Filozofia Nauki (*The Philosophy of Science*)
ISSN 1230-6894 e-ISSN 2657-5868
2020, vol. 28(3) [111]: 111-130
DOI: 10.14394/filnau.2020.0018

IDA MICZKE*

AUTORYTET PIERWSZOOSOBOWY I SAMOWIEDZA W KONCEPCJI CRISPINA WRIGHTA**

Abstract

CRISPIN WRIGHT'S ACCOUNT OF FIRST-PERSON AUTHORITY AND SELF-KNOWLEDGE

The aim of this paper is to analyze Crispin Wright's constitutivist account of self-knowledge and first-person authority. Wright offered an alternative to standard detectivist theories of self-knowledge and first-person authority. According to his proposal, the subject does not detect her mental states, but rather creates them. Wright offered his proposal as a result of considering the problem of rule-following. In the paper, I describe Wright's solution and analyze its problems. I claim that these problems render his theory unconvincing, and I try to uncover the sources of his failure. First of all, I claim that Wright did not get rid of picturing self-knowledge as a kind of perception, and I suggest that some problems within his theory are the same as those within perceptual theories of self-knowledge. I then turn to problems with the interpretation according to which Wright presents first-person authority as a product of our mental discourse only. Finally, I present an outline of a solution to the problems of Wright's theory in which I follow David Finkelstein's neo-expressivist proposal. I argue that an expressivist solution can be obtained by considering Fred Dretske's conciliatory skepticism and that investigating Dretske's account enables one to understand why Wright's question concerning the relationship between the subject and her mental states is ill-formulated.

Keywords: self-knowledge, first-person authority, Crispin Wright, constitutivism, epistemology

Celem artykułu jest analiza koncepcji samowiedzy[1] i autorytetu pierwszoosobowego zaproponowanej przez Crispina Wrighta. Stanowisko Wrighta,

* Wydział Filozofii, Uniwersytet Warszawski, ul. Krakowskie Przedmieście 3, 00-927 Warszawa, e-mail: i.miczke@student.uw.edu.pl.

** Za uwagi, które pozwoliły mi znacznie ulepszyć tę pracę, dziękuję serdecznie Joannie Komorowskiej-Mach.

[1] Choć będę zajmować się koncepcją samowiedzy Wrighta, w konkluzji artykułu pokażę, że skupienie na wiedzy uniemożliwia Wrightowi rozwiązanie problemu, który sobie postawił. Z tego względu być może trafniejsze byłoby mówienie o samopoznaniu niż o samowiedzy, by nie przesądzać, jaka jest wartość rezultatów tego poznania. Zdecydowałam się

zwane w literaturze konstytutywizmem, jest konkurencyjne wobec klasycznych modeli detektywistycznych. Na gruncie stanowiska Wrighta, które wyłania się z rozważań nad tzw. problemem kierowania się regułą, podmiot nie wykrywa własnych stanów mentalnych, lecz je konstruuje. W artykule pokażę, w jaki sposób konstytutywistyczna koncepcja stanów intencjonalnych wywodzi się z problemu reguł, a następnie omówię jej słabości, wskazując, że nie jest to przekonująca teoria i próbując wyjaśnić, skąd biorą się jej problemy. Pokażę, że Wright nie uwolnił się od myślenia o samowiedzy w kategoriach percepcyjnych i część problemów jego koncepcji to te same trudności, z którymi borykały się teorie przedstawiające introspekcję jako rodzaj percepcji. Rozważę także problemy z przedstawianiem autorytetu pierwszoosobowego jako produktu wyłącznie naszego dyskursu mentalnego. Na koniec przedstawię zarys rozwiązania problemów teorii Wrighta, rozwijając propozycję Davida H. Finkelsteina, który przyjmuje koncepcję neoekspresywistyczną jako rozwiązanie problemu reguł i samowiedzy. Pokażę, że tego rodzaju ujęcie można wyprowadzić z pojednawczego sceptycyzmu Freda Dretskego oraz że odwołanie do Dretskego pozwala zrozumieć, dlaczego pytanie o relację podmiotu do swoich stanów mentalnych, które Wright formułuje, jest źle postawione.

1. DETEKTYWIZM, MODEL PERCEPCYJNY I PROBLEMY TRADYCYJNYCH UJĘĆ INTROSPEKCJI

Uznaje się, że samoprzypisania stanów mentalnych cechują się specjalnym autorytetem (jest to tzw. „autorytet pierwszoosobowy"): w codziennej komunikacji nie kwestionujemy ich ani nie pytamy o dowody na ich rzecz[2]. Przeciwnie, powszechnie zakładamy prawdziwość zdań, w których podmiot przypisuje sobie pewien stan psychiczny. Równie powszechnie zakładamy — zarówno o sobie samych, jak i o innych ludziach — że każdy wie najlepiej, jaka jest treść jego umysłu. Co więcej, przyjmujemy, że nikt nie ma wątpliwości co do tego, jakie działania będą zgodne z jego aktualnym stanem mentalnym.

jednak na używanie terminu „samowiedza", ponieważ jest on bardziej rozpowszechniony w literaturze przedmiotu niż termin „samopoznanie".

[2] Terminu „samoprzypisanie" będę w tym tekście używać jako odpowiednika angielskiego *avowal*. Słowo *avowal* jest semantycznie skomplikowanym wyrażeniem opisującym wypowiedzenie, w którym podmiot przypisuje sobie samemu pewien aktualny stan mentalny w sposób nieinferencyjny i szczery. Z grubsza w takim znaczeniu używa się terminu *avowal* co najmniej od czasów Gilberta Ryle'a (por. Ryle 2009: 84-89, 162-166). Podobne rozumienie *avowals* można znaleźć u Donalda Davidsona (1987) i u Dorit Bar-On, która rozszerza *avowals* o samoprzypisania jedynie pomyślane (Bar-On 2004, 2009, 2012).

Kiedy mówię, że zamierzam obejrzeć nowy film Almodóvara, w mojej głowie nie rodzi się wątpliwość co do tego, jakie zachowanie będzie zgodne z tym zamierzeniem: wydaje się oczywiste, że pójście do kina i kupienie biletu będzie z nią zgodne, a pójście z psem na spacer już nie.

Klasyczne stanowiska filozoficzne w sporze na temat samowiedzy i samoprzypisań będę za Finkelsteinem (2003) nazywać detektywistycznymi. Finkelstein wyróżnia stary i nowy detektywizm (Finkelstein 2003, por. Komorowska-Mach 2015). Obie wersje detektywizmu łączy przekonanie, że samopoznanie polega na wykrywaniu własnych stanów mentalnych (Finkelstein 2003: 9). Stąd też akcent na introspekcyjny charakter samopoznania (Komorowska-Mach 2015: 41). Różnica między starym a nowym detektywizmem polega zaś na tym, że stary detektywizm przyjmuje nieomylność procesu detekcji (Finkelstein 2003: 12), natomiast nowy dopuszcza zawodność mechanizmu introspekcyjnego (Komorowska-Mach 2015: 42). Co jednak najistotniejsze, zarówno w starej, jak i w nowej wersji detektywizmu introspekcję ujmuje się przez analogię do percepcji (Finkelstein 2003: 10-11, Moran 2001: 2, 12). Zakłada się mianowicie istnienie stanów mentalnych jako pewnych rzeczywistych, oddzielnych od podmiotu obiektów, które ma on wykrywać przez zaglądanie w głąb siebie (por. Komorowska-Mach 2015: 42). Jako przykład starego detektywizmu Finkelstein (2003: 11-13) podaje wczesną koncepcję Bertranda Russella (1912). Do nowego detektywizmu można natomiast zaliczyć choćby koncepcję Davida Armstronga (1968) (Finkelstein 2003: 17).

