

MACIEJ TARNOWSKI*

CZY POSIADANIE SPRZECZNYCH PRZEKONAŃ JEST MOŻLIWE? OMÓWIENIE I KRYTYKA ARGUMENTÓW ZA PSYCHOLOGICZNĄ ZASADĄ NIESPRZECZNOŚCI¹

STRESZCZENIE: Celem tego tekstu jest rekonstrukcja i analiza argumentów przedstawianych za Psychologiczną Zasadą Niesprzeczności (PZN), stwierdzającą, że żaden podmiot nie może mieć sprzecznych przekonań lub być opisany jako posiadający sprzeczne przekonania. Poprzez rozróżnienie dwóch możliwych interpretacji PZN, deskryptywnej i normatywnej, oraz dokładne zbadanie argumentacji przedstawionej dla każdej z nich z osobna, wskazuję zawarte w nich błędy oraz problemy związane z uzgodnieniem ich z wynikami badań prowadzonych w psychologii poznawczej i klinicznej. Uzasadniam, dlaczego PZN nie może być wyprowadzona z żadnego ze stanowisk metafizycznych dotyczących nastawień sądzeniowych i że posiadanie sprzecznych przekonań powinno być uznane za możliwe. Następnie piszę, dlaczego zinterpretowanie niektórych podmiotów jako posiadających sprzeczne przekonania może być bardziej efektywne w wyjaśnianiu przypadków nieracjonalnego zachowania niż rozwiązania zgodne z PZN.

SŁOWA KLUCZOWE: zasada niesprzeczności, racjonalność, zasada życzliwości, strategia intencjonalna, przypisywanie przekonań, urojenia, psychologia potoczna.

* Uniwersytet Warszawski, Wydział Filozofii i Socjologii. E-mail: m.tarnowski3@student.uw.edu.pl. ORCID: 0000-0003-3824-4134.

¹ Chciałbym podziękować za cenne uwagi formułowane podczas pisania niniejszego tekstu Joannie Komorowskiej-Mach i Tadeuszowi Ciecierskiemu oraz anonimowym recenzentom lub recenzentkom za wnikliwą lekturę i zwrócenie mojej uwagi na kilka spraw, które nie zostały poruszone we wcześniejszej wersji tego artykułu.

Logiczna zasada niesprzeczności, zapisywana w klasycznym rachunku zdań jako $\neg(p \wedge \neg p)$, jest jednym z najbardziej stabilnych, podstawowych i w oczywisty sposób prawdziwych praw, obowiązujących w każdym procesie wnioskowania. Odrzucając je, w znaczącej większości systemów logicznych jesteśmy skazani na działanie tzw. zasady eksplozji (inaczej: zasady *ex contradictione quodlibet*), brzmiącej: ze sprzeczności wynika wszystko / można wydedukować wszystko. Wówczas w naszym systemie wnioskowania możliwe jest udowodnienie dowolnego twierdzenia, przez co staje się on bezużyteczny z praktycznego punktu widzenia i nie pozwala odróżnić twierdzeń prawdziwych od fałszywych. Zarówno logika klasyczna, jak i wiele spośród powszechnie używanych nieklasycznych systemów logicznych, uznają logiczną zasadę niesprzeczności.

Czy jednak zasada niesprzeczności obejmuje też nasze przekonania? Wielu filozofów, pomimo początkowej kontrintuicyjności, udzielało na to pytanie odpowiedzi pozytywnej. Celem tego artykułu jest dokładna analiza argumentacji popierającej tę tezę, określaną tutaj jako *psychologiczna zasada niesprzeczności*, przedstawienie rozbudowanej krytyki tego poglądu oraz wykazanie, że jego przyjęcie jest nieuzasadnione.

Twórcą psychologicznej zasady niesprzeczności i pierwszym argumentującym za nią filozofem był Arystoteles. W księdze IV *Metafizyki* znajdujemy następujący ustęp: „Niemożliwością istotnie jest uważać, że tym samym jest być i nie być, jak to zdaniem niektórych twierdził Heraklit. Nie musi przecież mówiący być przekonany o tym, co mówi” (*Met.*, 1005b)². Łukasiewicz (1987) uznał to sformułowanie za odrębne od pozostałych, na które składają się zasada ontologiczna (w późniejszych opracowaniach traktowana również jako zasada metalogiczna) i logiczna, nazywając je właśnie psychologiczną zasadą niesprzeczności. W jego ujęciu arystotelesowską tezę należy sformułować jako: „Dwa przekonania, którym odpowiadają sądy sprzeczne, nie mogą istnieć razem w tym samym umyśle” (Łukasiewicz, 1987, s. 13).

Zasadę tę, na potrzeby tego artykułu, przeformułuję w następujący sposób:

(PZN) Nie jest możliwe, aby jeden podmiot posiadał zarazem dwa sprzeczne przekonania.

Przeformułowanie to ma na celu wyeliminowanie niejasności występujących w sformułowaniu Łukasiewicza. Łukasiewicz przyjmuje, że przekonania, będąc bytami psychicznymi, nie mogą być sprzeczne – dlatego używa pojęcia sądu logicznego. Obecnie jednak mówienie o sprzecznych przekonaniach weszło do codziennego filozoficznego użycia, zaś z samym pojęciem sądu logicznego wiąże się wiele problemów dotyczących statusu ontologicznego i logicznego sądów. To ujęcie wydaje mi się klarowniejsze. Także ze względu na zmianę

² Tłumaczenie zostało zaczerpnięte z: Arystoteles (1996). *Metafizyka*. Lublin: Redakcja Wydawnictw KUL.

w popularnie przyjmowanej terminologii wymieniam pojęcie umysłu na pojęcie podmiotu. Stwierdzenie „nie jest możliwe”, użyte zamiast pierwotnego „nie mogą”, wyraźniej niż w oryginale zaznacza modalny charakter tej tezy.

Istnieją, pomimo kontrintuicyjności, istotne i powracające w filozoficznej dyskusji argumenty, służące uzasadnieniu PZN. Szczególnym przypadkiem takiej argumentacji jest stanowisko klasycznego interpretacjonizmu (Quine, 1999; Davidson, 1974) i pochodnej wobec niego teorii, jaką jest koncepcja strategii intencjonalnej Daniela Dennetta (1978; 1981a). Co stanie się jasne w dalszym ciągu analizy, metoda argumentacji za przyjęciem PZN zależy od kontekstu teoretycznego, w którym osadzimy pojęcie przekonania obecne w uwzględnianym tutaj sformułowaniu. W szczególności liczyć się będzie przyjęcie realizmu bądź instrumentalizmu w odniesieniu do przekonań.

W odczytaniu realistycznym PZN pozostaje deskryptywna. Można ją wówczas rozumieć jak każde inne zdanie dotyczące wykluczających się nawzajem konkretnych zjawisk: „nie jest możliwe, żeby zarazem padało i nie padało”, „nie jest możliwe, żeby zarazem panowała noc i dzień” itp. Przyjęcie takiej tezy wspiera realistyczne ujęcie zbioru przekonań jako „mapy, którą się kierujemy” (Ramsey, 1931); model świata, z którego czerpiemy uzasadnienie dla naszych działań, ze swojej natury nie może być według takiej tezy sprzeczny. Psychologiczną zasadę niesprzeczności rozumianą deskryptywnie będę od tej pory oznaczał jako PZN-D.

Można też, uznając instrumentalną interpretację przekonań, odczytywać PZN w sposób normatywny, jako wynik przyjmowanej metodologii, a nie ontologii. Według instrumentalistycznej interpretacji warunkiem uznania jakiegoś podmiotu za posiadającego przekonania jest użyteczność jego opisu w takich kategoriach. Można zatem bronić PZN w sposób alternatywny: dwa sprzeczne przekonania nie mogą istnieć zarazem w jednym umyśle, ponieważ takiego opisu umysłu zabrania nam teoria, wewnątrz której jest zdefiniowane pojęcie przekonania. To, czy można przypisać podmiotowi sprzeczne przekonania, zależy więc nie tyle od faktów dotyczących umysłu (co postulowałiby realisci), co od tego, czy takie przypisanie wykazuje przydatność i czy jest dozwolone przez psychologię potoczną (lub inną prawdziwą teorię przypisywania przekonań).

Sformułowanie, którego bronią zwolennicy takiej interpretacji, określam jako psychologiczną zasadę niesprzeczności w odczytaniu normatywnym. Można ją w prostej formie zapisać w taki sposób:

(PZN-N) Podmiotowi w jednym momencie nie można przypisać wzajemnie sprzecznych przekonań.

W tym artykule pragnę zająć się rekonstrukcją dwóch strategii argumentacyjnych wykorzystywanych przy uzasadnieniu psychologicznej zasady niesprzeczności. Pierwszą z nich określam jako argumentację ontologiczną, drugą jako metodologiczną – odpowiednio służą one uzasadnieniu zasady w jej sformułowaniu deskryptywnym i normatywnym. Omówię pokrótce ich historię oraz wer-

sje proponowane przez różnych autorów. Następnie poddam te argumenty krytyce, wskazując na błędy rozumowania oraz podając przykłady z zakresu psychologii klinicznej i poznawczej.

Przed przystąpieniem do analizy tych argumentacji, należy wyraźnie zaznaczyć, iż obydwie interpretacje PZN logicznie się nie wykluczają. Realista, uznający prawdziwość psychologii potocznej, może akceptować je obie: twierdzić zarówno, że posiadanie sprzecznych przekonań jest metafizycznie niemożliwe, jak i że niemożliwe jest ich przypisanie podmiotowi z punktu widzenia psychologii potocznej. Poglądy na „metodologię” psychologii potocznej (czy może ona być traktowana podobnie jak teoria naukowa i czy zabrania ona przypisania sprzecznych przekonań) są związane z poglądami na metafizykę umysłu lub filozofię nauki w sposób bardziej konwencjonalny niż logiczny. Stąd wydaje się, że czasem twierdzenia wspierające drugie sformułowanie PZN mają status milczącej przesłanki. Rozróżnienie tych dwóch strategii i sposobów rozumienia PZN podczas analizy poprawności tej zasady ma zatem charakter kluczowy dla jasnego przedstawienia debaty.

1. ARGUMENT ONTOLOGICZNY Z WZAJEMNEGO WYKLUCZANIA SIĘ WŁASNOŚCI

Pierwszym autorem argumentującym za PZN jest jej twórca, czyli Arystoteles. W księdze IV *Metafizyki* poświęca jej dużo miejsca, podając za przyjęciem zasady niesprzeczności co najmniej kilka osobnych argumentów – dowodząc, że zasadę niesprzeczności należy potraktować jako jedno z podstawowych praw myślenia (Łukasiewicz, 1987). Łukasiewicz i Gottlieb wyodrębniają rozdziały 4., 5. i 6. IV księgi *Metafizyki* jako zawierające argumentację za przyjęciem zasady niesprzeczności w jej psychologicznej wersji (w sformułowaniu przytoczonym w podrozdziale 2.1). Powiązane z tą linią argumentacji fragmenty pojawiają się również w *Hermeneutyce* i *Analitykach wtórych* (Łukasiewicz, 1987; Gottlieb, 2007).

Arystoteles stara się wykazać, że psychologiczna zasada niesprzeczności jest możliwa do dowiedzenia na podstawie zasad ontologicznej i logicznej (Łukasiewicz, 1987). Pisze on:

Jeżeli niemożliwe, by przeciwieństwa zarazem występowały w tym samym (przy tej przesłance trzeba także uwzględnić zwyczajne dopowiedzenia), a opinie, którym odpowiadają sądy sprzeczne, są sobie przeciwne, to oczywiste, że jeden i ten sam człowiek nie może być przekonany, że tym samym dla tego samego jest być i nie być, gdyż w takim razie miałby zarazem dwa przekonania przeciwne. (*Met. IV*, 1005b)

Jak zauważa Łukasiewicz, problematyczne w odniesieniu do tego fragmentu jest zrównanie ze sobą pojęć „przeciwieństwa” i „sprzeczności”: jedno bowiem przysługuje parom własności, drugie – sądów i przekonań. Odwołując się do fragmentów *Hermeneutyki* (*Hermeneutyka*, 14, 23, 27–39) wskazuje, że Arystoteles, by uniknąć tej trudności, traktuje przekonania jako własności umysłu –

wówczas sprzeczne wzajemnie przekonania odpowiadają przeciwnym sobie własnościom. Zatem gdyby jeden człowiek wierzył, że coś jest jakieś i zarazem o tym samym, że takie nie jest – wierzyłby w dwa sprzeczne sądy – posiadałby przeciwne przekonania, a więc własności przeciwne, co przeczyłoby ontologicznemu sformułowaniu zasady niesprzeczności³:

(OZN) „Żaden przedmiot nie może tej samej cechy zarazem posiadać i nie posiadać” (Łukasiewicz, 1987, s. 10).

