

ANITA PACHOLIK-ŻUROMSKA*

MODELE JAŻNI W SAMOWIEDZY

Abstract

MODELS OF SELF IN SELF-KNOWLEDGE

This article argues for an externalist self-model in self-knowledge, as an answer to the question of how to defend self-knowledge in the face of evident errors in self-cognition such as self-illusions. The externalist model is contrasted with an internalist model, although both are based on a predictive coding mechanism. The main thesis is that for preserving the status of self-knowledge as knowledge an external element is needed – namely, some feedback from the environment – which corrects not only cognitive prediction errors regarding the world but also errors in self-cognition by forcing the proper action of the subject (agent) and dynamic adaptation of the model of the embodied self. The mechanism of self-corrections is best described by externalist models.

Keywords: self-knowledge, self, predictive coding, externalism, internalism, enactivism, cybernetics

Celem artykułu jest analiza dwóch kognitywistycznych modeli jażni w kontekście problemu odporności samowiedzy na błąd poznawczy. Istotne przy tym będzie sprawdzenie, który z przedstawianych modeli pozwala na wykrywanie i naprawianie tego typu błędów. Zarówno przedstawiane tu ujęcie samowiedzy, jak i modele jażni nie mają klasycznego epistemologicznego charakteru, ponieważ odwołują się do koncepcji samowiedzy i rozumienia błędów samopoznania, które pojawiają się raczej w literaturze kognitywistycznej niż filozoficznej. Interdyscyplinarność kognitywistyki ma swoje zalety i wady, które ujawniają się też w podejściu do samowiedzy. Z jednej strony, łączenie dziedzin pozwala ująć empirycznie problem błędów samopoznania. Z drugiej strony, miesza podejścia filozoficzne i psychologiczne, co nie zawsze

* Katedra Kognitywistyki, Wydział Humanistyczny, Uniwersytet Mikołaja Kopernika w Toruniu, ul. Fosa Staromiejska 1a, 87-100 Toruń, e-mail: apz@umk.pl, ORCID: <https://orcid.org/0000-0002-8674-9412>.

skutkuje rozjaśnieniem problemu, a raczej przynosi odwrotny skutek. Widać to między innymi w samym sposobie rozumienia samowiedzy, a co za tym idzie i błędów samopoznania.

Wspólną dla psychologii i filozofii, bardzo ogólną definicję samowiedzy można by sformułować tak: jest to uprzywilejowana wiedza podmiotu na temat jego własnych stanów psychicznych (por. np. Gertler 2017, Neisser 1988, Peacocke 1999, Schooler i in. 2011). Należy przy tym zaznaczyć, że samo pojęcie wiedzy jest problematyczne, ponieważ o ile dla psychologów oznacza ono zazwyczaj zbiór przekonań podmiotu, o tyle dla filozofów owe przekonania muszą dodatkowo spełniać dalsze kryteria, np. być prawdziwe i uzasadnione. Niemniej, ogólnie ujęty problem samowiedzy można sformułować w postaci pytania, czy możemy się mylić co do własnych stanów mentalnych, a jeśli tak, to czy możemy taką pomyłkę samodzielnie skorygować. Innymi słowy, poszukujemy odpowiedzi na pytanie, jak silny pierwszoosobowy autorytet powinniśmy przyjąć (a w skrajnej wersji tego pytania: czy jest on w ogóle do utrzymania) w świetle badań empirycznych wskazujących pewne szczególne przypadki autoiluzji i deficytów poznawczych, takie jak somatoparafrenia, asomatognozja, pomijanie stronne, bóle fantomowe itp., w których podmiot fałszywie identyfikuje własny stan.

Na pierwsze pytanie można odpowiedzieć twierdząco. Nie tylko kognywiści interpretujący autoiluzje jako błędy samopoznania, lecz także filozofowie, którzy optują za klasycznym podejściem do samowiedzy, twierdzą, że „samowiedzy, czyli wiedzy pierwszoosobowej, słusznie odmawia się pewności” (Ziemińska 2008: 47) i „nie jest wcale jasne, na jakiej podstawie przypisujemy sobie wiedzę o własnych przekonaniach” (Piłat 1999: 77). Różnica polega na tym, że filozofowie „przypisują samowiedzy normatywny charakter, natomiast psychologowie podchodzą do niej empirycznie” (Piłat 2013: 221). W związku z tym, jak argumentuję gdzie indziej, tym, co *de facto* badają psychologowie, nie jest samowiedza, lecz jej psychologiczny aspekt tworzony przez samoświadomość (Pacholik-Żuromska 2016). Szczegółowe pytanie powinno dotyczyć tego, jakie w ogóle błędy w identyfikacji własnego stanu można uznać za błędy samowiedzy. A skoro błędna identyfikacja własnych stanów jest możliwa, to należy poszukiwać modelu jaźni, który uwzględniłby autokorektę takich błędów.

Pierwszy z przedstawianych w tym artykule modeli ma charakter internalistyczny i opiera się na proponowanym przez Thomasa Metzingera minimalnym modelu jaźni. Drugi jest ze swej istoty eksternalistyczny i nawiązuje do „ekologicznej” koncepcji jaźni Jamesa J. Gibsona¹. Będę przekonywać, że koncepcje

¹ Współcześnie do tego modelu odwołuje się m.in. Anil Seth, por. niżej, część 4.

samowiedzy budowane na internalistycznych modelach jaźni są nie do utrzymania. Swoją argumentację oprę na tezie, zgodnie z którą takie błędy, jak np. iluzje cielesne czy zaburzenia autopoznawcze, są dla podmiotu niewykrywalne ze względu na sposób ich powstawania. Tworzą się one bowiem już na poziomie neuronalnym, a następnie są przenoszone na kolejne poziomy przetwarzania informacji (fenomenalny i propozycjonalny). Dla podmiotu są to zatem realne i niepodważalne wrażenia, a jako takie nie mogą zostać sklasyfikowane przez sam podmiot, który ich doświadcza, jako iluzje czy inny rodzaj zaburzenia. Jeśli np. podmiot czuje ból, to wie, że go czuje, i nie przestanie go czuć, nawet jeśli zewnętrzny ekspert zapewni go, że jest to ból fantomowy. Istotne jest jednak to, że osoba taka, nie wiedząc, że jest to ból fantomowy, fałszywie identyfikuje jego przyczynę. Może się mylić co do źródła bólu i również ta pomyłka może nie zostać przez nią dostrzeżona. W takich wypadkach potrzeba zewnętrznego narzędzia dostarczanego przez środowisko podmiotu, z którym wchodzi on w interakcję². Teza ta odsyła do koncepcji usytuowanego poznania, a w jego ramach do eksternalizmu, który w odniesieniu do samowiedzy nazywam „eksternalizmem psychologicznym”³. W ujęciu tym podmiot i świat traktowane są jako jeden system, którego elementy nawzajem na siebie wpływają (Wilson 2002), a eksternalizm głosi, że nie tylko treść stanów mentalnych, lecz także jej nośniki mają charakter (przynajmniej częściowo) zewnętrzny wobec podmiotu (por. Ziemińska 2002).

W pierwszych częściach artykułu przedstawię szczegółową argumentację przeciwko używaniu internalistycznego modelu jaźni do objaśniania samowiedzy. Następnie zaproponuję model eksternalistyczny oraz poddam analizie problemy wynikające z jego przyjęcia.

² To, że podmiot niekiedy zmuszony jest do porównania wiedzy na swój temat z sądami trzecioosobowymi o nim samym czy też sprawdza swoją samowiedzę w interakcji ze środowiskiem, jest raczej bezsporne. W oddziaływaniu z otoczeniem wychodzą na jaw na przykład nasze postawy i dyspozycje do działania (por. Poczubut 2008: 7). Innymi słowy, podmiot sprawdza swoje możliwości w działaniu, czasem takim, o które by się nawet nie podejrzewał. Fakt ten rodzi jednak pytanie, w jaki sposób taka interakcja wpływa na samowiedzę i czy skoro podmiot może o swoich dyspozycjach czy postawach nie wiedzieć, to możemy w ogóle mówić o samowiedzy, a jeśli tak, to w jakim modelu jaźni można ją opisać, by zachowała cechy wiedzy pierwszoosobowej i uprzywilejowanej.

³ Za wersję eksternalizmu psychologicznego uważam koncepcję Lwa Wygotskiego, który interakcjonizm zastosował w psychologii rozwojowej. Najogólniej mówiąc, jego zdaniem dziecko nabywa wiedzę o sobie i sądy na temat siebie przez kontakt najpierw z rodzicem, a potem dalszym otoczeniem, internalizując sądy wypowiedziane przez osoby trzecie. W powstawaniu Ja ważny jest zatem kontekst społeczny. Koncepcję tę sam Wygotski ujmuje w postaci hasła: „Najpierw interpsychiczne, potem intrapsychiczne” (por. Wygotski 1989, 2006).