Stanowiska detektywistyczno-percepcyjne borykają się jednak z wieloma trudnościami. Po pierwsze, nie ma żadnego „narządu introspekcyjnego" analogicznego do narządów zmysłowych (Shoemaker 1994: 254, Moran 2001: 13)[3]. Po drugie, nie wiadomo, gdzie miałaby przebiegać granica między obserwatorem a przedmiotem obserwowanym. Jak wskazywał już Auguste Comte, obserwujący i obserwowany są w tym wypadku tym samym (Comte 2009: 11). Samoobserwacja zaś wydaje się niemożliwa: aby jej dokonać, obserwujący musiałby zrezygnować z innych aktywności mentalnych, wtedy jednak znikałby sam przedmiot badania (Comte 2009: 11). Aby percepcja była weredyczna, stany mentalne musiałyby być oddzielone od podmiotu poznania, nie mamy zaś gwarancji, że podmiot nie wpływa na obserwowany przedmiot, zmieniając go w trakcie obserwacji. I odwrotnie: nie wiemy, czy postrzegane stany mentalne nie zaburzają zdolności obserwacyjnych, jak może to mieć miejsce w przypadku emocji (Comte 2009: 11). To jednak nie jedyne problemy związane z domniemanym rozdzieleniem podmiotu i jego stanów mentalnych.

[3] Choć por. Armstrong 1968, gdzie jest propozycja odparcia tego zarzutu.

Należy bowiem zauważyć, że w takim wypadku stany mentalne mają być rzeczywistymi, wewnętrznymi obiektami, analogicznymi do przedmiotów postrzeganych w świecie zewnętrznym (Komorowska-Mach 2015: 42). To zaś sprawia, że trudno jest utrzymać autorytet pierwszoosobowy, ponieważ pojawia się możliwość błędnej identyfikacji: skoro w percepcji świata zewnętrznego mogę, na przykład, mylnie wziąć biegnącego przede mną owczarka niemieckiego za wilka, to podobnie powinnam móc błędnie zidentyfikować, na przykład, moje przekonanie jako wyraz wątpliwości (por. Shoemaker 1994: 260). Aby uniknąć tej konsekwencji, zwolennicy starego detektywizmu zastrzegają, że introspekcja jest specjalnym rodzajem percepcji (Finkelstein 2003: 13). To zaś prowadzi do przedstawiania jej jako nadnaturalnie sprawnego, nieomylnego mechanizmu i niesie zagrożenie dualistyczną ontologią (Finkelstein 2003: 13-14), nie mówiąc już o sceptycyzmie wobec istnienia innych umysłów i świata zewnętrznego (Finkelstein 2003: 14). Co więcej, opieranie się na nieomylności jest dziś trudne do utrzymania w świetle eksperymentów psychologicznych, które podają ją w wątpliwość (Schwitzgebel 2016). Rozsądniejsze byłoby zezwolenie na błędy w samopoznaniu przy jednoczesnym przyjęciu, że nie są to pomyłki polegające na błędnej identyfikacji jakiegoś obiektu wewnątrz naszego umysłu (Shoemaker 1994: 260) i podważające autorytet pierwszoosobowy. Model detektywistyczno-percepcyjny nie radzi sobie z tym zadaniem, ponieważ albo przedstawia samopoznanie jako nieomylne, albo upatruje pomyłki w błędnej identyfikacji. Okazuje się zatem, że model ten nie jest trafnym ujęciem samowiedzy i autorytetu pierwszoosobowego.

Propozycją konkurencyjną wobec detektywizmu jest konstytutywizm, czyli koncepcja, zgodnie z którą podmiot nie wykrywa swoich stanów mentalnych, lecz je wytwarza (Finkelstein 2003: 28). Stanowisko Wrighta, które będę analizować w tym artykule, jest uznawane za stanowisko konstytutywistyczne (Finkelstein 2003). Choć może się ono wydawać obiecującym rozwiązaniem problemów detektywizmu, okaże się, że w rzeczywistości stanowi równie niezadowalającą propozycję.

2. PROBLEM KIEROWANIA SIĘ REGUŁĄ I KONSTYTUTYWISTYCZNE ROZWIĄZANIE WRIGHTA

Intuicyjnie wydaje się, że wykonanie działania „68 + 57” polega na postępowaniu zgodnie z regułą dodawania (Kripke 2007: 22). Wyobraźmy sobie jednak, że nigdy nie wykonywałam działania „68 + 57”. Ponieważ w przeszłości dodawałam jedynie skończoną liczbę razy, można zawsze znaleźć taką parę

liczb, które będą większe od wszystkich, na których wykonywałam dotąd obliczenia. Kripke zauważa, że nie istnieje taki fakt, który determinowałby to, że odpowiedź brzmi „125", a nie „5". Nie mogę bowiem wiedzieć, czy używając znaku „+" w przeszłości, miałam na myśli funkcję dodawania, czy na przykład kwodawania, która dla liczb mniejszych od 57 zachowuje się dokładnie jak dodawanie, a powyżej tej granicy wskazuje „5" jako poprawną odpowiedź dla każdego działania (por. Kripke 2007: 22-23). Nic nie pomoże odwołanie się do interpretacji i stwierdzenie, że wiem, jak rozwiązać zadanie „68 + 57", ponieważ interpretuję symbol „+" w odpowiedni sposób. Taki sam problem pojawi się odnośnie do znaków, za pomocą których sformułowano interpretację, mielibyśmy więc do czynienia z regresem w nieskończoność (por. Dziobkowski 2016: 58). Według Kripkego *Dociekania filozoficzne* odkrywają nową, radykalną formę sceptycyzmu, w myśl którego nie istnieją fakty determinujące znaczenie używanych przez nas symboli. Słowa nie mają znaczeń, a podmiot nie może mieć o nich wiedzy. Niezależnie bowiem od tego, jaką liczbę podamy jako rozwiązanie działania „68 + 57", znajdzie się taka interpretacja pierwotnego polecenia, zgodnie z którą będzie to trafne rozwiązanie. Skoro jednak tak jest, to nie istnieje możliwość błędu — a więc także poprawnej odpowiedzi.

Paradoks sceptycki w takim sformułowaniu sprawia wrażenie problemu przede wszystkim semantycznego. W rzeczywistości jednak stanowi równie poważny problem dla teorii samowiedzy. Można bowiem przedstawić go także w odniesieniu do stanów mentalnych: nie istnieje fakt determinujący, co *miałam na myśli*, gdy w przeszłości używałam symbolu „+" albo słowa „dodawanie". Nie mogę wiedzieć, czy w przeszłości miałam *intencję*, żeby dodawać, czy raczej żeby kwodawać.

Kripke proponuje własne rozwiązanie paradoksu, które zasadza się na przyjęciu nonfaktualizmu (Dziobkowski 2016: 56). Zdanie „Rozwiązując działanie »68 + 57«, miałam na myśli dodawanie" nie ma warunków prawdziwości — ponieważ nie istnieją fakty, które determinowałyby jego znaczenie — natomiast ma warunki stwierdzalności: w określonych warunkach może zostać wypowiedziane i uznane w danej społeczności językowej (por. Dziobkowski 2016: 61).

