AVANT, wol. XII, nr 3
ISSN: 2082-6710 avant.edu.pl
DOI: 10.26913/avant.2021.03.02

Jak opisać nieracjonalność? Problem nieracjonalnych systemów przekonań a interpretacjonizm w filozofii umysłu

Maciej Tarnowski
Wydział Filozofii, Uniwersytet Warszawski
m.tarnowski3@student.uw.edu.pl

Przyjęto 7 listopada 2020; zaakceptowano 22 kwietnia 2021; opublikowano 23 września 2021.

Abstrakt

Jedną z najbardziej popularnych i kontrowersyjnych tez we współczesnej filozoficznej analizie psychologii potocznej i przypisywania przekonań jest założenie racjonalności podmiotu, zgodnie z którym każde przypisanie nastawienia sądzeniowego musi być poprzedzone założeniem racjonalności proceduralnej interpretowanego podmiotu. Wydaje się, że założenie to podważają badania w psychologii, socjologii i ekonomii, które dostarczają przykładów nieracjonalnych zachowań. Celem tego artykułu jest przedstawienie problemów napotykanych przez założenie o racjonalności i zaoferowanie alternatywnego modelu dla teorii interpretacji opartej na badaniach nad „szybkimi i oszczędnymi heurystykami" prowadzonymi w psychologii poznawczej. Identyfikuję dwie heurystyki: Heurystykę Podobieństwa i Heurystykę Uwłasnosłowienia, które wydają się grać kluczową rolę w przypisywaniu przekonań. Argumentuję, że podany model jest zarazem bardziej psychologicznie prawdopodobny i bardziej efektywny w wyjaśnianiu przypadków nieracjonalności.

Słowa kluczowe: racjonalność; heurystyki; interpretacja; zasada życzliwości; strategia intencjonalna; psychologia potoczna; przekonania

1. Wprowadzenie

Codzienna praktyka przewidywania, tłumaczenia, koordynowania czy oceniania cudzego zachowania wymaga od nas przypisywania stanów intencjonalnych. Powracającą i popularną w filozofii psychologii tezą jest twierdzenie, że warunkiem koniecznym przypisywania tych stanów jest założenie, iż opisywany podmiot jest proceduralnie racjonalny[1]. Popularność nadał jej Quine, zalecając tak zwaną zasadę życzliwości w interpretacji (1999), którą później rozwijał w swojej filozofii Donald Davidson (1974), jak również Daniel Dennett w swojej teorii „strategii intencjonalnej" (1987). W filozofii zajmującej się problemem interpretacji i pierwszych założeń psychologii potocznej podobne założenie jest często przyjmowane, zaś model oparty o nie, który określam w tym artykule zbiorczym mianem klasycznego interpretacjonizmu, jest przynajmniej jednym z najpopularniejszych na filozoficznym rynku.

Jednocześnie tocząca się w naukach społecznych – ekonomii, psychologii, socjologii – dyskusja nad ludzką racjonalnością każe nam zwątpić w to, czy faktycznie użyteczne jest charakteryzowanie człowieka jako „zwierzęcia racjonalnego". Szczególnie badania nad psychologią wnioskowania i podejmowania decyzji dostarczają nam danych świadczących o tym, że ludzie raczej rzadko bywają niż są proceduralnie racjonalni. Studia nad użyciem heurystyk czy błędów poznawczych występujących w procesie rozumowania nierzadko wskazują, iż ludzie często przyjmują pewne przesłanki lub akceptują wnioskowania niepoparte logicznym uzasadnieniem, zaś nierzadko – akceptują sądy wzajemnie sprzeczne. W badaniach nad racjonalnością coraz częściej odrzuca się założenie o wspólnej wszystkim proceduralnej racjonalności, przeciwstawiając temu próbę czystego opisu procesu wnioskowania oderwanej od porównywania jej z normatywną teorią racjonalności (patrz: Elqayam i Evans, 2011; Elqayam, 2012). Nie pociąga to jednak za sobą, wbrew analizie proponowanej przez interpretacjonistów, powściągliwości w używaniu pojęcia „przekonania".

Zwolennik założenia o racjonalności w procesie interpretacji stoi więc przed następującym dylematem: albo musi uzasadnić, dlaczego wyjaśnienie przypadków nieracjonalności proceduralnej opisywanej przez psychologię nie powinno wykorzystywać pojęcia „przekonanie", albo uzasadnić, dlaczego nie są to *de facto* przypadki nieracjonalności proceduralnej. Odpowiedzi formułowanych z pozycji interpretacjonistycznych, które podążają tymi drogami, nie uznaję za przekonujące[2]. Powody, dla których powyższe strategie nie są w stanie satysfakcjonująco wyjaśnić danych dostarczanych przez psychologię kliniczną i poznawczą, przedstawiam pokrótce w pierwszej części niniejszego artykułu. Wskazuję najważniejsze dane stanowiące trudność interpretacyjną dla klasycznego interpretacjonizmu i omówię dylematy z nimi związane. Drugim szczegółowym celem tego artykułu jest przedsta-

[1] Podążam tutaj za przyjmowanym przez wielu autorów (na przykład Bortolotti, 2010) podziale na racjonalność proceduralną (poprawność formalną przeprowadzanego wnioskowania), epistemiczną (racjonalną analizę i reakcję na dane ze świata) i sprawczą (spójność działań z przekonaniami i pragnieniami jednostki). Zwolennicy klasycznego interpretacjonizmu w filozofii umysłu kładą nacisk przede wszystkim na proceduralną racjonalność systemu przekonań – wobec czego poniższy artykuł polemizuje ściśle z założeniem o proceduralnej racjonalności interpretowanego podmiotu.

[2] Szczegółową krytykę tych stanowisk, opartą na szczególnym przypadku nieracjonalności, jakim jest sprzeczność przekonań, przedstawiam w innym tekście (Tarnowski, 2019). Ze względu na charakter niniejszego artykułu, część krytyczna została tutaj ograniczona do niezbędnego minimum.

wienie alternatywnego ujęcia przypisywania przekonań opartego o dobrze ugruntowane, skuteczne heurystyki poznawcze. Chciałbym pokazać, że takie podejście do psychologii potocznej, w przeciwieństwie do modelu klasycznego interpretacjonizmu, jest w stanie w bardziej intuicyjny sposób opisać przypadki nieracjonalności interpretowanych podmiotów oraz lepiej odpowiada faktycznym warunkom środowiskowym, w których interpretacja się odbywa.

Nie zamierzam jednocześnie zajmować stanowiska w sporze dotyczącym ontologii nastawień sądzeniowych. Za jedną z kluczowych zalet interpretacjonizmu w filozofii umysłu można uznać fakt, iż nie jest on zobowiązany do przyjęcia żadnego ze stanowisk: czy to realistycznego (przyjmowanego w ramach interpretacjonizmu, na przykład przez Donalda Davidsona i Davida Lewisa), czy instrumentalistycznego (przyjmowanego na przykład przez Daniela Dennetta). Pomimo tego, iż interpretacjonizm jest ogólną teorią przypisywania nastawień sądzeniowych (a więc, oprócz przekonań, także pragnień, intencji czy wyobrażeń[3]), w niniejszej pracy skupiam się przede wszystkim na przypisywaniu przekonań jako paradygmatycznym przypadku interpretacjonistycznej strategii wyjaśniania.

2. Interpretacjonizm a nieracjonalność

Wśród rozbudowanych koncepcji interpretacjonistycznych wymienić można dwie najbardziej wpływowe: klasyczną teorię interpretacji Davidsona, która stanowi bezpośrednie rozwinięcie teorii Quine'a, oraz strategię intencjonalną Dennetta. Obydwa modele dzieli oczywiście wiele w szczegółach – niezgodnych już choćby na poziomie statusu ontologicznego przekonań – jednak w obu w silny sposób przyjmuje się następującą tezę, którą określam tutaj jako Założenie Racjonalności Podmiotu:

> (ZRP) Podmiotowi P można przypisać nastawienia sądzeniowe (przekonania, pragnienia, intencje itd.) tylko wtedy, jeśli założy się, że P jest proceduralnie racjonalny.

W wizji interpretacjonistów ZRP jest warunkiem koniecznym rozpoczęcia procesu interpretacji i przypisywania przekonań, ponieważ – mówiąc zgrubnie – przypisywanie dużej liczby przekonań przy skromnych zazwyczaj danych początkowych wymaga normatywnego wzorca pozwalającego na tworzenie teorii konkretnego systemu przekonań. Bez założenia racjonalności nie mielibyśmy możliwości stworzenia modelu spójnego i efektywnego w wyjaśnianiu i przewidywaniu zachowania. Davidson pisze wprost:

> Przekonania, a także pragnienia, intencje i znaczenia, wyróżniamy i identyfikujemy na wiele sposobów. Jednakże decydującą rolę odgrywają relacje między przekonaniami: nie można zaakceptować zbyt dużych lub oczywistych odchyleń od racjonalności, jeśli przypisywanie przekonań ma być zrozumiałe. (…) W kontekście rozważanej teorii znaczy to, że na możliwe interpretacje zdań uznawanych za prawdziwe nakładamy warunek, iż są

[3] Szczegółową analizę typologii nastawień sądzeniowych, opartą na pracach Johna Searle'a (1983), przedstawia w swojej książce Tadeusz Ciecierski (2013, s. 127-185).

> one wzajemnie niesprzeczne logicznie. Inaczej mówiąc, zakłada się, że zbiór przekonań osoby mówiącej jest niesprzeczny (Davidson, 1980).

Nie do końca wiadomo, jaki zakres „zbyt dużych i oczywistych odchyleń od racjonalności" ma Davidson na myśli. Możemy jednak założyć, że teoria przekonań danego podmiotu w rozumieniu Davidsona musi zakładać pewną normatywność. Minimalnym warunkiem będzie logiczna niesprzeczność przypisanych przekonań oraz zgodność ich treści z podstawowym rachunkiem prawdopodobieństwa (Davidson, 1980). Zarówno Dennett, jak i Davidson zakładają, że natura przekonań jest holistyczna, zatem przypisując jedno przekonanie, musimy równocześnie założyć dla nich pewne „tło" innych przekonań. Innymi słowy, nigdy nie przypisujemy przekonań pojedynczo, a kluczową rolę w identyfikowaniu przekonań podmiotu będą odgrywać logiczne relacje między nimi.

