



Spór o ciągłość życia i umysłu Argumenty na rzecz kognitywizmu*

Michał PIEKARSKI**

ABSTRACT

The dispute over the continuity of life and mind. Arguments for cognitivism: The purpose of this paper is to discuss the position of non-cognitivism on the issue of the so-called dispute over the continuity / discontinuity of life and mind. In discussing the views of Michael Kirchhoff and Tom Froese, I will point out some difficulties related to their position. Next, I will formulate three arguments in favor of the cognitive alternative, emphasizing the need to resort to semantic information in explaining these phenomena. According to noncognitivist position, there is a continuity in the life-mind line, which can be justified by referring to the concept of Shannon's syntactic information. Opponents of this thesis, *i.e.* supporters of cognitivism, claim that the explanation of cognition requires the use of other tools than those used to explanation of life, because, first, the notion of syntactic information does not exhaust the complexity of these phenomena, and, second, the non-cognitive position raises many problems and ambiguities. According to cognitivists, when explaining life and mind, one should refer to the concept of semantic information that is rejected by noncognitivists. In the *Conclusion* I will analyze the ambiguities and assumptions related to the thesis about the continuity or discontinuity between life and cognitive processes.

KEYWORDS

life; mind; semantic information; syntactic information; free energy principle; normativity; body; content; Markov blankets

* W niniejszym artykule wykorzystuję zmodyfikowane i uzupełnione fragmenty mojej książki *Mechanizmy predykcyjne i ich normatywność* (Piekarski, 2020).

** Dr filozofii, adiunkt w Instytucie Filozofii Uniwersytetu Kardynała Stefana Wyszyńskiego w Warszawie. E-mail: m.piekarski@uksw.edu.pl.

WSTĘP

Natura relacji pomiędzy procesami życiowymi a mentalnymi należy do starych zagadnień filozoficznych¹. We współczesnej filozofii umysłu i kognitywistyce problem ten jest określany jako zagadnienie ciągłości bądź nieciągłości pomiędzy życiem i umysłem (*continuity and discontinuity between life and mind*). Kontekstem uzasadniającym podjęcie tej dyskusji przez badaczy jest sformułowana przez brytyjskiego neuronaukowca Karla Fristona zasada energii swobodnej (*free energy principle*, FEP). Zgodnie z tą zasadą każdy samoorganizujący się system, który nie pozostaje w stanie równowagi (*nonequilibrium steady-state*) energetycznej ze środowiskiem, musi minimalizować swoją energię swobodną w sensie informacyjno-teoretycznym (Friston, 2013; Friston & Stephan, 2007)². Oznacza to, że zasada ta ma zastosowanie do wszystkich systemów

¹ Problem ten poruszał już Arystoteles, wiążąc go ze stanowiskiem hylemorfizmu. W wiekach późniejszych (szczególnie w okresie nowożytności) pojawiły się takie koncepcje jak witalizm, mechanicyzm (sumatywizm), organicyzm (holizm), emergentyzm (Mayr, 2002). Współcześnie wskazuje się między innymi na nowy mechanicyzm, dynamizm czy teorię emergencji. Niektórzy (Poczobut, 2011) upatrują w sporze substancjalizm–funkcjonalizm podstawowe odróżnienie, które pozwala mówić o różnych koncepcjach życia na gruncie filozofii biologii. Inni zaś (Heller, Lubański, & Ślaga, 1997) wskazują na spór na osi redukcjonizm–antyredukcjonizm.