Tymczasem propozycja Wrighta jest tworzona z myślą o przeciwstawieniu się interpretacji Kripkego (Wright 2001a: 92-93). Jak zauważa Wright, Kripke dochodzi do sceptycznego wniosku o nieistnieniu faktów determinujących znaczenie słów i treść stanów mentalnych ze względu na podejście redukcjonistyczne: wyklucza spośród możliwych faktów te, które same mają treść (Wright 2001c: 176). Tymczasem faktem determinującym to, że w przeszłości miałam na myśli funkcję dodawania (inaczej mówiąc: miałam zamiar, żeby dodawać) jest po prostu to, że miałam ją na myśli (Finkelstein 2003: 35), a dokładniej

— mój obecny sąd, że rzeczywiście tak było. Należy porzucić naiwny obraz, zgodnie z którym nasze sądy na temat stanów intencjonalnych odpowiadają jakimś stanom rzeczy od nich niezależnym — rozważania Wittgensteina i Kripkego pokazały fiasko takich poszukiwań. Podobnie nie ma racji bytu założenie, że wiem, jaki miałam w przeszłości zamiar, i potrafię stwierdzić, czy moje aktualne działanie jest z nim zgodne, dzięki intuicyjnemu uchwyceniu treści stanu mentalnego. Musimy zatem przyjąć, że moje dawne zamiary i to, jakie działanie jest z nimi zgodne, są przedmiotem mojego aktualnego sądu (Wright 2001b: 142).

Rozwiązanie Wrighta w pierwszej chwili sprawia wrażenie chybionego. Wydaje się, jakby zrównywało poprawność z wrażeniem poprawności, a przecież, jak pisał Wittgenstein, jeśli „prawidłowe jest to, co mi się prawidłowe wyda", to „o »prawidłowości« nie ma tu co mówić" (Wittgenstein 2000: § 258). Wright jednak zdaje sobie sprawę z tego problemu. Odpowiada nań, posługując się analogią do sądów dotyczących własności wtórnych. Dobrym przykładem takiej własności jest kolor. Panuje w miarę powszechna zgoda co do tego, że twierdzenia na temat kolorów są zależne zarówno od postrzeganego przedmiotu, jak i od podmiotu poznającego. Nie ma sensu mówienie o sądzie niezależnym od obserwatora, ponieważ kolor uważa się za własność, która powstaje w wyniku interakcji między przedmiotem a układem poznawczym człowieka. A mimo to, jak twierdzi Wright, irrealistyczne interpretowanie takich sądów jest przesadą (Wright 2001c: 191). Możemy nadal interpretować zdania o kolorach obiektywnie, o ile tylko określimy poznawczo idealne warunki, w których musi znajdować się podmiot, żeby jego sąd miał wartość logiczną. W wypadku twierdzeń o kolorach zakładamy na przykład, że obserwowany przedmiot jest w pełni widoczny i w dobrym świetle, a podmiot obserwuje go uważnie i zna pojęcia, za pomocą których następnie go opisze (Wright 2001c: 192-193). Sądy wydane w takich okolicznościach zasługują na miano *best judgements* — najlepszych, najtrafniejszych sądów. Właśnie nasze *best judgements* determinują to, jak obiektywnie ma się sprawa z kolorami przedmiotów (Wright 2001c: 192-193).

Według Wrighta w wypadku stanów intencjonalnych sytuacja jest podobna[4]. Problem kierowania się regułą uświadamia nam, że nie istnieją niezależne od naszych sądów fakty, które determinowałyby treść stanów intencjonalnych. Pozostaje więc uznać, że to, czy w przeszłości miałam na przykład zamiar, żeby zrobić X, jest zależne od wydanego teraz sądu (w sensie *best judgement*), że w istocie go wtedy miałam. Zgodnie z przytoczoną analogią do

[4] Warto zaznaczyć, że teoria Wrighta dotyczy wyłącznie stanów intencjonalnych, a nie fenomenalnych. Dlatego też, jeśli kiedykolwiek w tym tekście piszę o „stanach mentalnych", mam na myśli wyłącznie stany intencjonalne.

kolorów powinniśmy jeszcze tylko zadbać o określenie odpowiednich warunków (*C-conditions*, Wright 2001c: 194), w których samoprzypisaniu można przyznać funkcję determinowania obiektywnej treści umysłu[5].

Tutaj ujawnia się pewna różnica w stosunku do przykładu kolorów. Rozważmy pojęcie zamiaru, którego Wright używa najczęściej jako przykładu. Mamy pewne wyobrażenie wymagań, które stawiamy podmiotowi godnemu zaufania w kwestii jego stanów mentalnych: zakładamy, że się sam nie oszukuje, że nie jest pod wpływem środków ograniczających poczytalność. W odróżnieniu jednak od kolorów, w przypadku samoprzypisań warunki nakładane na podmiot samoprzypisania mają w sposób domniemany status spełnionych (jak określa to Wright, są *positive-presumptive*, Wright 2001c: 202). Potoczne pojęcie zamiaru funkcjonuje bowiem w taki sposób, że wyjściowo zakłada się, że wszystkie warunki konieczne do uznania prawdziwości samoprzypisania zamiaru zostały spełnione. Innymi słowy, na co dzień nie mamy w zwyczaju podważać samoprzypisań zamiaru. Dopiero w szczególnych sytuacjach, w których mamy rozsądne powody, by uważać, że podmiot nie powinien zostać obdarzony zaufaniem, podważymy jego samoprzypisanie, stwierdzając niespełnienie któregoś z warunków (Wright 2001c: 206).

Wright nazywa swoją odpowiedź „nieporadną" (*flat-footed*, Wright 2001c: 177), podkreślając jej pozorną naiwność i prostotę. W rzeczywistości jest to bardzo złożona koncepcja, zawierająca ideę autorytetu pierwszoosobowego, która przynajmniej na pierwszy rzut oka odbiega od potocznych intuicji. Teza Wrighta stosuje się bowiem zarówno do przypisań przeszłych stanów intencjonalnych, jak i do przypisań aktualnych, a zatem tych, które zwykliśmy wiązać ze szczególnym autorytetem. Zarówno moje przeszłe, jak i aktualne stany intencjonalne są konstytuowane przez moje *best judgements*, stąd też stanowisko Wrighta jest nazywane konstytutywizmem. W następnych częściach pokażę, że koncepcja Wrighta prowadzi do wielu nieintuicyjnych rozwiązań i boryka się z problemami, które pozbawiają ją waloru atrakcyjnego wyjaśnienia autorytetu pierwszoosobowego i samowiedzy.

---

[5] Warto zauważyć, że gdyby trzymać się dokładnie analogii kolorystycznej, należałoby w tym miejscu mówić nie o determinowaniu, lecz o współdeterminowaniu. Niekiedy jednak Wright pisze tak, jak gdyby przyznawał samoprzypisaniom mocniejszą — determinującą, a nie tylko współdeterminującą — funkcję. W dalszej części tekstu omawiam dwa możliwe sposoby rozumienia tekstów Wrighta: jeden, w którym kluczowa jest analogia percepcyjna, i drugi, który jest od niej niezależny. Te dwie interpretacje traktuję jako osobne, w związku z tym w pierwszej z nich proponuję mówić o współdeterminowaniu, a w drugiej — o determinowaniu.

3. ANALOGIA PERCEPCYJNA RAZ JESZCZE — DLACZEGO *BEST JUDGEMENTS* NIE SĄ NAJLEPSZYM ROZWIĄZANIEM

W tej części skupię się na ontologicznych założeniach teorii Wrighta. W szczególności zajmę się ukrytymi w tej teorii elementami percepcyjnego myślenia o introspekcji. Zastanowię się nad analogią z sądami o kolorach i strukturą poznania introspekcyjnego, która powstaje w wyniku przeniesienia tej analogii na samopoznanie. W ten sposób pokażę, że Wright nie uwolnił się od myślenia o introspekcji w kategoriach percepcyjnych i że niektóre mankamenty jego teorii to te same problemy, z którymi borykały się modele samowiedzy jako percepcji.