Dla Davidsona racjonalność jest zatem niezbędna do stworzenia jasnej i klarownej procedury tworzenia teorii interpretacji. Jeśli nie będziemy mieli reguły, pozwalającej nam na określenie „tła", w którym dane konkretne przekonanie miałoby funkcjonować, nie będziemy w stanie przypisać poprawnie żadnego przekonania ani w ogóle rozpocząć procesu interpretacji. Można powiedzieć, że przypadki nieracjonalności, jeśli w ogóle jesteśmy w stanie je opisać, mogą pojawić się jedynie przy założeniu szerszego, racjonalnego tła powiązanych przekonań. W tym sensie „możliwość irracjonalności zależy od dużego stopnia racjonalności" (Davidson, 1982). Nie należy jej postrzegać jako braku racjonalności, ale jako odchylenie od normy, której dana osoba podlega.

Dennett postrzega ZRP w sposób silniejszy, często łącząc je z klasycznie pojmowaną definicją proceduralnej racjonalności zbioru przekonań jako jego niesprzeczności i domknięcia na modus ponens. Postulowana przez niego strategia intencjonalna to metoda przewidywania i wyjaśniania zachowań innych poprzez znajdowanie ich racjonalnych uzasadnień, która nadaje znaczenie terminom psychologii potocznej, takim właśnie jak „pragnienie" czy „przekonanie". ZRP jest dla niej, według Dennetta, zasadą konstytutywną – jedyną możliwą zarówno ze względów formalnych, jak i pragmatycznych, regułą organizującą proces przypisywania przekonań. Jak sam pisze:

> [Tworząc intencjonalną strategię danego systemu] zaczynamy od ideału doskonałej racjonalności i rewidujemy tak, jak dyktują nam to warunki. To znaczy, zaczynamy z założeniem, że ludzie są przekonani o wszystkich konsekwencjach swoich przekonań i nie posiadają żadnej sprzecznej pary przekonań (Dennett, 1981a, s. 21[4]).

Podobnie jak Davidson, uznaje on zatem „ideał racjonalności" za punkt wyjścia w procesie interpretacji zachowania innego podmiotu. W przeciwieństwie do niego uważa jednak, że nieracjonalności, według jego sformułowania, nie można w ogóle wyrazić w spójny sposób za pomocą „strategii intencjonalnej"[5].

[4] Przekłady fragmentów oryginalnej literatury obcojęzycznej są mojego autorstwa, o ile w bibliografii nie zaznaczono inaczej.

[5] Argument Dennetta odwołuje się do tego, że w naszych codziennych wyjaśnieniach nieracjonalnych działań bardzo często posługujemy się wyjaśnieniami, które Dennett uznaje za wyjaśnienia na niższym poziomie, przy użyciu „strategii funkcjonalnej". Patrz: Dennett, 1981b.

Każdy system intencjonalny – czy to ze względu na pragmatyczną charakterystykę teorii interpretacji, czy to w związku z wymogami strategii intencjonalnej – będzie dla klasycznego interpretacjonisty z konieczności proceduralnie racjonalny.

Jednymi z najczęściej podnoszonych argumentów przeciwko klasycznemu interpretacjonizmowi są wnioski z badań psychologicznych nad ludzkim wnioskowaniem, ukazujących wiele przypadków zachowań trudnych do zinterpretowania jako racjonalne proceduralnie. Najbardziej znane z tej grupy są przypadki opisane przez Petera Wasona oraz Daniela Kahnemana i Amosa Tversky'ego, które pokazują systematyczne i dominujące wśród populacji błędy w rozwiązywaniu prostych zadań logicznych (Wason, 1968; Johnson-Laird i Wason, 1970) lub dotyczą elementarnych zasad rachunku prawdopodobieństwa (Kahneman i Tversky, 1971; 1983). Prosty przykład tak zwanego błędu koniunkcji, zauważonego w pracach Kahnemana i Tversky'ego, może wskazać, dlaczego system przekonań osoby popełniającej ów błąd można uznać za niespójny. W badaniach Tversky'ego i Kahnemana badanych poproszono między innymi o uszeregowanie zestawu pewnych zdań od najbardziej do najmniej – w ich przekonaniu – prawdopodobnego. Pytanie oraz podane odpowiedzi wyglądały następująco:

> Załóżmy, że Bjorn Borg [szwedzki tenisista, który w roku poprzedzającym badanie czterokrotnie tryumfował w turniejach wielkoszlemowych] osiągnie finał Wimbledonu w 1981 roku. Uszereguj podane możliwości od najbardziej do najmniej prawdopodobnej:
>
> A. Borg wygra mecz
>
> B. Borg przegra pierwszy set
>
> C. Borg przegra pierwszy set, ale wygra mecz
>
> D. Borg wygra pierwszy set, ale przegra mecz (Tversky i Kahneman, 1983, s. 301-302)

Wielu badanych za bardziej prawdopodobne uznało wydarzenia opisywane przez odpowiedź C. niż odpowiedź B. – co jest niemożliwe, gdyż zdarzenie opisywane przez C. jest iloczynem zdarzeń opisywanych przez A. i B. Jednocześnie w osobnym badaniu sprawdzono, czy za ów błąd nie odpowiada interpretacja koniunkcji jako implikacji i czy badani interpretują w odpowiedni sposób pojęcie prawdopodobieństwa (Tversky i Kahneman, 1983, s. 302). Możemy zatem przynajmniej niektórym badanych przypisać następujące trzy przekonania:

> (1) To, że Borg przegra pierwszy set, ale wygra mecz, jest bardziej prawdopodobne niż to, że przegra pierwszy set (B). $(P(C) > P(B))$
>
> (2) Prawdopodobieństwo tego, że Borg przegra pierwszy set, ale wygra mecz (C), jest prawdopodobieństwem koniunkcji dwóch stwierdzeń: (A) że Borg wygra mecz oraz (B) że Borg przegra pierwszy set. $(P(C) = P(A \wedge B))$
>
> (3) Prawdopodobieństwo dwóch stwierdzeń jest niższe lub równe prawdopodobieństwu jednego z nich. (dla dowolnych A, B: $\big(P(A \wedge B) \leq P(A)\big) \wedge \big(P(A \wedge B) \leq P(B)\big)$)

Przekonania (1)-(3) stanowią sprzeczną trójkę przekonań. Powszechność występowania podobnego błędu została wielokrotnie potwierdzona eksperymentalnie (zazwyczaj podatnych jest na niego od 50 do nawet 90 % badanych, zob. Fisk, 2016) i chociaż jego źródła czy nawet uznanie prowadzonego do niego wnioskowania za błędne są sprawą kontrowersyjną (Sides, Osherson, Bonini i Viale, 2002), to wciąż trudno oprzeć się wrażeniu, że posiadanie podobnego zestawu przekonań stanowi wyraźny problem dla interpretacjonisty, postulującego, że zbiór przekonań przypisywanych danemu podmiotowi musi spełniać normy racjonalności proceduralnej. Przypisanie badanym przekonań (1)–(3) na podstawie ich odpowiedzi nie wydaje się nastręczać problemu, wbrew przewidywaniom interpretacjonisty.

Drugą grupą przypadków, która wzbudza wciąż żywe dyskusje pomiędzy zwolennikami interpretacjonistycznego podejścia do psychologii potocznej oraz ich krytykami, są urojenia badane przez psychiatrów i psychologów klinicznych. W zespołach takich jak urojenie Capgrasa (przekonanie, że osoba bliska została zastąpiona przez sobowtóra) czy Cotarda (przekonanie, że jest się martwym lub przestało się istnieć) mamy do czynienia z przypadkami szczerych deklaracji pacjentów przeczącym najbardziej elementarnym prawom logiki. Osoba badana może więc zarazem twierdzić, że nie żyje, przyznawać, że mówi, i zgadzać się, że osoby martwe nie mogą mówić (patrz: Nishio i Mori, 2012, McKay i Cipolotti, 2007). Mamy więc w tych wypadkach do czynienia nie tyle z potocznymi błędami w rozumowaniu, ile z jasnym i zauważalnym również dla potocznego obserwatora odstępstwem do racjonalności. Czy jednak oznacza to, że takie przypadki „wymykają się opisowi w zwykłych terminach przekonań i pragnień" (Dennett, 1981b, s. 67)?

Spośród różnych prób uzgodnienia tych przypadków z ramą klasycznego interpretacjonizmu dwie, wskazane we wstępie, są najbardziej popularne. Przedstawię je krótko, by następnie wskazać, dlaczego według mnie są one niewystarczające.

Pierwszą z odpowiedzi jest hipoteza odwołująca się do rozróżnienia Noama Chomsky'ego na „błędy kompetencji" oraz „błędy wykonania" (ang. *competence errors* i *performance errors*)[6], przyjmowana na przykład przez Jonathana Cohena (1981) i Daniela Dennetta (1981b). Według takiego wyjaśnienia wskazywane powyżej błędy we wnioskowaniach przynależą do tej drugiej kategorii: świadczą o kierowaniu się przez interpretowany podmiot nie błędną regułą wnioskowania (co byłoby „błędem kompetencji"), tylko raczej niedokładnością, która zostałaby przez podmiot zauważona po dłuższej refleksji. Nie możemy zatem odmówić interpretowanym racjonalności, a jedynie opisać te przypadki właśnie w kategorii chwilowej pomyłki – czy to na niższym poziomie wyjaśniania (na przykład „wypadło mu to z głowy", „przeoczyła to", patrz: Dennett, 1981b, s. 67-70), czy to wysnuwając hipotezę, że podobna pomyłka zostanie przez podmiot skorygowana w sytuacji, gdy będzie miała odpowiednie po temu warunki (patrz: Cohen, 1981).