Można spróbować zrekonstruować ten argument w następujący sposób:

- (1) (OZN) Nie jest możliwe, aby jeden przedmiot zarazem posiadał i nie posiadał jakiegś własności.
- (2) Przekonania wyrażające sądy wzajemnie sprzeczne są przeciwnymi sobie własnościami.
- (3) Przeciwnie sobie własności wykluczają się wzajemnie.
- (4) Przekonania wyrażające sądy wzajemnie sprzeczne wykluczają się wzajemnie (z [2]–[3]).
- (5) Nie jest możliwe, aby jeden podmiot posiadał wykluczające się wzajemnie własności (z [1]).
- (6) (PZN) Nie jest możliwe, aby jeden podmiot posiadał przekonania wyrażające sądy wzajemnie sprzeczne (z [4]–[5]).

Istotne z punktu widzenia analizy argumentu ontologicznego Arystotelesa będzie przyjęcie przez niego przesłanek (2) i (3). Autor wyraźnie zaznacza, że ten argument zależy od uznania OZN, którą uznaje za pierwotną zasadę ontologiczną: kwestionowanie tej przesłanki nie trafia więc w sedno rozumowania Stagiryty.

Przesłanka (2) uzasadniona jest, według Łukasiewicza, w następujący sposób: ponieważ Arystoteles interpretuje przekonania jako własności umysłu, niezbędne dla uzasadnienia przyjęcia tej przesłanki jest zrównanie ze sobą dwóch pojęć, sprzeczności (w odniesieniu do sądów i przekonań je wyrażających) oraz przeciwieństwa własności. Arystoteles pisze w *Hermeneutyce*:

³ Łukasiewicz, podając swoją rekonstrukcję argumentu Arystotelesa, uznaje, że Stagiryta wywodzi PZN z zasady logicznej, której przedmiotem są sądy. Pisze on: „Z logicznej zasady sprzeczności wynika jednak, że żaden przedmiot nie może posiadać zarazem cech przeciwnych” (Łukasiewicz, 1987, s. 24). Ponieważ logiczna zasada niesprzeczności w rekonstrukcji Łukasiewicza uzasadnia pogląd, że „żaden przedmiot nie może posiadać zarazem cech przeciwnych”, wcześniej zaś Łukasiewicz uznaje równoważność zasady logicznej i ontologicznej, zakładam, że rekonstrukcja podana wyżej, oparta na OZN, jest również uprawniona.

Jeżeli więc tak jest w dziedzinie przekonań, a wypowiedane twierdzenia i przeczenia są znakami rzeczy w duszy, to jasne, że przeczenie ogólne jest przeciwne twierdzeniu o tej samej rzeczy. (*Herm.*, 14, 24b)⁴

Sprzeczność pojawiająca się między sądami, które są wyrażone w przekonaniach, będzie w ujęciu Arystotelesa odpowiadała przeciwieństwu własności. Pewną trudność sprawia rozważenie, czym w tej argumentacji jest właściwie przeciwieństwo: Łukasiewicz uznaje, że „przeciwymi są przekonania, którym odpowiada sąd twierdzący i przeczący o tym samym przedmiocie, np. «Kallias jest sprawiedliwy» – «Kallias nie jest sprawiedliwy»” (Łukasiewicz, 1987, s. 21).

Łukasiewicz poddaje powyższe rozumowanie w podobnej rekonstrukcji gruntownej krytyce. Jego uwaga skupia się na dwóch kwestiach: zrównaniu pojęcia sprzeczności i przeciwieństwa własności w odniesieniu do przekonań oraz nieuprawnione wymieszanie pojęć logicznych i psychologicznych w przesłance (2). W tym fragmencie zrekonstruuję krytykę Łukasiewicza, poddam ją analizie i wyciągnę wnioski dotyczące statusu Arystotelesowej argumentacji za PZN.

W analizie uzasadnienia przesłanki (2) przyjmuje on zarówno założenie o tym, że przekonania mogą być traktowane jako własności, jak i to, że własności mogą pozostawać ze sobą w relacji przeciwieństwa. Czym jednak dla Arystotelesa jest przeciwieństwo własności lub cech – i w konsekwencji przeciwieństwo przekonań? Częściową odpowiedź na to pytanie znajduje Łukasiewicz we fragmentach *Hermeneutyki*:

Należy przeto zbadać, który rodzaj prawdziwego sądu jest przeciwny fałszywemu sądowi: czy ten, który przeczy fałszywemu sądowi, czy ten, który orzeka przeciwieństwo? [...] Jeżeli więc dobro jest dobre, a nie złe, i jeżeli jedna własność przysługuje istotnie, a druga tylko akcydentalnie (bo dobro akcydentalnie nie jest złe), i jeżeli ten sąd jest bardziej prawdziwy, który dotyczy istoty rzeczy, to także i ten sąd fałszywy jest bardziej fałszywy, który dotyczy istoty rzeczy. [...] Tak że bardziej fałszywym sądem dotyczącym dobra będzie raczej ten, który zaprzecza dobro, niż ten, który orzeka jakąś przeciwną własność dobra. (*Hermeneutyka*, 14, 24b)

Według Łukasiewicza następuje tutaj nieuprawnione przejście z domeny logiki do domeny psychologii, widoczne zwłaszcza w innym fragmencie z *Hermeneutyki*, w którym Arystoteles stwierdza, że „przekonanie, że dobre jest złe, jest złożone, bo przypuszczalnie ten, kto tak sądzi, musi równocześnie sądzić, że to co jest dobre, nie jest dobre” (*Hermeneutyka*, 14, 23b). Łukasiewicz zwraca uwagę, że podobny stosunek (logicznej konsekwencji) zachodzi pomiędzy sądami, ale między przekonaniem nie jest konieczny. Jeśli bowiem traktujemy przekonania jako własności, nie możemy jednocześnie przypisywać im praw-

⁴ Tłumaczenie zostało zaczerpnięte z: Arystoteles (1990). *Hermeneutyka*. Warszawa: PWN.

dziwości bądź fałszywości – te przysługują jedynie sądom lub zdaniom. Mówienie o prawdziwości lub fałszywości przekonań jest sensowne o tyle, o ile odwołuje się do ich reprezentacji językowej bądź wyrażanego przez nie sądu. Jest to kłopot dla Arystotelesowskiego rozumowania: niemożliwe jest proste przełożenie prawdziwości lub fałszywości przekonania, że p w podanym wyżej sensie na jakąkolwiek charakterystykę własności umysłu, jaką byłoby „bycie przekonany, że p ”. Ponieważ własności nie są ani prawdziwe, ani fałszywe, nie mogą być również sprzeczne.

W związku z tym nawet jeśli potraktujemy przekonania jako własności umysłu, nie jesteśmy w stanie pokazać, że sprzeczność przekonań potraktowanych jako nastawienia sądzeniowe pociąga za sobą przeciwieństwo przekonań interpretowanych jako własności – nie można zatem wywieść PZN z OZN. Należy więc odrzucić argumentację Arystotelesa na rzecz PZN.

Innym poglądem, który wyprowadza PZN z charakterystyki przekonań jako własności, jest dyspozycjonalizm uznający przekonania za dyspozycje do przejawiania określonych zachowań. Reprezentatywną dla tego nurtu analizą jest ta przeprowadzona przez Ruth Barcan Marcus. Autorka w swojej analizie pojęcia przekonania redukuje je do dyspozycji „do zachowywania się, jak gdyby S , aktualny lub nieaktualny stan rzeczy, zachodził” (Barcan Marcus, 1983, s. 241). Co miałyby znaczyć zachowywanie się, jak gdyby nie obowiązywało prawo niesprzeczności? Barcan Marcus wydaje się podążać tutaj za poglądami Wittgensteina z *Traktatu logiczno-filozoficznego* dotyczącymi statusu poznawczego tautologii i sprzeczności⁵. Nie można zachowywać się np. jak gdyby padało i nie padało zarazem, gdyż nie istnieją warunki (świat możliwy), w których podobny (niemożliwy) stan rzeczy mógłby być przyczyną naszego zachowania. W związku z tym, że nie możemy scharakteryzować dyspozycji do zachowywania się, jak gdyby zachodziło p i $nie-p$, niemożliwym jest posiadanie dwóch sprzecznych przekonań.

Nie wchodząc w dyskusję ontologiczną dotyczącej statusu przekonań, uważam, że dyspozycjonalizm nie pociąga za sobą logicznie PZN. Sprzeczność przekonań nie przekłada się bowiem (z tych samych, co w Arystotelesowskim przypadku powodów) bezpośrednio na wykluczanie się wzajemnie dwóch dyspozycji. Wszelkie zachowanie, będące podstawą do przypisania przekonania, musi być zachowaniem pozytywnie świadczącym o posiadaniu takiego przekonania. Przekonanie, że $nie-p$, nie może być po prostu niezachowywaniem się, jak gdyby p było faktem – wówczas nieposiadanie przekonania, że p byłoby równoznaczne z posiadaniem przekonania, że $nie-p$. Jeśli zaś istnieją wzorce zachowania odpowiednie dla bycia przekonany, że p i bycia przekonany, że $nie-p$, konsekwencja, która zabrania nam przypisania przekonania, że p i $nie-p$, przy

⁵ „4.461 Zdania pokazują, co mówią: tautologia i sprzeczność pokazują, że nie mówią nic. Tautologia nie ma warunków prawdziwości, gdyż jest prawdziwa bezwarunkowo; a sprzeczność nie jest prawdziwa pod żadnym warunkiem. Tautologia i sprzeczność są bezsensowne [...]. (Nie wiem np. nic o pogodzie, gdy wiem tylko, że pada lub nie pada.)” (Wittgenstein, 2000a, s. 38).

braku świadectw na odrzucenie któregoś z przekonań przez podmiot, jest dogmatyczna. Przynajmniej intuicyjnie istnieją też sposoby, w których wzorzec zachowania może być wytłumaczony przez bycie przekonanym o pewnej niemożliwości, jak opisuje to Wittgenstein w *Uwagach o podstawach matematyki*⁶:

106. [...] coś mnie kusi, by powiedzieć: nie można wierzyć, że $13 \times 13 = 196$, liczbę tę można tylko mechanicznie przyjąć od kogoś innego. Dlaczego jednak nie mam powiedzieć, że w to wierzę? Czyżby wierzenie w to było jakimś tajemniczym aktem, który niejako mocą podziemnych związków łączy się z poprawnym obliczeniem? W każdym razie mogę powiedzieć: „wierzę w to”, i zgodnie z tym działać. (Wittgenstein, 2000b, s. 53)

1.1 Argument ontologiczny z funkcji umysłu

Istnieje jednak również pewna wersja argumentu ontologicznego która, choć rzadko formułowana *explicite*, wydaje się mieć status milczącej przesłanki w wielu argumentacjach za przyjęciem PZN, formułowanych przez współczesnych filozofów. Postaram się odnieść się do niej w nadziei, że wpłynie to na oczyszczenie debaty na polu metodologicznym – nawet jeśli argument w poniższej wersji nie jest jako taki przyjmowany przez żadnego z filozofów.

Ta argumentacja, choć odbiega znacząco od opisywanej wcześniej, należy do argumentacji ontologicznej w stosowanym tu rozróżnieniu, gdyż stara się wywieść PZN w jej wersji deskryptywnej: faktem jest, że jeden podmiot nie może posiadać sprzecznych przekonań. Dotyczy jednak ona dużo węższej grupy przypadków. W jej myśl podmiot nie może posiadać przekonań, co do których wie (lub sądzi), że zachodzi między nimi sprzeczność.