² Fristonowskie pojęcie energii swobodnej może budzić skojarzenia z pochodzącym z termodynamiki pojęciem energii swobodnej — entalpii. Ta ostatnia oznacza wielkość fizyczną, która jest funkcją stanu mającą wymiar energetyczny. Innymi słowy jest ona równa sumie energii wewnętrznej danego układu oraz iloczynu ciśnienia i objętości tego układu. Różnica pomiędzy tymi pojęciami jest jednak zasadnicza: fristonowska energia swobodna (mimo że inspirowaną entalpią) jest miarą informacji i oznacza górną granicę zaskoczenia dla danego organizmu (wielkość ta jest dokładnie taka, jak dolna granica dowodu w uczeniu maszynowym lub dowodu modelu bayesowskiego w statystyce [Friston, 2019; Friston, Wiese, Hobson, 2021]). Technicznie rzecz biorąc, energia swobodna to teoretyczna ilość informacji, która ogranicza (będąc większa od niej) entropię wymiany sensorycznej między żywym systemem (na przykład mózgiem czy całym organizmem) a środowiskiem. Wspomniana minimalizacja tej energii jest zatem sposobem maksymalizacji prawdopodobieństwa sensorycznych danych, odbieranych przez system oraz jednocześnie uwarunkowanych modelem (statystycznym tego, w jaki sposób generowane są te dane), które te dane wejściowe odbiera. W tym ujęciu entropia odnosi się do (długoterminowego) średniego zaskoczenia lub ujemnego logarytmicznego prawdopodobieństwa danych sensorycznych napotkanych przez organizm (Friston, 2010). Zgodnie z tym dany organizm rozumiany jako system adaptacyjny, który jest częścią większego systemu, w którym jest osadzony, a jednocześnie od niego statystycznie niezależny, musi zawierać probabilistyczny model (tj. model generatywny) współzależności statystycznych i prawidłowości swojego otoczenia. Inspiracji dla energii swobodnej Fristona można szukać w zasadzie maksymalnej entropii Jaynesa, zasadzie najmniejszego działania Hamiltona, wykorzystywanych w mechanice statystycznej wariacyjnych metodach Feynmana (Feynman, 1980), uczeniu maszynowym oraz geometrii informacji (Caticha, 2015; zob. też Friston, 2019; Friston, Wiese, & Hobson, 2021; por. także Andrews, 2021). Wielkość, nazywana przez Fristona wariacyjną energią swobodną (ze względu na wykorzystywanie metod wariacyjnych), jest minimalizowana przez kolejne iteracje modelu, w ten sposób doprowadzając przybliżoną późniejszą gęstość prawdopodobieństwa określonych

biologicznych i jest związana z ich wrodzoną zdolnością do przeciwdziałania naturalnej tendencji do zaburzeń. Innymi słowy: to, co istnieje, musi (zgodnie z matematyką opartą na mechanice statystycznej i modelowaniu dynamiki stanów pozostających w nierównowadze energetycznej ze środowiskiem³ [Friston, 2019]) wykazywać takie własności, dzięki którym można je opisać tak, jakby optymalizowało określoną wielkość, tj. wariacyjną energię swobodną⁴. Oznacza to, że każdy żywy organizm jest ergodycznym, losowym układem dynamicznym (Friston, 2013), tj. takim, dla którego statystyki uśrednione w czasie są takie same, jak jego statystyki uśrednione w trakcie procesu. Ergodyczność oznacza, że średnia czasowa dowolnej mierzalnej funkcji systemu upodabnia się w pewnym skończonym czasie. Heurystycznie mówiąc: z jednej strony obserwacja ewolucji na przykład jednej wartości dostarcza informacji o całym systemie; z drugiej zaś strony dokładna znajomość całego systemu w danej chwili pozwala na przewidywanie średniego zachowania ewolucji określonej wartości w jakimś okresie czasu. Z tego powodu stany wewnętrzne danego systemu (komórki, ciała, organizmu itd.) można opisać za pomocą koców Markowa (Friston, 2013; Kirchhoff *et al.*, 2018)⁵, które dla zachowania integralności i autonomii muszą minimalizować swoją energię swobodną, czyli różnicę między oczekiwaniami dotyczącymi świata (jak jest on modelowany) a tym, jaki on jest. Minimalizacja wariacyjnej energii swobodnej jest zatem kluczowa dla organizmu, zmniejsza ona bowiem stopień niepewności i zaskoczenia⁶, jednocześnie zwiększając wiarygodność (*evidence*) jego modelu świata (Hohwy, 2014). Można powiedzieć, że

danych, w interpretacji statystyki bayesowskiej — do przybliżonej zgodności z rzeczywistą gęstością docelową tych danych. W tym sensie minimalizacja energii swobodnej jest przybliżonym (*approximate*) wnioskowaniem bayesowskim organizmu o swoim środowisku i samym sobie. Więcej na temat zależności pomiędzy FEP a fizyką statystyczną por. także Ramstead, Badcock, & Friston, 2017.

³ Mowa tutaj między innymi o równaniu Fokkera-Plancka czy równaniu Langevina.

⁴ Na temat pojęcia „tak, jakby” (*as if*) zob. McGregor, 2017.

⁵ W myśl omawianego tutaj stanowiska stany wewnętrzne danego organizmu to koce Markowa (Pearl, 1988), które oddzielają system od jego środowiska. Oznacza to, że stany te są niezależne (w sensie statystycznym) od stanów otoczenia. Mają one za zadanie minimalizować energię swobodną, aby zachować strukturalną i funkcjonalną integralność, która prowadzi do homeostazy i samoorganizacji. Warunkiem minimalizacji energii swobodnej jest zatem ścisła separacja wnętrza systemu od jego otoczenia, czyli zdolność tego systemu do odróżniania siebie od środowiska (Friston, 2013). Istnieje pewna rozbieżność między „instrumentalną” interpretacją koców Markowa Pearl’a a „realistyczną” Fristona (Bruineberg *et al.*, 2020). Interpretacja Pearl’a nie wykracza poza czysty formalizm, a same koce Markowa są po prostu matematycznym konstruktem służącym do wnioskowania na temat (na przykład) modeli generatywnych. W ujęciu Fristona właściwości koców są rzutowane na systemy docelowe, na przykład organizmy żywe. Takie podejście jednak nie jest, jak wykazuje Bruineberg *et al.*, „filozoficznie niewinne”.