Wright z pewnością jest świadomy pułapek myślenia percepcyjnego i nie chciałby powtarzać błędów starych teorii. Niestety, posługując się analogią z sądami dotyczącymi kolorów, proponuje model, który niebezpiecznie zbliża się do percepcyjno-detektywistycznego[6]. Jak bowiem dokładnie wygląda sytuacja poznawcza, w której znajduje się podmiot, kiedy wygłasza sądy na temat kolorów przedmiotów? Własności wtórne, za które uznawane są kolory, są współdeterminowane przez przedmiot, któremu przysługują, i podmiot postrzegający — innymi słowy, kolor rzeczy fizycznej powstaje w wyniku interakcji obserwatora z tą rzeczą. Jak taka rama interpretacyjna wyglądałaby w przypadku stanów intencjonalnych? Przede wszystkim należałoby wskazać podmiot i oddzielone od niego stany intencjonalne jako przedmioty, które podmiot wewnątrz siebie „wykrywa". Dzięki współdziałaniu podmiotu z tymi przedmiotami powstawałyby ich własności wtórne, analogiczne do koloru. W koncepcji Wrighta tą własnością najpewniej musiałaby być treść intencjonalna. Już jednak na pierwszy rzut oka widać, jak niezadowalające byłoby takie rozwiązanie. Zaczynając od klasycznych problemów detektywizmu percepcyjnego: taka ontologia zmuszałaby nas do wizji stanów mentalnych jako obiektów, które są oddzielone od obserwującego je podmiotu (nie wiadomo przy tym, jak miałoby wyglądać to oddzielenie), a kończąc na bardzo nieintuicyjnej wizji relacji między stanem intencjonalnym a jego treścią. Tak jak istnieje w sensie onto-

[6] Sugestię, że analogia kolorystyczna może narzucać Wrightowi percepcyjny model introspekcji, formułuje Richard Moran (2001: 25), jednak jej nie rozwija. Wprowadza taką sugestię, gdy opisuje równoważności, które miałyby wyrażać konstytutywistyczne warunki przypisywania zamiarów (np. *S* ma zamiar, żeby robić ϕ wtedy i tylko wtedy, gdy *S* wydaje sąd, że ma zamiar zrobić ϕ — Moran 2001: 24). Jak pisze, „even if the relevant biconditionals for intention could be specified nontrivially and their a priori status were secured, this would not serve to show that first-person authority was not based on some kind of genuine cognitive advantage. Nor would it even serve the purpose of ruling out a perceptual model of introspection, as the color analogy shows" (Moran 2001: 25).

logicznym przedmiot pozbawiony koloru (nawet jeśli dla ludzkiego oka przedmioty są widziane zawsze w kolorze, to odpowiednia teoria fizyczna informuje nas o tym, że to jedynie złudzenie), tak zgodnie z omawianą ramą interpretacyjną powinien istnieć stan intencjonalny bez intencjonalnej treści. Ale co to miałoby znaczyć — być stanem intencjonalnym i nie mieć treści? Taka konstrukcja sprawia wrażenie sprzeczności pojęciowej[7].

Oczywiście analogia Wrighta jest jedynie analogią: jej filozoficzne znaczenie jest z pewnością mniejsze niż znaczenie standardowej argumentacji filozoficznej. Jednak niezależnie od tego, jaką wagę jej przyznamy, ontologia Wrighta, bazująca na zależności od sądu, okazuje się niewystarczająca do ujęcia autorytetu pierwszoosobowego i samowiedzy. W trafny sposób tę nieadekwatność uchwycił Moran, pisząc, że zależność stanów mentalnych od sądu nie jest niczym swoiście pierwszoosobowym. Taka zależność cechuje wiele pojęć i nic w analizie Wrighta nie wskazuje na swoiście pierwszoosobowy charakter tej relacji w przypadku przypisań stanów mentalnych (Moran 2001: 25). Co więcej, moglibyśmy w analogiczny sposób sprecyzować odpowiednie warunki (*C-conditions*) rządzące przypisaniami trzecioosobowymi i precyzujące, kiedy sąd innej osoby determinowałby[8] stan mentalny podmiotu tego stanu. W istocie zresztą nasze funkcjonowanie jako racjonalizujących interpretatorów w sensie Donalda Davidsona czy Daniela Dennetta działa w podobny sposób (Moran 2001: 25). Koncepcja Wrighta nie daje nam żadnych argumentów za tym, że to pierwszoosobowe przypisania miałyby mieć specjalny status. Jak dodaje Moran, teoria Wrighta nie mówi nam też nic o swoiście asymetrycznej roli dowodów w uznawaniu samoprzypisań. W przypadku pierwszoosobowego przypisania zwykle pytanie o dowód na jego rzecz nie ma sensu — inaczej niż gdy mamy do czynienia z sądem osoby trzeciej (Moran 2001: 26)[9]. W sytuacji Wrightowskiego stwarzania stanów mentalnych nie ma zatem nic swoiście pierwszoosobowego.

---

[7] Można by replikować, że sytuacja percepcji kolorów dostarcza pewnego rozwiązania: tak jak przedmioty zewnętrzne mają już pewną własność, która sprawia, że w kontakcie z układem wzrokowym człowieka zyskują kolor, tak można by tego rodzaju własności protointencjonalności doszukiwać się w stanach umysłowych, które dopiero po obserwacji podmiotu zyskiwałyby rzeczywistą intencjonalność. Problem polega jednak na tym, że trudno wyobrazić sobie, czym miałaby być owa własność protointencjonalności, która jeszcze intencjonalnością nie jest. Nie umiem znaleźć przekonującej koncepcji tego rodzaju.

[8] Moran wyraża się w tym miejscu w sposób, który sugerowałby determinowanie, jednak gdyby chcieć pozostać w zgodzie z analogią kolorystyczną, należałoby to nieco osłabić i napisać raczej o współdeterminowaniu (por. przypis 5).

[9] Moran dodaje, że koncepcja Wrighta nie wyjaśnia również, dlaczego autorytet pierwszoosobowy jest związany z pewnymi racjonalnymi wymaganiami w stosunku do podmiotu (Moran 2001: 26). To jednak jest w tym miejscu mniej istotne niż asymetrie wymienione w tekście głównym.

Warto zwrócić uwagę, że ten sam problem dotyka także teorii, które ujmują samowiedzę na wzór percepcji. Teorie percepcyjne mają bowiem trudność z wyjaśnieniem asymetrii między pierwszoosobowym a trzecioosobowym dyskursem psychologicznym. Percepcja jest omylna, tymczasem w przypadku przypisań stanów mentalnych zakładamy możliwość błędu znacznie mniej chętnie w przypisaniach pierwszoosobowych niż trzecioosobowych. Ponadto każdy obiekt może być potencjalnie postrzeżony przez większość ludzi, o ile znajdą się w tych samych warunkach, tymczasem nie jesteśmy skłonni uznać samoprzypisań za posiadające taki sam status jak twierdzenia innych osób na nasz temat[10]. Jak się zatem okazuje, Wright nie uwolnił się od problemów koncepcji detektywistyczno-percepcyjnych.

Wskazane trudności podważają model samowiedzy zaproponowany przez Wrighta. Myślę jednak, że uznanie ich za decydujące nie oddałoby sprawiedliwości konstytutywizmowi. Można bowiem odczytać teksty Wrighta nieco inaczej. W kolejnych częściach artykułu przedstawię taką alternatywną interpretację, a następnie pokażę, że nie chroni ona omawianego poglądu przed zarzutami.

4. AUTORYTET PIERWSZOOSOBOWY JAKO KWESTIA „GRAMATYCZNA"

Dorit Bar-On proponuje wyróżnienie dwóch rodzajów konstytutywizmu (Bar-On 2009: 60). Pierwszy z nich to konstytutywizm „gramatyczny". Na gruncie tej koncepcji autorytet pierwszoosobowy to po prostu element naszego dyskursu mentalnego. Konwencje rządzące dyskursem wyznaczają kategorię bytów mentalnych i ustanawiają specjalny status samoprzypisań. Natomiast konstytutywizm „ontologiczny" uznaje autorytet pierwszoosobowy za konsekwencję naszej szczególnej relacji do własnych stanów mentalnych, dzięki której trafnie ujmujemy ich naturę.