W *Logice* Wilfrida Hodgesa znaleźć możemy poniższe sformułowanie, prawdopodobnie najbliższe omawianej tu tezie:

Po prostu niemożliwym jest [...] być przekonanym o dwóch rzeczach, o których wiesz, że są sprzeczne ze sobą. Wydaje się, że jesteśmy zobligowani do bycia przekonanym jedynie o tym, co uważamy za niesprzeczne. (Hodges, 1977, s. 15)

Podobny fragment możemy znaleźć w artykule Willarda Quine'a i Johna Ulliana *The Web of Belief*:

Nie można być przekonanym o czymś, co uważa się za niemożliwe. [...] uznajemy to za podstawowy powód podjęcia procesu myślenia, analizy świadectw i rewizji systemu przekonań. Kiedy powstają konflikty, tworzące niemożliwe kombi-

⁶ Barcan Marcus interpretuje ten fragment jako wprowadzenie różnicy pomiędzy „byciem przekonanym”, a „twierdzeniem, że jest się przekonanym”. Jednak przyjmując dyspozycjonalistyczną analizę przekonań, to rozróżnienie jest arbitralne (gdy „twierdzenie, że jest się przekonanym” uznamy także za rodzaj zachowania); także sam Wittgenstein wydaje się być skłonny do jego odrzucenia, o ile je wcześniej przyjął (patrz: uwagi I-106 do I-119).

nacje, nie możemy ich zachować; musimy je rozwiązać. (Quine, Ullian, 1978, s. 37)

Sposoby argumentacji wspierające się na podobnych przesłankach wymagają przyjęcia pewnego założenia dotyczącego celu działania systemu formowania przekonań – jest nim mianowicie formowanie prawdziwych przekonań. Jak słusznie zauważa Ruth Barcan Marcus, komentując uwagi Hodgesa: „Czemu skupiać się na sprzeczności? Czy możliwe jest bycie przekonanym, że p , gdy wie się, że p jest fałszywe?” (Barcan Marcus, 1983, s. 145). Można zatem tezę Hodgesa i Quine’a, jakobyśmy byli „zobligowani do bycia przekonanym jedynie o tym, co uważamy za niesprzeczne”, zinterpretować w następujący sposób. Przyjmijmy, że naturalną skłonnością czy celem ludzkiego systemu poznawczego jest posiadanie prawdziwych przekonań: w wypadku posiadania informacji świadczącej przeciw pewnemu przekonaniu lub wprost mu przeczących – czy będą to dane empiryczne, czy też wynik wnioskowania dedukcyjnego – podmiot jest zmuszony odrzucić przekonanie, które okazuje się fałszywe. Tylko jeśli założymy, że celem naszego systemu poznawczego jest podtrzymywanie prawdziwych przekonań, możemy uznać, że miałyby on jakąś motywację do eliminacji przekonań fałszywych. Ponieważ zaś sprzeczność jest oczywistą oznaką fałszywości, nie można posiadać sprzecznych przekonań, będąc świadomym ich sprzeczności.

1.2 Krytyka argumentu z funkcji umysłu

Twierdzenie, że maksymalizacja ilości prawdziwych przekonań jest systemową funkcją naszego umysłu, wydaje się być intuicyjne, jednak rzadko było ono broniące wprost. Jedyną znaczącą próbą jego obrony jest argument odwołujący się do zasad selekcji naturalnej, jednak wydaje się, że wynika on z niedostatecznego rozważenia konkurencyjnych strategii ewolucyjnych⁷. Jeśli jednak nawet bez uzasadnienia mielibyśmy przyjąć to twierdzenie, wciąż można podać w wątpliwość, na ile twierdzenie to może pełnić jakąkolwiek rolę w uzasadnieniu PZN.

Po pierwsze, wydaje się, że argumentując w ten sposób za przyjęciem PZN, popadamy w błędne koło. Zasadę eliminacji fałszywych przekonań, którą przyjmujemy jako przesłankę w naszym rozumowaniu, można spróbować wyrazić w taki sposób:

(ZE) Jeśli podmiot posiada przekonanie, że p , i dowiaduje się, że p jest fałszywe, to pozbywa się przekonania, że p .

⁷ Rozbudowaną krytykę tego argumentu, którego najpełniejsze sformułowanie stanowi praca Dennetta (1978), można znaleźć u Sticha (1985) oraz u Lewisa i Coopera (1979).

ZE oznacza zatem, że wraz z nabyciem uzasadnionego, prawdziwego przekonania, że *p jest fałszywe*, podmiot musi się pozbyć przekonania, że *p*. Z ZE wynika więc bezpośrednio, że nie jest możliwe, aby podmiot zarazem podtrzymał przekonanie, że *p* i przekonanie, że *p jest fałszywe*. To drugie przekonanie nie wydaje się niczym różnić od przekonania, że *nie-p* – wówczas zaś w naszym rozumowaniu założyliśmy już PZN wśród przesłanek.

Innym wartym odnotowania w tym argumencie założeniem jest uznanie, że wewnętrzna sprzeczność jakiegoś przekonania (postaci *p* i *nie-p*) jest oczywistym dowodem na jego fałszywość. Należałoby się zastanowić, czym właściwie miałyby być stwierdzana w tym argumencie oczywista fałszywość sprzecznych zdań i przekonań. Nie każde zdanie wewnętrznie sprzeczne jest takie w sposób oczywisty: dowodem na to może być choćby historia matematyki, gdzie nieraz zdarzało się społeczności matematyków akceptować nieuprawnione dowody twierdzeń, które później okazywały się fałszywe. Większość z nas za w oczywisty sposób sprzeczne (i w związku z tym fałszywe) uzna pewne kontrtautologie rachunku zdań: zdania o postaci *p* i *nie-p* lub *nie-nie-p* i *nie-p*, jednak na pierwszy rzut oka nie powie tego samego np. o twierdzeniu geometrii algebraicznej „udowodnionym” przez Francesco Severiego w 1934 roku (a którego fałszywości w ogólnym przypadku dowiedziono 34 lata później). Gdzie więc zaczyna się oczywista sprzeczność jakiegoś zdania? Czy uznanie jakiegoś zdania za sprzeczne w naszym rozumieniu nie jest zatem tym samym, czym uznanie go za fałszywe (i w związku z tym zakłada PZN)? Czy dialetyści, tacy jak Graham Priest (którzy wierzą, że istnieją zdania zarazem prawdziwe i fałszywe), mylą się w swojej ocenie niektórych zdań, czy źle interpretują pojęcie sprzeczności?

Nie mam zamiaru odpowiedzieć tutaj na te pytania, lecz raczej zaznaczyć, że przeprowadzenie tego z pozoru niewinnego argumentu wymaga precyzyjnej odpowiedzi na każde z nich. Wydaje się raczej, że aby uznać sprzeczność za oczywisty dowód fałszywości, należy już działać i wnioskować w zgodzie z ZE, a zatem ta przesłanka również nie może brać udziału w uzasadnieniu PZN. Możliwe więc, że warunek dotyczący świadomości sprzeczności (tak jak zdajemy się ją rozumieć), który musi posiadać podmiot, zakłada już działanie w zgodzie z PZN – zatem nie może nam pomóc w jej uzasadnieniu⁸.

⁸ Inną jeszcze, choć również interesującą w tym kontekście grupą, są przypadki sprzecznych przekonań, których sprzeczności podmiot nie może spoznać nawet przy założeniu, że jest w pełni racjonalny. Przykładem takim będą tzw. zagadka Kripkego (Kripke, 1979) oraz problem Richarda (Richard, 1983): w obydwu przypadkach, mówiąc ogólnie, mamy, do czynienia z przekonaniem nabytym w izolowanych od siebie kontekstach językowych bądź epistemicznych, które jednak dotyczą tego samego przedmiotu, przedstawionego w tych kontekstach w różny sposób. Wówczas, jak wyjaśnia Kripke, „żadna doza czystej logiki bądź semantycznej introspekcji nie jest wystarczająca aby [podmiot] odkrył swój błąd” (Kripke, 1979, s. 451): spostrzeżenie sprzeczności przekonań przez podmiot może dokonać się jedynie na drodze nabycia nowego przekonania (np. „London = Londres” w przykładzie Kripkego, bądź „ty [osoba z którą rozmawiam przez

1.3 Problem aposterioryczności

Jak zostało wyżej pokazane, wskazane próby udowodnienia PZN w jej deskryptywnym odczytaniu zawodzą ze względu na nieuprawnione mieszanie pojęć logicznych i psychologicznych albo ciche przyjęcie konkluzji między przesłankami. Pragnąłbym jeszcze zwrócić uwagę na bardziej ogólny argument, który w podobnej formie wymierzony był przez Łukasiewicza przeciw samej PZN w jej Arystotelesowym ujęciu (Łukasiewicz, 1987, s. 30–34).

Jak zauważył Lewis Carroll (1995), próba uzasadnienia wnioskowania poprzez odwołanie się do samych reguł (czyli uzasadnienie wnioskowania logicznego za pomocą narzędzi czysto logicznych) prowadzi do regresu w nieskończoność. Oprócz zasad – uznania za prawdziwe określonych zdań lub formuł – potrzebujemy jeszcze przyjęcia metody wnioskowania. Jak ujmuje to Penelope Maddy:

Nie byłoby [podobnego problemu] gdybym podważała prawdziwość wszystkich aksjomatów ZFC, lecz gdy próbuję podważyć prawdziwość samej logiki, okazuje się, że nasze ustalone konwencje muszą być ogólne, a zatem pozostają bez swojej zamierzonej mocy, o ile logika nie jest już faktycznie dostępna – by zobaczyć, w jaki sposób poszczególne prawdy logiki pochodzą od ogólnych zasad. (Maddy, 2012, s. 495)

Tego, czy podmiot wnioskuje w zgodzie z zasadami logiki, nie da się w związku z tym ustalić poprzez odwołanie się do ogólnych praw formowania przekonań. Nawet gdybyśmy (hipotetycznie) odkryli w ludzkim umyśle reprezentację prawa niesprzeczności, to aby w ten sposób uzasadnić PZN, trzeba założyć, że podmiot wnioskuje logicznie, stosując ogólne prawo do poszczególnych jego przypadków. Podobnie musielibyśmy postąpić w wypadku zasady eliminacji lub innych psychologicznych praw dotyczących formowania przekonań. Żadna ogólna zasada psychologiczna również nie posiada więc wystarczającej siły, by dowieść PZN-D.

Jeśli zatem nie da się wywieść PZN-D z ogólnych praw formowania przekonań, jedyną możliwą formą uzasadnienia tej zasady w jej deskryptywnym odczytaniu jest zinterpretowanie jej jako dobrze potwierdzonej hipotezy empirycznej. Można to ująć jako problem aposterioryczności: PZN w jego wersji deskryptyw-

telefon] = on [osoba, którą widzę na ulicy]” w przykładzie Richarda) drogą doświadczenia. Jeśli argument wykorzystujący podane problemy jako dowód na możliwość posiadania sprzecznych przekonań jest poprawny (czego nie mam zamiaru w tym artykule rozstrzygać), to nie tylko (jak staram się pokazać wyżej) jest możliwe, że dla niektórych podmiotów sprzeczność przekonań może nie być oczywistym dowodem fałszywości, ale również istnieją pewne pary przekonań, których sprzeczności nigdy nie będziemy w stanie jedynie poprzez logiczną analizę sobie uświadomić, wówczas zaś nawet założenie silnej racjonalności podmiotu nie pozwoli nam wywieść PZN-D z jego osłabionej wersji.

nej nie może być uzasadnione apriorycznie, a jedynie jako wynik badań empirycznych.

Takie postawienie sprawy wydaje się dość problematyczne w filozoficznej dyskusji – nawet jeśli PZN byłaby dobrze potwierdzoną hipotezą, wielu filozofów uznałoby za niepożądane przyznawać taki status zasadzie, która wyjściowo została opisana jako jedna z naczelných reguł ludzkiego myślenia. Zakładana stu procentowa zgodność przebadanych przypadków z PZN nie świadczyłaby przeciw jeszcze o tym, że posiadanie sprzecznych przekonań jest niemożliwe.

Jednak jeśli taki byłby obecny stan badań psychologii, mógłby on w przekonujący sposób uzasadnić przyjęcie PZN lub przynajmniej wystarczająco ją uprawdopodobnić. Chciałbym, kończąc rozważania nad deskryptywnym odczytaniem PZN, podważyć interpretację PZN-D jako pozytywnie zweryfikowanej hipotezy empirycznej.

O ile nie podważymy konkluzywności wyników tych badań, tak jak będą to usiłowali robić filozofowie interpretujący PZN w sposób normatywny, nie będziemy mieć powodu, aby uznać PZN za dobrze potwierdzone twierdzenie psychologii. Podam dwie grupy przykładów: powszechnie znane badania nad heurystykami poznawczymi, które pokazują, na ile często badani nieświadomie przyjmują sprzeczne przekonania, oraz przypadki z zakresu psychologii klinicznej, będące radykalnymi przykładami nieracjonalnych i wzajemnie sprzecznych przekonań.

Pierwszą interesującym nas zestawem przykładów, które mogą podważyć prawdziwość PZN jako hipotezy empirycznej, są badania pochodzące z tzw. nurtu badań nad ludzką nieracjonalnością, prowadzonych przez psychologów poznawczych i kognitywistów od lat 60. Badania te starają się wykazać, że większość ludzi (nawet do 87% badanych [Tversky, Kahneman, 1983]) w codziennym wnioskowaniu logicznym i ocenie prawdopodobieństwa określonych zdarzeń⁹ posługuje się prostymi heurystykami, a nie zasadami logicznego i probabilistycznego wnioskowania – i że te dwie strategie wchodzi z sobą w konflikt. Przedstawię krótko dwa badania pokazujące dwa popularne błędy wnioskowania: błąd koniunkcji i błąd alternatywy, które wydają się przy tym najbliższej spokrewnione z wnioskowaniem logicznym i jako takie mogą być dowodem na podtrzymywanie przez podmiot ewidentnie sprzecznych przekonań.