⁶ Zaskoczenie to ujemny logarytm prawdopodobieństwa wystąpienia zdarzenia. Innymi słowy, zaskoczenie mówi o tym, jak mało prawdopodobne jest określone zdarzenie w odniesieniu do danego modelu świata.

w ten sposób organizmy opierają się drugiej zasadzie termodynamiki (Friston & Stephan, 2007; Ramstead, Badcock, & Friston, 2017)⁷. Chodzi o to, że ze wszystkich możliwych stanów, w których organizm może się znaleźć, musi on wybrać taki odpowiedni podzbiór stanów, jaki pozwoli mu przetrwać i skutecznie dokonywać wymiany energetycznej z otoczeniem. Istnieje zatem określony rozkład prawdopodobieństwa dla wszystkich możliwych stanów, w których organizm może się znaleźć. W dowolnym momencie rozkład ten (określony przez fenotyp danego organizmu) jest gwałtownie osiągany wokół pewnych wartości określających warunki organizmu niezbędne dla jego funkcjonowania i przeżycia.

Teoria zaproponowana przez Fristona pozwala na nowo postawić pytanie o relację procesów życiowych oraz procesów konstytuujących to, co filozofowie nazywają umysłem. Należy jednak podkreślić, że nie ma zgody co do tego, jak najlepiej interpretować właściwości wariacyjnej energii swobodnej w zastosowaniu do tego problemu. Istnieją dwie główne możliwości interpretacji tej zasady w tym kontekście — ujęcie kognitywistyczne i nonkognitywistyczne⁸. Zwolennicy pierwszego głoszą niezależność życia i umysłu. Łączą energię swobodną z informacją semantyczną, ograniczając tym samym zakres organizmów, którym można przypisać proste funkcje poznawcze (*mentality*) (Hohwy, 2015; Hohwy, 2017). Druga interpretacja ma trzy odmiany. Pierwsza odmiana, termodynamiczna, łączy FEP z fizyką statystyczną i stosowana jest w wyjaśnianiu dynamiki samoorganizacji w organizmach unikających równowagi termodynamicznej (Friston, 2009; Friston, 2010). Druga odmiana, ewolucyjna, opiera się na stwierdzeniu braku ciągłości pomiędzy systemami żyjącymi a systemami poznawczymi. W tym ujęciu procesy poznawcze wymagają istnienia złożonych mechanizmów neuronalnych, które nie występują u prostych organizmów. Oznacza to, że żywy organizm nie jest z konieczności systemem poznawczym (Clark, 2017). Trzecia zaś odmiana, określona mianem enaktywistycznej, podkreśla silną ciągłość zachodzącą pomiędzy życiem a umysłem. Analizy przeprowadzone w tym artykule poświęcę dyskusji z tym ostatnim stanowiskiem (sformułowanym w wersji zaproponowanej przez Michaela Kirchhoffa i Toma Froeseego [Kirchhoff & Froese, 2017]) ze względu na jego teoretyczną doniosłość oraz oddziaływanie na innych badaczy.

⁷ Według drugiej zasady termodynamiki entropia, czyli miara stopnia nieuporządkowania układu, rośnie z czasem. Zasadniczo druga zasada odnosi się do układów izolowanych bądź zamkniętych, ale w tym ujęciu chodzi przede wszystkim o to, że entropia zwiększa się wraz z przekazywaniem energii (Clark, 2017: 3).

⁸ Są to stanowiska skrajne. Należy jednak podkreślić, że istnieje wiele różnych propozycji pośrednich, które w mniejszym lub większym stopniu zakładają istnienie takiej ciągłości bądź argumentują za jej brakiem (Allen, 2018; Allen & Friston, 2018; Korbak, 2021; Villalobos & Dewhurst, 2018).