Jednak czy faktycznie przytoczone przykłady dają się wpisać w ramy „błędów kompetencji"? Z dwóch powodów wydaje mi się to nieprawdopodobne. Po pierwsze, hipoteza, zgodnie z którą każdy podmiot dokona korekty swojego osądu po wskazaniu mu jego omyłki, jest empirycznie fałszywa bądź w ogóle niefalsyfikowalna. Dobrą ilustracją tego jest replikacja wyników oryginalnego testu selekcji kart Wasona (1969). Badani na kilku etapach, przy coraz

[6] Terminów tych w odniesieniu do koncepcji Cohena i Dennetta używam za Stephenem Stichem (1985).

wyraźniejszym wskazywaniu przez eksperymentatora popełnionego błędu w rozumowaniu (aż po otwarte zanegowanie odpowiedzi badanego i wyczerpujące wyjaśnienie poprawnego rozwiązania), byli proszeni o ponowne przemyślenie swojej odpowiedzi. Choć niektórzy z nich zmieniali po jakimś czasie swoje stanowisko, istotna część badanych nie potrafiła nawet na ostatnim etapie badania dostrzec w swoim wnioskowaniu niczego nieracjonalnego. Czy zatem badanych, twardo obstających przy swojej odpowiedzi, nie można opisać jako posiadających sprzeczną parę przekonań? Uważam, że jest to co najmniej nieoczywista konkluzja. Podobne wyniki można zaobserwować w przypadkach urojeń, gdzie analogiczna wobec sytuacji eksperymentalnej Wasona metoda „sokratycznej dyskusji", mającej na celu wykazanie wewnętrznej sprzeczności wypowiedzi pacjentów, nie przynosi zawsze pożądanych rezultatów (Bortolotti, 2010, s. 86-96).

Bardziej zasadniczy problem tej odpowiedzi leży jednak, jak sądzę, gdzie indziej. Nawet jeśli bowiem przyjmiemy na wiarę przesłankę, zgodnie z którą wymienione przypadki są przypadkami „błędów wykonania", nie wynika z niej, że przypisywanie stanów intencjonalnych wymaga ZRP. Jeśli bowiem wierzyć, że błędy te są systematyczne wśród większości ludzi (jak świadczyłyby o tym przywoływane wyżej badania Wasona czy Tversky'ego i Kahnemana wskazujące na ich powszechność), to strategia przewidująca ich zachowanie w zgodzie z zasadami uzasadniającymi takie przypadki (a więc sprzecznej z normami proceduralnej racjonalności) jest skuteczniejsza od modelu interpretacjonistycznego podtrzymującego ZRP. Każda interpretacjonistyczna strategia powinna więc przynajmniej dopuszczać nieracjonalność, aby być skuteczna w przewidywaniu ludzkich zachowań. Na podobnej zasadzie wydają się funkcjonować reguły gramatyki: systematycznie popełniane lapsusy językowe po jakimś czasie zostają włączone do obowiązujących reguł (jeśli tylko są wystarczająco powszechne), nawet jeśli popełniający je użytkownicy języka są w stanie uświadomić sobie, że poprawna konstrukcja jest odmienna. Nie ma żadnego pragmatycznego powodu, aby przypadki nieracjonalności opisywać jako „błędy wykonania" i w związku z tym odmawiać podmiotom przypisywania nieracjonalnych przekonań, *nawet jeśli* są oni w stanie swoją nieracjonalność dostrzec.

Drugą z najpopularniejszych strategii poradzenia sobie przez interpretacjonizm z kłopotliwymi danymi empirycznymi jest postulat, który bierze swój rodowód z uwag dotyczących reguł przekładu Quine'a: jeśli podmiot deklaruje uznawanie zbioru zdań niezgodnych z normami proceduralnej racjonalności, należy to potraktować jako świadectwo nieadekwatności przekładu z języka lub idiolektu podmiotu na nasz język. Według takiej hipotezy nie jest zatem tak, że osoby cierpiące na zespół Cotarda jednocześnie są przekonane o tym, że żyją i że są martwe, zaś osoby badane przez Kahnemana i Tversky'ego o tym, że prawdopodobieństwo iloczynu dwóch zdarzeń jest wyższe niż prawdopodobieństwa jednego z nich. Należy raczej powiedzieć, jak twierdzą zwolennicy tej strategii, że takie osoby inaczej niż my *rozumieją* pojęcia „prawdopodobieństwa" czy „śmierci". Tylko wówczas możemy mieć wgląd w sposób rozumowania drugiej osoby, jeżeli – zamiast przypisywać nieracjonalny system przekonań – przypisujemy niestandardowe użycie pojęć znanych z języka publicznego.

Taki postulat, będący konsekwencją przyjmowanej przez Davidsona i Quine'a zasady życzliwości, nie daje nam dużo lepszego obrazu nieracjonalności; jeśli potraktujemy go jako rzeczywistą zasadę metodologiczną, to prędko dojdziemy do absurdu. Nie jest bowiem tak, że

dane osoby nie są w stanie wnioskować w zgodzie z rachunkiem prawdopodobieństwa czy poprawnie przypisywać bądź zdefiniować śmierci. Ich nieracjonalność jest ograniczona tylko do pewnych przypadków bądź konkretnych sformułowań problemu (patrz: Bayne i Pacherie, 2005, Bortolotti, 2010). Nieracjonalność nie wydaje się tutaj na tyle systematyczna, by można ją było w adekwatny sposób zracjonalizować, określając mianem „pomieszania pojęć". Jeśli jednak spróbujemy, prędko dotrzemy do absurdalnej konsekwencji, w zgodzie z którą każde sprzeczne z poprzednim zestawem reguł użycie jakiegoś pojęcia ustanawia nową regułę jego użycia. Oznaczałoby to wysoką niestabilność hipotez bazowych, jeśli próbowalibyśmy podobny postulat wprowadzić w naukach empirycznych. Dużo bardziej użyteczną – ze względu na swoją stabilność i zgodność z danymi, i stosowaną w tychże naukach – hipotezą wyjściową byłoby przypisanie interpretowanym podmiotom sprzecznych i nieracjonalnych przekonań.

Jeśli klasyczny interpretacjonizm nie jest w stanie podać przekonującego opisu teoretycznego tego, co nazywamy przypadkami nieracjonalności, powyższa krytyka skłania nas do odrzucenia Założenia Racjonalności Podmiotu zarówno jako twierdzenia deskryptywnego (wynikającego z natury przekonań lub danych empirycznych), jak i normatywnego (użytecznego lub koniecznego w metodologii psychologii).

3. Teza o heurystycznej naturze procesu interpretacji

Przy odrzuceniu Założenia Racjonalności Podmiotu interpretacjonizm w swojej klasycznej postaci wymaga więc przeformułowania swoich pierwotnych założeń. Czy w związku z odrzuceniem ZRP należy odrzucić cały interpretacjonizm? Uznaję, że taka konkluzja byłaby nieuprawniona. Żeby wyjaśnić możliwość nieracjonalności, należy jednak zaproponować daleko idące modyfikacje tego stanowiska.

W tym rozdziale przedstawię propozycję alternatywnego modelu przypisywania przekonań w ramach psychologii potocznej. Szkicując ją, stawiam sobie dwa cele. Po pierwsze: wyjaśnienie intuicyjności rozwiązań proponowanych w ramach interpretacjonizmu, przy jednoczesnym utrzymaniu fałszywości ZRP, czyli zaoferowanie wyczerpującego wyjaśnienia sposobu przypisywania nieracjonalnych proceduralnie zbiorów przekonań na poziomie przypisań indywidualnych[7], dokonywanych przez nas w codziennym życiu. Po drugie: zademonstrowanie, w jaki sposób przypisanie nieracjonalnych przekonań może być użyteczne naukowo, czyli dawać prostsze i bogatsze wyjaśnienie problematycznych przypadków oraz bardziej kompleksowe przewidywania na poziomie przypisań naukowych. Model tutaj zarysowany jest rozwinięciem uwag dotyczących charakteru teorii interpretacji Lisy Borto-

[7] Podążam tutaj za rozróżnieniem zaoferowanym przez Richarda Duba (2015) na przypisania indywidualne i naukowe: „[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). Przypisania indywidualne będą zatem tymi, które biorą się z intuicyjnych zastosowań reguł psychologii potocznej i to właśnie rządzący nimi mechanizm jest w niniejszym artykule najszerzej opisywany. Do kwestii relacji między nimi a przypisaniami naukowymi, które biorą się z określonych reguł metodologicznych zastosowanych w praktyce naukowej, odnoszę się pod koniec tego tekstu.

lotti – z nadzieją, że zachowana została towarzysząca im myśl. Nie rości on sobie również prawa do kompletności[8]; uważam, że zaproponowany tutaj sposób myślenia o przypisywaniu przekonań jest prosty i otwarty na dalsze, bardziej szczegółowe modyfikacje.

Charakterystycznym elementem teorii interpretacjonistycznych jest to, że przypisywanie przekonań opiera się w nich na specyficznej kalkulacji. Przypisuje się zestaw podstawowych przekonań, racjonalnych (zgodnie ze stosowanym tu rozróżnieniem) epistemicznie bądź sprawczo, następnie zaś, przy założeniu racjonalności proceduralnej podmiotu, niejako „oblicza się" przekonania, które powinien posiadać dany podmiot przy podanych założeniach. Ten sposób myślenia, poza krytyką zaoferowaną wyżej, narażony jest na pewien zarzut natury praktycznej. Podany przez interpretacjonistów sposób kalkulacji przekonań interpretowanego podmiotu na poziomie przypisań indywidualnych okazuje się co najmniej trudny w zastosowaniu w sytuacjach, w których dochodzi do interpretacji. Sytuacje takie, jak zaznaczono wyżej, cechuje konieczność podejmowania decyzji interpretacyjnych pod presją czasu i w stanie niepełnej lub nawet szczątkowej informacji. Przypisywanie przekonań w sposób podany przez klasyczne teorie interpretacjonistyczne wymagałoby zatem umiejętności niezwykle szybkiego i precyzyjnego wnioskowania logicznego, której większość ludzi prawdopodobnie nie posiada. Byłoby ono również narażone na poważne ryzyko błędu, gdyż informacja, do której nie mieliśmy dostępu w stanie wyjściowym, mogłaby znacząco wpłynąć na wynik naszego wnioskowania. Podana strategia nie wyjaśnia, które z tych informacji powinniśmy zdobyć za wszelką cenę bądź pominąć w toku interpretacji. Odpowiednia strategia interpretacji powinna zatem przy stosunkowo niewielkim wysiłku poznawczym, w środowisku, o którym nie mamy kompletnej informacji, oraz pod presją czasu prowadzić do uzyskania informacji prawdziwych, niezbędnych do przewidzenia zachowania innego podmiotu – czyli maksymalizować nasz zysk poznawczy.