1. CHARAKTERYSTYKA SAMOWIEDZY I BŁĘDÓW SAMOPOZNANIA W UJĘCIU KOGNITYWISTYCZNYM

Z filozoficznego punktu widzenia samowiedza to oparta na introspekcji wiedza podmiotu na temat własnych stanów mentalnych (por. np. Descartes 2009, Boghossian 2008, Peacocke 1999). Ma ona formę propozycjonalną, tj. wyrażana jest w języku w postaci sprawozdań pierwszoosobowych typu „Jestem przekonana, że jestem przekonana, że *p*” (gdzie *p* wyraża treść doświadczanego stanu).

Jak pisze Joanna Komorowska-Mach:

Klasyczne teorie samowiedzy zakładają, że przypisywanie sobie samemu stanów psychicznych jest możliwe dzięki zdolności do obserwacji lub wykrywania aktualnie przeżywanego stanu. Zdolność tę nazywa się najczęściej introspekcją (Komorowska-Mach 2015b: 41, por. też 2015a: 75).

W podejściu klasycznym introspekcja polega na uprzywilejowanym dostępie do własnych stanów mentalnych i gwarantuje sądowi samowiedzy niepodważalny status wiedzy pewnej (por. np. Chisholm 1994: 31-32). Charakterystyczna dla tego rodzaju wiedzy, w jej kartezjańskim ujęciu, jest odporność na błąd autopoznawczy. Podmiot zachowuje zatem autorytet pierwszoosobowy w kwestii własnych stanów mentalnych (por. np. Guttenplan 1994: 91).

Istnieją trzy powody, dla których podmiot nie może się mylić co do własnych stanów mentalnych. Po pierwsze, są one mu dane bezpośrednio, to znaczy ich treść ma charakter samoprezentujący (Chisholm 1994: 32). Po drugie, są niefalsyfikowalne przez osoby trzecie, ponieważ nie mamy bezpośredniego wglądu w stany mentalne innych podmiotów (Brentano 1999⁴, Guttenplan 1994). Po trzecie, są niekorygowalne, ponieważ skoro tylko my mamy bezpośredni wgląd we własne stany mentalne, to stanowi to podstawę bycia pewnym, że jest się w danym stanie (Chisholm 1994: 32, Wittgenstein 2003)⁵. Te cechy autorytetu pierwszoosobowego we współczesnych analizach

⁴ Brentano (1999: 30) pisze na przykład, że fenomeny spostrzeżenia wewnętrznego „są prawdziwe same w sobie. Oczywiście, z jaką je spostrzegamy, świadczy o tym, że jakimi się nam jawią, takie też są w rzeczywistości”.

⁵ Wittgenstein (2003: 166) twierdzi, że „mogę jedynie wierzyć, że ktoś odczuwa ból, ale kiedy mnie boli, wiem o tym. Istotne w prywatnym przeżyciu jest właściwie nie to, że każdy ma tu swój własny egzemplarz, lecz to, że nikt nie wie, czy drugi też ma właśnie to, czy coś innego”. Podaję tu własne tłumaczenie ze względu na konieczność przełożenia słowa *glau-ben* jako „wierzyć”, a nie jak tłumaczy to Bogusław Wolniewicz (Wittgenstein 2000: § 303) „sądzić”. Wierzenie oznacza akceptację danego stwierdzenia bez możliwości jego oceny pod względem prawdziwości, co ze względu na brak obiektywnych kryteriów ma szczególnie miejsce w wypadku prywatnych doznań.

osadzone są jednak na różnych podstawach, np. na pamięci epizodycznej (Shoemaker 1970), na szczególnym rodzaju samoświadomości: cielesnej (Bermúdez 2011), działania (Peacocke 2008, O'Brien 2007) czy przedrefleksyjnej, w której treść ma charakter prezentacyjny, a nie reprezentacyjny (Metzinger 2011), lub też ostatecznie na introspekcji (Shoemaker 1994)⁶.

W pracach psychologicznych zamiast o samowiedzy mówi się raczej o metaświadomości (Schooler i in. 2011) definiowanej jako zbiór przekonań o naszych doświadczeniach introspekcyjnych (Schooler, Schreiber 2004: 17). W wypadku dyskusji na temat metaświadomości częściej pojawia się sceptycyzm co do możliwości nieomylnego samopoznania. Bierze się to z przyjęcia odmiennej metody badawczej: podczas gdy w filozofii częściej stosuje się analizę pojęciową i wnioskowanie dedukcyjne, w psychologii mamy do czynienia z konkretnymi przypadkami, z których wyprowadza się uogólnienia, a więc stosuje się metody empiryczne i analizę statystyczną. Charakterystyka samowiedzy ma zatem różne podstawy, ponieważ zależy od tego, czy odwołamy się do jej filozoficznej definicji, która opiera się na analizie pojęciowej, czy do definicji psychologicznej opartej na badaniu zdolności poznawczych podmiotu. Tym, co wspólne dla samowiedzy i metaświadomości, jest przede wszystkim przywołanie introspekcji jako metody samopoznania polegającej na bezpośrednim wglądzie w zawartość świadomości. Rola introspekcji w samowiedzy może być jednak różnie rozumiana, szczególnie w analizach filozoficznych. Bardzo uogólniając, jest to swoista percepcja, pewnego rodzaju zmysł wewnętrzny, na którym opiera się samowiedza (Speaks 2004: 2). Introspekcja nie zawsze musi przy tym zakładać percepcyjny model samowiedzy (Gertler 2017). Niezależnie jednak od tego, jaki jej model przyjmujemy, sądy samowiedzy powinny mieć wyjątkowy status: jeśli X jest przekonany, że p , to X wie, że jest przekonany, że p (por. np. Peacocke 1999).

Problem polega na tym, że nauka zna przypadki, w których bezpośredni dostęp do własnych stanów mentalnych nie gwarantuje ich prawdziwości, a przez to sądy samowiedzy tracą swoją wartość epistemiczną. Sytuacje te pokazują, że nasz pierwszoosobowy autorytet może zostać zakwestionowany. Błędy samowiedzy rozumiem jako błędy w reprezentacji odnoszącej się do właściciela danego stanu oraz w reprezentacji treści danego stanu. Przykła-

⁶ Należy zaznaczyć, że wymienione fundamenty samowiedzy są kategoriaalnie zróżnicowane. Mówiąc o osadzeniu samowiedzy na samoświadomości, stawiamy albo tezę ontologiczną, ponieważ rozpatrujemy fenomen, który stanowi jej podstawę, albo tezę epistemologiczną lub psychologiczną, ponieważ mówimy o zdolności niezbędnej do samopoznania. Natomiast kiedy mówimy o introspekcji, mamy na myśli albo metodę poznania siebie, albo zdolność do introspekcji jako szczególnego rodzaju percepcji. Stawiamy więc albo tezę metodologiczną, albo epistemologiczno-psychologiczną.

dem błędów pierwszego rodzaju są zaburzenia występujące na przykład w somatoparafrenii (chorzy nie rozpoznają kończyn jako własnych, a nawet zaprzeczają posiadaniu części ciała, tzn. widzą daną część, ale uważają, że nie jest to ich własna część ciała, i konfabulują, twierdząc np., że owa część została do ich ciała „doczepiona”) oraz asomatognozji (chorzy czasowo mylą własne części ciała z częściami ciał innych osób, np. uznają, że widziana właśnie ręka należy do kogoś innego). Badacze Timothy Lane i Caleb Liang (2009, 2010) opisujący te przypadki twierdzą, że świadomość stanu umysłowego nie gwarantuje jego pierwszoosobowego posiadania.

Z kolei błędy w reprezentacji własnego stanu co do treści opisywane są też jako błędna identyfikacja przyczyny danego stanu. Przykład stanowią złudzenia cielesne, takie jak iluzja gumowej ręki (*rubber hand illusion*), iluzja całego ciała (*full body illusion*) czy iluzja zamiany ciał (*body swap illusion*). Ulegający tego typu iluzjom podmiot, choć przypisuje sobie dany stan i poprawnie identyfikuje jego modus, np. odczuwanie bólu czy łaskotania, to nie odczytuje trafnie treści danego stanu. Oznacza to, że uważa na przykład, iż czuje łaskotanie we własnej ręce, choć stymulowana jest sztuczna ręka leżąca obok (Ionta, Heydrich, Lenggenhager 2011, Lenggenhager i in. 2007, Blanke, Metzinger 2009, Aspell, Lenggenhager, Blanke 2012). Jeszcze innym rodzajem błędów w samowiedzy są te dotyczące rodzaju czy modusu danego stanu, czyli np. mylenie przekonania z wierzeniem a złości ze smutkiem (por. Pacholik-Żuromska 2015). Powstaje więc pytanie, jak rozumieć samowiedzę, by była ona do utrzymania w świetle wymienionych rodzajów błędów i przywołanych wyników badań.