TEZA NONKOGNITYWISTYCZNA

Argumentacja zwolenników enaktywistycznego podejścia do FEP opiera się na dwóch przesłankach. Pierwsza z nich dotyczy natury umysłu, druga zaś analizowania umysłu w kategoriach autopoezy i adaptacji. Zaczniemy od pojęcia umysłu. Michael Kirchoff i Tom Froese wykorzystują zaproponowane przez Daniela Hutto i Erika Myina pojęcie umysłu prostego (*basic mind*) (Hutto & Myin, 2013). Zdaniem tych autorów umysł w swoim najbardziej podstawowym wymiarze (tj. na poziomie wczesnej onto- i filogenezy) jest pozbawiony jakiegokolwiek treści. Można go wyjaśnić jedynie w kategoriach złożonych i dynamicznych interakcji sensomotorycznych ze środowiskiem, nie powołując się przy tym na żadne reprezentacje. Takie pojęcia jak „treść” czy „reprezentacja” nie mają wartości eksplanacyjnej. Treść w umyśle jest nabywana jedynie w kontekście praktyk społeczno-lingwistyczno-kulturowych. Czym zatem jest? Zdaniem Hutto i Myina treścią jest wszystko to, co ma określone warunki spełniania (*conditions of satisfaction*), czyli musi być albo prawdziwe, albo fałszywe, poprawne bądź nie. Tutaj pojawia się tak zwany trudny problem treści (*hard problem of content*), który podważa możliwość naturalistycznego uzasadnienia zachodzenia relacji pomiędzy własnościami, które mają treść, a własnościami fizycznymi (Hutto & Myin, 2013: 69). Omawiani autorzy twierdzą, że nie ma takiego naturalistycznego pojęcia treści, jakie rozwiązywałoby jej trudny problem, ponieważ wszelkie zaproponowane w literaturze pojęcia informacji oparte są na kowariancji, czyli wielkości, która charakteryzuje wspólne zmiany dla dwóch zmiennych. Pojęcie informacji semantycznej z tego powodu również nie pozwala wyjaśnić sposobu, w jaki umysł zyskuje treść. Postulowanie istnienia takiego rodzaju informacji jest raczej założeniem niż wyjaśnieniem istnienia własności semantycznych albo zawierających treść (Hutto & Myin, 2013: 67). Z tego też powodu autorzy ci posługują się pojęciem prostego poznania, które nie ma charakteru fenomenalnego, tj. odnoszącego się do tych własności stanów mentalnych (jak wrażenia czy dane zmysłowe), dzięki którym są one „o czymś” i mają naturę przekonaniową, ale ma charakter intencjonalny, czyli wykazuje celowe ukierunkowanie. Owa celowość nie implikuje jednak intencjonalności w sensie osobowym, tj. świadomego odnoszenia się do jakiejś przedmiotowości, ale raczej sposób, w jaki organizm aktywnie angażuje się w swoje środowisko. Można tutaj, jak sądzę, mówić o minimalnej intencjonalności biologicznej, ukonstytuowanej ewolucyjnie (Hutto & Myin, 2013: 30).

W tym kontekście nawet najbardziej prymitywne organizmy żywe, na przykład bakterie, można scharakteryzować jako mające proste umysły. Dowiedzeniu tego ma służyć uwaga Petera Godfrey-Smitha i Kima Sterelnego, że jedyną charakterystyką informacji biologicznej ma być ta, którą podał Claude E. Shannon (Shannon, 1948), a która pojawia się zawsze wtedy, gdy istnieje relacja przyczynowo-skutkowa czy określona korelacja (Godfrey-Smith & Sterelny, 2016).

Froese i Kirchoff nie zwracają jednak uwagi na pełny kontekst tej wypowiedzi. Mianowicie: Godfrey-Smith i Sterelny stwierdzają, że informacja Shannona spełnia pewne minimalne warunki, jakie można nałożyć na jakąkolwiek informację. W tym sensie przykładowo dym niesie informację o pożarze. Oznacza to, że sygnał przynosi informację o źródle w tym sensie, że możemy przewidzieć stan źródła na podstawie sygnału. Nie powiemy jednak, że z tego powodu dym jest żywy. Shannon zaproponowane pojęcie informacji wiązał z technologiami komunikacyjnymi. I to jest ważna wskazówka, bowiem jeżeli uznać wymagania nakładane na informację przez Shannona za kryteria informacji biologicznej, to zanika różnica pomiędzy organizmami żywymi, takimi jak bakterie, zwierzęta czy nawet rośliny, a sztucznym życiem, zjawiskami fizycznymi (jak dym) lub wszelkimi urządzeniami przetwarzającymi informację, tj. prędkościomierzami, termometrami czy kalkulatorami.