W książce Lisy Bortolotti *Delusions and Other Irrational Beliefs* (2010), pod koniec rozdziału dotyczącego relacji między wymogiem proceduralnej racjonalności a statusem teoretycznym urojeń, napotykamy interesujący pod tym względem fragment:

> Założenia interpretatorki są elastycznymi i podatnymi na rewizję heurystykami, nie ograniczeniami. Powinny one prowadzić interpretatorkę i pomóc jej przypisywać stany intencjonalne ze zdeterminowaną treścią różnym podmiotom w różnych sytuacjach, lecz nie są one kryteriami, zgodnie z którymi decyduje ona, czy dana wypowiedź podmiotu wyraża przekonanie. (...) gdy już mamy tę prostą ramę pojęciową, nie musimy szukać warunków koniecznych dla intencjonalności w racjonalnym zachowaniu danego podmiotu, skoro posiadamy już pewne źródła do wyjaśnienia relacji między racjonalnością a intencjonalnością, bez konieczności uznawania twierdzenia, które jest odrzucane przez nasze codzienne doświadczenie jako interpretatorów (Bortolotti, 2010, s. 107).

[8] Z pewnością olbrzymią, a często pomijaną w filozofii analitycznej, rolę w naszej codziennej praktyce przypisywania przekonań pełni na przykład przypisywanie innym podmiotom cech charakteru, emocji czy prywatnych upodobań. Celem tej pracy nie jest zidentyfikowanie wszystkich możliwych heurystyk biorących udział w procesie interpretacji – jest to z pewnością także zadanie niewykonalne; myślę jednak, że jest to owocny kierunek rozważań dotyczących psychologii potocznej.

Błąd, który dostrzega Bortolotti w klasycznych teoriach interpretacjonistycznych, to opisywanie heurystyk, którymi posługujemy się przy interpretacji, jako warunków koniecznych przypisywania stanów intencjonalnych. Tezę, którą formułuje przy tej okazji, można określić jako tezę o heurystycznym charakterze teorii interpretacji. Zgodnie z nią teoria interpretacji, którą posługujemy się na co dzień w przypisywaniu przekonań, ma być zbiorem odpowiednich heurystyk: prostych w zastosowaniu i skutecznych strategii służących do określania zbioru przekonań interpretowanego podmiotu.

Za przyjęciem takiej ramy dla interpretacjonizmu stoi teza, zgodnie z którą większość ludzkich decyzji opiera się na heurystykach poznawczych. Nurt w psychologii, przyjmujący podobny paradygmat w modelowaniu ludzkiego wnioskowania, ma swój początek w cytowanych w tej pracy badaniach Wasona oraz Kahnemana i Tversky'ego mieszczących się w tak zwanym programie badań nad heurystykami i błędami poznawczymi (ang. *heuristics and biases*); w pracach tych zwracano uwagę najczęściej na pomyłki w osądzie lub wnioskowaniu, powodowane przez stosowanie heurystyk. Heurystyki w rozumieniu Kahnemana stosowane są głównie ze względu na swoją obliczeniową prostotę i szybkość w zastosowaniu; nie mogą one jednak w swojej dokładności zastąpić skomplikowanej analizy i kanonów proceduralnej racjonalności (Wójtowicz i Winkowski, 2018, s. 258-260)[9]. Mówiąc mniej zawile, można powiedzieć, że stosowane są przez nas w celu szybkiego znalezienia optymalnego rozwiązania, jedynie racjonalna analiza zaś, przeprowadzona w drugiej kolejności, może nas upewnić w tym, że jest to faktycznie rozwiązanie optymalne. W latach dziewięćdziesiątych ten sposób widzenia heurystyk został zakwestionowany przez badaczy pracujących w tak zwanym programie „szybkich i oszczędnych" heurystyk (ang. *fast and frugal heuristics*)[10], inspirowanych pracami Herberta Simona. Według nich uzasadnieniem dla użycia heurystyk jest nie tylko obliczeniowa prostota, ale również skuteczność niemożliwa do zapewnienia przez inne strategie wnioskowaniam która wynika z dostosowania do konkretnego środowiska. Odpowiedni zbiór adekwatnych środowiskowo strategii określa się w tym programie badawczym „racjonalnością ekologiczną" (patrz: Gigerenzer i Sturm, 2011). „Ekologicznie racjonalne" heurystyki mogą zatem stanowić nie tyle narzędzie zastępujące nam skomplikowane obliczenia, co podstawową metodę analizy problemów występujących w środowisku.

Niezależnie od przyjętej przez nas tradycji psychologicznej, użycie heurystyk prawdopodobnie odpowiada za większość z naszych wnioskowań dotyczących przewidywania zjawisk i podejmowania decyzji w zgodzie z nimi. Kompetencja przypisywania przekonań

[9] Jak zauważył jeden z recenzentów tego artykułu, podobne wyjaśnienie użycia heurystyk jako „prostszych obliczeniowo" jest obecnie poddawane w wątpliwość (patrz np. Van Rooij, Wright i Wareham, 2012). Nie należy tego jednak traktować jako argumentu przeciwko wyjaśnieniom naszych procesów poznawczych w kategoriach heurystyk, a raczej przeciwko określonej wizji heurystyk jako alternatywnych metod rozwiązywania problemów obliczeniowo trudnych (ang. *intractable*). Iris Van Rooij i współpracownicy (2012, s. 481-483) jako przykład dobrego wyjaśnienia czyniącego użytek jedynie z obliczeniowo prostych, dostosowanych środowiskowo heurystyk podają paradygmat „ekologicznej racjonalności" Gigerenzera, na którym opieram się dalej w tekście.

[10] Dyskusję między tymi stanowiskami można odnaleźć w artykułach: Gigerenzer, 1991, 1996, Kahneman i Tversky, 1996.

jest zatem bezpośrednio związana z ich użyciem; w istocie większość potocznie uznawanych komunałów, które współtworzą „prawa" psychologii potocznej, można by uznać za heurystyki.

Jak wskazano wyżej, interpretując jakąś regułę jako heurystykę poznawczą, musimy sprawić, aby spełniała ona kilka warunków. Gigerenzer definiuje heurystykę jako „strategię, która ignoruje część informacji w celu szybszego, oszczędniejszego i/lub trafniejszego od bardziej skomplikowanych metod podejmowania decyzji" (Gigerenzer i Gaissmaier, 2011, s. 454). Aby uznać, że dana strategia jest heurystyką faktycznie wykorzystywaną przez interpretatorów, musi ona oczywiście być też zgodna z naszymi potocznymi przypisaniami. Za Lisą Bortolotti uznaję, że ZRP w odpowiednim sformułowaniu jest właśnie tego typu heurystyką, pozwalającą nam „szybko i oszczędnie" przypisywać przekonania, nie zaś warunkiem koniecznym, pozwalającym na takie przypisanie.

Wróćmy więc do sformułowania ZRP przytaczanego w tej pracy:

> (ZRP) Podmiotowi P można przypisać nastawienia sądzeniowe tylko wtedy, jeśli założymy, że P jest racjonalny.

Psychologiczne przypadki nieracjonalności, opisywanej za pomocą pojęć intencjonalnych, jak i wady propozycji wysuwanych przez zwolenników ZRP skłaniają do uznania, że ZRP w przytoczonym tu brzmieniu jest fałszywe. W jaki sposób wytłumaczyć jednak intuicyjną trafność argumentów Dennetta i Davidsona? Podana tu propozycja, mam nadzieję, jest w stanie pogodzić te intuicje z niepowodzeniami zastosowania interpretacjonistycznego modelu, w szczególności z przypadkami przypisywania nieracjonalnych przekonań.

Uznaję, że źródeł większości wątpliwości, towarzyszących nam podczas rozważań dotyczących zjawiska przypisywania nieracjonalnych przekonań, można szukać w zastosowaniu heurystyki, którą możemy nazwać Heurystyką Podobieństwa:

> (HP) Przypisuj podmiotowi te przekonania, które *uznajesz za racjonalne*, biorąc pod uwagę własną wiedzę na temat środowiska, preferencji oraz inne (przypisane uprzednio) pragnienia i przekonania podmiotu.

Strategia, którą oferuje nam ta heurystyka, jest bardzo prosta. Zamiast zakładania perfekcyjnej proceduralnej racjonalności podmiotu, przypisujemy te przekonania, które albo wydają nam się racjonalne *per se* (chociaż nie muszą takie być według normatywnego wzorca), albo są w naszym mniemaniu racjonalne, jeśli weźmiemy pod uwagę resztę przekonań i pragnień podmiotu, środowisko, w którym się znajduje, itp. Co ważne, strategia ta nie dostarcza nam danych o sposobie wnioskowania, który doprowadził dany podmiot do takich, a nie innych przekonań. Jest to tylko i wyłącznie strategia pozwalająca nam na dość dokładne określenie przekonań podmiotu w sytuacji, gdy brakuje nam wielu podstawowych danych, które mogłyby pozwolić nam równie dokładnie opisać podmiot z punktu widzenia „strategii intencjonalnej" lub Davidsonowskiej „interpretacji radykalnej".