Błąd koniunkcji można zilustrować klasycznym badaniem Tversky'ego i Kahnemana (1983)¹⁰. Grupie 93 badanych przedstawiono następujące zadanie:

⁹ Tego typu przekonania często są pomijane w filozoficznych rozważaniach, gdzie często uważa się je za nieznaczącą podklasę lub atypowe przypadki przekonań; ich powszechność można jednak zauważyć w często używanych w języku potocznym frazach takich jak: „W wypadku, jeśli...”, „Musimy być przygotowani na to, że...”, „To bardzo prawdopodobne, że...”, „To się raczej nie zdarzy, że...” itp.

¹⁰ Klasycznym przykładem z tego badania jest oczywiście „Linda, kasjerka w banku i feministka”; ten przykład jest jednak na tyle szeroko omawiany w literaturze i poddany z wielu stron krytyce (także ze względu na sformułowanie problemu), że zdecydowałem się użyć innego przykładu obecnego w tym badaniu, a ilustrującego identyczne zjawisko.

Załóżmy, że Bjorn Borg [szwedzki tenisista, który w roku poprzedzającym badanie czterokrotnie triumfował w turniejach wielkoszlemowych] osiągnie finał Wimbledonu w 1981 roku. Uszereguj podane możliwości od najbardziej do najmniej prawdopodobnej:

- A. Borg wygra mecz (1.7)
- B. Borg przegra pierwszy set (2.7)
- C. Borg przegra pierwszy set, ale wygra mecz (2.2)
- D. Borg wygra pierwszy set, ale przegra mecz (3.5) (Tversky, Kahneman, 1983, s. 301–302)

Liczby podane w nawiasach reprezentują średnie miejsce przyznane tej możliwości spośród pozostałych czterech. Duża część badanych uznała, że możliwość C jest bardziej prawdopodobna od możliwości B, co z punktu widzenia rachunku prawdopodobieństwa jest niemożliwe (możliwość C jest iloczynem dwóch zdarzeń, z których jedno jest reprezentowane w możliwości B – nie może być zatem bardziej prawdopodobne). Badacze tłumaczyli to zjawisko występowaniem tzw. heurystyki reprezentatywności – badani, znając reputację Borga jako wspaniałego tenisisty, uznawali od razu za wysoce prawdopodobne każde zdanie, które przewidywało jego zwycięstwo. Podobne badania (Bar-Hillel, Neter, 1993) dotyczyły również rozumowania w sytuacji, gdy mamy do czynienia z sumą dwóch zdarzeń, wyrażaną za pomocą alternatywy – ze względu na zastosowanie heurystyki reprezentatywności badani często uznawali też, że jedno ze zdarzeń jest bardziej prawdopodobne od jego sumy z innym zdarzeniem.

Warto podkreślić, że znaczącej większość badanych w obydwu badaniach nie można opisać jako nieświadomych lub przeczących w swoich deklaracjach podstawowym prawom rachunku prawdopodobieństwa. W pierwszym z wymienionych badań sprawdzono nawet w osobnym badaniu, czy za błąd nie jest odpowiedzialna potoczna interpretacja koniunkcji jako implikacji (Tversky, Kahneman, 1983, s. 302); wydaje się, że większość z osób, które popełniły w jednym lub drugim badaniu błąd, jednocześnie zna zasady rachunku prawdopodobieństwa i potrafi je zastosować w niektórych przypadkach (potwierdzają to między innymi badania korzystające z innego, statystycznego ujęcia problemu Borga [Fiedler, 1988]).

Można zatem zarówno badanym, jak i dużej części ludzkiej populacji, przypisać przynajmniej możliwość posiadania sprzecznych przekonań. Jeśli nasze wnioskowanie jest oparte w większości na heurystykach, które w przeciwieństwie do praw logiki czy matematyki mogą produkować sprzeczne rezultaty, wydaje się, że PZN-D nie jest dobrym opisem naszego systemu formowania przekonań.

Innym, znacznie bardziej bezpośrednim przykładem na posiadanie przez jeden podmiot dwóch sprzecznych przekonań, są kliniczne przypadki urojeń¹¹.

¹¹ Podobną interpretację przedstawia w swoim tekście Tadeusz Ciecierski (2017) – jemu też zawdzięczam zwrócenie uwagi na tę grupę przypadków.

Występowanie podobnych zaburzeń – odpornych na kontrprzykłady, niepodążających za normami racjonalnego wnioskowania przekonań – może świadczyć o tym, że empiryczna hipoteza stwierdzająca ludzką racjonalność może być gruntownie fałszywa: nie chodzi bowiem o przypadki niewielkich lub wytłumaczalnych odchyłeń od racjonalności jak w badaniach nad heurystykami, ale o bardzo poważne zaburzenie zdolności do logicznego myślenia i oceny sytuacji. Czy jednak dotyczy ono również zachwiania PZN?

Najbardziej bezpośrednim przykładem¹² będą niektóre przypadki urojenia Cotarda polegającego na żywieniu przez chorego przekonania, że nie posiada organów wewnętrznych, jest martwy, nieśmiertelny, znajduje się w piekle bądź nie istnieje. Pacjentka badana przez Ryana McKaya i Lisę Cipolotti, LU (2007, s. 353), zapytana o to, skąd wie, że jest martwa i czy widziała kiedyś martwą osobę, odpowiedziała, że widziała ciało swojej babci po jej śmierci i wiedziała, że jest martwa, ponieważ miała ona zamknięte oczy i nie poruszała się. Przykład podobnej sprzeczności możemy zobaczyć w przypadku pacjenta opisanego w badaniu Nishio i Mori:

[Pacjent] powiedział swojemu lekarzowi (Y.N.), „Wydaje mi się, że jestem martwy. Chciałbym prosić o pańską opinię”. Później, jego przekonanie co do śmierci stało się wyraźniejsze. Powiedział, „Mój certyfikat zgonu został zarejestrowany. Spaceruje pan z martwym człowiekiem”, oraz „Jestem martwy. Otrzymam certyfikat zgonu dla siebie od mojego doktora i muszę zanieść go do urzędu miasta niedługo w przyszłym tygodniu”.

Omawianie jego zgonu nie było powiązane z depresyjnym nastrojem lub uczuciem strachu. Kiedy jego lekarz zapytał go, czy martwy człowiek może mówić, zrozumiał, że jego słowa przeczyły logice, ale nie mógł zmienić swojego sposobu myślenia. (Nishio, Mori, 2012, s. 217–218)

Niewiele potrzeba, aby obydwu wymienionym pacjentom przypisać przekonania sprzeczne: choćby (1) że mówią, (2) że nie żyją, i (3) że osoby martwe nie mogą mówić. W drugim przypadku mamy nawet do czynienia, jak się wydaje, z rozpoznaniem tej sprzeczności przez sam podmiot, połączonej z nieumiejętnością odstąpienia od wyraźnie niezgodnych ze sobą przekonań.

Nie twierdzą przy tym, że podanych powyżej przypadków nie da się podważyć. Można próbować twierdzić, że osoby chore na zespół Cotarda tak naprawdę co innego rozumieją przez „śmierć”, a przypadki wadliwego wnioskowania w badaniach Tverskiego i Kahnemana nie są przykładami sprzecznych przekonań. Te twierdzenia jednak przynależą już do argumentacji za normatywnym odczytaniem PZN, bowiem dają wyraźne wskazówki, jak należy interpretować podane wyniki. Co więcej, pokazanie, że podana tu interpretacja wyników badań jest błędna, będzie obowiązkiem każdej koncepcji podtrzymującej PZN-N –

¹² Szerokiego wachlarza przykładów dotyczących innych urojeń dostarczają Breen i in. (2000), zaś omówienie ich filozoficznych implikacji można znaleźć w pracy Bortolotti (2010).

w dalszej części tej pracy pokażę, dlaczego nie spełniają tego wymogu. Jeśli jednak zinterpretujemy PZN jedynie jako hipotezę empiryczną, jesteśmy zmuszeni uznać ją w świetle podanych wyżej przykładów za co najmniej wątpliwą.

2. ARGUMENTACJA METODOLOGICZNA ZA PRZYJĘCIEM PZN

W tej części skupię się na argumentacji metodologicznej za przyjęciem PZN. Zgodnie z wprowadzonym wyżej rozróżnieniem będzie to argumentacja wspierająca zasadę w jej odczytaniu normatywnym:

(PZN-N) Podmiotowi w jednym momencie nie można przypisać wzajemnie sprzecznych przekonań.

Przyjęcie takiej tezy, jak zostało wyżej zarysowane, osadzone jest w innym kontekście filozoficznym niż argumentacja za PZN-D. Przede wszystkim, najczęściej wiąże się ona z instrumentalizmem lub przynajmniej agnostycyzmem wobec statusu ontologicznego przekonań. Istotą przytoczonych tutaj argumentów jest uznanie „przekonania” za pojęcie teoretyczne psychologii potocznej lub naukowej i wykazanie, że przyjęcie PZN-N jest niezbędne właśnie z punktu widzenia teorii pozwalających na przypisywanie przekonań. Takie stanowisko wydaje się w mniejszym lub większym stopniu zakładać interpretacjonizm w odniesieniu do przekonań – i nie zamierzam tego założenia z początku podważyć.

Użytecznym podziałem – przed przejściem do właściwego omówienia argumentów – jest odróżnienie indywidualnych i naukowych przypisań przekonań. Wprowadzam je za Richardem Dubem (2015)¹³, który używa go do wyróżnienia dwóch poziomów argumentacji za założeniem o racjonalności przeprowadzonych przez Daniela Dennetta. Rozróżnienie to uznaję jednak za znacznie bardziej podstawowe i kluczowe do wyłuszczenia konkretnych założeń i celów, jakie stawiają sobie poszczególne teorie i argumenty.

Przypisanie indywidualne to, krótko mówiąc, przypisanie przekonania, którego dokonuje każdy kompetentny użytkownik języka w codziennych sytuacjach. Są to najbardziej potoczne użycia fraz takich jak: „On/ona jest przekonany, że...”, „Chyba sądzi, że...”, „Myśli, że...”. Przypisania te są dość oszczędne, jeśli chodzi o wykorzystywane dane: dokonujemy ich pod presją czasu, często nie posiadając dużej wiedzy o cudzym życiu, zachowaniach, zwyczajach itp. Są one formułowane niejako automatycznie i nie podlegają zaawansowanej refleksji¹⁴. Podtrzymywanie PZN-N w odniesieniu do przypisań indywidualnych oznaczałoby, że nasze codzienne posługiwanie się psychologią potoczną wymaga założenia niesprzeczności przekonań podmiotu, którego dotyczy przypisanie.

¹³ w oryginale: individual and scientific ascription.

¹⁴ Sam Dennett w odpowiedzi Dubowi określa je jako “the time-pressured quick-and-dirty attributions of folk psychology” (Dennett, 2015, s. 206).

Tradycyjnie uzasadnienie takiego twierdzenia będzie się opierało na opisanu potocznego użycia języka i pokazanie, że wynika ono z „metody” i natury pojęć psychologii potocznej.

Zupełnie innym kontekstem użycia przytoczonych wyżej konstrukcji językowych jest dokonanie przypisania naukowego. „[Przypisania indywidualne i naukowe] odróżnia to, kto dokonuje przypisania: pierwsze jest wykorzystywane przez indywidualne osoby w sytuacjach z prawdziwego życia, drugie są wykorzystywane przez naukowców i filozofów w rozwijaniu teorii” (Dub, 2015, s. 98). Argumentacja dowodząca PZN-N w odniesieniu do przypisań naukowych będzie się skupiać nie na tym, w jaki sposób pojęcie przekonania jest, ale jak powinno być używane, kiedy terminy wywodzące się z psychologii potocznej przyjmujemy w psychologii naukowej. Nawet jeśli potoczne użycie dopuszcza przypadki złamania PZN-N, to w żadnym wypadku nie można do tego dopuścić na gruncie „dorosłej” teorii przekonań – zdają się mówić niektórzy teoretycy. Argumenty idące podobną linią będą dowodzić PZN-N, powołując się na metodologiczne podstawy psychologii, antropologii lub lingwistyki. Postaram się pokazać, że nawet przy przyjętych przez autorów założeniach, uznanie PZN-N zarówno wobec przypisań indywidualnych, jak i naukowych jest przynajmniej problematyczne.

2.1 Argument ze strategii intencjonalnej Daniela Dennetta

Głównym zwolennikiem PZN-N wobec przypisań indywidualnych można określić Daniela Dennetta z okresu rozwoju teorii strategii intencjonalnej i w późniejszych dziełach¹⁵. Dennett przyjmuje, że pojęcie przekonania jest elementem szerszej strategii przewidywania i opisu zachowania innych systemów poznawczych, nazywanej przez niego właśnie „strategią intencjonalną”. Polega ona na przyjęciu, że opisywany podmiot jest racjonalny i przypisywaniu mu przekonań, pragnień i intencji wyjaśniających jego działanie oraz ich konsekwencji w zgodzie z przyjętym kanonem racjonalności („przypisuj te przekonania, które podmiot powinien mieć”) oraz dalszym przewidywaniu jego działań jako zgodnych z przypisanymi przekonaniem¹⁶. Strategia intencjonalna

¹⁵ „W *Content and Consciousness* Dennett jasno uznaje za istotne przypisania stanów mentalnych drugiego [naukowego] typu. [...] Przesunięcie nastąpiło mniej więcej, gdy Dennett rozwinął swoją Strategię Intencjonalną. Strategia Intencjonalna stała się częścią przypisania indywidualnego: interpretacja była teraz uznana za coś, co wszyscy w naturalny sposób robimy” (Dub, 2015, s. 98).