Tych słabości omawiani autorzy zdają się nie spostrzegać i raczej sugerują, że z trudnym problemem treści dobrze sobie radzi pojęcie informacji wzajemnej (*mutual information*) (Kirchoff & Froese, 2017: 7)⁹. Ma ono w sposób satysfakcjonujący ujmować intuicje zawarte we fristonowskiej FEP: wzajemność jest kowariancją w tym sensie, że dwie zmienne są ze sobą skorelowane w taki sposób, że informacja niesiona przez jedną redukuje poziom zaskoczenia u drugiej. Tak rozumiana informacja optymalizuje model. Nie potrzeba się tutaj odwoływać do żadnych stanów wewnętrznych organizmu, manipulacji reprezentacjami czy treści semantycznych¹⁰. Należy jednak zwrócić uwagę na fakt, że informacja wzajemna jest uogólnioną kowariancją w takim samym sensie, jak informacja semantyczna. Niemniej od informacji semantycznej wymaga się spełnienia dodatkowych kryteriów.

Różnica pomiędzy umysłem a życiem, twierdzą Kirchoff i Froese, jest jak różnica pomiędzy informacją semantyczną a niesemantyczną. Ta pierwsza jest ukonstytuowana na drugiej za sprawą procesów zewnętrznych względem organizmu. Proste umysły są silnie powiązane z biologią danego organizmu. Są one intencjonalne i celowo skierowane na swoją niszę biologiczną, na którą reagują (Hutto & Myin, 2013: x). Owa reakcja nie jest jednak zależna od jakiegokolwiek interpretacji, rozumienia czy jakkolwiek rozumianej aktywności kierowanej reprezentacjami. Ową celową intencjonalność można wyjaśnić przez odniesienie do historii doboru naturalnego, ponieważ jego jedynym zadaniem

⁹ Informacja wzajemna jest miarą zależności pomiędzy dwiema zmiennymi losowymi. Inaczej mówiąc: informacja wzajemna mierzy, ile informacji o danej zmiennej można poznać, znając inną, czyli o ile poznanie jednej z tych zmiennych zmniejsza niepewność co do drugiej. Jeśli zmienne są od siebie niezależne, to ich wzajemna informacja jest zerowa (znajomość jednej nie mówi niczego o drugiej). Jeśli są identyczne, to każda zawiera pełną wiedzę o drugiej.

¹⁰ Interesujące w tym kontekście jest to, że podobieństwo w reprezentacjach opartych na podobieństwie (na przykład S-reprezentacjach) można mierzyć w kategoriach informacji wzajemnej. Za tę uwagę dziękuję Marcinowi Miłkowskiemu.

jest wzbudzenie określonych reakcji na takie, a nie inne rzeczy czy zdarzenia (Hutto, 2006). Z tego to powodu ktoś mógłby stwierdzić, że intencjonalne odniesienie czy ukierunkowanie ma wymiar normatywny (Kirchhoff & Froese, 2017: 13), to znaczy, że można mówić tutaj o poprawnym bądź niepoprawnym reagowaniu na środowisko. Przykładowo: organizm reaguje na zmianę temperatury termoregulacją, na przykład poci się albo trzęsie, gdy temperatura zmienia się odpowiednio powyżej lub poniżej pewnej średniej. Zaburzenie mechanizmów termoregulacji można zatem traktować jako dysfunkcję (Millikan, 1984). Kirchhoff i Froese nie chcą jednak mówić o normatywności funkcji albo mechanizmów. W zamian proponują ujęcie normatywności, które odnosi się do procesów samoorganizacji i autopojezy realizowanych przez organizm jako pewną inherentną całość. Ujawnia się tutaj różnica pomiędzy systemami żywymi a artefaktami, takimi jak termostat. Te ostatnie realizują funkcje normatywne za sprawą swoich projektantów. Żywe systemy autopojetyczne, jako autonomiczne, są źródłem swojej wewnętrznej normatywności, którą należy wiązać z utrzymywaniem stanów wewnętrznych przez ich własne koce Markowa, na drodze aktywnej ingerencji w środowisko. Życie jest procesem, dzięki któremu organizm unika stanu równowagi termodynamicznej, inwestując energię w minimalizowanie niepewności. W ten sposób definiuje on (statystyczną) granicę pomiędzy tym, co należy do systemu, a tym, co należy do jego środowiska. Owa granica określa autonomię jednostki.

Powiązanie jednostki ze środowiskiem można opisać w kategoriach holistycznej normatywności. Bezpośrednie intencjonalne odniesienie opiera się na procesach autopojezy oraz adaptacji w środowisku opartej na FEP, które prowadzą do indywiduacji. Ciągłość życia i umysłu jest zatem wynikiem bezpośredniej normatywnej relacji zachodzącej pomiędzy podstawowymi wymiarami życia (autopojetycznymi) a podstawowymi wymiarami umysłowości (bezpośrednio intencjonalnej, pozbawionej treści i semantyki) (Kirchhoff & Froese, 2017: 16). Za pośrednictwem swojego koca Markowa system jest operacyjnie domknięty, to znaczy jego dynamika, oddzielająca stany wewnętrzne i zewnętrzne, umożliwia jego jednostkowienie.