HP posiada również kilka oczywistych zalet heurystyki, które sprawiają, że bardziej prawdopodobne jest stosowanie przez interpretatorów w codziennych sytuacjach właśnie jej, nie zaś ZRP postulowanego przez Dennetta czy Davidsona. HP nie wymaga od interpretatora podążania za jednym, często nieintuicyjnym, kanonem idealnej racjonalności, a jedynie

przypisywania przekonań, które *samemu uznaje się za racjonalne*. Jest to pierwsza kluczowa różnica pomiędzy HP a ZRP. HP odstępuje bowiem od nieintuicyjnej konsekwencji interpretacjonizmu, zgodnie z którą każdy interpretator posiada umiejętność idealnego wnioskowania logicznego. Podobnie jak podmioty, których zachowanie interpretujemy, często sami korzystamy z heurystyk podczas wnioskowania, zatem wykorzystanie takiej analogii lepiej pozwala nam przewidzieć nieracjonalne z klasycznego punktu widzenia zachowania podmiotu. Drugą kluczową różnicą jest sam sposób, w który dokonujemy przypisania. Stosowanie HP czyni bowiem zadość Lewisowskiemu kryterium racjonalności treści (Lewis 1986): wielu przekonań czy pragnień, których nie uznajemy za racjonalne ze względu na ich treść, niezależnie od możliwej ścieżki wnioskowania, które mogłoby do nich prowadzić, w ogóle nie rozpatruje się jako możliwych wyników użycia HP. Do przykładów podawanych przez Lewisa należy chociażby silne pragnienie, aby zjeść spodek pełen błota, oraz przekonanie, że sól w Australii bywa czasem słodka. Nie możemy odrzucić przypisania żadnego z tych przekonań, jeśli posługujemy się wyłącznie kryterium racjonalności proceduralnej. Tym niemniej nie robimy tego, uznając, że przekonania i pragnienia interpretowanego podmiotu są „zdroworozsądkowe" – co przecież nie znaczy „proceduralnie racjonalne". HP ignoruje zatem część informacji i możliwych wyników: jest oszczędna i skuteczna, bowiem ignorowane informacje faktycznie najczęściej nie mają wpływu na wynik, który potrzebujemy uzyskać; nie byłoby to jednak możliwe, gdybyśmy chcieli podążać za kanonem racjonalności proceduralnej. Zamiast więc zakładać, że opisywany podmiot formuje przekonania w pewien sposób (i na tej podstawie przypisywać konkretne przekonania), przypisujemy mu przekonania, wykorzystując nasz własny system formowania przekonań, który również nie musi podlegać normom proceduralnej racjonalności.

Jednak takie wytłumaczenie intuicyjności tez interpretacjonistów i wątpliwości, które rodzą w nas przypadki drastycznego odejścia od norm racjonalności, nie wystarcza do opisania wszystkich sytuacji, w których przypisujemy podmiotom nieracjonalne przekonania. Niektóre przekonania są bowiem na tyle nieracjonalne, że bardzo rzadko znalazłyby się wśród przekonań przypisywanych za pomocą HP; tym niemniej dokonujemy takich przypisań, jak w klinicznych przypadkach urojeń. HP pozwala nam intuicyjnie wyjaśnić przypadki lokalnych odstępstw od norm racjonalności proceduralnej. Aby jednak poradzić sobie z drugą zarysowaną w pierwszej części artykułu trudnością za pomocą HP, musimy zastanowić się, jaka heurystyka mogłaby tłumaczyć przypisywanie przekonań stanowiących przypadki skrajnej nieracjonalności w formułowaniu sądów.

Pomocne w znalezieniu podobnej heurystyki może być rozważenie, jakiego rodzaju świadectwa zazwyczaj prowadzą nas do przypisania podobnych przekonań. Można zauważyć, iż w znaczącej większości przypadków podstawą dla takich przypisań będzie zachowanie werbalne. Chociaż faktem jest, iż na przykład osoby cierpiące na urojenie Cotarda często przejawiają inne zachowania, mogące płynąć z przekonania o własnej śmierci (na przykład głodzą się, gdyż uznają, iż martwe osoby nie muszą jeść), to podstawą przypisania im tego przekonania w większości wypadków będzie wywiad kliniczny przeprowadzany z pacjentem przez lekarza.

Za punkt wyjścia niech posłuży nam zatem uznawana często za jeden z niezbędnych elementów poprawnej teorii interpretacji Zasada Uwłasnosłowienia (ang. *disquotation principle*). Sformułowana została w słynnym artykule Saula Kripkego „A Puzzle About Belief" (1979) i wciąż (np. Ciecierski, 2013; Puczyłowski, 2014) jest żywo dyskutowana na polu filozofii psychologii i języka. Jak czytamy w oryginalnym artykule:

> Wypowiedzmy zatem wyraźnie *zasadę uwłasnosłowienia* zakładaną tutaj, łączącą szczerą akceptację i przekonanie. Może być sformułowana następująco, gdzie 'p' może zostać zastąpione przez każde poprawne zdanie języka polskiego: „Jeśli zwykły użytkownik języka polskiego, po namyśle, szczerze akceptuje 'p', to jest przekonany, że p". Zdanie zastępujące 'p' powinno nie zawierać okazjonalizmów lub wyrażeń zaimkowych, lub dwuznaczności, które byłyby wbrew intuicyjnemu znaczeniu zasady (…). Kiedy przyjmujemy, że mamy do czynienia ze zwykłym użytkownikiem polskiego, uznajemy, że używa on wszelkich słów w zdaniu w sposób standardowy, układając je zgodnie z regułami poprawnej gramatyki, etc.: w skrócie, rozumie zdanie tak, jak rozumie je zwykły użytkownik języka (Kripke, 1979, s. 248-249).

Po odpowiednim przeformułowaniu (za Ciecierski, 2013, s. 255), możemy ją zapisać jako:

> (ZU) Jeśli szczery, racjonalny i kompetentny użytkownik języka J potwierdza po namyśle zdanie 'p' języka J, to, o ile zdanie nie zawiera wyrażeń okazjonalnych, jest przekonany, że p.

Kripke używa tej zasady, by pokazać, że korzystając z niej w połączeniu z Millowską koncepcją bezpośredniego odniesienia nazw własnych, docieramy do pewnego paradoksu: postępując w zgodzie z tymi koncepcjami, interpretator lub interpretatorka może przypisać w pełni racjonalnemu i kompetentnemu użytkownikowi języka parę przekonań sprzecznych lub przekonanie wewnętrznie sprzeczne. Jedną z podanych przez Kripkego ilustracji tej zagadki jest następująca sytuacja: Pierre, kompetentny użytkownik języka francuskiego, mieszka w Paryżu, i wchodząc w kontakt z wizerunkiem medialnym Londynu (Big Benem, Tamizą, Tower of London itp.), dochodzi do szczerego uznania zdania *Londres est jolie* (z fr. „Londyn jest ładny"). Następnie, po przeprowadzce do ubogiej, robotniczej dzielnicy stolicy Wielkiej Brytanii i opanowaniu (w kontakcie z „niewyedukowanymi sąsiadami") podstaw języka angielskiego, dochodzi do uznania zdania w języku angielskim: *London is not pretty* (z ang. „Londyn nie jest ładny"). Poprzez użycie ZU zmuszeni jesteśmy przypisać Pierre'owi parę przekonań:

(1) Londres jest ładny.

(2) London nie jest ładny.

Zgodnie zaś z teorią bezpośredniego odniesienia, nazwy własne o tym samym odniesieniu mają również to samo znaczenie; w związku z tym jesteśmy zobowiązani przypisać Pierre'owi parę przekonań sprzecznych, jednocześnie nie mogąc odmówić przypisania mu racjonalności:

(1') Londyn jest ładny.

(2') Londyn nie jest ładny.

Zagadka Kripkego stanowi faktyczny problem dla normatywnych teorii przekonań i wiedzy. Jeśli zgodzimy się na realność przedstawionego paradoksu, dopuszczam, że wnioskujący podmiot może posiadać przekonania logicznie sprzeczne, jednocześnie jednak (mówiąc słowami Kripkego) żadna doza czystej logiki lub introspekcji semantycznej nie jest w stanie sprawić, że odrzuci jedno z nich. Często wysnuwaną z tego paradoksu konkluzją jest to, że fałszywe jest jedno z użytych w wywodzie założeń: teoria bezpośredniego odniesienia nazw własnych bądź sama ZU. Rozwiązania proponowane przez różnych autorów wskazywały na konieczność wzmocnienia ZU: albo poprzez restrykcje modalne (Marcus, 1983), dodające warunek, zgodnie z którym przypisywane przekonanie musi wyrażać logicznie możliwy sąd, albo poprzez zrelatywizowanie przekonań do języka, którego zdaniem jest ‘p’ (Zabłudowski, 1986), lub sytuacji, w której zdanie zostało wypowiedziane (Ciecierski, 2013).

Wydaje się, że zagadka Kripkego i próby jej rozwiązania niejako z językowej strony odzwierciedlają problemy, które dotykają ZRP w argumentacji interpretacjonistów i oczekują, według mnie, analogicznego rozstrzygnięcia. Analogicznie Zasadę Uwłasnosłowienia można uznać za regułę wypływającą z naszej codziennej praktyki. Teoria bezpośredniego odniesienia zaś jest wciąż cieszącą się największą popularnością teorią semantyczną dla nazw. W poniższych rozważaniach uznaję całość argumentu Kripkego, nie uważając jednak przedstawionej zagadki za paradoks. Sprzeczność przekonań Pierre’a wydaje mi się realną, lecz nie dowodzącą wadliwości jego użycia, konsekwencją przyjętego aparatu teorii interpretacji.

Skupię się zatem na podanym tutaj sformułowaniu ZU i zastanowię się nad warunkami, w których mogłaby być wykorzystana w potocznym wnioskowaniu i przypisywaniu przekonań zgodnie z modelem heurystycznym. Przypomnijmy ją:

> (ZU) Jeśli szczery, racjonalny i kompetentny użytkownik języka J potwierdza po namyśle zdanie ‘p’ języka J, to o ile zdanie nie zawiera wyrażeń okazjonalnych, jest on przekonany, że p.

Tym, co sprawia, że akceptujemy Zasadę Uwłasnosłowienia, jest jej intuicyjność, czyli zgodność z codziennym doświadczeniem interpretatorów. Jednak czy faktycznie takie jej sformułowanie odzwierciedla regułę, którą kierujemy się w potocznej praktyce? Podobnie jak w wypadku ZRP, chciałbym postawić hipotezę wypływającą z tezy Bortolotti, zgodnie z którą takie jej ujęcie jest nadinterpretacją stosowanej przez nas w indywidualnych przypisaniach heurystyki.

Tym, co zwraca uwagę w warunkach nałożonych na ZU przez Kripkego oraz jego następców i następczynie, jest „namysł” i „szczerość”; są to jednocześnie (jak w wypadku ZRP) warunki, które czynią tę zasadę trudną do zastosowania w sytuacjach, w których stosowane są przypisania indywidualne. Kryteria te, co oczywiste, są nieobserwowalne i w związku z tym niemożliwe jest ustalenie przez zwykłego interpretatora tego, czy zostały one spełnione, czy nie. Jeśli chcemy zatem odczytywać ZU jako fragment naszej potocznej teorii interpretacji innych podmiotów, musimy przyjąć, że spełnienie przez podmiot tych kryteriów zostaje przez interpretatora założone w procesie przypisywania przekonań w sposób analogiczny do założenia racjonalności w modelu przypisywania przekonań klasycznego interpretacjonizmu.

