niezwykłą wartość. Uważam, że nauka powinna być dostępna dla wszystkich zainteresowanych praktyków, a dane i materiały powinny być dostępne bez ograniczeń dla wszystkich zainteresowanych badaczy. Postulaty Otwartej Nauki wcielam w życie publikując najczęściej jak się da w otwartych czasopismach (Open Access), a mój profil w Open Science Framework zawiera wersje autorskie oraz dane i materiały wszystkich publikacji po 2017 roku. Ponadto, w ramach promocji otwartej nauki i replikowalności badań uczestniczyłem w projekcie masowych replikacji Many Labs 5, oraz w projekcie Pipeline Project. Ten drugi koncept badań wpływ wyboru procedury eksperymentalnej (oraz

*Prace* 24

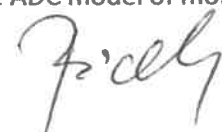


metody analizy danych) spośród kilku równoważnych i potencjalnie o podobnej jakości na uzyskane wyniki. Jestem członkiem oraz recenzentem dla Psychological Science Accelerator, instytucji mającej na celu organizację wielozespołowych, międzynarodowych projektów badawczych.

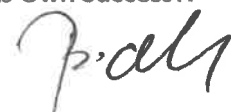
Ponieważ nauka powinna być samokorygująca, uważam że jednym z moich obowiązków jest podejmowanie aktywnej debaty nad najnowszymi badaniami. Robię to, często publikując komentarze do publikacji naukowych innych autorów. Uważam, że „kto nie milczy ten przyzwala” i ignorowanie błędów w publikacjach innych badaczy skutkuje trudnością dla innych naukowców w konstruowaniu spójnych teorii i przewidywań. Uważam, że Otwarta Nauka jest kierunkiem, w jakim powinna iść polska i światowa nauka, a przyjęcie Planu S przez rządy niektórych krajów Unii Europejskiej jest znakomitą krokiem w tę stronę.

### Literatura cytowana

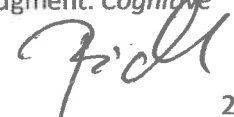
- Appelt, K. C., Hardisty, D. J., & Weber, E. U. (2011). Asymmetric discounting of gains and losses: A query theory account. *Journal of Risk and Uncertainty*, 43(2), 107-126.
- Białek, M., (2015) Przegląd badań współczesnej kognitywistyki nad efektem przekonania. *Przegląd Filozoficzny. Nowa seria*, 95, 91-107.
- Białek, M. (2017). Not that neglected! Base rates influence related and unrelated judgments. *Acta Psychologica*, 177, 10-16.
- Białek, M. (2017). Mechanika moralności. dylematy moralne i intuicyjne rozumienie dobra i zła. *Czasopismo Psychologiczne*, 23(1), 9-19.
- Białek, M. (2018). Replications can cause distorted belief in scientific progress. *Behavioral and Brain Sciences*, 41, e122
- Białek, M., & De Neys, W. (2016). Conflict detection during moral decision-making: Evidence for deontic reasoners' utilitarian sensitivity. *Journal of Cognitive Psychology*, 28(5), 631-639.
- Białek, M., & De Neys, W. (2017). Dual processes and moral conflict: Evidence for deontological reasoners' intuitive utilitarian sensitivity. *Judgment and Decision Making*, 12(2), 148-167.
- Białek, M., & Domurat, A. (2018). Cognitive abilities, analytic cognitive style and overconfidence: a commentary on Duttler (2016). *Bulletin of Economic Research*, 70(1), E119-E125.
- Białek, M., & Fugelsang, J. (2019) No evidence for decreased foreign language effect in highly proficient and acculturated bilinguals. A commentary on Čavar and Tytus (2018). *Journal of Multilingual and Multicultural Development*, advance online publication
- Białek, M., Fugelsang, J., & Friedman, O. (2018). Choosing victims: Human fungibility in moral decision-making. *Judgment and Decision Making*, 13(5), 451-458.
- Białek, M., Markiewicz, Ł., & Sawicki, P. (2015). Introducing conjoint analysis method into delayed lotteries studies: its validity and time stability are higher than in adjusting. *Frontiers in psychology*, 6, 23.
- Białek, M., & Pennycook, G. (2018). The Cognitive Reflection Test is robust to multiple exposures. *Behavior Research Methods*, 50, 1953-1959.
- Białek, M., & Sawicki, P. (2014). Can taking the perspective of an expert debias human decisions? The case of risky and delayed gains. *Frontiers in Psychology*, 5, 989
- Białek, M., & Sawicki, P. (2018). Cognitive Reflection Effects on Time Discounting. *Journal of Individual Differences*, 39(2), 99-106.
- Białek, M., & Terbeck, S. (2016). Can cognitive psychological research on reasoning enhance the discussion around moral judgments? *Cognitive Processing*, 17(3), 329-335.
- Białek, M., Terbeck, S., & Handley, S. (2014). Cognitive psychological support for the ADC model of moral judgment. *AJOB Neuroscience*, 5(4), 21-23.