¹⁶ Dennett odróżnia ją od dwóch innych strategii opisu: „strategii funkcyjnej” (odwołującej się do funkcji, jaką ma pełnić określony przedmiot) i „strategii fizycznej” (odwołującej się do własności fizycznych i praw fizyki). Strategie te odróżnia od siebie poziom skomplikowania i dokładność przewidywań: działanie budzika możemy opisać i przewidywać za pomocą modelu fizycznego, odwołania się do jego funkcji (np. dzwonenie o godzinie, na którą jest nastawiony) bądź przypisania mu przekonań i pragnień (np.

według Dennetta jest elementem naszej codziennej praktyki: czymś, co robimy, kiedy posługujemy się pojęciem przekonania w potocznym języku, interpretując zachowanie innych. Stąd główny nacisk w argumentacji Dennetta położony jest na przypisania indywidualne (jego argumentacja za uznaniem PZN-N w psychologii naukowej jest zbieżna z uwagami poczynionymi przez Quine'a i Davidsona, których poglądy omówię później).

Trudno określić z początku, czy Dennett uznaje PZN-N za nienaruszalną. Z pewnością istotny jest fragment, w którym rozważa włączenie przypadków interpretacji nieracjonalnego zachowania do zasad strategii intencjonalnej:

Jakie powody moglibyśmy jednak mieć, by ustawić jakiś zbiór [zdań] pomiędzy ekstremami [racjonalności i nieracjonalności] i nazwać go zbiorem przekonań [...]? To inny sposób pytania, czy moglibyśmy zastąpić normatywną teorię przekonań Hintikki empiryczną teorią przekonań i, jeśli tak, jakich świadectw byśmy użyli. „W zasadzie”, ktoś może chcieć powiedzieć, „ludzie są czasami przekonani o sprzecznościach, jak dowodzą ich wypowiedzi; zatem każda adekwatna logika przekonań lub analiza pojęcia przekonania musi wziąć pod uwagę ten fakt”. Ale jakakolwiek próba legitymizacji ludzkiej omyłności w teorii przekonań poprzez ustawienie dozwolonego poziomu błędu byłaby jak dodanie dodatkowej reguły do szachów: Oficjalnej Reguły Tolerancji, zgodnie z którą każda gra w szachy zawierająca nie więcej niż k ruchów niezgodnych z innymi regułami gry jest prawidłową grą w szachy. (Dennett, 1978, s. 21)

Czy jednak również w takich wypadkach, jak urojenia lub choroba psychiczna, czy też codziennych przypadkach działań sugerujących odstępstwo od normatywnego wzoru szukamy wyjaśnienia podążając za idealną racjonalnością – lub powstrzymujemy się od osądu? Wydaje się, że chociaż możemy nasz proces interpretacji rozpoczynać, odnosząc się do modelu w pełni racjonalnego podmiotu, z czasem porzucamy to założenie, dostosowując nasz model do wyjaśnienia nowych świadectw. W dyskusji ze Stichem (1981), podnoszącym podobne zarzuty, Dennett (1981) podaje argument, zgodnie z którym wyjaśnienie takich przypadków jest dostępne jedynie poprzez strategię z niższego poziomu.

W interpretacji Dennetta opis nieracjonalności w języku „strategii intencjonalnej” jest niemożliwy: wyrażenia takie jak „wypadło mi to z głowy”, „pomyliłem się” itp. odwołują się do strategii wyjaśniającej moje zachowanie poprzez złe działanie jednej z funkcji (pamięci, wzroku etc.) podmiotu. Przykłady nieracjonalności możemy tłumaczyć zatem jedynie jako *błędy wykonania*, nie zaś *błędy kompetencji*^{17, 18}. Nie jest więc tak, że w podanej wyżej sytuacji jestem

pragnienia obudzenia nas o określonej porze i przekonania dotyczącego pory dnia) (Dennett, 1981a).

¹⁷ w oryginale: *performance errors* i *competence errors*; rozróżnienie to zapożyczam od Sticha (1985). Oryginalnie pochodzi ono (w odniesieniu do pomyłek i kompetencji gramatycznych) od Noama Chomsky'ego.

¹⁸ Podobną linię argumentacji i traktowanie wszelkich przykładów nieracjonalności jako „pomyłek” czy właśnie „błędów wykonania” podają autorzy szukający ugruntowania

przez użycie strategii intencjonalnej zobowiązany do wyjaśnienia jej poprzez odwołanie się do rozbudowanej racjonalizacji mojego zachowania albo odrzucenia możliwości interpretacji. Po prostu odwołuję się wówczas do błędu na niższym poziomie wyjaśnienia, podobnie jak wówczas, gdy za błąd w działaniu budzika (a więc działanie wbrew przewidywaniom strategii funkcyjnej) obwiniamy przetarcie zębátky lub rozładowanie baterii (czyli zdarzenia opisywane za pomocą strategii fizycznej).

Podanie szerokiego wachlarza kontrprzykładów nie wystarczy zatem, aby podważyć argumentację Dennetta. Można w niej wyróżnić dwie kluczowe tezy. Po pierwsze, że wzorcem, według którego konstruujemy szczególnie pojęcie systemu intencjonalnego (podmiotu), jest idealny podmiot racjonalnie formułujący przekonania (między innymi unikając sprzeczności) i działający w zgodzie z nimi. Po drugie, że w momencie, w którym dany system intencjonalny działa wbrew przewidywaniom, zobowiązani jesteśmy do wyjaśnienia jego zachowania poprzez odwołanie się do błędu na poziomie funkcyjnym, nie intencjonalnym (i zazwyczaj to właśnie robimy). Osoby, które czynią inaczej, popełniają zatem błąd w posługiwaniu się potocznym użyciem pojęć psychologii potocznej. Tezy te mają zarazem charakter opisowy (tak właśnie posługujemy się pojęciem przekonania) i normatywny (w przypadku niezgodności przewidywań z rzeczywistością powinniśmy preferować wyjaśnienie na poziomie funkcyjnym). Poniżej postaram się rozwinąć krytykę tej linii argumentacji zaofierowaną przez Sticha (1981; 1985), najpierw odnosząc się do strony opisowej przesłanek Dennetta, później zaś do normatywnej.

Stich (1981), polemizując z Dennettem, zarzuca mu, że jego koncepcja, niezależnie od przyjętej interpretacji, nie pozwala wyjaśnić najprostszyc przypadków odstępstw od racjonalności. Dzieli analizy i sugestie Dennetta na zgodne z „twardą” i „miękką” linią¹⁹, zarzucając mu jednocześnie niekonsekwencję w swojej argumentacji. Twarda linia obejmuje w perspektywie Sticha założenie racjonalności podmiotu, której konsekwencją jest przyjęcie PZN-N. Linia miękka natomiast to dająca się wywnioskować z niektórych fragmentów pism Dennetta²⁰ wizja podobnego założenia i ujęcia PZN-N jedynie jako warunków koniecznych rozpoczęcia procesu interpretacji; warunki te nie obowiązują nas później, po nabyciu większej wiedzy dotyczącej zachowania danego podmiotu. Stich uznaje twardą linię za niemożliwą do obrony. Nie chodzi jedynie o intuicyjną absurdalność twierdzenia, że każdy, kto zna podstawy klasycznego rachunku zdań, wierzy również we wszystkie tautologie. Strategia „twardej linii” zawodzi, jeśli chodzi o opisy najbardziej powszechnych przypadków nieracjonalności:

poprawności zasad logiki w psychologii, patrz: Cohen (1981), Sober (1978). Krytyka także tych stanowisk podobna do zaofierowanej poniżej zawarta jest w: Stich (1985); Thagard i Nisbett (1983).

¹⁹ Ang. *soft line* i *hard line* (Stich, 1981, s. 49).

²⁰ Jak choćby z cytowanego już wyżej: „[Tworząc intencjonalną strategię danego systemu], zaczynamy od ideału doskonałej racjonalności i rewidujemy tak, jak dyktują nam to warunki” (Dennett, 1981a, s. 21).

Gdy chłopiec z sąsiedztwa daje mi złą resztę, kiedy kupuję od niego lemoniadę, nie zakładam, że jest przekonany, że ćwierćcentówki warte są 23 centy, albo że chce mi ukraść 2 centy. Moim *pierwszym* założeniem jest, że nie jest jeszcze zbyt dobry w wykonywaniu dodawania w pamięci. (Stich, 1981, s. 50)

„Miękką linią” cierpi natomiast, w ujęciu Sticha, na inny mankament – jeśli ją przyjmujemy, trudno zrozumieć, dlaczego wizerunek „idealnej racjonalności” miałby być punktem wyjścia dla naszej teorii i czemu nie posługujemy się od początku zmodyfikowanym pojęciem racjonalności opartej na tym, jak standardowo (w większości przypadków) przebiega nasz proces wnioskowania, choćby opierając się na badaniach dotyczących heurystyk poznawczych. „Miękką linią” prowadzi zatem do uznania, że „ideał”, który uznajemy za pierwszy model w procesie interpretacji, różni się od przyjętego przez Dennetta rozumienia racjonalności.

Jak zarysowałem wyżej, koncepcja Dennetta wymyka się jednak prostemu podziałowi na „miękką” i „twardą linię”. Sam autor pisze: „Te różne linie są wynalazkami Sticha, narodzonymi z frustracji wynikającej z próby zrozumienia mojego wysłowienia przyjętej przeze mnie perspektywy, która jest zarazem twarda i miękka – można powiedzieć, elastyczna” (Dennett, 1981b, s. 73). Dennett postrzega przypadki podobne do przytoczonego wyżej jako możliwe do ujęcia jedynie za pomocą strategii funkcyjnej. Wszak wyjaśnienie, że chłopiec „nie jest jeszcze zbyt dobry w wykonywaniu dodawania w pamięci” wydaje się pochodzić z tego właśnie poziomu. Przy takiej interpretacji Dennett jest w stanie zachować pełną moc „twardej linii”, jednocześnie tłumacząc jej hipotetyczną nieskuteczność.

Trudno jednak nie zauważyć pewnych kłopotów z takim sformułowaniem stanowiska Dennetta. Czy z a w s z e preferujemy wyjaśnienie na poziomie funkcyjnym? I czy, przede wszystkim, te wyjaśnienia faktycznie muszą wynikać z naszej awersji do złamania PZN-N? Moim zdaniem, na obydwa te pytania odpowiedź brzmi: nie.

Po pierwsze, strategia funkcyjna nie zawsze jest nam dostępna – często też nasze potoczne intuicje na ten temat wydają się być niejednoznaczne. Interesujący materiał do rozpatrzenia stanowią badania Wasona (1969), w których badano wpływ tłumaczenia i wskazywania sprzeczności na poprawę w rozwiązywaniu zadania selekcji kart (Wason, 1968). Badanym, którzy źle wykonali zadanie selekcji kart, nie stosując się do zasad elementarnej logiki (w tym wypadku rozumowania modus tollens)²¹, próbowano przedstawić pomyłkę w ich rozumo-

²¹ Oryginalne badanie (Wason, 1968) polegało na przedstawieniu badanemu czterech dwustronnych kart, z zestawami dwóch liter i dwóch cyfr (np. „D”, „3”, „B”, „7”), gdzie po jednej stronie karty zapisana była litera, po drugiej zaś cyfra, i prośbie o wytypowanie kart, które należy odwrócić, aby sprawdzić, czy spełniona jest pewna implikacja (np. „Jeśli D jest po jednej stronie karty, to po drugiej stronie karty jest 3”). Jedynie niewielka grupa badanych była w stanie rozwiązać to zadanie poprawnie (wytypować karty „D”

waniu tak, by przekonać ich do zmiany poprzedniej odpowiedzi. Najpierw upewniano się, że badany dobrze rozumiał pytanie i wie, że podana reguła, której prawdziwość miał sprawdzić, może być również fałszywa. Eksperymentator zaczynał od pytania: „Gdyby po drugiej stronie znalazł się [bodziec wspomniany przez badanego], czy mógłbyś powiedzieć cokolwiek o prawdziwości lub fałszywości tej reguły?” (Wason, 1969, s. 474). Gdy zaś zawodziły coraz bardziej perswazyjne, lecz wciąż hipotetyczne rozważania, w których badani pozostawali przy deklarowaniu odpowiedzi logicznie sprzecznych (utrzymywali, że koniunkcja przesłanki i negacji wniosku nie falsyfikuje danej implikacji), badacze prosili badanych o odkrycie kart. Jeśli badany wciąż nie był w stanie wybrać odpowiedniej karty, eksperymentator wprost oznajmiał mu, że się myli i prosił go o przemyslenie swojej odpowiedzi. Badanie wykazało, że 12% przebadanych osób na żadnym z etapów rozważań nie potrafiło zmienić swojego zdania.