2. JAŻŃ JAKO PRZEDMIOT SAMOWIEDZY

Zacznijmy od pytania, czy podmiot może sam z siebie dojść do wniosku, że jego samowiedza jest fałszywa. Problem z autorytetem pierwszoosobowym bierze się między innymi z tego, że w percepcyjnym modelu samowiedzy zachodzi relacja podmiot–przedmiot, w której przedmiotem są stany wewnętrzne podmiotu. W takim modelu, odwołującym się do metafory wewnętrznego oka umysłu (Locke 1955), łatwo o błąd, ponieważ jaźń raz jest tym, co postrzega, a raz tym, co postrzegane (por. np. Musholt 2015). W modelu percepcyjnym samowiedzy dobrze można jednak wyjaśnić problemy z autorytetem pierwszoosobowym, tłumaczone tym, że albo dostęp podmiotu do przedmiotu (czyli stanów wewnętrznych podmiotu) jest zaburzony, albo owe stany nie są klarowne, a zatem sam przedmiot jest jaźni niedostępny.

Percepcyjny model samowiedzy krytykowany był między innymi przez Sydneya Shoemakera, który w swoich polemikach specyficznie rozumie introspekcję. Shoemaker nie krytykował bowiem introspekcji jako metody, czyli sposobu dostępu do własnych stanów mentalnych, lecz właśnie jako szczególnego rodzaju percepcję polegającą na postrzeganiu własnych stanów umysłowych (Shoemaker 1994, Speaks 2004). Według Shoemakera samowiedza nie może być po prostu wiedzą percepcyjną. Nawet jeśli stany mentalne są przedmiotem samowiedzy, to ich identyfikacja wymaga uprzedniego odniesienia się do siebie jako „ja”, a to znaczy, że samoidentyfikacja jest czymś pierwotnym wobec identyfikacji własnego stanu. W związku z tym wiedza „Czuję ból” nie dotyczy treści danego stanu, lecz tego, że to „ja czuję ból, a nie ktoś inny” (Lane 2011). Introspekcja jest zatem sposobem samoidentyfikacji, ponieważ podmiot sobie, a nie komuś innemu przypisuje ból.

Pomysł Shoemakera oparty jest na propozycji pochodzącej jeszcze od Williama Jamesa, który rozróżniał dwa rodzaje ja: obecne (aktualne) i narracyjne. Pierwsze z nich to ja doświadczające siebie — ja podmiotowe. Drugie to ja jako obiekt, któremu przypisywane są predykaty mentalne i fizyczne — ja przedmiotowe (James 1950). Jaźń obecna w samowiedzy to ja podmiotowe, doświadczające, a nie ja, do którego się odnosimy przez określone deskrypcje. Stanowi ono podstawę samowiedzy i gwarantuje odporność na błąd autopoznawczy. Nie możemy się bowiem mylić co do tego, że to my jesteśmy podmiotami doznającymi danego stanu. Takie rozumienie samowiedzy koncentruje się na podmiocie, a nie na doznawanym stanie. Jak pokazują jednak przypadki takie jak somatoparafrenia, można się mylić również co do bycia właścicielem danego stanu, co podważa argumentację Shoemakera (Lane 2011, Lane, Liang 2009).

Współcześnie krytycy percepcyjnego modelu samowiedzy (por. np. Metzinger 2003, Musholt 2015) odwołują się przede wszystkim do argumentacji Hume’a na rzecz tego, że nie postrzegamy naszej jaźni: nasza jaźń nie może być przedmiotem naszej wiedzy, ponieważ tym, co postrzegamy, są jedynie percepcje. Jaźń nie jest zatem niczym więcej jak tylko wiązką percepcji (Hume 1963). Dochodzi się przy tym do różnych wniosków, np. do twierdzenia, że nikt nigdy nie miał jaźni ani nią nie był (Metzinger 2003) czy też że należy skupić się na świadomości przedrefleksyjnej jako źródle bezpośredniości samowiedzy (Musholt 2015).

W artykule skupię się na stanowisku Thomasa Metzingera, które od 2003 r. przeszło kilka przeobrażeń. W książce *Being No One* (2003) Metzinger wyrażał przedstawiony pogląd, zgodnie z którym substancjalne jaźnie nie istnieją, a także zgodnie z tezami enaktywizmu i ucieleśnionych teorii umysłu twierdził, że tym, co konstytuuje rzekomą jaźń jest przede wszystkim środowisko i oto-

czenie, o czym świadczą przypadki pokazujące, jak można manipulować postrzeganiem siebie przez podmiot. Od 2015 r. Metzinger swoje dociekania kieruje w stronę pytania, jak powstaje nasza samowiedza, i buduje model oparty na kodowaniu predykcyjnym.

3. INTERNALISTYCZNY MODEL JAŻNI W SAMOWIEDZY

Internalistyczny model samowiedzy to taki, który nie zawiera elementu zewnętrznego jako niezbędnego składnika. W wypadku samowiedzy mamy do czynienia z modelem przynajmniej prawdopodobnym, gdy po pierwsze odzwierciedla on, jak powstaje samowiedza, ukazując na przykład mechanizm przetwarzania informacji od poziomu nieświadomego do poziomu treści propozycjonalnych, a po drugie, gdy spełnia wymagania nakładane na samowiedzę, tzn. pokazuje, w jaki sposób zachowany zostaje autorytet pierwszoosobowy, który z kolei gwarantuje, że wiedza to rzeczywiście prawdziwe uzasadnione przekonanie. Model taki powinien oddawać relację intencjonalności zachodzącą między jaźnią a jej treścią – taki model proponuje choćby Christopher Peacocke (1992, 1999) – lub pokazywać, że samowiedza nie ma charakteru relacyjnego, lecz jest na przykład rezultatem przekształcenia jakiejś formy samoświadomości (O'Brien 2007).

Proponowany przez Metzinger model samowiedzy jest internalistyczny właśnie dlatego, że nie zawiera elementu zewnętrznego związanego ze środowiskiem fizycznym czy społecznym, to znaczy, jak to ujmie Metzinger, nie wymaga odniesienia do jakiegoś nieneuronalnego obiektu (Metzinger 2017). Mechanizm działania tego modelu oparty jest na przetwarzaniu czy kodowaniu predykcyjnym, zwanym również aktywnym wnioskowaniem (*active inference*). Jego podstawę stanowi wnioskovanie bayesowskie.

W takim modelu poznania procesy predykcyjne zachodzące w mózgu umożliwiają przewidywanie najbardziej prawdopodobnych sytuacji na podstawie bodźca zmysłowego, pozwalając na wykonanie odpowiedniego działania (Friston, Daunizeau, Kiebel 2009). Mówiąc metaforycznie, mózg formułuje hipotezy na temat oczekiwanych sytuacji i jeśli zewnętrzny bodziec odpowiada hipotezie, to zostaje ona wzmocniona, a jeśli jej nie odpowiada, to następuje modyfikacja hipotezy. Należy przy tym zauważyć, że mechanizm bayesowski opisuje, jak działa mózg, a nie jaźń. „Hipotezy na temat świata” to po prostu metafora. Podmiot nie jest świadomy rzeczywistych obliczeń predykcyjnych. Aktywne wnioskovanie to sposób przetwarzania informacji w relacji mózg–świat.

Modele oparte na przetwarzaniu predykcyjnym nawiązują do zasady energii swobodnej. Karl Friston w odniesieniu do poznania interpretuje ją jako dążenie do minimalizacji błędu predykcyjnego (Friston, Daunizeau, Kiebel 2009, Hohwy 2013). W wypadku układów żywych (czyli organizmów) zasada ta głosi, że minimalizują one swoją energię całkowitą dzięki przewidywaniu warunków zewnętrznych i dostosowywaniu stanów wewnętrznych do tych przewidywań.

W internalistycznym modelu samowiedzy proponowanym przez Metzingera aktywne wnioskowanie nie tylko może służyć do przewidywania stanów świata, lecz także do wyjaśnienia procesów modelowania jaźni. Innymi słowy, mózg reprezentuje nie tylko zewnętrzne stany rzeczy, lecz także stany jego właściciela (Metzinger 2017). Jest to model podmiotu epistemicznego (*epistemic agent model*, EAM) mający odzwierciedlać, w jaki sposób dochodzi do poznania siebie. Podmiot epistemiczny rozumiany jest tu jako aktywny podmiot poznający (*active cognizer*).