Interpretacja Wrighta, którą jak dotąd zaproponowałam, przedstawia jego tezy jako twierdzenia ontologiczne na temat natury stanów intencjonalnych. Jak jednak zauważa Bar-On, Wright waha się między interpretacją ontologiczną a „gramatyczną" (Bar-On 2009: 60-61). Z jednej strony analogia z kolo-

[10] Należy zatem zaznaczyć, że różnice między samowiedzą a percepcją przedmiotów zewnętrznych są zarówno jakościowe (np. błędy w samopoznaniu nie polegają na mylnej identyfikacji przedmiotu wewnętrznego), jak i ilościowe (np. zakładamy, że w samoprzypisaniach podmiot myli się rzadziej, niż kiedy wygłasza sądy na temat przedmiotów zewnętrznych, które postrzega).

rami i fragmenty na temat zależności od *best judgements* sugerują, że próbuje się tu zarysować wiarygodne ujęcie statusu ontycznego stanów intencjonalnych. Z drugiej strony, gdy trudno o wiarygodną argumentację ontologiczną, Wright częściej sięga po argumenty „gramatyczne", broniąc się niejako przed odpowiedzialnością za pomocą twierdzenia, że mówi jedynie o gramatyce pojęć mentalnych. Sugeruje, że to „gramatyka" naszych przypisań intencji uprawnia nas do zakładania autorytetu podmiotu w kwestii własnego umysłu (Wright 2001c: 202). Pojęcie intencji działa bowiem w taki sposób, że opinia[11] samego podmiotu odgrywa istotną rolę w determinowaniu jego stanów mentalnych (Wright 2001c: 203). Można zatem powiedzieć, że zgodnie z takim ujęciem mamy po prostu zwyczaj innego traktowania zdań, w których pewna osoba przypisuje sobie pewien aktualny stan mentalny, niż zdań, w których na przykład osoba trzecia dokonuje przypisania tej osobie tego stanu. Gdy mówię, że zamierzam pójść na najnowszy film Almodóvara, moje samoprzypisanie jest akceptowane bez wątpliwości ze względu na to, że tak funkcjonuje w kontekście naszych praktyk językowych pojęcie intencji.

Na pierwszy rzut oka wydaje się, że przyjęcie stanowiska „gramatycznego" może pomóc rozwiązać przynajmniej niektóre ze wspomnianych już problemów stanowiska Wrighta. Pozwala na przykład częściowo oddalić ostatni zarzut na temat braku odpowiedniej asymetrii pierwszej i trzeciej osoby. Wright mógłby na niego odpowiedzieć twierdzeniem, że asymetrie są zapewnione przez gramatykę naszych pojęć. Tak po prostu działa nasze pojęcie zamiaru, że sądy pierwszoosobowe traktujemy inaczej niż trzecioosobowe, i na tym polega całe wyjaśnienie. Takie podejście okazuje się jednak niezadowalające. Przede wszystkim rozczarowujące jest to, co za Barrym C. Smithem można nazwać deflacjonizmem teorii Wrighta (Smith 2012: 141). Dla Smitha rozwiązanie Wrighta jest deflacyjne, ponieważ mimo że uwzględnia pewne aspekty samowiedzy — takie jak autorytet pierwszoosobowy — to jego autor nie ma odwagi, by podać wyjaśnienie swoiście pierwszoosobowej perspektywy (Smith 2012: 141-142). W tym kontekście można powiedzieć, że deflacjonizm Wrighta zasadza się na założeniu pozostawania na powierzchni dyskursu mentalnego: poprzestaniu na opisie zjawiska gramatycznego bez jego głębszego wyjaśnienia. Jak zauważa Bar-On, takie podejście jest rozczarowujące, ponieważ oczekujemy czegoś więcej od wyjaśnienia filozoficznego (Bar-On 2012: 178). W zasadzie można powiedzieć, że pytanie o gramatykę naszych stanów mentalnych jest jedynie innym sformułowaniem pytania o autorytet pierwszoosobowy: wszak autorytet pierwszoosobowy to zjawisko językowe, polegające na

[11] W przytaczanym fragmencie mowa jest o opinii, sugeruję jednak, by rozumieć to jako sąd podmiotu, tak jak wyrażałam się we wcześniejszych partiach tekstu.

tym, że pewne zdania uznajemy za prawdziwe bez pytania o dowody na ich rzecz. To jest właśnie zjawisko gramatyczne i dlatego gramatyka naszych stanów mentalnych jest raczej tym, co chcemy wyjaśnić, niż tym, za pomocą czego chcemy wyjaśniać. Kiedy stawiamy pytanie o autorytet pierwszoosobowy, pytamy: dlaczego nasze pojęcia funkcjonują w taki sposób? Dlaczego ich gramatyka jest właśnie taka, że uprzywilejowuje samoprzypisania względem przypisań trzecioosobowych? Odpowiedź, która zatrzymuje się na konstatacji, że taki jest stan rzeczy, nie jest żadnym rozwiązaniem filozoficznego problemu rozważanego w tej pracy.

Na ten zarzut można by odpowiedzieć, że niekiedy Wright proponuje pewne wyjaśnienie gramatyki autorytetu pierwszoosobowego. Łączy ją mianowicie z naszym rozumieniem tego, co to znaczy być podmiotem racjonalnym. Twierdzi, że autorytet pierwszoosobowy jest prawem, które przyznajemy każdemu, kogo bierzemy za racjonalny podmiot (Wright 2001b: 137-138), natomiast całkowite podważenie samoprzypisań stanów intencjonalnych danej osoby kłóci się z pojmowaniem jej jako posiadającej stany intencjonalne (Wright 1998: 17-18). Członkowie mojej społeczności są bowiem w stanie wchodzić ze mną w interakcje tylko pod warunkiem założenia mojej racjonalności, to zaś wymaga od nich przyznania autorytetu moim samoprzypisaniom. Takie wyjaśnienie nie przybliża nas jednak do rozwiązania problemu. Po pierwsze, jest wysoce niejasne, w szczególności: nie wiadomo, gdzie jest granica, przed którą jeszcze kwestionowanie samoprzypisań nie musi wiązać się z kwestionowaniem racjonalności (mimo autorytetu, samoprzypisania nie są przecież niekwestionowalne). Po drugie, taka interpretacja pociąga daleko idące konsekwencje epistemologiczne i ontologiczne. Skoro bowiem treść moich stanów intencjonalnych i moja wiedza na jej temat mają być determinowane przez moje sądy, a te zyskują sankcję tylko dzięki innym ludziom, to w rezultacie należy uznać również, że bez trzecioosobowej gwarancji moje stany mentalne nie miałyby treści, a ja nie byłabym ich świadoma. Nie jestem przekonana, czy tego rodzaju konsekwencję jesteśmy gotowi przyjąć. Tym bardziej że istnieje inne wyjaśnienie autorytetu pierwszoosobowego, które przedstawię — za Finkelsteinem — w następnej części artykułu.

5. DRETSKE, EKSPRESYWIZM I ŹLE POSTAWIONE PYTANIE WRIGHTA

Przypomnijmy: Wright wyszedł od problemu kierowania się regułą, pochodzącego od Wittgensteina i Kripkego. Finkelstein, który krytycznie ocenia

konstytutywistyczną propozycję Wrighta, proponuje neoekspresywistyczne rozwiązanie problemu reguł i samowiedzy (Finkelstein 2003). Podążając za jego propozycją, pokażę, że ekspresywistyczne z ducha rozwiązanie problemów teorii Wrighta można zaczerpnąć z pojednawczego sceptycyzmu Freda Dretskego. Według mnie przywołanie pojednawczego sceptycyzmu pozwala zrozumieć, że błąd Wrighta wynika ze skupienia się na poszukiwaniu samowiedzy, podczas gdy pierwotna wobec wiedzy jest ekspresja bezpośredniej świadomości treści własnych stanów mentalnych.