Trudność, jaką dostrzegam w takiej możliwości, bierze się z analizy przyjmowanego w ZU warunku szczerości: jego użycie zahacza o kolistość w sformułowaniu tej zasady. Zgodnie ze standardowo przyjętym stanowiskiem Searle'a (1969) możemy uznać, że warunkiem koniecznym i wystarczającym szczerości wypowiedzi jest istnienie stanu intencjonalnego odpowiadającego jej treścią. Wówczas, zakładając szczerość interpretowanego podmiotu, musielibyśmy założyć również, że posiada on przekonanie, że p. Jeśli przyjmiemy zatem takie kryterium, wówczas ZU staje się trywialna, gdyż aby zadecydować, czy wypowiedź podmiotu spełnia warunki opisane w przesłankach, musimy przesądzić o spełnieniu przez podmiot warunku opisanego we wniosku. Przyjęcie zaś, że warunki szczerości nie obejmują posiadania odpowiedniego stanu intencjonalnego, wymaga zazwyczaj wskazania, że istnieją szczere wypowiedzi, którym nie towarzyszy adekwatne przekonanie[11] – a takich przypadków nie dopuszcza sama ZU. Oznacza to zatem, że sformułowanie ZU albo wymaga przyjęcia konkluzji w przesłankach, co czyni tę zasadę nieużyteczną w praktyce interpretacyjnej, albo też wymaga użycia argumentu, którego prawdziwość podważona zostaje przez sformułowanie zasady[12].

Jeśli więc chcemy uzyskać prostą w zastosowaniu i oszczędną heurystykę, powinniśmy odwrócić kierunek dotychczasowych poszukiwań i zamiast nakładać kolejne restrykcje na ZU, opuścić zbędne warunki obecne w jej pierwotnym sformułowaniu. Takie jej sformułowanie możemy określić jako Heurystykę Uwłasnosłowienia:

> (HU) Jeśli podmiot uznaje lub wypowiada zdanie 'p' w języku J, to przypisz mu przekonanie, że p.

Strategia wyznaczana przez tę heurystykę jest, podobnie jak wszystkie heurystyki, bardzo prosta i skuteczna w zastosowaniu, minimalizując przy okazji ilość informacji wymaganej do jej zastosowania. Nie nakłada ona na nas obowiązku sprawdzenia bądź założenia, że interpretowany podmiot jest szczerym i kompetentnym użytkownikiem języka. Posłużę się przykładem rzeczonej strategii: chociaż nie pretenduję w żaden sposób do bycia kompetentnym użytkownikiem języka rumuńskiego, wciąż moje pragnienia czy przekonania komunikowane za pomocą ułożonych przy pomocy słownika zdaniami w tym języku mogą zostać rozpoznane przez innych jego użytkowników. By to zrobić, nie potrzebują oni ani rozpoznania we mnie kompetentnego użytkownika rumuńskiego, ani sprawdzenia bądź założenia mojej szczerości. Gdyby zapytać interpretatora o szczerość mojej wypowiedzi, nie otrzymalibyśmy prawdopodobnie jasnej, twierdzącej odpowiedzi. Udzielenie jej wymagałoby oddzielnego sprawdzenia lub wykorzystania innej heurystyki.

[11] Przykład takiej polemiki ze stanowiskiem Searle'a można znaleźć m.in. w: Ridge, 2006.

[12] Wydaje mi się, że przedstawiona krytyka jest odzwierciedleniem znanego problemu „kolistości intencjonalnej": behawiorystyczne definicje terminów mentalnych muszą albo same zawierać predykaty mentalne, albo są podatne na bardzo proste kontrprzykłady (nie sposób w behawiorystycznych terminach zdefiniować choćby bólu zęba w sposób, który nie używa pojęcia „szczerości" i jednocześnie jest odporny na przykład symulowania przez kogoś takiego bólu; jak więc w języku czysto obserwacyjnym zdefiniować symulowanie bądź udawanie?). Patrz: Chisholm, 1957.

Istnieją, oczywiście, także sytuacje, w których odstępujemy od wykorzystania HU. Jest to grupa przypadków, w których na przykład podejrzewamy, że naruszony zostaje warunek szczerości, albo dochodzi do rażących konfliktów z Heurystyką Podobieństwa, gdzie mamy informacje, które pozwalają nam zwątpić w kompetencje językowe interpretowanego podmiotu bardziej niż w podobieństwo do nas samych pod względem innych przekonań, racjonalnych według naszych standardów. Gdy idę do teatru obejrzeć „Hamleta", nie przypiszę grającemu tytułową rolę aktorowi przekonania, że Dania jest więzieniem, na podstawie jego wypowiedzi. Gdy dowiem się, że historia, którą opowiedziała mi wcześniej pewna osoba, jest zmyślona, z rezerwą podejdę do dalszego stosowania HU wobec tej osoby, podejrzewając u niej skłonność do kłamstwa. Podobnie będzie w klasycznym przykładzie Davidsona: jeśli jakaś osoba twierdzi w rozmowie ze mną, że Statua Wolności zlokalizowana jest na Placu Trafalgarskim w Londynie i że wielkie wrażenie robią na tej osobie lwy u jej stóp, uznaję raczej w pierwszej chwili, że podobne przekonania żywi, podobnie jak ja, o Kolumnie Nelsona – priorytetyzując użycie HP ze względu na podejrzenie o niekompetencję językową.

Zauważmy jednak, że czym innym jest zawieszenie działania strategii pod pewnymi warunkami, a czym innym uzależnienie jej działania od spełnienia pewnych warunków. Według przedstawionego modelu, pierwszym naszym działaniem podczas interpretacji jest zastosowanie heurystyki, nie zaś upewnienie się (na wątpliwej podstawie) lub nawet dogmatyczne założenie, że nasz rozmówca jest szczerym i kompetentnym użytkownikiem języka polskiego oraz dysponuje czasem do namysłu. Proponowany model jak najdalszy jest od oferowania zestawu założeń, które musimy przyjąć o danym podmiocie podczas interpretacji. Proponuje on interpretację zorientowaną na konkretny cel poprzez zastosowanie skutecznych heurystyk.

Jak więc proponowany sposób odczytania głównych założeń interpretacjonizmu wpływa na dyskusję dotyczącą Założenia Racjonalności Podmiotu? Jeśli przytoczone wyżej heurystyki podobieństwa i uwłasnosłowienia faktycznie odgrywają rolę w indywidualnym przypisywaniu przekonań, wówczas przypadki przeczące ZRP można w sposób użyteczny opisać przy użyciu terminów psychologii potocznej. Wyjaśnić adekwatność takich przypisań może krótkie opisanie wzajemnej relacji tych dwóch heurystyk, jak i opisanie ich samodzielnego działania.

Po pierwsze, samo zastosowanie HU pozwala nam przypisywać sprzeczne i nieracjonalne proceduralnie przekonania na podstawie zdań uznawanych przez interpretowany podmiot. Dobry przykład tego typu stanowi sama zagadka Kripkego. Żaden z elementów tej historii nie jest wystarczającym sygnałem, aby odstąpić od wykorzystania HU w interpretacji zachowania Pierre'a: nie otrzymujemy od Kripkego jasnych dowodów na to, że Pierre jest w swoich sądach rażąco nieracjonalny, kłamie, udaje lub że nie posiada odpowiedniej (minimalnej) kompetencji językowej. Nie ma zatem żadnego powodu, aby zawiesić działanie HU, przez co można uznać za słuszne przypisanie mu dwóch sprzecznych przekonań. Czy godzi to w jakiś sposób w nasze możliwości interpretacji i przewidywania jego zachowania? Nie. Przyjęcie, że Pierre uważa Londyn zarówno za ładny, jak i nieładny, pozwala nam adekwatnie przewidywać pewne jego zachowania: gdy zaproponujemy mu wizytę w pięknym *Londres*, zgodzi się bez wahania; będziemy spodziewać się, że będzie narzekał na brzydotę miasta, w którym mieszka, i w końcu – że zdziwi się, jeśli kiedyś odwiedzi metrem okolice, które kojarzył dotychczas z odległym

Londres. Tych faktów nie jesteśmy w stanie wytłumaczyć, jeśli nie przypiszemy Pierre'owi wspomnianych przekonań, podążając za wskazaniami HU.

Drugą sytuacją tłumaczącą przypisania nieracjonalnych proceduralnie zbiorów przekonań jest możliwy konflikt HU i HP, czyli sytuacja, w której strategie te dają wzajemnie sprzeczne wyniki. Warunki i posiadane informacje mogą nas wówczas skłonić do przyznania priorytetu jednej z dwóch strategii: jeśli uznamy, że możemy mieć do czynienia ze znacząco odmiennym systemem przekonań, odrzucimy HP i będziemy przypisywać nasze przekonania, wspierając się głównie na HU. Jeśli natomiast uznamy, że dana osoba błędnie używa języka – na przykład wyrażając się niegramatycznie, albo (jak w przytoczonym przykładzie Davidsona) *systematycznie* myląc nazwy – lub też gdy nie możemy zaufać jej prawdomówności, będziemy priorytetyzować użycie HP. Nie można jednak uznać, że zawsze mamy dobre podstawy do tego, aby odrzucić którąś z tych strategii i priorytetyzować użycie drugiej z nich. Taką sytuację wydają się przedstawiać przytoczone tutaj wypadki systematycznej nieracjonalności proceduralnej opisywane przez Kahnemana i Tversky'ego czy Wasona, albo przypadki urojeń klinicznych. Zgodnie z Heurystyką Uwłasnosłowienia zmuszeni jesteśmy przypisać tym podmiotom przekonania rażąco sprzeczne ze zdroworozsądkowymi warunkami poprawności użycia terminów (których znajomość przypisujemy im na podstawie HP), którymi się posługują (śmierć, identyczność, prawdopodobieństwo itp.). Jednocześnie nie istnieją dane, które świadczyłyby o tym, że mamy do czynienia z problemem na gruncie językowym.