Czy w momencie, w którym dany podmiot nie chce, pomimo usilnych starań eksperymentatora, przyznać się do popełnienia pomyłki, wciąż jesteśmy w stanie uznać to za „błąd wykonania” i opisać na poziomie funkcyjnym? Wydaje się, że nie – nielogiczność nie jest tutaj jedynie kwestią chwilowego zaburzenia kompetencji wnioskowania, gdyż pomimo długich prób pomyłka ta nie może zostać skorygowana. Choć nie należy wykluczyć, że często owe błędy mogą być (i są) korygowane, zaś nasz język umożliwia nam „zracjonalizowanie” ich w sposób podany przez Dennetta, jednakże „co jest naprawdę istotne w tym i innych eksperymentach [...] nie są poprawy w postępowaniu, a raczej ilość badanych, która nigdy, nie ważne co im robiono, nie wybrała [dobrej odpowiedzi]” (Manktelow, 1981, s. 259).

Podobnie sprawy się mają, gdy powrócimy do przykładów osób cierpiących na urojenia. Metoda „sokratycznej dyskusji”, zgodnie z którą poprzez wykazywanie sprzeczności ukrytej w słowach osoby chorej można ją nakłonić do zmiany zdania, a przez to uleczyć, również jest często nieskuteczna (Bortolotti, 2010, s. 86–96). Przykładem może być choćby pacjent z cytowanego wyżej badania, który po wyleczeniu i opuszczeniu szpitala psychiatrycznego dalej mówił „Teraz jestem żywy. Ale byłem w tym czasie martwy” oraz „Widziałem Kim Jong-Ila w szpitalu, w którym byłem” (Nishio, Mori, 2012, s. 218). Jednocześnie niczym absurdalnym lub niezgodnym z potocznym użyciem języka w świetle powyższych danych nie jest powiedzenie, że pacjent Nishio jest przekonany, że był kiedyś martwy, albo że badani w teście Wasona są przekonani, że karta zawierająca na sobie przesłankę i zaprzeczenie wniosku nie falsyfikuje reguły stwierdzającej ich implikację. Oznacza to, że w sytuacji, w której uzyskujemy jedynie trochę więcej danych dotyczących podmiotu, opisanie sprzeczności przekonań w ramach strategii intencjonalnej jest jak najbardziej możliwe i preferowane

i „7”). W późniejszych badaniach modyfikowano treść kart i sformułowanie reguły tak, aby wyeliminować niejasności.

wobec Dennettowskiej alternatywy: użycia strategii funkcyjnej lub powstrzymania się od opisu w jakichkolwiek kategoriach.

Innym deskryptywnym elementem propozycji Dennetta jest uznanie, że użycie strategii funkcyjnej wynika z naszej niechęci do opisanie innych lub siebie samych jako nieracjonalnych lub posiadających sprzeczne przekonania. To twierdzenie wymaga jednak uzasadnienia: aby dowieść, że owe sposoby mówienia czy konstrukcje językowe pochodzące ze strategii funkcyjnej są „racjonalizacjami”, musimy wykazać, że to właśnie na racjonalności i niesprzeczności zależy nam, kiedy ich używamy. To jednak nie zachodzi. Można dla takiego użycia języka znaleźć wiele innych uzasadnień, nie mających wiele wspólnego z racjonalnością²². Nie wydaje się to jednak nawet niezbędne. Jak już zostało powiedziane, przypisania błędów na gruncie funkcyjnym dokonujemy automatycznie, jest ono naszym *pierwszym* założeniem, a nie racjonalizacją, która dochodzi do głosu, gdy wyczerpane zostają możliwości spójnego wytłumaczenia naszego zachowania. Co zatem istotne w argumencie Sticha, Dennett myli się w wyjaśnieniu przebiegu naszej potocznej interpretacji tej sytuacji, nie zaś w konkluzji, do której zobowiązuje go jego teoria.

Należy zatem, by przeprowadzić krytykę do końca, zwrócić się ku normatywnemu aspektowi koncepcji Dennetta. Przyjmując interpretacjonistyczną postawę, dalej bowiem musimy rozważyć, dlaczego to właśnie silnie pojmowana racjonalność i niesprzeczność miałyby być ideałem, którym kierujemy się w przypisaniach indywidualnych. Jeśli weźmiemy pod uwagę przytoczone wyżej badania Wasona, Kahnemana i Tversky’ego, Bar-Hillel czy przypadki urojeń, powinno zdziwić nas, jak bardzo strategia intencjonalna odbiega w swoich przewidywaniach od faktycznych ludzkich zachowań – jak wiele przypadków taka teoria wyklucza. Jeśli uważamy, że strategia intencjonalna powinna pozwalać nam na jak najlepsze przewidywanie cudzego zachowania, musimy założyć, że przynajmniej statystycznie najczęściej najużyteczniejszym opisem naszego zachowania jest jego opis w kategoriach racjonalnie działającego i formującego swoje przekonania podmiotu. To jednak, jak pokazują powyższe przykłady, jest co najmniej dalekie od pewności: ludzki system wnioskowania i formułowania przekonań zwyczajnie wydaje się za tymi normami nie podążać. Przykładem może być tu choćby słynny błąd hazardzisty – niepoprawne wnioskowanie, zgodnie z którym prawdopodobieństwo zdarzenia maleje, jeśli wcześniej owo zdarzenie często się zdarzało (np. że prawdopodobieństwo wypadnięcia orła w rzucie rzetelną monetą jest mniejsze niż $\frac{1}{2}$, jeśli wcześniej wypadł on dwudziestokrotnie). Popęłnianie tego błędu jest stosunkowo intuicyjne dla większości badanych i częste wśród nich (Tversky, Kahneman, 1971), często wykorzystują oni podobną zasadę w przewidywaniu faktów zależących od prawdopodobieństwa naprzemiennie z dającym sprzeczne z nim wyniki błędem „dobrej pas-

²² Przydatne w tworzeniu takich wyjaśnień może być choćby odróżnienie niesprzeczności i potocznie rozumianej spójności: nie ma bowiem nic sprzecznego czy nielogicznego w wielu naszych zachowaniach, które tłumaczymy w ten sposób (np. przejęzyczenia czy „społecznie niezręczne” lub niepożądane zachowania).

sy”²³, zgodnie z którym prawdopodobieństwo zdarzenia rośnie, gdy powtórzy się ono wystarczająco wiele razy (Konold i in., 1993). Nie jest tutaj istotne, jak w sytuacjach przywoływanych wcześniej, czy ludzie są w stanie rozpoznać takie postępowanie jako błędne, ale to, że często postępują w zgodzie z błędnymi zasadami. Jeśli zatem strategia intencjonalna miałaby zależeć od modelu najczęściej dającego prawidłowe przewidywania, nie powinna ona zakładać, że dany podmiot jest proceduralnie racjonalny, ale raczej, że formuje swoje przekonania w oparciu o pewne heurystyki zgodne z błędnymi rozumowaniami hazardzisty i dobrej passy – a więc użyteczna strategia intencjonalna powinna dopuszczać posiadanie sprzecznych przekonań.

Wskazane problemy z koncepcją Dennetta można uogólnić na wszystkie koncepcje traktujące pojęcie przekonania jako pojęcie psychologii potocznej, które postulują PZN-N jako element praktyki indywidualnych przypisań przekonań. Jeśli bowiem faktycznie istnieją przypadki przypisań indywidualnych, które faworyzują przypisanie przekonań sprzecznych zamiast opisanie danego zachowania jako „zwykle odstępstwo” od PZN-N, to teza o jej powszechnym obowiązywaniu jest empirycznie fałszywa. Nawet jeśli jednak przymkniemy oko na te przypadki lub zanegujemy intuicyjność takich przypisań indywidualnych, istnieje jednak dla każdej z tych teorii problem dużo poważniejszy. Skoro w tak wielu przypadkach ludzie nawet powierzchownie mają skłonność do działania w zgodzie z proceduralnie nieracjonalnymi regułami zezwalającymi na wywnioskowanie wzajemnie sprzecznych informacji, zasady rządzące „podejmowanymi pod presją czasu, szybkimi i brudnymi przypisaniami psychologii potocznej” powinny zawierać takie reguły, a nie sztywny kanon proceduralnej racjonalności.

Utrzymanie PZN-N w odniesieniu do przypisań indywidualnych i uznanie jej, w wersji słabej lub silnej, za metodologiczną zasadę psychologii potocznej w jej codziennym życiu jest zatem nieuzasadnione, a sama PZN-N przedstawiona w takim kontekście prawdopodobnie fałszywa. Wobec tego należy odnieść się do argumentów, które bronią PZN-N w odniesieniu do przypisań naukowych.

2.2 Niesprzeczność a znaczenie

Donald Davidson, jeden z głównych zwolenników PZN-N wśród współczesnych filozofów, dzieli z Danielem Dennettem zestaw intuicji dotyczących pochodzenia i warunków poprawnego użycia terminów psychologii potocznej. Koncepcja Davidsona jednak w jasny sposób odnosi się do teorii interpretacji jako do teorii naukowej, pozwalającej nam wytworzyć „jednolitą teorię znaczenia i działania” inspirowaną modelami przypisywania przekonań na podstawie preferencji zaproponowanymi przez Franka Ramseya w teorii decyzji (Ramsey, 1926; Davidson, 1980). Przypisania, o których mówi Davidson, będą zatem przypisaniami naukowymi, wynikającymi z odpowiedniego teoretycznego na-

²³ Błąd ten jako pierwszy został wykryty i opisany w słynnym badaniu o postrzeganiu przez kibiców rzutów osobistych w koszykówce: Gilovich i in. (1985).

mysłu, objaśniającego nam w jak najbardziej zbliżony do prawdy sposób werbalne i niewerbalne zachowania ludzi.

Inspiracją dla obydwu koncepcji są spostrzeżenia Willarda Quine'a, którego Davidson i Dennett byli studentami²⁴, zwłaszcza zaś teoria niezdeterminowania przekładu. Choć głównym obiektem zainteresowań Quine'a jest zagadnienie przekładu i pojęcie znaczenia językowego, wiele z jego uwag dotyczy również przypisywania przekonań i koresponduje z poglądami jego następców. W słynnym fragmencie ze *Słowa i przedmiotu* Quine uzasadnia pogląd, zgodnie z którym każdy przekład musi zachowywać podstawowe prawa logiki:

Zauważamy, że w praktyce uczciwy przekład zachowuje prawa logiki, i to nawet wtedy, gdy używając paradoksalnego określenia, nie mamy do czynienia z żadnym obcym językiem. I tak, kiedy pytamy o coś w języku polskim, użytkownik języka polskiego odpowiada zaś „Tak i nie”, to przyjmujemy, że inaczej rozumie nasze pytanie w kontekście twierdzącej, a inaczej w kontekście przeczącej odpowiedzi; raczej wybierzemy tę możliwość, niż uznamy go za osobę na tyle niemądrą, by uznawała i odrzucała jedno i to samo. Poza tym, kiedy ktoś akceptuje logikę o prawach wyraźnie sprzecznych z naszymi prawami, jesteśmy gotowi przypuszczać, że nadaje po prostu pewnym znanym, starym wyrazom („i”, „lub”, „nie”, „każdy” itp.) nowe znaczenia. (Quine, 1999, s. 74–75)

Według Quine'a jesteśmy zobowiązani do takiej interpretacji wypowiedzi innych, która będzie pozostać w zgodzie z prawami logiki – w tym z zasadą niesprzeczności. Tezę tę można przedstawić również w taki sposób: akceptacja przez podmiot zdań wzajemnie sprzecznych jest dowodem na to, że nasz przekład obcego języka lub idiolektu danego podmiotu jest błędny. Spójniki logiczne są bowiem zdefiniowane funkcjonalnie (poprzez tabelę prawdziwościową) i czymś z definicji niemożliwym jest rozumieć koniunkcję czy negację tak, jak rozumiemy je w logice i jednocześnie uznawać zdanie postaci *p i nie-p*²⁵.

Podobna motywacja wydaje się stać za Davidsonowską zasadą życzliwości. Według niej, aby w ogóle rozpocząć proces interpretacji, należy założyć, że jak najwięcej przekonań danego podmiotu jest prawdziwych, jak i że podmiot ten nie jest skłonny do posiadania przekonań jawnie fałszywych. – np. logicznie sprzecznych. Tam, gdzie Quine szuka przekładu, czyli, by użyć jego terminologii, bodźcowej synonimii między zdaniami dwóch języków, Davidson stara się wskazać równoważność na poziomie warunków prawdziwości zdań obydwu języków – by zaś mówić o znajomości warunków prawdziwości przez użytkowników języka, musimy przypisać im pewne przekonania. W niektórych odczytaniach myśli Davidsona często uważa się, że na zasadę życzliwości składają się dwie osobne zasady: zasada, zgodnie z którą jak najwięcej przekonań i zdań

²⁴ Analizę podobieństw tych stanowisk do stanowiska Quine'a zawiera (Dub, 2015, s. 94–98).