Dretske (2012) przedstawia koncepcję pojednawczego sceptycyzmu (*conciliatory skepticism*) w kwestii samopoznania. W myśl tej koncepcji mamy bezpośrednią świadomość i autorytet odnośnie do tego, *co* myślimy, czujemy, *czego* doświadczamy i w ogólności: jaka jest treść naszego umysłu, nie mamy jednak żadnego specjalnego dostępu ani uprzywilejowanej pozycji, by stwierdzać, *że* myślimy, czujemy, doświadczamy i w ogólności: że mamy umysł (Dretske 2012: 49).

Dretske ilustruje swoją tezę historyjką o Sarze, trzyletniej dziewczynce, która myśli, że jej tata właśnie przyjechał do domu (Dretske 2012: 55). Sara nie wie jeszcze, co to znaczy „myśleć", nie wie też więc, że myśli. W jej głowie nie pojawia się myśl drugiego rzędu o treści: „Mam myśl, że tata wrócił do domu". Jednocześnie jednak, jak celnie zauważa Dretske, pragnienia i przekonania Sary wyjaśniają jej zachowania równie dobrze jak w przypadku osób intelektualnie dojrzałych i samoświadomych myślenia (Dretske 2012: 55). Dlatego Sara biegnie, by otworzyć drzwi, kiedy słyszy dźwięk samochodu, i mówi do mamy „Tata wrócił". Zachowanie dziewczynki jest celowym i w pewnym sensie świadomym zachowaniem, którego przyczyną była myśl o treści „tata wrócił do domu". To tej myśli dziewczynka daje wyraz, wykrzykując w stronę matki informację o przybyciu ojca (por. Dretske 2012: 57). Sens, w jakim Sara jest świadoma treści swojej myśli, to właśnie owa bezpośrednia świadomość tego, co jest w naszym umyśle. Taka świadomość jest czymś innym niż wiedza w sensie propozycjonalnym — wiedza, że mamy pewnego rodzaju myśli. Świadomość treści naszych myśli jest nawet czymś innym niż wiedza *o* treści naszych myśli. Sara nie ma jeszcze zdolności pojęciowych pozwalających zrozumieć, co to jest myślenie, a więc do uzyskania wiedzy, *że myśli*, która jest potrzebna do wiedzy *o tym, co myśli* (Dretske 2012: 54). Na razie dostępna jest jej tylko świadomość w sensie jakiejś epistemicznej relacji, która nie jest wiedzą. Dziewczynka dysponuje więc czymś, co Dretske nazywa *unwitting authority*, a co można przetłumaczyć jako mimowolny czy nieświadomy autorytet (Dretske 2012: 57).

Należy podkreślić, że koncepcja Dretskego nie wyklucza możliwości zdobycia samowiedzy. Istotne jest jedynie, by odróżnić od siebie trzy aspekty życia

umysłowego: świadomość treści własnych myśli, wiedzę, że ma się określone myśli, i wiedzę o treści własnych myśli. Sara dysponuje jedynie tym pierwszym. Świadomość treści własnych myśli daje Sarze mimowolny autorytet odnośnie do nich, ale nie jest źródłem wiedzy, że ma ona takie właśnie myśli. Tego rodzaju wiedza musi przyjść skądinąd. Kiedy zaś Sara będzie już wiedzieć, że ma określone myśli, świadomość ich treści przekształci się w wiedzę o ich treści. Aby zrozumieć relację między tymi pojęciami, warto przywołać jeszcze jeden przykład omawiany przez Dretskego. Wyobraźmy sobie, że „The Philosophical Gazette" daje Dretskemu ustne zapewnienie, że opublikuje wszystko, co filozof napisze. Dretske pisze więc *p*, a periodyk publikuje *p*. W takiej sytuacji wiedza o tym, co gazeta opublikuje, ma inne źródło niż wiedza o tym, że gazeta to właśnie publikuje. Wiedzę o tym, co gazeta opublikuje, Dretske zdobywa za pomocą wzroku — obserwując, co sam pisze. Aby jednak ta wiedza była możliwa, musi on najpierw wiedzieć, że gazeta opublikuje jego tekst — o tym zaś zapewnia go inne źródło: usłyszana obietnica wydawców (Dretske 2012: 54). Sama świadomość *p* — czyli tego, co gazeta opublikuje — nie jest w stanie zapewnić Dretskemu wiedzy, że gazeta opublikuje właśnie *p*.

Podobnie Dretske każe nam myśleć o samowiedzy. Choć mamy uprzywilejowaną świadomość własnych myśli, nie jest ona źródłem wiedzy o tym, że myślimy (Dretske 2012: 54), i nie wystarcza sama do zapewnienia nam wiedzy o tym, co myślimy. Co jednak istotne, historyjka o Sarze pokazuje, że wiedza wcale nie jest potrzebna, żeby działać zgodnie z regułą. Należy bowiem zauważyć, że problem określenia, jakiego działania wymaga od dziewczynki jej stan mentalny, w ogóle się dla niej nie pojawia. Działanie jest tu niejako automatycznie sprzężone z posiadaniem stanu intencjonalnego, a luka między treścią stanu a zachowaniem, które jest z nim zgodne, nie ma jak powstać. Być może jest tak dlatego, że działanie dziewczynki to naturalna ekspresja jej stanu mentalnego, jak powiedzieliby zwolennicy jakiejś formy ekspresywizmu w kwestii stanów mentalnych. W tym kontekście warto przywołać uwagę 244 *Dociekań filozoficznych*, w której Wittgenstein sugeruje ekspresywistyczną genezę samoprzypisań. Według Wittgensteina zdania na temat bólu zastępują naturalne reakcje i przez to są tak samo naturalnym wyrazem doznania (Wittgenstein 2000: § 244). Dorośli uczą dziecko mówić „Boli mnie głowa" w zastępstwie skrzywienia twarzy czy jęku. Tak jak nie powstaje w ogóle luka między stanem mentalnym a jękiem, czyli jego ekspresją, tak też nie powstaje żadna luka między samoprzypisaniem a stanem intencjonalnym. I choć przykład z bólem dotyczy stanu fenomenalnego, historyjka Dretskego sugeruje, że o stanach intencjonalnych można myśleć podobnie. Sara mogła równie dobrze po prostu pobiec do drzwi zamiast mówić, że tata wrócił do domu — każda z tych czynności byłaby jakąś ekspresją jej stanu mentalnego.

Ten przykład pokazuje, że problem Wrighta jest źle postawiony. Teoria Wrighta jest budowana w odpowiedzi na pytania: „Skąd mogę wiedzieć, w jakim stanie mentalnym się znajduję? Skąd mogę wiedzieć, jakiego działania wymaga ode mnie w tym miejscu reguła?". Wright kładzie nacisk na wiedzę, w rezultacie czego poszukuje warunków prawdziwości dla zdań, w których samoświadomy podmiot *explicite* stwierdza bycie w pewnym stanie mentalnym[12]. Tymczasem być może na problem reguł należy spojrzeć z innej perspektywy: nie z perspektywy pytania o wiedzę i interpretację, lecz od strony, od której „widzi" swoje stany mentalne trzyletnia Sara. Z jej punktu widzenia bieg w kierunku drzwi jest naturalną ekspresją przekonania, że tata wrócił do domu. Z tym przekonaniem Sara jest w szczególnej relacji epistemicznej „bycia świadomym", dziewczynka nie ma jednak wiedzy w sensie przekonania, że ma przekonanie, że tata wrócił do domu.