Kirchhoff (Kirchhoff, 2018), powołując się na analizy Ezequiela Di Paolo (Di Paolo, 2005) i Evana Thompsona (Thompson, 2011), dodaje, że sama rama wyznaczana przez teorię autopojetyczną nie pozwala w pełni wyjaśnić istnienia umysłów. Dopiero wzbogacenie jej przez FEP czyni zadość temu oczekiwaniu. Należy się zgodzić, powiada Kirchhoff, że autopojeza jest koniecznym i wystarczającym warunkiem dla życia. Z tego jednak nie wynika, że jest wystarczającym warunkiem dla pojawienia się umysłu. Dzięki temu, że w ramach FEP poznanie jest zawsze zorientowane na przyszłość, zyskuje ono charakter adaptacyjny: „systemy, które minimalizują swoją energię swobodną, odpowiadają na zmiany w środowisku w sposób adaptacyjny” (Friston & Stephan, 2007: 428, przekład własny). Znaczy to, że cel, jakim jest minimalizacja zaskoczenia

i niepewności, wymaga regulacji, która opiera się na ciągłych przewidywaniach. Dzieje się to na drodze realizacji wniosku aktywnego, czyli takiego próbkowania środowiska, które ma charakter epistemiczny. Dobrym przykładem są eksploracyjne ruchy sakkadowe gałki ocznej (Friston *et al.*, 2012). W ten sposób życie wiąże się z poznaniem: procesy homeostatyczne są współregulowane przez procesy epistemiczne. Zdaniem Kirchhoffa takie powiązanie FEP i teorii autopojęzy pozwala w satysfakcjonujący sposób wykazać ciągłość zachodzącą pomiędzy życiem a umysłem. Pogląd ten można określić mianem „biopsychizmu” (Kirchhoff, 2018: 2531–2532).

KRYTYKA UJĘCIA ENAKTYWISTYCZNEGO

Koncepcji proponowanej przez Kirchhoffa i Froeseego można postawić zarzuty. Pierwszy z nich dotyczy zastosowania takich pojęć jak życie czy prosty umysł oraz postulatu głoszącego konieczność powiązania ze sobą desygnatów tych pojęć. Trudno jednoznacznie stwierdzić, co ci autorzy rozumieją przez życie. W ich tekstach próżno szukać innego określenia niż biologiczna samoorganizacja. Jest to jednak ujęcie czysto opisowe i nie można przypisać mu specjalnej mocy eksplanacyjnej. Co więcej, samo mówienie o życiu jako takim jest mocno problematyczne, ponieważ prowadzi do urzeczowienia procesów, które przebiegają w organizmach. „Nie istnieje ono [życie — przyp. M.P.] jako niezależny byt. Naukowo można zajmować się procesami życiowymi, lecz nie abstrakcyjnym «życiem»” (Mayr, 2002: 20). Brak jasnej eksplikacji tego, co kryje się pod pojęciem życia, prowadzi do wielu niejasności i budzi pytania dotyczące tego, co decyduje o zakwalifikowaniu danego systemu do świata ożywionego. Dodatkowo mówienie o życiu wyłącznie w kategoriach autopojęzy nie pozwala wyjaśnić oczywistych różnic poznawczych zachodzących pomiędzy organizmami oraz różnic w ich funkcjonowaniu, co w sposób satysfakcjonujący może zrobić na przykład psychologia porównawcza (Miłkowski, 2018). Pytanie o definicję życia jest ważne, ponieważ (1) można wskazać na istnienie form ewolucyjnych (na przykład wirusy) oraz nieewolucyjnych (sztuczne życie, roboty, sztuczna inteligencja itd.), których nie da się jednoznacznie zakwalifikować do układów żywych (Latawiec, 1992; Heller, Lubański, & Ślaga, 1997: 299, 308); zaś (2) dyskutowana tutaj propozycja zbliża się do koncepcji funkcjonalistycznych, które w pewien sposób abstrahują od substratu materialnego życia (cielesności, morfologii itd.) oraz ujmują życie przede wszystkim w kategoriach przetwarzania informacji (Poczobut, 2011: 233) lub energii. Zwolennicy teorii autopojetycznej oczywiście *explicite* wskazują na istotny wpływ ciała oraz usytuowania w środowisku, jednak proponowane przez nich ujęcie życia *implicitie* nie korzysta z wartości eksplanacyjnej tych pojęć. Mianowicie: nie jest jasno określone, w jaki sposób ucieleśniony jest koc Markowa charakteryzujący żywy

organizm. Na czym miałyby polegać faktycznie domknięcie operacyjne układu, który w pewnym sensie jest odgraniczony od środowiska przez granice własnego ciała? Jaką funkcję pełni morfologia ciała w aktywnym wnioskowaniu? Na te pytania omawiani autorzy nie dają jasnej odpowiedzi¹¹.