Predykcje powstają w EAM na poziomie świadomego przetwarzania i reprezentują właściciela mózgu jako podmiot zdolny do autonomicznej samokontroli epistemicznej, tj. monitorowania i wolicjonalnej modyfikacji jego własnych stanów mentalnych (Metzinger 2017: 8). Fundament EAM stanowi minimalna jaźń fenomenologiczna – subiektywne doświadczenie bycia jaźnią – ponieważ, aby przypisać sobie dany stan, podmiot musi się odnieść do samego siebie jako ja (mieć reprezentację ja). EAM powstaje zatem jako efekt relacji między ja a intencjonalnym celem rozumianym jako wewnętrzny przedmiot mentalny.

EAM nie wymaga zewnętrznego (nieneuronalnego) elementu. Mimo że zdarzenia mentalne są wynikiem całkowicie wewnętrznego subosobowego (tj. neuronalnego) przetwarzania informacji (Metzinger 2017: 2), co nie jest niczym nadzwyczajnym – skoro podłożem stanów umysłowych jest fizyczny mózg, to na poziomie EAM mają one charakter wolicjonalny. Przykładem mogą tu być czynności uwagowe (przy czym mowa tu o uwadze endogennej, a więc kontrolowanej przez podmiot⁷). Choć EAM zbudowany jest na podstawie fenomenalnej, którą stanowi minimalna sobość (jaźń fenomenologiczna), to jest on o wiele bardziej złożony, ponieważ jaźń jest w nim reprezentowana nie tylko jako poczucie bycia podmiotem, lecz przede wszystkim jako aktywny podmiot poznający siebie (Metzinger 2017: 8). *De facto* jest to więc model samowiedzy, ponieważ modelowane są w nim przyszłe stany umysłowe pod-

⁷ Sam Metzinger tego nie podkreśla, a powinien, ponieważ nie każda uwaga ma charakter wolicjonalny. Uwaga egzogenna nie pojawia się jako stan w EAM, ponieważ wzbudzana jest automatycznie – dopiero po jej nieświadomym nakierowaniu na źródło bodźca, który ją przyciągnął, podmiot zdaje sobie z tego sprawę.

miotu (Metzinger 2017: 15). Zdarzenia czy też działania mentalne są nakierowane na epistemiczne stany o charakterze intencjonalnym, przy czym intencjonalność w tym wypadku rozumiana jest jako nakierowanie na cel (Metzinger 2017: 3). Owo „bycie nakierowanym na” nie jest niczym innym jak tylko postawą propozycjonalną drugiego rzędu, czyli przekonaniem o byciu przekonanym. Odpowiada więc charakterystyce samowiedzy. Tak rozumiana samowiedza gwarantuje autonomię mentalną (*M-autonomy*) w tym sensie, że podmiot jako poznawczo aktywny zdolny jest do wolicjonalnej kontroli swoich stanów i sterowania nimi, np. może wybierać jedne przekonania, a inne odrzucać, choć oczywiście zdarzają się przypadki takie jak zamyślenie czy też „błądzenie myślami” (*mind wandering*), w których kontrola epistemiczna jest bardzo słaba (Metzinger 2015, 2017)⁸.

Autonomia mentalna równa jest zatem epistemicznej kontroli stanów mentalnych, która polega na predykcyjnej kontroli epistemicznej wartości owych stanów wpisanych w fenomenalny model jaźni (Metzinger 2017: 1):

Dla świadomego systemu poznawczego mieć EAM oznacza wiedzieć, że wie. Zgodnie z subiektywnym doświadczeniem niektóre z jego stanów mają wartość epistemiczną (Metzinger 2017: 7).

Owa kontrola nie jest niczym innym jak tylko samowiedzą opartą na przekonaniach drugiego stopnia:

Każda reprezentacja działania musi zawierać trzy logiczne składowe: model bytu sprawującego kontrolę (jaźń), model warunków spełniania danego działania (cel) i asymetryczną relację dynamicznie integrującą pierwsze dwie składowe („strzałka intencjonalności”) (Metzinger 2017: 4).

Choć brzmi to skomplikowanie, Metzinger opisuje tu zwykły model postawy propozycjonalnej w postaci: „*X* (jaźń) jest przekonany, że *p*”. Postawa propozycjonalna ma charakter intencjonalny wyrażający nastawienie podmiotu wobec danego sądu. Są to zazwyczaj asercje, a zatem uznanie przez podmiot danego sądu za prawdziwy, czy też, używając języka psychologii potocznej, przekonania.

⁸ Badania empiryczne wskazują, że w wypadku błądzenia myślami podmiot jest tylko częściowo samoświadomy, tzn. tylko częściowo zdolny do refleksji nad czynnościami własnego umysłu. Przy czym należy zauważyć, że samoświadomość (*meta-awareness*) definiowana jest tak samo jak samowiedza. Można więc powiedzieć, że przy błądzeniu myślami wiedza o tym, że właśnie błądzimy myślami, jest mglista, o czym miałyby świadczyć choćby to, że badany, jeśli zostanie zapytany o to, czy właśnie błądził myślami, z łatwością odpowie tak lub nie, lecz kiedy sam ma stwierdzić (złapać się na błądzeniu myślami), może tego w ogóle nie zauważyć (por. np. Schooler i in. 2011, Christoff i in. 2009, Pacholik-Żuromska 2016).

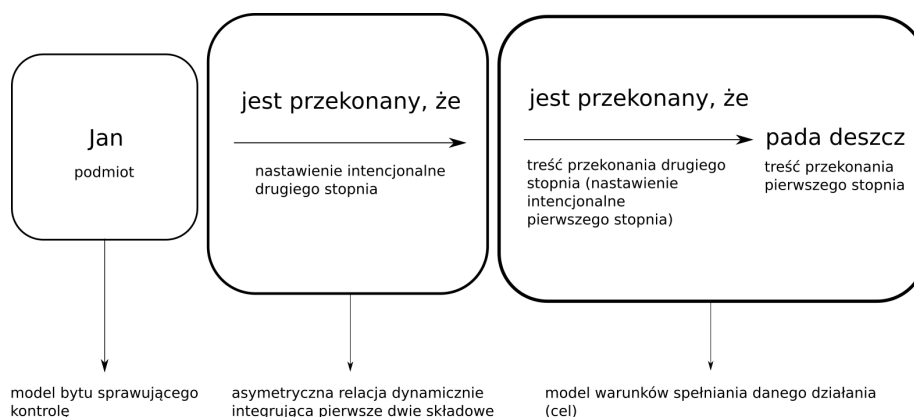
Na język tradycyjnej epistemologii terminy Metzingera można przełożyć tak:

1. model bytu sprawującego kontrolę (jaźń) – ja (podmiot),
2. model warunków spełniania danego działania (cel) – warunki prawdziwości danego sądu (uznanie sądu za prawdziwy),
3. asymetryczna relacja dynamicznie integrująca pierwsze dwie składowe („strzałka intencjonalności”) – relacja intencjonalności.

Normalnie, gdy mówimy o typowych postawach propozycjonalnych, warunki prawdziwości określane są np. przez zgodność sądu z rzeczywistością, choć oczywiście można stosować także inną teorię prawdy. Załóżmy jednak, że przyjmujemy realizm. W postawie propozycjonalnej drugiego stopnia, a więc ugruntowującej samowiedzę, mamy do czynienia z postawą propozycjonalną na temat postawy propozycjonalnej, czyli: „*X* jest przekonany, że jest przekonany, że *p*”. Problemem pozostają warunki prawdziwości sądów stanowiących treść przekonania drugiego rzędu, a więc sądów następujących po pierwszym „że”. Nie jest więc jasna wartość epistemiczna takiego przekonania drugiego stopnia. Kwestia ta jest sednem wątpliwości wysuwanych wobec samowiedzy.

Część teoretyków uważa, że sądy samowiedzy uznawane są za prawdziwe na mocy samego się ich pojawiania (Peacocke 1992, 1999). Prekursorem tego stanowiska jest oczywiście Kartezjusz. Inni twierdzą, że nie są to sądy w zwykłym sensie, lecz raczej deklaracje (*avowals*) mające charakter ekspresji bezpośrednich doznań i wyrażające uprzywilejowaną wiedzę na temat własnych stanów mentalnych (Bar-On 2004a, b). Dla Metzingera „wiedza w tym sensie jest posiadaniem wiarygodnych informacji” (Metzinger 2017: 7).

Model EAM jest modelem samowiedzy, w którym jaźń (podmiot) jest nakierowana intencjonalnie na własne stany umysłowe, tzn. wie, że wie (jak pisze Metzinger), czy też jest przekonana, że jest przekonana (jak pisze Peacocke). Możemy mówić o intencjonalności drugiego stopnia, ponieważ nastawienie intencjonalne jest tu podwójne, mamy bowiem do czynienia z nakierowaniem na sąd na temat własnych przekonań i na sąd o odniesieniu przedmiotowym (por. Rysunek 1).