Przywołanie pojednawczego sceptycyzmu Dretskego pozwala zauważyć, że nie trzeba wiedzieć, w jakim stanie mentalnym się jest, by działać zgodnie z nim. Dzięki temu zaś okazuje się, że pewien aspekt problemu reguł zostaje rozwiązany: wystarczy bezpośrednia świadomość, by wyrażać swój umysł. A podejście Wrighta, które koncentruje się na poszukiwaniu odpowiedzi na pytania o wiedzę, pomija ten aspekt.

Jak już wspomniałam, moja propozycja idzie za rozważaniami Finkelsteina, który interpretuje na sposób neoekspresywistyczny[13] problem kierowania się regułą na gruncie samych *Dociekań filozoficznych* i upatruje klęskę Wrighta między innymi w błędnym podejściu do problemu reguł (Finkelstein 2003: 42). Jak zauważa Finkelstein, według Wittgensteina problem reguł w ogóle nie powstanie, jeśli spojrzymy na regułę z perspektywy konkretnych praktyk i sposobów życia (Finkelstein 2003: 86). Nikt na co dzień nie zastanawia się nad znaczeniem symboli, za pomocą których się porozumiewamy, ponieważ napotykamy symbole w kontekście konkretnego sposobu życia (por. Finkelstein 2003: 82). Znak nie wymaga od nas interpretacji, ponieważ w kontekście naszych praktyk widzimy go od razu jako wyposażony w znaczenie — tak jak patrząc na słowa, czytamy je od razu jako słowa, a nie jako nieznaczące plamy farby drukarskiej (Finkelstein 2003: 82). Na tej samej zasadzie nie pojawia się też wątpliwość, jaka jest treść mojego stanu intencjonalnego i jakie za-

[12] Niewykluczone, że Wright częściowo „dziedziczy" taki sposób zadania pytania po tradycji rozważań nad problemem kierowania się regułą, w szczególności po Kripkem.

[13] Neoekspresywistycznymi nazywa się koncepcje, które rozwijają podstawową myśl ekspresywizmu, że samoprzypisania są ekspresjami stanów mentalnych, ale przyjmują jednocześnie, że samoprzypisania mają wartość logiczną (Gertler 2020). Neoekspresywiści dopuszczają więc, że podmiot może mieć wiedzę na temat własnych stanów mentalnych. Do tej ostatniej kwestii odniosę się w podsumowaniu.

chowanie będzie z nim zgodne. Finkelstein podaje przykład szefa wydającego polecenie asystentowi. W obrębie określonego sposobu życia asystent widzi zachowanie szefa od razu jako wyrażenie jego pragnienia i chwyta rozkaz bez potrzeby interpretacji (Finkelstein 2003: 90). Dlatego, jak twierdzi Finkelstein, Wright popełnia błąd, zakładając, że luka między zachowaniem a ekspresją pragnienia w takiej sytuacji w ogóle powstaje (Finkelstein 2003: 90).

Najistotniejszym aspektem koncepcji neoekspresywistycznych, który uniemożliwia pojawienie się problemu reguł, jest bezpośredni charakter ekspresji: między moim stanem mentalnym a działaniem, które je wyraża, nie pośredniczy żaden sąd, w którym rozpoznawałabym ten stan (por. Gertler 2020). Bar-On — również znana neoekspresywistka — jako paradygmatyczny przykład zachowania ekspresywnego omawia sięganie przez małą dziewczynkę po pluszowego misia, strukturalnie analogiczne do zachowania Sary opisywanej przez Dretskego. Bohaterka Bar-On, Jenny, chce misia, a jej sięganie jest po prostu wyrazem tego pragnienia. Jak pisze Bar-On, między stanem mentalnym Jenny a jej zachowaniem nie ma żadnego epistemicznego dystansu: Jenny po prostu daje upust pragnieniu (Bar-On 2004: 241). Rozważenie przykładów Dretskego i Bar-On pokazuje, że nawet podmioty, które nie dysponują wiedzą na temat własnych stanów mentalnych, potrafią działać zgodnie z regułą, gdy wyrażają bezpośrednio swoje stany mentalne.

Można jednak wysunąć wątpliwość wobec skuteczności takiego rozwiązania. Nasuwa się bowiem pytanie, czy odwołanie do ekspresji i bezpośredniej świadomości pozwala ująć wszystkie istotne zjawiska — nie chodzi nam wszak jedynie o sytuacje, w których podmiot nieznający jeszcze języka wyraża swoje stany mentalne. Chodzi również o przypisywanie sobie stanów mentalnych za pomocą artykułowanych sądów, które są czymś więcej niż ekspresja niejęzykowa[14]. Być może nie powstaje epistemiczna luka między pragnieniem misia a sięgnięciem po niego, ale to nie znaczy, że rozwiązany został problem luki między stanami mentalnymi a samoprzypisaniami w sądach, a zatem nie posunęliśmy się do przodu w rozwiązywaniu głównego problemu. Na taki zarzut należy odpowiedzieć dwojako. Po pierwsze, nie jest prawdą, że samo zwrócenie uwagi na ekspresję i bezpośrednią świadomość nie przybliża nas do rozwiązania problemu. Jednym z pytań, które powstają na gruncie rozważania problemu reguł, jest pytanie: jakiego działania wymaga ode mnie reguła? Omówione przykłady pokazują, że można działać zgodnie z regułą, nie wiedząc jednocześnie, że się zgodnie z nią działa. Po drugie, w teoriach neoekspresywistycznych można odnaleźć propozycje wyjaśnienia procesu, w wyniku którego naturalne zachowania ekspresywne przekształcają się w samoprzypi-

[14] Dziękuję anonimowemu recenzentowi za zwrócenie mi uwagi na tę kwestię.

sania mające formę sądów wyrażonych w języku naturalnym. Zgodnie z koncepcją Bar-On w toku rozwoju jednostki samoprzypisania artykułowane językowo zastępują naturalne sposoby ekspresji, stając się nowym narzędziem wyrażania stanów mentalnych (Bar-On 2004: 288). Dorośli uczą Jenny mówić „Chcę misia" zamiast wyciągania rąk w kierunku zabawki lub wykrzykiwania pojedynczych słów (Bar-On 2004: 288). Co jednak istotne, samoprzypisanie Jenny będzie tak samo bezpośrednie jak naturalne formy ekspresji — będzie wypływać bezpośrednio z jej pragnienia, nie będąc zapośredniczone przez żaden sąd, w którym zdawałaby sprawę z udanego rozpoznania stanu mentalnego (Bar-On 2004: 241, 262)[15]. Ekspresja nie musi zatem wykluczać samoprzypisań artykułowanych w sądach.

PODSUMOWANIE

Na podstawie analizy koncepcji samowiedzy Wrighta pokazałam, jak powstała w odpowiedzi na problem kierowania się regułą i jako alternatywa dla koncepcji detektywistyczno-percepcyjnych. Następnie przedstawiłam problemy związane z takim myśleniem o samowiedzy i autorytecie pierwszoosobowym. Po pierwsze, argumentowałam, że teoria Wrighta pojmowana jako konstytutywizm ontologiczny (zgodnie z przywoływanym podziałem Bar-On) wcale nie uwalnia się od percepcyjnego myślenia o samowiedzy, co podważa jej antydetektywistyczny charakter. Elementy myślenia percepcyjnego w konstytutywizmie wprowadzają ponadto nieintuicyjną wizję relacji między stanem intencjonalnym a jego treścią. Pojawia się też problem braku asymetrii między przypisaniem pierwszoosobowym i trzecioosobowym, z którym borykały się także teorie percepcyjne. W dalszej kolejności rozważyłam teorię Wrighta jako konstytutywizm „gramatyczny" i omówiłam jego nieadekwatność jako propozycji ujęcia autorytetu pierwszoosobowego i samowiedzy.