Wielu badaczy jest również zgodnych co do tego, że nie da się podać jednej ostatecznej (monoatrybutywnej) definicji życia; w zamian ograniczają się oni do formułowania definicji poliatributywnych, które wskazują na wiele różnych właściwości układów żywych. Do tych własności można zaliczyć: (1) istnienie programów genetycznych; (2) występowanie w strukturze organizmów żywych specyficznych makrocząstek (hormony, enzymy itd.), które odpowiadają za ich rozwój i funkcjonowanie; (3) określoną organizację funkcjonalną; (4) określony rząd wielkości; (5) istnienie cykli życiowych i rozrodczych; (6) występowanie złożonych mechanizmów regulacyjnych, w tym sprzężeń zwrotnych; oraz (7) otwartości energetycznej i informacyjnej na oddziaływanie ze środowiskiem. Zwolennicy ujęcia autopojetycznego zdają się literalnie ograniczać do tych dwóch ostatnich własności, co pozwala zaliczyć ich podejście do monoatrybutywnego traktowania życia, przy oczywistym ignorowaniu pozostałych własności, na które zwracają uwagę zwolennicy ujęć poliatributywnych.

Wiele wątpliwości budzi pojęcie umysłu wykorzystywane przez Kirchoffa i Froese. Jak wspomniałem wcześniej, termin ten został zaczerpnięty przez tych badaczy z prac Hutto i Myina oraz wzbogacony o pojęcie informacji w rozumieniu Shannona. Hutto i Myin twierdzą, że prosty umysł to ten, który jest pozbawiony treści, czyli nie zawiera żadnych reprezentacji. Jednak w przeciwieństwie do badaczy układów dynamicznych (Chemero, 2014) albo robotyki (Brooks, 1991) nie pokazują, w jaki sposób można modelować albo wyjaśnić funkcje umysłowe,

¹¹ Co prawda, wspominają to, że ciało współkonstruuje model generatywny (Kirchhoff & Froese, 2017: 14), ale poza samą tą deklaracją nie wyjaśniają, na czym owa konstytucja miałyby polegać. A jest to istotny brak dla stanowiska, które chce uwypuklić znaczenie życia dla zrozumienia funkcjonowania systemów ergodycznych. Chcę przez to powiedzieć, że brak konkretnych analiz roli ciała w minimalizowaniu energii swobodnej, kluczowej dla utrzymania życia, powoduje, że stanowisko Kirchoffa i Froesego nieznacznie różni się od stanowisk, które ignorowałyby zupełnie rolę cielesności w tych aspektach. Można tutaj postawić analogiczny zarzut do tego, który sformułował William Ramsey (Ramsey, 2007) wobec stanowisk określanych jako reprezentacjonistyczne. Jego zdaniem zbyt często i łatwo w kognitywistyce przypisuje się reprezentacjom moce eksplanacyjne. Kluczowe jest to, czy przyjęcie reprezentacjonizmu pozwala lepiej wyjaśnić dane zjawisko lub jego klasę. Chodzi zatem o wykazanie, że używane w danej koncepcji czy modelu pojęcie reprezentacji nie jest trywialne, czyli że spełnia „wymóg opisu zadania” (*job description challenge*), który polega na pokazaniu, w jakim sensie postulowane w ramach danej koncepcji czy modelu pojęcie reprezentacji rzeczywiście służy jako reprezentacja (Ramsey, 2007: xv). Sądzę, że wymóg sformułowany przez Ramseya można zastosować także do innych pojęć (na przykład ciała), którymi posługują się kognitywiści. W kontekście prowadzonych tutaj analiz należałoby zatem wykazać, czy pojęcie ciała pełni jakąś określoną eksplanacyjną funkcję, której nie dałoby się zredukować na przykład do funkcji realizowanych przez pojęcia organizmu albo systemu poznawczego. Twierdzę, że nie.