 25

- Białek, M., Turpin, M., & Fugelsang, J. (w druku). What is the right question for moral psychology to answer? *Psychological Science*.
- Bostyn, D. H., Sevenhant, S., & Roets, A. (2018). Of Mice, Men, and Trolleys: Hypothetical Judgment Versus Real-Life Behavior in Trolley-Style Moral Dilemmas. *Psychological Science*, 0956797617752640.
- Čavar, F., & Tytus, A. E. (2018). Moral judgement and foreign language effect: when the foreign language becomes the second language. *Journal of Multilingual and Multicultural Development*, 39(1), 17-28.
- Cipolletti, H., McFarlane, S., & Weissglass, C. (2016). The moral foreign-language effect. *Philosophical Psychology*, 29(1), 23-40.
- Cohen, D. J., & Ahn, M. (2016). A subjective utilitarian theory of moral judgment. *Journal of Experimental Psychology: General*, 145(10), 1359-1381.
- Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: a process dissociation approach. *Journal of Personality and Social Psychology*, 104(2), 216-235.
- De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, 20(2), 169-187.
- De Neys, W., & Białek, M. (2017). Dual processes and conflict during moral and logical reasoning: A case for utilitarian intuitions? W: B. Trémolière & J. F. Bonnefon (Red.), *Moral Inferences* (pp. 123–136). Hove, UK: Psychology Press
- Domurat, A., & Białek, M. (2016). Dowodzenie hipotez za pomocą zynnika bayesowskiego (bayes factor): przykłady użycia w badaniach empirycznych. *Decyzje*, (26), 109-141.
- Evans, J. S. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255-278.
- Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3), 223-241.
- Farsides, T., Sparks, P., & Jessop, D. (2018). Self-reported reasons for moral decisions. *Thinking & Reasoning*, 24(1), 1-20.
- Fiske, S. T., & Taylor, S. E. (1991). *Social Cognition*, (Wyd. 2). New York: McGraw-Hill.
- Fiske, S. T., & Taylor, S. E. (2013). *Social cognition: From brains to culture*. Sage.
- Gawinkowska, M., Paradowski, M. B., & Bilewicz, M. (2013). Second language as an exemptor from sociocultural norms. Emotion-related language choice revisited. *PloS one*, 8(12), e81225.
- Geipel, J., Hadjichristidis, C., & Surian, L. (2015). How foreign language shapes moral judgment. *Journal of Experimental Social Psychology*, 59, 8-17.
- Geipel, J., Hadjichristidis, C., & Surian, L. (2016). Foreign language affects the contribution of intentions and outcomes to moral judgment. *Cognition*, 154, 34-39.
- Green, L., & Myerson, J. (2004). A Discounting Framework for Choice With Delayed and Probabilistic Rewards. *Psychological Bulletin*, 130(5), 769-792.
- Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, 11(8), 322-323.
- 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.
- Gürçay, B., & Baron, J. (2017). Challenges for the sequential two-system model of moral judgement. *Thinking & Reasoning*, 23(1), 49-80.
- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316(5827), 998-1002.
- Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog?. *Journal of Personality and Social Psychology*, 65(4), 613-628.
- Haigh, M. (2016). Has the Standard Cognitive Reflection Test Become a Victim of Its Own Success?. *Advances in Cognitive Psychology*, 12(3), 145-149.