²⁵ Podobny w duchu zarzut przedstawiony wobec logik parakonsystentnych (jako „zmieniających przedmiot”) można znaleźć w (Slater, 1995).

wypowiedzianych przez interpretowany podmiot powinno być prawdziwych oraz zasada, zgodnie z którą wypowiedzi i przekonania tego podmiotu powinny być zgodne z kanonem racjonalności²⁶. Te dwie zasady wydają się jednak mieć wspólny rodowód: racjonalność rozumiana jest w nich jako zasada działania mająca na celu niedoprowadzenie do podtrzymywania poglądów jawnie fałszywych, w tym i wewnętrznie sprzecznych – a więc do maksymalizacji liczby poglądów prawdziwych.

Powyższe rozumowanie prowadzi do przyjęcia jako jednego z naczelných praw metodologicznych w procesie interpretacji wypowiedzi poniższej zasady:

(PC) Jeśli podmiot akceptuje bądź wygłasza zdania wzajemnie sprzeczne według dotychczasowej interpretacji jego języka lub idiolektu, zinterpretuj jego wypowiedzi jako niesprzeczne w języku lub idiolektie podmiotu.

Ponieważ jednym z główných metodologicznych zaleceń wystosowanych przez Davidsona jest potraktowanie wszystkich wypowiedzi jako szczerych i prawdziwych dla interpretowanego podmiotu – a więc pociągających za sobą bycie przekonaniem o ich treści – PC możemy uznać za konsekwencję przyjęcia PZN-N.

Czy zatem PC jest dobrą i zawsze obowiązującą zasadą metodologiczną? Argument najczęściej przedstawiany za odpowiedź twierdzącą przyjmuje postać równi pochyłej. Zgodnie z nim, jeśli raz zawiesimy działanie PC, wówczas zmuszeni jesteśmy przyjąć odmienną zasadę dla interpretacji podmiotu, która, pozostając w zgodzie ze zmodyfikowanymi regułami wnioskowania, da nam w efekcie przekonania i wypowiedzi „tak dziwaczne, jak tylko sobie tego życzymy” (Quine, 1999, s. 74). Podobne rozumowanie nie może być jednak przyjęte jako uzasadnienie dla powszechnego obowiązywania PC i PZN-N. Podział na przekłady i interpretacje zgodne z prawami klasycznej logiki i te, które pozostają w zgodzie z odmiennymi prawami wnioskowania, nie jest zupełny, nie obejmuje bowiem różnych stopni zgodności; interpretacje mogą też różnić się w zależności od podmiotu i nie muszą od razu obejmować całej społeczności i języka.

Analizę dwóch odmienných interpretacji: zgodnej i niezgodnej z PC, można przeprowadzić na przykładzie debaty dotyczącej statusu urojeń. Ich ogólna charakterystyka została zarysowana już wyżej. Wielu autorów, idąc za sugestią Dennetta i Davidsona, odmawiało urojeniom statusu przekonań, tłumacząc je jako wyobrażenia, co do których podmiot jest przekonany, że są przekonaniem (Currie, 2000), albo jako przypadki osobnych nastawień sędzienných określaných jako „między-przekonania” (Schwitzgebel, 2010) lub „bimaginacje” (Egan, 2008)²⁷. Te rozwiązania, choć zgodne z PC (nie nakładające na podmiot „odpowiedzialności za słowo”), nie wydają się być jednak użyteczne naukowo,

²⁶ Patrz: artykuł Joseph (2004, s. 62–64).

²⁷ Ang. *in-betweenish* lub *grey-area cases of belief* oraz *bimagninations*.

a stanowią raczej wybiegi wobec niepowodzenia hipotezy bardziej zasadniczej, według której pojęciem „bycia martwym” lub „identyczności” chory przypisuje inne niż nasze znaczenie. Jednym z najbardziej obszernych omówień różnych rozwiązań sformułowanych w tym duchu stanowi praca Johna Campbella (2001). Podstawową intuicję Campbella i innych wydaje się przedstawiać poniższy fragment:

Pacjentka może twierdzić, że jest martwa pomimo zdolności do chodzenia i mówienia. W rzeczy samej, pacjentka może mówić, że jest martwa, chociaż zdaje sobie sprawę z tego, że nikt inny nie uznałby tego twierdzenia. Problemem jest, w jaki sposób pacjenci mogą być rzeczywiście uznani za posiadających znajomość znaczenia ich wypowiedzi, kiedy używają słów w tak dziwny sposób? (Campbell, 2001, s. 91)

Campbell rozpoznaje dwie zgodne z PC możliwe strategie opisanie przypadków urojeń: pierwszą określa jako „empiryczną”, drugą jako „racjonalistyczną”. Empiryczna strategia stara się wyjaśnić zachowanie pacjenta jako wynikające z danych, które zaczyna w pewnym momencie otrzymywać (np. w wyniku uszkodzenia ośrodków odpowiedzialnych za percepcję twarzy w mózgu, patrz: Ellis, Young, 1990). Można wówczas próbować wyjaśnić zachowanie pacjenta jako racjonalne w szerokim sensie – jeśli z osobą bliską, choć identycznie wyglądającą, przestaje się wiązać subiektywne uczucie podobieństwa, pacjent dochodzi do wniosku, że w takim razie bliski został zastąpiony przez sobowtóra (urojenie Capgrasa) lub że sam pacjent jest martwy (urojenie Cotarda), co jest przyczyną braku emocji związanych z percepcją identyczną z wcześniejszą²⁸. Jak zauważa sam Campbell, nie jest to jednak zadowalająca strategia „racjonalizująca” urojeń: istnieją bowiem osoby o podobnych problemach neurologicznych, które nie wyciągają podobnych wniosków. Nie ma także nic w naszym doświadczeniu, co mogłoby być racjonalnym uzasadnieniem dla przyjęcia twierdzenia o własnym nieistnieniu – podane zaburzenia neurologiczne, przynajmniej w wypadku specyficznie urojeń Cotarda, mogą być więc korelatami, ale raczej nie źródłem urojeń.

Strategia racjonalistyczna wydaje się iść bardziej bezpośrednio za wskazaniami (PC), interpretując zachowanie pacjentów z urojeniami jako wynik przyjęcia innych przekonań ramowych²⁹ (Campbell, 2001, s. 96). Koncepcja przekonań ramowych zapożyczona jest z *O pewności* Wittgensteina, który opisuje je jako zdania niewzruszalnie pewne, tworzące ramę, wewnątrz której dokonuje się procesu wnioskowania i oceny prawdziwości bądź fałszywości innych zdań. Wszelkie uzasadnienie musi zachodzić wewnątrz tak pojmowanej ramy (Wittgenstein, 2014).

²⁸ Dalsze omówienie podobnych modeli można znaleźć w artykule McKay i Cipolotti (2007, s. 351–352).

²⁹ Ang. *framework propositions and beliefs*.

Choć koncepcja Wittgensteina nigdy nie została szczegółowo dopracowana, możemy za Campbellem przyjąć, że przekonania szczególne dla konkretnego urojenia: „Jestem martwy”, „Ta [postrzegana teraz] kobieta nie jest tą [postrzeżaną kiedyś] kobietą” stanowią dla pacjentów przekonania ramowe w proponowanym sensie. Idąc dalej, można uznać, że niektórzy pacjenci posługują się „dziwaczną logiką”³⁰, która pozwala uzgodnić dane ze świata z treścią przekonań ramowych. Zgodnie z nią np. prawo Leibniza (nieodróżnialności identycznych) funkcjonuje jako prawo dopuszczające identyczność przedmiotów które mają ze sobą kilka własności różnych, co pozwala uzasadnić pogląd, zgodnie z którym np. pacjentka jest identyczna z Maryją Dziewicą (Evnine, 1989, s. 7–8). Nie mamy tu jednak do czynienia z prawdziwym naruszeniem zasad logiki i sprzecznością – osoba rozumie wówczas w inny sposób (w innym kontekście teoretycznym czy ramie) pojęcia takie jak „negacja”, „identyczność” czy „bycie martwym”.

Problemy z przytoczonymi koncepcjami są dwa – zacznę od mniej zasadniczego, by później pokazać bardziej ogólny kłopot metodologiczny tkwiący w (PC) widoczny w tych przykładach. Po pierwsze, jak zauważają Bayne i Pacherie (2005) oraz Bortolotti (2010), nie wszystkie urojenia mogą być określone jako przekonania ramowe, bowiem w wielu wypadkach chorzy odczuwają nieprawdopodobieństwo treści swoich urojeń; nie wszystkie urojenia stawiają nas przed podobnym dylematem zastosowania (PC) (np. urojenia prześladowcze), choć je też moglibyśmy opisać jako przypadki wprowadzenia do systemu przekonań kolejnych przekonań ramowych. W wielu innych dziedzinach rozumowanie osób z urojeniami pozostaje niezmiennie poprawne, zdają się też rozpoznawać rozdźwięk między logiką a ich słowami (jak w przytoczonym wyżej badaniu Nishio i Mori, 2012).

Po drugie, z tego, że osoby chore inaczej niż logicy czy lekarze używają pojęcia bycia martwym lub negacji nie wynika, że inaczej je rozumieją i co innego dla nich znaczy. Dlaczego mielibyśmy przyjąć, że rozumieją je tak samo? Jak już zaznaczyłem, pacjenci, gdy nie pytamy ich o treść ich urojeń, wydają się rozumować w sposób klasyczny, a nie zgodny z jakkolwiek „dziwaczną logiką”. PC wymagałoby w tym kontekście postulowania posiadania przez podmiot systemu wnioskowania wyjaśniającego zmiany przez podmiot jednego sposobu rozumowania na inny w zależności od treści wnioskowania. Przypomnijmy sobie jednak na chwilę pytanie dotyczące prawdopodobieństwa pewnych zdarzeń i faktów w badaniach nad heurystykami. Istnieją kontekstualizacje³¹, które znacząco podnoszą ilość poprawnych odpowiedzi wśród badanych – i faktycznie badacze opisują je poprzez użycie różnych sposobów wnioskowania. Czy jednak znaczy to, że dla badanych słowo „prawdopodobieństwo” zna-

³⁰ Za Quinem przez „dziwaczną logikę” rozumiem tutaj system rozumowania, w którym terminy takie jak *negacja*, *identyczność* czy *koniunkcja* mają odmiennie własności od ich odpowiedników w logice klasycznej (sposób wnioskowania pozostaje jednak ten sam).

³¹ Patrz choćby wspomniany artykuł Fiedlera (1988).

czy co innego w jednym, a co innego w drugim kontekście? Nie – raczej, że posiadają sprzeczne intuicje dotyczące interpretacji konkretnych sytuacji i co więcej nie są to intuicje uświadomione, a zmiana rozumowania nie jest wynikiem decyzji.

Hipotezą znacznie bardziej płodną jest uznanie, że w tych wszystkich przypadkach mamy do czynienia z obiektywnie fałszywymi zdaniami również w idiolekcie podmiotu, które jednak uznaje za prawdziwe. Problem z pacjentami z urojeniem Cotarda, czy kłopotami badanych z interpretacją prawdopodobieństwa i zasad logiki w badaniach Kahnemana, Tversky'ego i Wasona, nie jest problemem czysto językowym. Korzystając z PC, zadajemy wciąż pytanie: w jaki sposób racjonalnie ktoś może być przekonany o rzeczy podobnie niedorzecznej i każda kolejna odpowiedź na nie wydaje się być wrażliwa na kontrprzykłady, których dostarczają nam dane empiryczne. Prowadzi to zatem do trudnej do utrzymania w naukach empirycznych niestabilności hipotez. Dziwne wydaje się zresztą preferowanie hipotezy o obcym i podobnie nieracjonalnym sposobie używania języka nad hipotezą o posiadaniu podobnie obcego nam i nieracjonalnego systemu przekonań. Użyteczniejsze naukowo jest przyjęcie, że dana osoba wierzy w niedorzeczność i próba docieczenia: jakie mogą być niezależne od podmiotu przyczyny³² powstania takiego systemu formowania przekonań? To pytanie bowiem jest, na gruncie przyjętego założenia, rozstrzygalne empirycznie i stanowi krok do przodu w wyjaśnianiu naukowym.