Rysunek 1. Model jaźni w EAM

Ostatecznie model jaźni proponowany przez Metzingera okazuje się klasycznym kartezjańskim modelem samowiedzy internalistycznej zbudowanej z przekonań pierwszego rzędu, do których podmiot (jaźń) odnosi się intencjonalnie przez przekonania drugiego rzędu. Do takiego stwierdzenia uprawnia, po pierwsze, struktura tego modelu oparta na relacji intencjonalności w rozumieniu Brentana. Metzinger pisze o intencjonalności jako byciu nakierowanym przez podmiot na treść danego stanu. Po drugie, nic nowego nie wnoszą również składowe owej relacji: podmiot i treść przekonania drugiego stopnia, które Metzinger nazywa modelem warunków spełnienia danego działania mentalnego (*mental action*), charakteryzowanego jako stan umysłowy nakierowany na cel, gdzie cel ujęty jest w treści danego stanu mentalnego i osiągnięty jest przez zachowanie zgodne z żywionym przekonaniem (Metzinger 2017: 4)⁹.

Nie jest jednak jasne, czy taki model gwarantuje zwiększenie wartości epistemicznej, rozumianej jako posiadanie wiarygodnych informacji. Byłoby tak, gdyby model ten ukazywał, jak możliwa jest autokorekta błędów w samopoznaniu. Tego jednak nie czyni, ponieważ zamyka jaźń w wewnętrznym modelu neuronalnym. Każda neuronalna stymulacja takiego mózgu będzie wywoływała stan wewnętrzny, który mógłby być dowolnie iluzoryczny (jak na przykład w somatoparafrenii, kiedy podmiot jest przekonany, że jego ręka należy do kogoś innego), i którego nie sposób zweryfikować bez zewnętrznego odniesienia (bez udziału nieneuronalnego przedmiotu), czyli bez interakcji z rzeczywistym środowiskiem zewnętrznym. Najlepszymi kandydatami do takiej zewnętrznej konfrontacji są osoby trzecie, które, jak się okazuje, czasem

⁹ Fodor podaje przykład działania opartego na racjonalnym charakterze postaw propozycjonalnych: „Chcę się dostać do domu/do pracy/do cioci i mam powody, by sądzić, że tam – lub gdzieś blisko tego miejsca – jedzie autobus” (Fodor 1999: 18).

wiedzą lepiej niż sam podmiot, że myli się on co do własnych stanów i mogą powiedzieć: „To twoja ręka, nie moja”.

Takimi osobami są eksperci, dokładnie tak jak w eksternalizmie semantycznym przyjmowali Tyler Burge (1979) czy Hilary Putnam (1998), choć w kontekście innego problemu, mianowicie identyczności stanów mentalnych przy różnym odniesieniu ich treści. W wypadku przekonań samowiedzy problem polega na tym, że podmiot nie jest w stanie odróżnić fałszywej treści danego stanu od prawdziwej, ponieważ obydwie są mu dane tak samo. Dla niego są to stany, których doznaje. Takim wykwalifikowanym ekspertem może być natomiast psycholog, neurolog czy terapeuta, który uświadomi pacjentowi, że dany stan, choć dla niego realny, jest fałszywy ze względu na swoje odniesienie, na przykład do obiektu, który jak w przypadku bólów fantomowych nie istnieje.

Czy jednak naprawdę nie ma możliwości, by sam podmiot dokonał takiej właściwej identyfikacji własnego stanu? Odpowiedzi na to pytanie dostarczają eksternalistyczne modele jażni, które optymistyczniej podchodzą do autopozytywnych możliwości podmiotu, przyjmując, że może on zweryfikować swój stan w działaniu¹⁰.

4. EKSTERNALISTYCZNY MODEL JAŻNI W SAMOWIEDZY

Na tym samym mechanizmie co model Metzinger'a, a zatem na kodowaniu predykcyjnym, oparty został eksternalistyczny model jażni Anila Setha. Model ten został wybrany przeze mnie jako przykład tego, że kodowanie predykcyjne nie pociąga za sobą z konieczności internalizmu, a proponowany mechanizm jest wydajny eksplanacyjnie również wtedy, gdy odrzuca się model jażni jako zamkniętego w sferze neuronalnej systemu. Model Setha jest interesujący również dlatego, że inaczej niż EAM nie odwołuje się do introspekcji jako metody dostępu do własnych stanów mentalnych (co model Metzinger'a ostatecznie implikuje).

Seth (2015) łączy teorie ucieleśnionego poznania i enaktywizm (tzw. e-teorie) z inżynierią komputerową, przekonując, że polegająca na działaniu interakcja ze światem służy jako tester predykcji sensorycznych. W ten sposób buduje model predykcyjnej percepcji spójności sensomotorycznej (*predictive perception of sensorimotor contingencies*, PPSMC). Jego zdaniem:

Operacyjne kryterium udanego modelu jażni to zdolność przewidywania bodźców sensorycznych w ramach repertuaru działań. To znaczy, że modele predykcyjne powstają

¹⁰ Pisałam o tym również w innym artykule (Pacholik-Żuromska 2019), w którym skupiam się jednak na przykładach wpływu otoczenia podmiotu na samowiedzę.

w celu kontroli zachowania (jak zakłada się w cybernetyce), a nie po to, by dostarczać ogólnych reprezentacji na temat świata czy ciała (Seth 2015: 14).

Seth uważa, że eksternalistyczne e-teorie mają przewagę nad internalistycznymi, ponieważ poza zwróceniem szczególnej uwagi na rolę środowiska w poznaniu, usuwają również problem reprezentacji, eliminując je z modelu jaźni:

Enaktywizm jest wyraźnie antyreprezentacjonistyczny. Teorie enaktywne mogą się zgadzać, że zachowania adaptacyjne wymagają tego, by organizmy i struktury kontroli były systematycznie wrażliwe na statystyczne struktury w ich środowisku, ale większość będzie zaprzeczać, że owa wrażliwość implikuje istnienie zastosowania jakiegoś wewnętrznego opisu czy modelu tych probabilistycznych wzorców (Seth 2015: 16).

Zwrot w kierunku enaktywizmu w robotyce i kognitywistycznych badaniach nad sztuczną inteligencją nastąpił, ponieważ tradycyjne modele poznania powstające w ramach teorii reprezentacjonistycznych przestały być wydajne. W związku z tym podjęto próby zastąpienia reprezentacji w pętli interakcji mózg–ciało–świat elementem pełniącym rolę pośrednika między podmiotem (systemem) a światem (Seth 2015: 15). Zwrot w kierunku cybernetyki umożliwił budowę modeli poznania bez wprowadzania dodatkowych sztucznych tworców (reprezentacji) jako nośników informacji, zastępując je po prostu samą informacją realizowaną za pomocą odpowiednich struktur neuronalnych (naturalnych w wypadku organizmów i sztucznych w wypadku sztucznej inteligencji).

Zamiast do reprezentacji siebie czy wytwarzanego przez system modelu jaźni Seth odwołuje się do percepcji jako aktywności poznawczej, która daje „poczucie” bycia tu i teraz jako percepcyjnej obecności (*perceptual presence*). Koncepcja Setha jeszcze bardziej zbliża się do Hume’owskiego podejścia, w którym poczucie bycia jaźnią powstaje dzięki świadomości tego, że się postrzega – jestem tym, kto postrzega. „Percepcyjna obecność to poczucie subiektywnej rzeczywistości treści percepcyjnej” (Seth 2015: 3, Nöe 2006). Aktywny podmiot poznający, aby skutecznie działać w środowisku, musi mieć do dyspozycji modele spójnego działania (model predykcyjnej percepcji spójności sensomotorycznej), co możliwe jest wtedy, gdy potrafi symulować sytuacje kontrfaktyczne. Fundamentalna różnica między EAM a PPSMC polega na tym, że w wypadku PPSMC percepcja konieczna jest do aktualizacji informacji już posiadanych przez system, żeby doprecyzować kolejne predykcje. W ten sposób system tworzy model, który testuje w działaniu, a zaangażowanie uwagi umożliwia optymalizację tego modelu przez nakierowanie na informacje najbardziej relewantne (Seth 2015, Hohwy 2013). Odpowiada to za fundamentalną różnicę między internalistycznym a eksternalistycznym po-