Istotą proponowanego tu rozwiązania jest odrzucenie założenia, które można uznać za wspólne wszystkim teoriom interpretacjonistycznym: że proces interpretacji wymaga przyjęcia normatywnego zespołu założeń odnośnie do formy przypisywanego zbioru przekonań, czy też posłużenia się jakiegoś rodzaju logiką przekonań. Wskazane tu heurystyki nie mają tutaj wymiaru normatywnego. Służą nam one do *opisu* podmiotu w kategoriach przekonań, jednak ich skuteczność *nie jest warunkiem koniecznym* przypisywania przekonań. Osoba korzystająca z niej naturalnie nie musi uznawać, że nieskuteczność danej heurystyki w pewnym przypadku świadczy o niemożliwości opisu danego podmiotu w kategoriach przekonań (podobnie jak nieskuteczność pewnych heurystyk matematycznych w przypadku danego problemu nie świadczy o jego nierozwiązywalności). W zarysowanym tu modelu interpretacja zorientowana jest przede wszystkim na uzyskanie możliwie najmniejszym kosztem jak największej ilości informacji na temat podmiotu. Innymi słowy, przypisując podmiotowi przekonanie, że p, nie jesteśmy zmuszeni założyć równocześnie, że na przykład nie jest on przekonany, że nie-p – podobne założenie nie jest bowiem w ogóle potrzebne do tego, aby skutecznie zinterpretować czy przewidzieć działanie podmiotu. Przypisanie braku takiego przekonania jest oczywiście możliwe, ale dopiero poprzez skorzystanie ze specyficznej heurystyki (zidentyfikowanej tutaj jako HP) – i, jak bywa z rozwiązaniami podsuwanymi nam przez heurystyki, może być fałszywe. Z pewnością zaś nie jest ono warunkiem koniecznym przypisania przekonania, że p.

Tylko w sytuacji, w której uznajemy konieczność nałożenia na interpretację silnych normatywnych ograniczeń, opis nieracjonalności może stanowić jakikolwiek problem. Celem, który pragniemy osiągnąć poprzez przypisywanie przekonań, jest opisanie możliwie najpełniejszego, użytecznego zbioru przekonań, którymi kieruje się interpretowany podmiot w swoim działaniu. Wzajemne relacje między tymi przekonaniami dopiero należy ustalić, nie zaś zakładać przed samym aktem przypisania. Zarysowany tutaj model heurystyczny jest więc

z natury deskryptywny, a nie normatywny: heurystyki są jedynie wykorzystywanymi przez nas sposobami określenia, jakie przekonania posiada interpretowany podmiot, nie zaś jedynymi kanonami sensownego dyskursu psychologii potocznej.

Aby zakończyć prowadzone w tym artykule rozważania, warto wspomnieć o relacji łączącej model heurystyczny z ogólną metodologią psychologii oraz to, w jaki sposób wpływa ona na status ZRP. Heurystyki Uwłasnosłowienia i Podobieństwa omówiono powyżej jako narzędzia służące skutecznym przypisaniom indywidualnym – zawartym w naszym codziennym doświadczeniu interpretatorów, dokonywanym pod presją czasu i w stanie ograniczonej informacji. W jaki sposób ich wykorzystanie ma się do przypisań naukowyc[13], dokonywanych w naukach społecznych? Aby odpowiedzieć na to pytanie, należy wcześniej rozstrzygnąć dwie kwestie. Pierwszą z nich jest kwestia samego kształtu i celu nauk społecznych, których dotyczy problem, oraz statusu ZRP w ich obrębie, a więc tego, jak operacjonalizowane jest pojęcie przekonania w psychologii, socjologii czy ekonomii. W formułowanych w swoich ramach wyjaśnieniach i przewidywaniach nauki społeczne często korzystają z pojęcia przekonania zaczerpniętego z psychologii potocznej (Egan, 1986). Jeśli ZRP przyjmowane jest *implicite* w wymienionych naukach, wówczas praktyka tych nauk – kierowana na przykład wynikającą z ZRP Davidsonowską zasadą życzliwości – powinna to jasno odwzorowywać w swoich wynikach i operacjonalizacjach pojęcia przekonania. Nie jest to jednak prawdą, z powodów wskazanych w pierwszej części artykułu. Badania Tversky'ego i Kahnemana nie spotkały się z odrzuceniem wśród społeczności psychologów i psycholożek, ale zapoczątkowały owocne badania nad ludzkim wnioskowaniem i podejmowaniem decyzji w psychologii poznawczej, socjologii czy ekonomii behawioralnej. Definicja urojeń jako *nieracjonalnych* (fałszywych, odpornych na dane empiryczne bądź logicznie sprzecznych z innymi przekonaniami podmiotu) *przekonań*, stosowana w DSM-V, również nie podlega dyskusji i pozostaje użyteczna w praktyce klinicznej[14]. Można więc założyć, że ZRP nie stanowi wcale podstawy operacjonalizacji pojęcia przekonania, lecz właśnie wyżej wymienione heurystyki. Jako przykłady zgodnych z tymi heurystykami operacjonalizacji możemy podać chociażby wykorzystanie skali AB (Fishbein i Raven, 1962, zobacz też: Egan, 1986, s. 320-321) stosowanej w kwestionariuszach w psychologii społecznej i poznawczej czy „poprawną odpowiedź" na pytanie eksperymentalne w „teście fałszywych przekonań" (Wimmer i Perner, 1983). Pierwsza z nich opiera się na oszacowaniu stopnia zgody z pewnym twierdzeniem przez osobę badaną pewnym kwestionariuszem – a więc w jasny sposób podąża

[13] Rozróżnienie na przypisania indywidualne i naukowe za: Dub, 2015, patrz: przypis 6.

[14] Jak słusznie zauważył jeden z recenzentów tego artykułu, treść tej definicji jak i jej użyteczność w praktyce klinicznej bywa podważana (patrz: Spitzer, 1990; Bortolotti, 2010, s. 23-27). Jak jednak zauważa Bortolotti, największym problemem tej definicji nie jest charakterystyka urojeń jako przekonań, ale nie zwrócenie uwagi na potencjalną szkodliwość żywienia takiego przekonania przez pacjenta dla jego zdrowia lub funkcjonowania społecznego. Inne proponowane definicje starają się uwzględnić ten warunek (McKay, Langdon i Coltheart, 2005). Wykorzystanie pojęć epistemicznych („nieuzasadnione", „niepodatne na zmianę") czy sama charakterystyka urojeń jako przekonań nie jest raczej poddawana w wątpliwość i jest użyteczna w ich diagnozowaniu. Inną interesującą kwestią pozostaje relacja pomiędzy taksonomią urojeń a ich treścią (patrz: Clutton, Gadsby i Klein, 2017).

za wytycznymi HU. Natomiast w powszechnie stosowanym w psychologii rozwoju poznawczego „teście fałszywych przekonań" zakłada się, iż umiejętność przypisywania przekonań i posługiwania się psychologią potoczną może zostać sprawdzona poprzez zasymulowanie sytuacji, w której inna osoba *powinna* mieć określone przekonanie, zważywszy na jej inne przekonania, dostęp do danych z otoczenia itp., *inne* od tego, którym dysponuje dziecko. Widać zatem, iż pojęcie przekonania jest w psychologii operacjonalizowane w zgodzie z HP i HU, nie zaś – w zgodzie z ZRP i zasadą życzliwości.

Docieramy tutaj do drugiej kwestii: czy *powinniśmy* utrzymywać HU i HP jako elementy naszej praktyki badawczej i czy wykorzystanie tych heurystyk przynosi naukom społecznym korzyści? Odpowiedź na te pytania zależy z pewnością od przyjmowanej optyki samych heurystyk: czy widzimy w nich raczej wadliwe reguły wynikające z ograniczeń naszego aparatu poznawczego, czy też oszczędne i efektywne uproszczenia. Uważam jednak, że należy rozstrzygnąć ten spór na korzyść heurystyk i zaakceptować je jako jasno sformułowane zasady metodologiczne. Ogólnym celem nauk społecznych nie jest bowiem zazwyczaj jak najbardziej precyzyjne określenie przekonań badanych, ale raczej wykorzystanie opisu tych przekonań do przewidywania ich zachowania w sytuacjach eksperymentalnych – ten cel dzieli zatem z psychologią potoczną i naszymi indywidualnymi przypisaniami. Jeśli więc heurystyki te okazały się efektywne wobec alternatywnych strategii przypisywania przekonań i jednocześnie, wbrew staraniom eliminatywistów, nauki społeczne wciąż chętnie korzystają z terminów zapewnianych przez psychologię potoczną, nie widać przeciwwskazań, aby heurystyki te nadal wykorzystywano w przypisywaniu przekonań w praktyce naukowej lub nadano im postać jasno wyartykułowanych zasad metodologicznych.

4. Zakończenie

W artykule tym starałem się porównać dwie odmienne charakterystyki teorii interpretacji innych podmiotów pod kątem efektywnego opisu przypadków nieracjonalności. Pierwszą z nich jest model proponowany przez klasyczny interpretacjonizm, postulujący racjonalność podmiotu jako warunek konieczny przypisywania przekonań. Próbowałem pokazać, że posługiwanie się tym modelem nie pozwala nam przekonująco wytłumaczyć przypadków ludzkiej nieracjonalności dokumentowanych przez psychologię poznawczą i kliniczną, w których intuicyjnie przypisujemy ludziom skrajnie nieracjonalne proceduralnie przekonania. Obydwie odpowiedzi, uzgadniające te przypadki z założeniem racjonalności podmiotów, narzucające opis tych przypadków jako korygowalnych „błędów wykonania" bądź postulujące zasadę metodologiczną polegającą na reinterpretacji pojęć używanych przez podmiot, wydają mi się nieadekwatne empirycznie bądź szkodliwe metodologicznie. Zamiast tego zaproponowałem ujęcie opierające się na programie badań nad „szybkimi i oszczędnymi" heurystykami, uznające potoczną teorię interpretacji za zbiór efektywnych heurystyk. Wydaje mi się ono bardziej prawdopodobne z punktu widzenia psychologii ludzkiego wnioskowania, pokazałem też, że bardziej skutecznie może ono radzić sobie z opisem ludzkiej nieracjonalności. W tym celu zaproponowałem opis tych przypadków jako wyników działania dwóch interpretacyjnych heurystyk: heurystyki podobieństwa (HP), będącej przeformułowaniem postulowanego przez interpretacjonistów warunku racjonalności, i heurystyki uwłasnosłowienia (HU), będącej przeformułowaniem Zasady Uwłasnosłowienia

zaproponowanej przez Saula Kripkego. W mojej ocenie model heurystyczny może być bardziej użyteczny nie tylko przy opisie ludzkiego przypisywania przekonań i posługiwania się psychologią potoczną, ale też z punktu widzenia praktyki nauk społecznych.