Na temat tego, na ile postępowanie w sposób zgodny z PC i PZN-N (to nie to samo, co powiedzieć: według PC i PZN-N) jest niezbędne w interpretowaniu języka i zachowania całych społeczności, nie chcę wnikać. Jeśli wierzyć niektórym badaniom antropologicznym (pisali o tym np. Rudiak-Gould [2010] czy Thagard i Nisbett [1983, s. 253–255]) przypisanie całym społecznościom posiadania pewnych konkretnych sprzecznych przekonań, podobnie jak treść urojeń uznawanych zarazem za sprzeczne i prawdziwe przez osoby je podtrzymujące³³, nie jest problematyczne i może prowadzić do interesujących wniosków. Dane empiryczne są jednak zbyt nieliczne, a metodologia badań antropologicznych zbyt niejednorodna, aby wyciągać rozstrzygające wnioski.

³² Warto zaznaczyć, że (odwracając Davidsonowskie hasło) chodzi tu o przyczyny, które nie są racjami.

³³ Rudiak-Gould (2010) przypisuje tubylcom zamieszkującym Wyspy Marshalla sprzeczne przekonania dotyczące przeszłości, która jest przez nich przedstawiana jednocześnie jako idylla, którą zniszczyło przybycie kolonizatorów, i jako stan wojny wszystkich ze wszystkimi, której kres przyniosło dopiero przyniesienie przez kolonizatorów chrześcijańskiej moralności. Badacz tłumaczy ten fakt ugruntowaniem tych przekonań w różnych systemach tożsamości: narodowej czy wspólnotowej (Marshalliańskiej) i religijnej (chrześcijańskiej). Tubylcy zauważali sprzeczność w swoich wizjach przeszłości, jednak nie potrafili porzucić żadnej z nich.

3. ZAKOŃCZENIE

W artykule starałem się przeanalizować dokładnie argumentację przytaczaną za przyjęciem psychologicznej zasady niesprzeczności. Wyszczególniłem jej dwa różne odczytania: deskryptywne i normatywne, odpowiadające z grubsza podejściu realistycznemu bądź instrumentalistycznemu wobec psychologii potocznej. Przytoczyłem argumentację za odczytaniem deskryptywnym PZN, zarówno odwołując się do interpretacji przekonań jako własności, jak i do systemowych funkcji ludzkiego umysłu, jaką miałyby być podtrzymywanie prawdziwych przekonań. Pokazałem, że obydwa te argumenty są niewystarczające dla przyjęcia PZN. Następnie wskazałem bardziej ogólny argument pokazujący, dlaczego PZN w odczytaniu deskryptywnym może być interpretowana jedynie jako hipoteza empiryczna i przedstawiłem badania z zakresu psychologii poznawczej i klinicznej, które jej przeczą. Później wskazałem dwie zasadnicze linie argumentacji za PZN w odczytaniu normatywnym: argument ze strategii intencjonalnej Daniela Dennetta oraz argumentację Donalda Davidsona i Willarda Quine'a, zgodnie z którą naukowe przypisanie przekonań wymaga założenia niesprzeczności przekonań. Na przykładzie debaty dotyczącej interpretacji przypadków urojeń w psychologii klinicznej pokazałem, dlaczego strategia metodologiczna oferowana przez Davidsona i Quine'a prowadzi do wysokiej niestabilności hipotez wyjściowych i w związku z tym nie jest bardziej użyteczna naukowo, niż przypisanie choremu sprzecznych przekonań.

W świetle powyższej argumentacji należy uznać, że psychologiczna zasada niesprzeczności w przyjętym tu sformułowaniu nie znajduje satysfakcjonującego uzasadnienia ontologicznego ani metodologicznego. Jako że zasada niesprzeczności wydaje się być jednym z kluczowych elementów racjonalności proceduralnej, wątpliwe w związku z tym jest również przyjmowane przez wielu teoretyków twierdzenie, zgodnie z którym przypisywanie przekonań i innych stanów intencjonalnych wymaga założenia racjonalności interpretowanego podmiotu. Otwarte pozostaje pytanie, czy zaoferowana tutaj krytyka wymaga przeformułowania założeń interpretacjonizmu najsilniej postulującego podobne założenie, czy też przyjęcia innego modelu przypisywania przekonań – uważam jednak, że konstrukcja takiego modelu jest możliwa i pozwoli na bardziej adekwatne wyjaśnienie danych empirycznych pokazujących, że racjonalność jest dla człowieka cechą znacznie bardziej przygodną, niż sugerują to niektórzy filozofowie.

BIBLIOGRAFIA

- Bar-Hillel, M., Neter, E. (1993). How Alike Is It Versus How Likely Is It: A Disjunction Fallacy in Probability Judgments. *Journal of Personality and Social Psychology*, 65(6), 1119–1131.
- Bortolotti, L. (2010). *Delusions and Other Irrational Beliefs*. Oxford: Oxford University Press.

- Breen, N., Caine, D., Coltheart, M., Hendy, J., Roberts, C. (2000). Towards an Understanding of Delusions of Misidentification: Four Case Studies. *Mind & Language*, 15(1), 74–110.
- Campbell, J. (2001). Rationality, Meaning and the Analysis of Delusion. *Philosophy, Psychiatry, & Psychology*, 8(2–3), 89–100.
- Carroll, L. (1995). What the Tortoise Said to Achilles. *Mind*, 104(416), 691–693.
- Ciecierski, T. (2017). Attitudes and Normativity. *Axiomathes*, 27(3), 265–283.
- Cohen, L. J. (1981). Can Human Irrationality Be Experimentally Demonstrated? *The Behavioral and Brain Sciences*, 4(3), 317–370.
- Currie, G. (2000). Imagination, Delusion and Hallucinations. *Mind & Language*, 15(1), 168–183.
- Davidson, D. (1992). Przekonania a podstawy znaczenia. W: D. Davidson (red.), *Eseje o prawdzie, języku i umyśle* (s. 118–140). Warszawa: PWN.
- Davidson, D. (1992). Ku jednolitej teorii znaczenia i działania. W: D. Davidson (red.), *Eseje o prawdzie, języku i umyśle* (s. 141–159). Warszawa: PWN.
- Dennett, D. (1978). *Brainstorms*, Cambridge: MIT Press.
- Dennett, D. (1987). True Believers: The Intentional Strategy and Why It Works. W: D. Dennett (red.), *The Intentional Stance* (s. 13–36). Cambridge: MIT Press.
- Dennett, D. (1981). Making Sense of Ourselves. *Philosophical Topics*, 12(1), 63–81.
- Dennett, D. (2015). Not Just a Fine Trip Down Memory Lane: Comments on the Essays on Content and Consciousness. W: C. Muñoz-Suárez, F. De Brigard (red.), *Content and Consciousness Revisited* (s. 199–218). New York: Springer.
- Dub, R. (2015). The Rationality Assumption. W: C. Muñoz-Suárez, F. De Brigard (red.), *Content and Consciousness Revisited* (s. 93–111). New York: Springer.
- Egan, A. (2008). Imagination, Delusion, and Self-Deception. W: T. Bayne, J. Fernandez (red.), *Delusion and Self-Deception: Affective and Motivational Influences on Belief Formation (Macquarie Monographs in Cognitive Science)* (s. 263–280). Londyn: Psychology Press.
- Ellis, H. D., Young, A. W. (1990). Accounting for Delusional Misidentifications. *British Journal of Psychiatry*, 157(2), 239–248.
- Evnine, s. J. (1989). Understanding Madness? *Ratio*, 2(1), 1–18.
- Fiedler, K. (1988). The Dependence of the Conjunction Fallacy on Subtle Linguistic Factors. *Psychological Research*, 50(2), 123–129.
- Gottlieb, P. (2007). Aristotle on Non-Contradiction. *The Stanford Encyclopedia of Philosophy*, pobrane z: <https://plato.stanford.edu/archives/sum2015/entries/aristotle-noncontradiction/>
- Hodges, W. (1977). *Logic*. Harmondsworth: Penguin Books.
- Joseph, M. (2004). *Donald Davidson*. Chesham: Acumen.
- Kahneman, D., Tversky, A. (1971). Belief in the Law of Small Numbers. *Psychological Bulletin*, 76(2), 105–110.

- Kahneman, D., Tversky, A. (1983). Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review*, 90(4), 293–315.
- Konold, C., Pollatsek, A., Well, A., Lohmeier, J., Lipson, A. (1993). Inconsistencies in Students' Reasoning about Probability. *Journal for Research in Mathematics Education*, 24(5), 392–414.
- Kripke, s. (1979). A Puzzle About Belief. W: A.Margalit (red.), *Meaning and Use* (s. 239–283). Dordrecht: D. Reidel.
- Lewis, H. A., Cooper, D. (1979). The Argument from Evolution. *Proceedings of the Aristotelian Society*, 53, 207–237.
- Łukasiewicz, J. (1987). *O zasadzie sprzeczności u Arystotelesa*. Warszawa: PWN.
- Maddy, P. (2012). The Philosophy of Logic. *The Bulletin of Symbolic Logic*, 18(4), 481–504
- Manktelow, K. I. (1981). Recent Developments in Research on Wason's Selection Task. *Current Psychological Reviews*, 1(3), 257–268.
- Marcus, R. B. (1983). Rationality and Believing the Impossible. W: R.B.Marcus (red.), *Modalities: Philosophical Essays* (s. 143–162). New York: Oxford University Press.
- McKay, R., Cipolotti, L. (2007). Attributional style in a case of Cotard delusion. *Consciousness and Cognition*, 16(2), 349–359.
- Nishio, Y., Mori, E. (2012). Delusions of Death in a Patient with Right Hemisphere Infraction. *Cognitive and Behavioural Neurology*, 25(4), 216–223.
- Pacherie, E., Bayne, T. (2005). In Defence of the Doxastic Conception of Delusions. *Mind & Language*, 20(2), 163–188.
- Quine, W. V. O. (1999). *Słowo i przedmiot*. Warszawa: Aletheia.
- Quine, W. V. O., Ullian, J. s. (1978). *The Web of Belief*. New York: Random House.
- Ramsey, F. (1926). Truth and Probability. W: R.B. Braithwaite, *The Foundations of Mathematics and other Logical Essays* (s. 156–198). New York: Harcourt Brace.
- Ramsey, F. (1931). *General Propositions and Causality*. W: D. H. Mellor (red.), *The Foundations of Mathematics* (s. 145–163), New York: Harcourt, Brace and Company.
- Richard, M. (1983). Direct Reference and Ascriptions of Belief. *Journal of Philosophical Logic*, 12(4), 425–452.
- Rudiak-Gould, P. (2010). Being Marshallese and Christian: a case of multiple identities and contradictory beliefs. *Culture and Religion: An Interdisciplinary Journal*, 11(1), 69–87.
- Schwitzgebel, E. (2010). Acting Contrary to Our Professed Beliefs or The Gulf Between Occurrent Judgment and Dispositional Belief. *Pacific Philosophical Quarterly*, 91(4), 531–553.
- Slater, B. H. (1995). Paraconsistent Logics? *Journal of Philosophical Logic*, 24(4), 451–454.
- Sober, E. (1978). Psychologism. *Journal for the Theory of Social Behaviour*, 8(2), 165–191.

- Stich, s. (1981). Dennett on Intentional Systems. *Philosophical Topics*, 12(1), 39–62.
- Stich, s. (1985). *Could Man Be Irrational Animal?* *Synthese*, 64(1), 115–135.
- Thagard, P., Nisbett, R. (1983). Rationality and Charity. *Philosophy of Science*, 50, 250–267.
- Wason, P. C. (1968). Reasoning About a Rule. *The Quarterly Journal of Experimental Psychology*, 20(3), 273–281.
- Wason, P. C. (1969). Regression in Reasoning? *British Journal of Psychology*, 60(4), 471–480.
- Wittgenstein, L. (2000a). *Traktat logiczno-filozoficzny*. Warszawa: PWN.
- Wittgenstein, L. (2000b). *Uwagi o podstawach matematyki*. Warszawa: Wydawnictwo KR.
- Wittgenstein, L. (2014). *O pewności*. Warszawa: Aletheia.

IS HAVING CONTRADICTORY BELIEFS POSSIBLE? DISCUSSION AND CRITIQUE OF ARGUMENTS FOR THE PSYCHOLOGICAL PRINCIPLE OF NON- CONTRADICTION

SUMMARY: The aim of this paper is to present and analyze arguments provided for the Psychological Principle of Non-Contradiction which states that one cannot have, or cannot be described as having, contradictory beliefs. By differentiating two possible interpretations of PNC, descriptive and normative, and examining arguments (ontological and methodological) provided for each of them separately I point out the flaws in reasoning in these arguments and difficulties with aligning PNC with the empirical data provided by research done in cognitive and clinical psychology. I claim that PNC cannot be derived from any metaphysical stance regarding the mental phenomena and that having contradictory beliefs should be regarded as possible. Furthermore, I argue that interpreting a subject as having contradictory beliefs, and therefore abandoning PNC, can be more effective in explaining the phenomena of contradictory beliefs and irrational behaviour than solutions consistent with the PNC.

KEYWORDS: principle of non-contradiction, rationality, principle of charity, intentional stance, belief ascription, delusion, folk psychology.