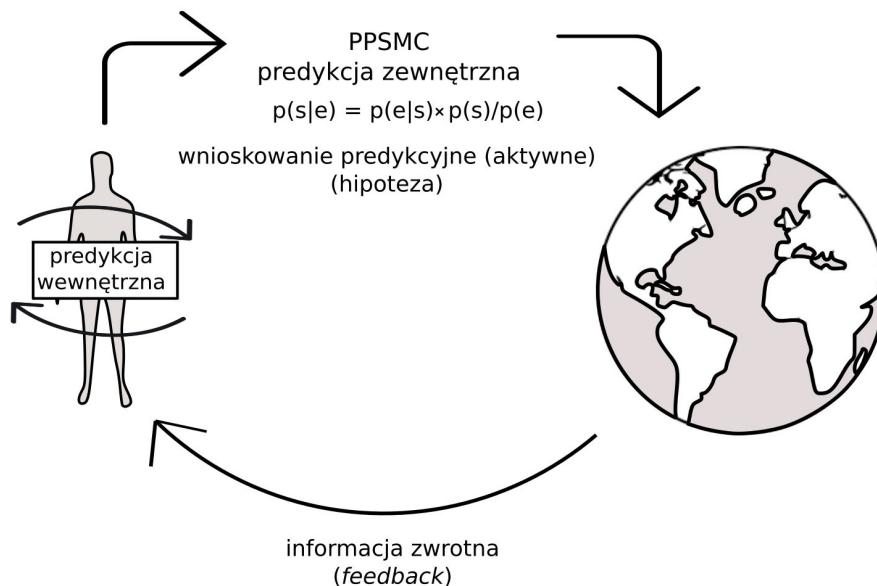
dejsiem do kodowania predykcyjnego. W obydwu koncepcjach tym, co niezmiennie, jest teza, że system przewiduje stany świata i owe przewidywania „nakłada” na to, co ostatecznie otrzymuje jako informację zwrotną. Jednak w internalizmie system nie koryguje hipotez, lecz raczej wpływa na zmianę informację zwrotną (*feedback*), natomiast w eksternalizmie system po prostu zmienia predykcję. Innymi słowy, w internalizmie to nie predykcje są korygowane, ale warunki zewnętrzne dostosowywane do predykcji, a w eksternalizmie to właśnie warunki zewnętrzne korygują predykcje¹¹.

Seth stawia tezę, że dla systemu właściwie istotniejsza jest wiedza na temat jego własnych stanów wewnętrznych niż na temat świata, ponieważ to dzięki niej dostosowuje się do bieżących warunków środowiska (Seth 2015) czy też afordancji (by użyć terminologii Gibsona)¹². Stany fenomenalne organizmu byłyby wtedy reakcją wywołaną pośrednio przez środowisko zewnętrzne, a bezpośrednio przez „predykcyjne wnioski o przyczynach interoceptywnych sygnałów sensorycznych” (Seth 2015: 9). Seth w zgodzie z e-teoriami uważa, że głównym punktem w interakcji między systemem (organizmem, podmiotem) a światem jest ciało. Aby interakcja była skuteczna, organizm musi dysponować odpowiednimi zdolnościami umożliwiającymi kontrolę ciała. Należą do nich m.in. poczucie własności ciała (*ownership*) i samoidentyfikacja (Seth 2015: 11). Potrzebna do tego jest integracja cielesnych informacji w postaci świadomości cielesnej, ponieważ w ten sposób mózg tworzy model ciała jako całości (Seth 2013, 2015: 11). Jażń jako ja jest tu wypadkową interoceptywnych, eksteroceptywnych i proprioceptywnych bodźców sensorycznych (Seth 2015: 12). Istotna jest tu interakcja między sygnałami interoceptywnymi i eksteroceptywnymi (Seth 2015: 13). Tego typu model jażni musi zawierać jako niezbędny składnik również zewnętrzny wobec ciała element, którego obecność dopiero pozwala na konstytucję tego modelu (por. Rysunek 2)¹³.

¹¹ Gdyby jednak rzeczywiście tak miało być, to internalistyczne ujęcie kodowania predykcyjnego nie opierałoby się na wnioskowaniu bayesowskim.

¹² W skrócie afordancje to możliwości działania wysyłane przez środowisko w postaci odpowiednich fizycznych struktur (por. Gibson 1950, 1978, 2002).

¹³ W zobrazowaniu tego modelu sugerowałam się cudzymi propozycjami (Seth 2015: 10, Christoff i in. 2011: 106), jednak w przedstawionym kształcie jest on moim pomysłem. Rysunki człowieka i globu na licencji Creative Commons.



Rysunek 2. Model jaźni w PPSMC

Seth wyraźnie pisze, że choć proces przetwarzania informacji dokonuje się w ciele podmiotu i to mózg jest „maszyną predykcyjną”, to jednak do postawienia właściwych hipotez na temat oczekiwanego stanu świata niezbędna jest mu informacja o „ukrytych przyczynach informacji sensorycznej” (Seth 2015: 4). Stąd należy wnioskować, że niezbędnym elementem modelu jaźni, będącej efektem integracji multisensorycznej, jest zewnętrzne otoczenie podmiotu, ponieważ to ono stanowi ową „ukrytą przyczynę bodźców”.

Z przypadków autoiluzji, jak iluzja gumowej ręki czy iluzja całego ciała, Seth wnioskuje, że:

Aktywne wnioskowanie nie polega jedynie na potwierdzaniu przewidywań, lecz także na poszukiwaniu działań „zakłócających”, które są najbardziej informatywne w odniesieniu do testowania bieżących prognoz i/lub rozróżniania predykcji konkurencyjnych (Seth 2015: 13).

W eksternalistycznym modelu jaźni to właśnie działanie jest testerem samowiedzy i ukazuje, że podmiot w konkretnych przypadkach może się mylić co do tego, że postrzegany obiekt należy do jego ciała, choć pierwsza informacja mówi o jego zintegrowaniu w obrębie ciała. Zilustrujmy to przykładem iluzji gumowej ręki. Polega ona na tym, że prosi się osobę biorącą udział w eksperymencie, by usiadła przy stole i położyła na nim obie dłonie. Również z dłonią na stole leży sztuczna gumowa ręka oddzielona od prawdziwej dłoni przegro-

dą ograniczającą pole widzenia badanego. Badany widzi zatem jedną dłoń własną, a drugą gumową. Następnie obserwuje, jak eksperymentator synchronicznie stymuluje pędzelkiem kilka minut dłoń prawdziwą i gumową. W wyniku takiej stymulacji osoba badana czuje swoją dłoń w miejscu tej gumowej, a nie tam, gdzie *de facto* jej dłoń się znajduje. Kontynuując eksperyment, prosi się badanego, żeby zamknął oczy i wskazał, gdzie znajduje się jego ręka. W rezultacie osoba wskazuje jako własną rękę gumową, a nie organiczną. Jest to wynik przemieszczonej propriocepcji, czyli zmysłu identyfikującego położenie własnego ciała i kończyn. Weryfikatorem tego iluzorycznego stanu będzie konkretna czynność, o którą proszony jest badany, np. „Unieś obydwie ręce” (a nie „Wskaż swoją rękę”). Odpowiednie działanie przywraca zatem właściwą percepcję badanego.

Charakterystyczne dla modeli opartych na cybernetyce jest to, że są one proste w opisie, choć skomplikowane w konstrukcji. Bazują na złożonych algorytmach służących do programowania zachowania robotów i androidów. Sama charakterystyka modelu jest jednak prosta, ponieważ modelowane są w nim konkretne zachowania, które następnie są implementowane w danej maszynie (robocie), a więc łatwo sprawdzić, czy działają. Opis takich modeli zdaje się nie zawierać elementów spekulatywnych; inaczej jest w EAM, który jest raczej hipotezą, że ludzki system poznawczy w wypadku samowiedzy mógłby działać tak a tak. W wypadku PPSMC mamy do czynienia z już sprawdzonymi hipotezami. Ostatecznie jednak on też oparty jest tylko na przypuszczeniu, że model działania robota można przełożyć na model jaźni w samowiedzy u człowieka.

Innym modelem opartym na osiągnięciach cybernetyki jest model proponowany przez zespół Luca Steelsa. W laboratoriach Sony stworzył roboty wchodzące ze sobą w interakcje i potrafiące się od siebie uczyć. Nie zaprogramowano im zatem całego repertuaru wiedzy, lecz dysponowały programem umożliwiającym uczenie się od siebie nawzajem. Wchodząc ze sobą w interakcje, miały nauczyć się komunikacji i współpracy (Steels, Spranger 2008). Chodziło o szczególny rodzaj działania, jakim jest mowa ciała (*body language*). Aby roboty zaczęły współpracować, musiały najpierw „poznać siebie”, obserwując swoje zachowanie w lustrze. Uczyły się w ten sposób na przykład, że uniesiona ręka to ich ręka i że to one są sprawcami danego ruchu. Informacja wzrokowa (*feedback*) była konieczna do zbudowania własnego obrazu ciała. Daleko stąd jeszcze do samowiedzy, ale niewątpliwie jest to już jakaś namiastka świadomości cielesnej.