- Hayakawa, S., Costa, A., Foucart, A., & Keysar, B. (2016). Using a foreign language changes our choices. *Trends in Cognitive Sciences*, 20(11), 791-793.
- Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., & Keysar, B. (2017). Thinking more or feeling less? Explaining the foreign-language effect on moral judgment. *Psychological Science*, 28(10), 1387-1397.
- Kahane, G. (2015). Sidetracked by trolleys: Why sacrificial moral dilemmas tell us little (or nothing) about utilitarian judgment. *Social Neuroscience*, 10(5), 551-560.
- Kahane, G., Everett, J. A., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond Sacrificial Harm: A Two-Dimensional Model of Utilitarian Psychology. *Psychological Review*, 125(2), 131-164.
- Kahane, G., Everett, J. A., Earp, B. D., Farias, M., & Savulescu, J. (2015). 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition*, 134, 193-209.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237-251.
- Koenigs, M., Kruepke, M., Zeier, J., & Newman, J. P. (2011). Utilitarian moral judgment in psychopathy. *Social Cognitive and Affective Neuroscience*, 7(6), 708-714.
- Körner, A., & Volk, S. (2014). Concrete and abstract ways to deontology: Cognitive capacity moderates construal level effects on moral judgments. *Journal of Experimental Social Psychology*, 55, 139-145.
- Kuhberger, A. (2002). The rationality of risky decisions: A changing message. *Theory & Psychology*, 12(4), 427-452.
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 2515245918770963.
- Lammers, J., Galinsky, A. D., Dubois, D., & Rucker, D. D. (2015). Power and morality. *Current Opinion in Psychology*, 6, 15-19.
- Lickel, B., Hamilton, D. L., Wierzchowska, G., Lewis, A., Sherman, S. J., & Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology*, 78(2), 223-246.
- McPhetres, J., Conway, P., Hughes, J. S., & Zuckerman, M. (2018). Reflecting on God's will: Reflective processing contributes to religious peoples' deontological dilemma responses. *Journal of Experimental Social Psychology*, 79, 301-314.
- Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience*, 6(10), 799-809.
- Muda, R., Niszczota, P., Białek, M., & Conway, P. (2018). Reading dilemmas in a foreign language reduces both deontological and utilitarian response tendencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(2), 321-326.
- Osman, M., & Wiegmann, A. (2017). Explaining moral behavior: A minimal moral model. *Experimental Psychology*, 64(2), 68.
- Patil, I., & Silani, G. (2014). Alexithymia increases moral acceptability of accidental harms. *Journal of Cognitive Psychology*, 26(5), 597-614.
- Patil, I., & Trémolière, B. (2018). I think, therefore I forgive: The role of reasoning in forgiving accidental harms. *PsyArxiv*, doi: 10.31234/osf.io/8gjt
- Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2014). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Social neuroscience*, 9(1), 94-107.
- Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, 36(1), 163-177.



- Pennycook, G., Trippas, D., Handley, S. J., & Thompson, V. A. (2014). Base rates: Both neglected and intuitive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2), 544.
- Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, 15(2), 165-184.
- Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought*. Yale University Press.
- Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, 119(3), 454-458.
- Terbeck, S., Kahane, G., McTavish, S., Savulescu, J., Cowen, P. J., & Hewstone, M. (2012). Propranolol reduces implicit negative racial bias. *Psychopharmacology*, 222(3), 419-424.
- Thaler, R. H. (1990). Anomalies: Saving, fungibility, and mental accounts. *Journal of Economic Perspectives*, 4(1), 193-205.
- Thompson, V. A., Turner, J. A. P., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, 63(3), 107-140.
- Tomaszewski, T. (1975). Człowiek w sytuacji. *Psychologia*, 17-36.
- Trémolière, B., & Bonnefon, J. F. (2014). Efficient kill-save ratios ease up the cognitive demands on counterintuitive moral utilitarianism. *Personality and Social Psychology Bulletin*, 40(7), 923-930.
- Turpin, M., Meyers, E., Fugelsang, J., Friedman, O., Białek, M. (2019). Sunk cost and withdrawal aversion. *American Journal of Bioethics*, 19(2), 38-39.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297-323.
- Waldmann, M. R., & Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science*, 18(3), 247-253.
- Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, 16(10), 780-784.
- Wilkinson, D., Butcherine, E., & Savulescu, J. (2019). Withdrawal Aversion and the Equivalence Test. *American Journal of Bioethics*, 19(2), 16-32.
- Wojciszke, B. (2011). Psychologia władzy. *Nauka*, 2, 51-69.
- Żukowski, T. (2015). Autowizerunek po katastrofie. Zofia Kossak-Szczucka i Jerzy Andrzejewski: dwa polskie świadectwa Zagłady z lat 40. *Poznańskie Studia Polonistyczne. Seria Literacka*, (25), 165-186.
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, 41, 1-50.

*Andreas P. Bailly*