Na koniec zaś sięgnęłam do pojednawczego sceptycyzmu Dretskego, żeby pokazać, że źródłem problemów Wrighta jest źle zadane pytanie. Wright próbuje rozwiązać problem reguł, poszukując faktu, który umożliwiałby podmiotowi wiedzę na temat własnych stanów intencjonalnych. Tymczasem problem reguł znika, kiedy przyjmie się, że pierwotna wobec wiedzy jest ekspresja.

[15] Ściślej rzecz biorąc, Bar-On nie wyklucza tego, że kiedy podmiot wyraża swój stan mentalny za pomocą samoprzypisania, jednocześnie wyraża sąd wyższego rzędu dotyczący tego, że znajduje się w pewnym stanie mentalnym — istotne jest to, że wyjaśnienie uprzywilejowanego statusu samoprzypisań, które nie odwołuje się do takiego sądu, jest wystarczające (Bar-On 2004: 306-307).

W tym kierunku idzie Finkelsteina rozwiązanie problemu reguł, ja zaś pokazałam, że przekonujące ekspresywistyczne rozwiązanie można wywieść także z pojednawczego sceptycyzmu Dretskego.

Nasuwa się oczywiście pytanie, czy rozwiązanie, które omówiłam, jest w stanie wyjaśnić także samowiedzę. Wszak czym innym jest wyrażanie bezpośrednio swojego stanu mentalnego, a czym innym wiedza, że określony stan mentalny się wyraża. Analogicznie można powiedzieć, że jest różnica między działaniem zgodnie z regułą a podporządkowywaniem się regule, gdzie to drugie zakłada wiedzę na temat reguły. Czy w takim razie udało się pokazać również możliwość samowiedzy wobec problemu reguł?

Odpowiedź na to pytanie jest dwuaspektowa. Z jednej strony koncepcje, które przedstawiłam, nie wykluczają możliwości samowiedzy. Dretske wprost pisze, że nie neguje istnienia samowiedzy, a jedynie podaje w wątpliwość popularny pogląd na temat jej źródła (Dretske 2012: 50). Jednocześnie jednak nie proponuje teorii, która pokazywałaby, jak samowiedza rzeczywiście powstaje, a jedynie sugeruje, że wiedza, iż myślimy, pochodzi od innych ludzi — od rodziców, nauczycieli — od których uczymy się, co to znaczy „myśleć" (Dretske 2012: 60)[16]. Dlatego też nie jest jasne, jak z jego koncepcji wyprowadzić teorię samowiedzy. W związku z tym należy przyznać, że przedstawione rozwiązania nie są kompletne: nie wynika z nich jasna odpowiedź na pytanie o samowiedzę wobec problemu reguł. Problem ten wymaga osobnych badań. Jednocześnie jednak — jak pisałam w części 5 — przesunięcie ciężaru rozważań na ekspresję pozwala na przynajmniej częściowe poradzenie sobie z problemem reguł: można bowiem wyrażać swoje stany mentalne, nie wiedząc, że się je rzeczywiście wyraża.

BIBLIOGRAFIA

Armstrong D. M. (1968), *A Materialist Theory of the Mind*, London: Routledge & K. Paul.

Bar-On D. (2004), *Speaking My Mind: Expression and Self-Knowledge*, Oxford: Clarendon Press. https://doi.org/10.1093/0199276285.001.0001

Bar-On D. (2009), *First-Person Authority: Dualism, Constitutivism, and Neo-Expressivism*, „Erkenntnis" 71(1), 53-71. https://doi.org/10.1007/s10670-009-9173-y

[16] Z kolei Bar-On (2004) przedstawia kilka koncepcji samowiedzy, które można by pogodzić z jej neoekspresywizmem, nie jestem jednak pewna, czy możliwych do pogodzenia z pojednawczym sceptycyzmem Dretskego. Finkelstein rezygnuje zaś z zajmowania jasnego stanowiska, twierdząc, że spory na ten temat nie mają autentycznej wagi filozoficznej (Finkelstein 2003: 148-152).

Bar-On D. (2012), *Expression, Truth, and Reality: Some Variations on Themes from Wright* [w:] *Mind, Meaning, and Knowledge: Themes from the Philosophy of Crispin Wright*, A. Coliva (ed.), Oxford: Oxford University Press, 162-192. https://doi.org/10.1093/acprof:oso/9780199278053.001.0001

Comte A. (2009), *The Positive Philosophy of Auguste Comte* (Cambridge Library Collection — Religion), vol. 1, tr. H. Martineau, Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511701450

Davidson D. (1987), *Knowing One's Own Mind*, „Proceedings and Addresses of the American Philosophical Association" 60(3), 441-458. https://doi.org/10.2307/3131782

Dretske F. (2012), *Awareness and Authority: Skeptical Doubts about Self-Knowledge* [w:] *Introspection and Consciousness*, D. Smithies, D. Stoljar (eds.), Oxford: Oxford University Press, 49-64. https://doi.org/10.1093/acprof:oso/9780199744794.003.0002

Dziobkowski B. (2016), *Teorie znaczenia* [w:] *Przewodnik po filozofii języka*, J. Odrowąż-Sypniewska (red.), Kraków: Wydawnictwo WAM, 19-66.

Finkelstein D. H. (2003), *Expression and the Inner*, Cambridge, MA: Harvard University Press.

Gertler B. (2020), *Self-Knowledge* [w:] *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), E. N. Zalta (ed.), https://stanford.io/34SJqJY.

Komorowska-Mach J. (2015), *Przypisywanie sobie emocji — wyrażanie czy wykrywanie? Elementy detektywistyczne w modelu neoekspresywistycznym*, „Filozofia Nauki" 23(4) [92], 41-54.

Kripke S. (2007), *Wittgenstein o regułach i języku prywatnym*, tłum. K. Posłajko, L. Wroński, Warszawa: Aletheia.

Moran R. (2001), *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9781400842971

Ryle G. (2009), *The Concept of Mind*, London–New York: Routledge. https://doi.org/10.4324/9780203875858

Russell B. (1912), *The Problems of Philosophy*, London: Williams and Norgate.

Schwitzgebel, E. (2016), *Introspection* [w:] *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), E. N. Zalta (ed.), https://stanford.io/31k5IDB.

Shoemaker S. (1994), *Self-Knowledge and "Inner Sense": Lecture I. The Object Perception Model*, „Philosophy and Phenomenological Research" 54(2), 249-269. https://doi.org/10.2307/2108488

Smith B. C. (2012), *The Publicity of Meaning and the Interiority of Mind* [w:] *Mind, Meaning, and Knowledge: Themes from the Philosophy of Crispin Wright*, A. Coliva (ed.), Oxford: Oxford University Press, 127-161. https://doi.org/10.1093/acprof:oso/9780199278053.001.0001

Wittgenstein L. (2000), *Dociekania filozoficzne*, tłum. B. Wolniewicz, Warszawa: Wydawnictwo Naukowe PWN.

Wright C. (1998), *Self-Knowledge: The Wittgensteinian Legacy* [w:] *Knowing Our Own Minds*, C. Wright, B. C. Smith, C. Macdonald (eds.), Oxford: Oxford University Press, 13-45. https://doi.org/10.1093/0199241406.001.0001

Wright C. (2001a), *Kripke's Account of the Argument against Private Language* [w:] *Rails to Infinity: Essays on Themes from Wittgenstein's* Philosophical Investigations, Cambridge, MA: Harvard University Press, 91-115.

Wright C. (2001b), *On Making Up One's Mind: Wittgenstein on Intention* [w]: *Rails to Infinity: Essays on Themes from Wittgenstein's* Philosophical Investigations, Cambridge, MA: Harvard University Press, 116-142.
Wright C. (2001c), *Wittgenstein's Rule Following Considerations and the Central Project of Theoretical Linguistics* [w:] *Rails to Infinity: Essays on Themes from Wittgenstein's* Philosophical Investigations, Cambridge, MA: Harvard University Press, 170-213.