Jeśli przyjąć założenie leżące u podstaw koncepcji samowiedzy, że budowana jest ona na podstawie informacji neuronalnej przekształcanej następnie w doznanie na poziomie fenomenalnym, a ostatecznie i docelowo na struktu-

rę propozycjonalną, to problem dziedziczenia błędu autopozytywczego przez kolejne poziomy może zostać rozwiązany właśnie przez jego korektę na poziomie neuronalnym i kolejne przekształcenia. Informacja zwrotna ze świata jest zatem tym elementem, który koryguje dane samowiedzy, a modele oparte na cybernetyce zakładają mechanizm autokorekty. Nie są to jednak jedyne modele ukazujące niezbędność zewnętrznego elementu jako konstytutywnego dla właściwego działania samowiedzy.

Jeszcze inny eksternalistyczny model jaźni odwołujący się nie do cybernetyki, lecz do neurologicznych uwarunkowań człowieka, zbudowała ze swymi współpracownikami Kalina Christoff. Zespół Christoff przekonuje, że w modelu tym, choć jest on neurologiczny, niezbędny jest również zewnętrzny nieneurologiczny element: jaźń jako doświadczenie siebie konstytuuje się na gruncie podstawowych procesów neuronalnych przebiegających w ośrodkowym układzie nerwowym i polegających na określeniu spodziewanej informacji zmysłowej po wykonaniu danego ruchu, gdzie kopie eferentne monitorują i pozwalają na korektę działania, natomiast reaferecja jako informacja płynąca z zewnętrznego świata sugeruje ową korektę. Ta pętla sygnałów eferentnych i reaferecyjnych determinuje rozróżnienie na jaźń i nie-jaźń: sygnały aferentne stanowią bowiem rezultat własnego działania poznawczego podmiotu, czyli jaźni, a sygnały zewnętrzne w stosunku do organizmu, informujące o poziomie konfliktu na podstawie własności bodźców pochodzących z otoczenia podmiotu, stanowią informację o nie-ja (Christoff i in. 2011: 109). Badania zespołu Christoff pokazały, że to uwaga zewnętrzna, a nie introspekcja czy zdolność do refleksji ma znaczenie dla powstawania poczucia bycia odrębnym podmiotem. Innymi słowy, działania, które przez afordancje „wymusza” na podmiocie świat, mają znaczenie dla kształtowania się jaźni.

WNIOSKI

W artykule odpowiedziałam na dwa pytania: i) czy podmiot jest w stanie wykryć i skorygować błędy samowiedzy oraz ii) jaki model najlepiej ten mechanizm opisuje. Twierdzę, że podmiot może wykryć i skorygować błędy w samowiedzy, ale jedynie za pomocą czynnika zewnętrznego, a zatem najlepszy model jaźni opisujący ten mechanizm autokorekty to model eksternalistyczny. Internalistyczny kartezyjański (choć we współczesnej postaci rozwijany np. przez Metzinger) model samowiedzy należy odrzucić. Pokazałam, że model EAM, nie uznając zewnętrznego nieneuronalnego elementu jako jednej ze swoich składowych, nie pozwala na korektę błędów powstających na po-

ziomie neuronalnym. Oznacza to, że podmiot zamknięty jest w pętli przetwarzania informacji na temat wewnętrznych stanów swojego organizmu. W EAM słusznie podkreśla się rolę, którą w tworzeniu samowiedzy odgrywa minimalna sobość powstająca na podstawie integracji multisensorycznej. Również słusznie model ten ukazuje, w jaki sposób samowiedza konstryuuje się na poziomie propozycyjalnym. Nie wykracza się jednak poza to, co już proponowano, definiując samowiedzę jako przekonania na temat własnych stanów mentalnych, które powstają na podstawie tzw. treści wąskiej. Ta propozycja, ze względu na rodzaj przyjętego rodzaju treści nadaje EAM charakter internalistyczny.

EAM przeciwstawiłam modelom eksternalistycznym odwołującym się bądź do cybernetyki (szczegółowo opisałam model autorstwa Setha), bądź do procesów neuronalnych, które jednak mogą zostać zalgorytmizowane. Te ostatnie modele pokazują, że dla zachowania statusu samowiedzy jako wiedzy konstytutywny jest element zewnętrzny, czyli informacja zwrotna ze środowiska, która dzięki wymuszaniu właściwego działania i dynamicznemu dostosowywaniu modelu ucieleśnionej jaźni pozwala na korektę nie tylko błędów poznawczych na temat świata, lecz także błędów autopoznawczych. Modele takie jak PPSMC nie odrzucają twierdzenia, że konstytucja podmiotu odbywa się na podstawie informacji o jego stanach wewnętrznych, ale dodają do niej również informacje płynące ze środowiska. W tych modelach dobrze sprawdza się kodowanie predykeyjne jako mechanizm leżący u podstaw samowiedzy, ponieważ pokazuje on, w jaki sposób następuje aktualizacja wiedzy podmiotu o nim samym. Dzieje się to po korekcie stawianej uprzednio hipotezy o danym stanie dzięki otrzymaniu informacji zwrotnej pochodzącej z otoczenia podmiotu.

BIBLIOGRAFIA

- Aspell J. E., Lenggenhager B., Blanke O. (2012), *Multisensory Perception and Bodily Self-Consciousness: From Out-of-Body to Inside-Body Experience* [w:] *The Neural Bases of Multisensory Processes*, M. M. Murray, M. T. Wallace (eds.), Boca Raton, FL: CRC Press, 467-481.
- Bar-On D. (2004a), *Speaking My Mind. Expression and Self-Knowledge*, Oxford: Clarendon Press. <https://doi.org/10.1093/0199276285.001.0001>
- Bar-On D. (2004b), *Externalism and Self-Knowledge: Content, Use, and Expression*, „Noûs” 38(3), 430-455. <https://doi.org/10.1111/j.0029-4624.2004.00477.x>
- Bermúdez J. L. (2011), *Bodily Awareness and Self-Consciousness* [w:] *The Oxford Handbook of the Self*, S. Gallagher (ed.), Oxford: Oxford University Press, 157-179. <https://doi.org/10.1093/oxfordhb/9780199548019.003.0007>

- Blanke O., Metzinger T. (2009), *Full-Body Illusions and Minimal Phenomenal Selfhood*, „Trends in Cognitive Sciences” 13(1), 7-13. <https://doi.org/10.1016/j.tics.2008.10.003>
- Boghossian P. A. (2008), *Content and Justification: Philosophical Papers*, Oxford: Clarendon Press.
- Brentano F. (1999), *Psychologia z empirycznego punktu widzenia*, tłum. W. Galewicz, Warszawa: Wydawnictwo Naukowe PWN.
- Burge T. (1979), *Individualism and the Mental*, „Midwest Studies in Philosophy” 4(1), 73-121. <https://doi.org/10.1111/j.1475-4975.1979.tb00374.x>
- Chisholm R. M. (1994), *Teoria poznania*, tłum. R. Ziemińska, Lublin: Daimonion.
- Christoff K., Gordon A. M., Smallwood J., Smith R., Schooler J. W. (2009), *Experience Sampling during fMRI Reveals Default Network and Executive System Contributions to Mindwandering*, „Proceedings of the National Academy of Sciences of the U.S.A.”, 106, 8719-8724. <https://doi.org/10.1073/pnas.0900234106>
- Christoff K., Cosmelli D., Legrand D., Thompson E. (2011), *Specifying the Self for Cognitive Neuroscience*, „Trends in Cognitive Science” 15(3), 104-112. <https://doi.org/10.1016/j.tics2011.01.001>
- Descartes R. (2009), *Rozprawa o metodzie właściwego kierowania rozumem i poszukiwania prawdy w naukach*, tłum. T. Boy-Żeleński, Kęty: Marek Derewiecki.
- Fodor J. A. (1999), *Jak grać w reprezentacje umysłowe – poradnik Fodora*, tłum. A. Putko, [w:] *Modele umysłu*, Z. Chlewiński (red.), Warszawa: Wydawnictwo Naukowe PWN.
- Friston K. J., Daunizeau J., Kiebel S. J. (2009), *Reinforcement Learning or Active Inference?*, „PLoS ONE” 4(7), e6421. <https://doi.org/10.1371/journal.pone.0006421>
- Gertler B. (2017), *Self-Knowledge* [w:] *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), E. N. Zalta (ed.), <https://stanford.io/20DfTm5>.
- Gibson J. J. (1950), *The Perception of the Visual World*, Cambridge, MA: The Riverside Press.
- Gibson J. J. (1978), *The Ecological Approach to the Visual Perception of Pictures*, „Leonardo” 11(3), 227-235. <https://doi.org/10.2307/1574154>
- Gibson J. J. (2002), *A Theory of Direct Visual Perception* [w:] *Vision and Mind: Selected Readings in the Philosophy of Perception*, A. Noë, E. Thompson (eds.), London: MIT Press, 77-91.
- Guttenplan S. (1994), *First Person Authority* [w:] *A Companion to the Philosophy of Mind*, S. Guttenplan (ed.), Oxford: Blackwell, 291.
- Hohwy J. (2013), *The Predictive Mind*, Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199682737.001.0001>
- Hume D. (1963), *Traktat o naturze ludzkiej*, tłum. C. Znamierowski, Warszawa: Państwowe Wydawnictwo Naukowe.
- Ionta S., Heydrich L., Lenggenhager B. (2011), *Multisensory Mechanisms in Temporoparietal Cortex Support Self-Location and First Person Perspective*, „Neuron” 70(2), 263-374. <https://doi.org/10.1016/j.neuron.2011.03.009>
- James W. (1950), *The Principles of Psychology*, t. I, New York: Dover.
- Komorowska-Mach J. (2015a), *Pojęcie introspekcji w psychologii potocznej*, „Przegląd Filozoficzny – Nowa Seria”, 24(3) [95], 75-89.
- Komorowska-Mach J. (2015b), *Przypisywanie sobie emocji – wyrażanie czy wykrywanie? Elementy detektywistyczne w modelu neoekspresywistycznym*, „Filozofia Nauki” 23(4) [92], 41-54.
- Lane T. (2011), *Self-Consciousness and Immunity*, „The Journal of Philosophy” 108(2), 78-99. <https://doi.org/10.5840/jphil201110824>