Bibliografia

Bortolotti, L. (2010). *Delusions and Other Irrational Beliefs.* Oxford: Oxford University Press.

Chisholm, R. (1957). *Perceiving: A Philosophical Study.* Ithaca: Cornell University Press.

Ciecierski, T. (2013). *Nastawienia sądzeniowe. Wykłady z filozofii psychologii.* Warszawa: PWN.

Clutton, P., Gadsby, S. i Klein, C. (2017). Taxonomising delusions: content or aetiology?. *Cognitive Neuropsychiatry*, 22(6), 508-527. https://doi.org/10.1080/13546805.2017.1404975

Cohen, L. J. (1981). Can Human Irrationality Be Experimentally Demonstrated?. *The Behavioral and Brain Sciences*, 4, 317-370. https://doi.org/10.1017/S0140525X00009092

Davidson, D. (1974). Przekonania a podstawy znaczenia, W: Davidson, D. (1992). *Eseje o prawdzie, języku i umyśle*(s. 118-140), (tłum. B. Stanosz). Warszawa: PWN.

Davidson, D. (1980). Ku jednolitej teorii znaczenia i działania. W: Davidson, D. (1992). *Eseje o prawdzie, języku i umyśle*, (s. 141-162), (tłum. B. Stanosz). Warszawa: PWN.

Davidson, D. (1982). Zwierzęta racjonalne. W: Davidson, D. (1992). *Eseje o prawdzie, języku i umyśle* (s. 234-250), (tłum. C. Cieśliński). Warszawa: PWN.

Dennett, D. (1981a). True Believers: The Intentional Strategy and Why It Works. W: Dennett D. (1987), *The Intentional Stance* (s. 13-43). Cambridge: MIT Press.

Dennett, D. (1981b). *Making Sense of Ourselves. Philosophical Topics* 12, 63-81.

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, O. (1986). The Concept of Belief in Cognitive Theory. W: L. P. Mos (red.), *Annals of Theoretical Psychology. Annals of Theoretical Psychology* 4. Springer: Boston. https://doi.org/10.1007/978-1-4615-6453-9_23

Elqayam, S. i Evans, J. S. B. (2011). Subtracting "ought" from "is": Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences*, 34(5), 233-290. https://doi.org/10.1017/S0140525X1100001X

Elqayam, S. (2012). Grounded rationality: Descriptivism in epistemic context. *Synthese*, 189(1), 39-49. https://doi.org/10.1007/s11229-012-0153-4

Fishbein, M., Raven, B.H. (1962). The AB scales: An operational definition of belief and attitude. *Human Relations*, 15, 35-44. https://doi.org/10.1177/001872676201500104

Fisk, J. E. (2016). Conjunction fallacy. W: R. F. Paul (red.), *Cognitive Illusions: Intriguing Phenomena in Judgement, Thinking and Memory.* London: Psychology Press.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology*, 2, 83-115. https://doi.org/10.1080/14792779143000033

Gigerenzer, G. (1996). On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky. *Psychological Review*, 103, 592-596. https://doi.org/10.1037/0033-295X.103.3.592

Gigerenzer, G. i Gaissmaier, W. (2011). Heuristic Decision Making, *Annual Review of Psychology*, 62, 451-482. https://doi.org/10.1146/annurev-psych-120709-145346

Gigerenzer, G. i Sturm, T. (2011). How (far) can rationality be naturalized?. *Synthese*, 187, 243–268. https://doi.org/10.1007/s11229-011-0030-6

Johnson-Laird, P. N. i Wason, P. C. (1970). A Theoretical Analysis of Insight Into a Reasoning Task. *Cognitive Psychology*, 1, 134-148. https://doi.org/10.1016/0010-0285(70)90009-5

Kahneman, D. i Tversky, A. (1971). Belief In the Law of Small Numbers. *Psychological Bulletin*, 76, 105-110. https://doi.org/10.1016/0010-0285(70)90009-5

Kahneman, D. i Tversky, A. (1983). Extensional Versus Intuitive Reasoning: The Conjunction Fallacy In Probability Judgment. *Psychological Review*, 90, 293-315. https://doi.org/10.1037/0033-295X.90.4.293

Kahneman, D. i Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review*, 103, 582-591. https://doi.org/10.1037/0033-295X.103.3.582

Kripke, S. (1979). A Puzzle About Belief. W: A. Margalit (red.), *Meaning and Use*, Dordrecht: D. Reidel. https://doi.org/10.1007/978-1-4020-4104-4_20

Lewis, D. (1986). *On the Plurality of Worlds.* London: Blackwell.

Marcus, R. B. (1983). *Rationality and Believing the Impossible.* W: Marcus R. B. (1983). *Modalities. Philosophical Essays.* New York: Oxford University Press.

McKay, R., Langdon, R. i Coltheart, M. (2005). 'Sleights of mind': Delusions, defences, and self deception. *Cognitive Neuropsychology* 10(4), 305–326. https://doi.org/10.1080/13546800444000074

McKay, R. i Cipolotti, L. (2007). Attributional style in a case of Cotard delusion. *Consciousness and Cognition*, 16, 349–359. https://doi.org/10.1016/j.concog.2006.06.001

Nishio, Y., Mori, E. (2012). Delusions of Death in a Patient with Right Hemisphere Infraction. *Cognitive and Behavioural Neurology*. 25, 216–223. https://doi.org/10.1097/WNN.0b013e31827504c7

Pacherie, E., Bayne, T. (2005). In Defence of the Doxastic Conception of Delusions. *Mind & Language*, 20, 163-188. https://doi.org/10.1111/j.0268-1064.2005.00281.x

Puczyłowski, T. (2014). Czy można być przekonanym, że Sherlock Holmes jest detektywem? O sensowności zdań z nazwami fikcjonalnym. *Studia z Kognitywistyki i Filozofii Umysłu*, 8 (1), 34–50.

Quine, W. V. O. (1999). *Słowo i przedmiot* (tłum. C. Cieśliński). Warszawa: Aletheia.

Ridge, M. (2006). Sincerity and Expressivism. *Philosophical Studies*, 131, 487–510. https://doi.org/10.1007/s11098-005-2218-4

Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language.* Cambridge: Cambridge University Press.

Searle, J. (1983). *Intentionality. An Essay in the Philosophy of Mind.* Cambridge: Cambridge University Press.

Sides, A., Osherson, D., Bonini, N. i Viale, R. (2002). On the reality of the conjunction fallacy. *Memory & Cognition*, 30(2), 191-198. https://doi.org/10.3758/BF03195280.

Spitzer, M. (1990). On Defining Delusions. *Comprehensive Psychiatry*, 31, 377-397, https://doi.org/10.1016/0010-440X(90)90023-L.

Stich, S. P. (1985). Could man be an irrational animal? Some notes on the epistemology of rationality, *Synthese,* 64(1), 115-135. URL: https://www.jstor.org/stable/20116149

Tarnowski, M. (2019). Czy posiadanie sprzecznych przekonań jest możliwe? Omówienie i krytyka argumentów za Psychologiczną Zasadą Niesprzeczności. *Studia Semiotyczne*, XXXIII (2), 323-353. https://doi.org/10.26333/sts.xxxiii2.11

Van Rooij, I., Wright, C. D. i Wareham, T. (2012). Intractability and the use of heuristics in psychological explanations. *Synthese*, 187(2), 471-487. https://doi.org/10.1007/s11229-010-9847-7

Wason, P. C. (1968). Reasoning About a Rule. *The Quarterly Journal of Experimental Psychology*, 20, 273-281. https://doi.org/10.1080/14640746808400161

Wason, P. C. (1969), Regression in Reasoning?. *British Journal of Psychology*, 60, 471-480. https://doi.org/10.1111/j.2044-8295.1969.tb01221.x

Wimmer, H. i Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Wójtowicz, A. i Winkowski, J. (2018). Heuristics: Daniel Kahneman vs Gerd Gigerenzer, W: M. Hetmański (red.), *Rationality and Decision Making. From Normative Rules to Heuristics.* Leiden: Brill Rodopi.

Zabłudowski, A. (1986). On Marcus's Solution to Kripke's Puzzle and a Few Related Issues. *Pacific Philosophical Quarterly*, 67, 279-296. https://doi.org/10.1111/j.1468-0114.1986.tb00279.x

How to Describe Irrationality? Irrational Belief Systems and Interpretationism in Philosophy of Mind

Abstract: One of the most popular and controversial theses in contemporary philosophical analysis of folk psychology and belief ascription is that every propositional attitude ascription must be preceded by the assumption of procedural rationality of the agent whose behaviour we are trying to interpret. This assumption seems to be challenged by research done in psychology, sociology and economics, which provides the examples of blatantly irrational behaviour. The aim of this paper is to briefly show the problems that such rationality assumption encounters and to offer an alternative view of interpretation theory based on approach given by „fast and frugal heuristics" programme in cognitive psychology. I identify two heuristics – Similarity and Disquotation Heuristic – that seem to play crucial role in our

everyday belief ascriptions. I argue that this model is both more psychologically probable and more effective in explaining the cases of irrationality.

Keywords: rationality; interpretation; principle of charity; intentional stance; heuristics; folk psychology; belief ascription

Maciej Tarnowski (ur. 1997) – doktorant Szkoły Doktorskiej Nauk Humanistycznych na Wydziale Filozofii Uniwersytetu Warszawskiego. Zajmuje się naukowo epistemologią formalną, a także analityczną filozofią umysłu i języka oraz filozofią eksperymentalną.