- Lane T., Liang C. (2009), *Higher-Order Thought and Pathological Self: The Case of Somatoparaphrenia*, „Analysis” 1. <https://doi.org/10.1093/analysis/anq094>
- Lane T., Liang C. (2010), *Mental Ownership and Higher-Order Thought: Response to Rosenthal*, „Analysis” 70(3), 496-501. <https://doi.org/10.1093/analysis/anq022>
- Lenggenhager B., Tadjji T., Metzinger T., Blanke O. (2007), *Video Ergo Sum: Manipulating Bodily Self-Consciousness*, „Science” 31, 1096-1099. <https://doi.org/10.1126/science.1143439>
- Locke J. (1955), *Rozważania dotyczące rozumu ludzkiego*, tłum. B. Gawecki, t. I, Warszawa: Państwowe Wydawnictwo Naukowe.
- Metzinger T. (2003), *Being No One: The Self-Model Theory of Subjectivity*, Cambridge, MA–London: MIT Press.
- Metzinger T. (2011), *Treść prezentacyjna*, tłum. A. Pacholik-Żuromska [w:] *Niemiecka analityczna filozofia umysłu*, A. Pacholik-Żuromska (red.), Toruń: Wydawnictwo Naukowe UMK.
- Metzinger T. (2015), *M-Autonomy*, „Journal of Consciousness Studies” 22(11-12), 270-302.
- Metzinger T. (2017), *The Problem of Mental Action – Predictive Control without Sensory Sheets* [w:] *Philosophy and Predictive Processing* 19, T. Metzinger, W. Wiese (eds.), Frankfurt am Main: MIND Group, 1-26. <https://doi.org/10.15502/9783958573208>
- Musholt K. (2015), *Thinking about Oneself: From Nonconceptual Content to the Concept of Self*, Cambridge, MA–London: The MIT Press.
- Neisser U. (1988), *Five Kinds of Self-Knowledge*, „Philosophical Psychology” 1(1), 35-59. <https://doi.org/10.1080/09515088808572924>
- Noë A. (2006), *Experience without the Head* [w:] *Perceptual Experience*, T. Gendler, J. Hawthorne (eds.), Clarendon, NY: Oxford University Press, 411-433.
- O'Brien L. (2007), *Self-Knowing Agents*, New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199261482.001.0001>
- Pacholik-Żuromska A. (2015), *On Confusing Propositional Attitudes*, „ProtoSociology: An International Journal and Interdisciplinary Project”, <https://bit.ly/2oQ75JK>, 1-15.
- Pacholik-Żuromska A. (2016), *Intencjonalność – rozumienie – samowiedza. Filozoficzny problem poznania w świetle nauk kognitywnych*, Seria „Umysł. Prace z Filozofii i Kognitywistyki”, Toruń: Wydawnictwo Naukowe UMK.
- Pacholik-Żuromska A. (2019), *Self-Knowledge as a Result of the Embodied and Social Cognition*, „Frontiers in Psychology”, 17 July, <https://doi.org/10.3389/fpsyg.2019.01679>
- Peacocke C. (1992), *A Study of Concepts*, Cambridge, MA: MIT Press.
- Peacocke C. (1999), *Being Known*, Oxford: Clarendon Press. <https://doi.org/10.1093/0198238606.001.0001>
- Peacocke C. (2008), *Mental Action and Self-Awareness (II): Epistemology* [w:] *Mental Action*, L. O'Brien, M. Soteriou (eds.), Oxford: Oxford University Press, 192-214. <https://doi.org/10.1093/acprof:oso/9780199225989.003.0010>
- Piłat R. (1999), *Umysł jako model świata*, Warszawa: IFiS PAN.
- Piłat R. (2013), *Aporie samowiedzy*, Warszawa: IFiS PAN.
- Poczobut R. (2008), *Rodzaje samoświadomości*, „Analiza i Egzystencja” 7, 5-31.
- Putnam H. (1998b), *Znaczenie wyrazu „znaczenie”* [w:] *Wiele twarzy realizmu i inne eseje*, tłum. A. Grobler, Warszawa: Wydawnictwo Naukowe PWN.
- Schooler J. W., Schreiber C. A. (2004), *Experience, Meta-Consciousness, and the Paradox of Introspection*, „Journal of Consciousness Studies” 11(7-8), 17-39.

- Schooler J. W., Smallwood J., Christoff K., Handy T. C., Reichle E. D., Sayette M. A. (2011), *Meta-awareness, Perceptual Decoupling and the Wandering Mind*, „Trends in Cognitive Sciences” 15(7), 319-326. <https://doi.org/10.1016/j.tics.2011.05.006>
- Seth A. K., Critchley H. D. (2013), *Interoceptive Predictive Coding: A New View of Emotion?*, „Behavioral and Brain Sciences” 36(3), 227-228. <https://doi.org/10.1017/S0140525X12002270>
- Seth A. K. (2015), *The Cybernetic Bayesian Brain: From Interoceptive Inference to Sensorimotor Contingencies* [w:] *Open MIND*, T. Metzinger, J. M. Windt (eds.), MIT Press, 1-24. <https://doi.org/10.15502/9783958570108>
- Shoemaker S. (1970), *Persons and their Pasts*, „American Philosophical Quarterly” 7(4), 269-285.
- Shoemaker S. (1994), *Self-Knowledge and “Inner Sense”*, „Philosophy and Phenomenological Research” 54(2), 249-314. <https://doi.org/10.2307/2108490>
- Speaks J. (2004), *A Guide to Shoemaker’s “Self-Knowledge and Inner Sense”*, „Philosophy” 519, March 5.
- Steels L., Spranger M. (2008), *The Robot in the Mirror*, „Connection Science” 20(4), 337-358. <https://doi.org/10.1080/09540090802413186>
- Wilson M. (2002), *The Six Views of Embodied Cognition*, „Psychonomic Bulletin & Review” 9(4), 625-636. <https://doi.org/10.3758/BF03196322>
- Wittgenstein L. (2000), *Dociekania filozoficzne*, tłum. B. Wolniewicz, Warszawa: Wydawnictwo Naukowe PWN.
- Wittgenstein L. (2003), *Philosophische Untersuchungen*, Frankfurt am Main: Suhrkamp Verlag.
- Wygotski L. S. (1989), *Myślenie i mowa*, tłum. E. Flesznerowa, J. Fleszner, Warszawa: Państwowe Wydawnictwo Naukowe.
- Wygotski L. S. (2006), *Narzędzie i znak w rozwoju dziecka*, tłum. B. Grell red. nauk. M. Klimkowski, Warszawa: Wydawnictwo Naukowe PWN.
- Ziemińska R. (2002), *Eksternalizm we współczesnej epistemologii*, Szczecin: Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.
- Ziemińska R. (2008), *Samoświadomość i samowiedza z punktu widzenia epistemologii*, „Analiza i Egzystencja” 7, 33-51.