nie korzystając jednocześnie ze słownika reprezentacjonistycznego. Pewne trudności wynikają także z utożsamienia poznawania z posiadaniem umysłu. Wystarczy przypisać pewnym artefaktom realizację procesów przetwarzania informacji (Dretske, 2004), aby uznać je za istoty mające proste umysły. Tutaj już dzieli nas krok od przypisania tym struktutom życia, co wiązałoby się z wykazaniem ciągłości pomiędzy posiadaniem prostego umysłu a byciem żywym organizmem. Wystarczyłoby zatem odwrócić rozumowanie Kirchoffa i Froese'go. Zabieg ten jest dopuszczalny, ponieważ epistemiczny wymiar FEP, który podkreślają omawiani autorzy, wskazuje na konstytutywną zależność zachodzącą pomiędzy procesami życiowymi a procesami mentalnymi. Trudno się zatem zgodzić z Froesem (Froese, 2014), który twierdzi, że rozróżnienie pomiędzy żywym insektem a jego robotycznym modelem opiera się na posiadaniu prostego umysłu i autonomii¹². Różnica pomiędzy tymi dwoma układami jest różnicą w przypisywaniu normatywności: normatywność, którą można przypisać żywemu owadowi, jest biologiczna; normatywność modelu jest zaś zależna od producenta czy architekta. Hutto i Myin nie wywodzą jednak normatywności weredycznej, czyli tej, która jest istotna z punktu widzenia wyjaśnienia procesów poznawczych, z normatywności biologicznej (Hutto & Myin, 2013: 112)¹³.

Należy stwierdzić, że ujęcie enaktywistyczne oparte na teorii autopojezy ostatecznie proponuje (niezbyt) nowy słownik dla już opisanych problemów, zaś w aspekcie spekulatywnym, który uważa się za charakterystyczny dla tego nurtu (na przykład problem ciągłości na linii życie–umysł), wciąż w pewnym stopniu tkwi na poziomie pojęciowych analiz, konstruowania definicji projektujących¹⁴ oraz prób coraz precyzyjniejszego wypowiedzenia swoich założeń ba-

¹² W inny, bardziej przekonujący, sposób argumentuje Margaret Boden, twierdząc, że różnica pomiędzy życiem biologicznym a życiem sztucznym zasadza się na realizacji procesów metabolicznych przez to pierwsze (Boden, 2000). Jednak i takie postawienie sprawy może budzić wątpliwości, bowiem można wykorzystywać mikrobiologiczne ogniwa paliwowe do projektowania sztucznych systemów metabolicznych (Ieropoulos, Melhuish, & Greenman, 2003).

¹³ Problematyczne jest także odniesienie się do (syntaktycznej) informacji Shannona jako odpornej na trudny problem treści (Miłkowski, 2015).

¹⁴ Jako wyraźny przykład tej tendencji można podać pracę Chauncey'ego Mahera dotyczącą umysłów u roślin (Maher, 2017). Autor, aby uzasadnić hipotezę, która zakłada posiadanie przez rośliny prostych umysłów, konstruuje argument przeciwników tej hipotezy oraz odrzuca go, konstruując tym samym kontrargument wykorzystujący rozważania Hutto, Myina i zwolenników autopojezy. Argument prowadzący do odrzucenia życia mentalnego u roślin wygląda następująco (Maher, 2017: 106):

- posiadanie umysłu wymaga posiadania reprezentacji;
- rośliny nie reprezentują niczego; zatem
- rośliny nie mają umysłu.

Kontrargument Mahera (Maher, 2017: 120):

- aby mieć umysł, wystarczy być systemem autopojetycznym i adaptacyjnym;
- rośliny są takimi systemami; zatem
- rośliny mają umysł.

zowych. Tym można uzasadniać potrzebę uzupełnienia teorii Francisca Vareli i Humberta Maturany o FEP Fristona. Ten stan rzeczy może być powodem, dla którego, jak twierdzą niektórzy badacze (Miłkowski, 2018; Villalobos, 2013), ujęcie autopojetyczne nie zrewolucjonizowało nauk poznawczych ani nie dało mocnych podstaw eksperymentalnych, które pozwoliłyby w nowy sposób wyjaśnić wcześniej przebadane w innych podejściach zjawiska.

Ostatnim zarzutem, który można postawić temu ujęciu, jest pytanie o zawarty w nim *implicite* redukcjonizm metodologiczny. Zakłada on, że w miarę postępu badawczego prawa poszczególnych nauk da się zredukować do praw bardziej podstawowych. W tym przypadku prawa czy wyjaśnienia z zakresu nauk psychologicznych miałyby się sprowadzić do pojęć i wyjaśnień z zakresu teorii autopojetycznych czy mówiąc bardziej ogólnie — biologii. Po pierwsze nie jest jasne, dlaczego zjawiska badane przez psychologię lub neuronauki miałyby być wyjaśniane przez jedno autopojetyczne ujęcie; po drugie zaś samo to podejście, jak dotąd, nie może się pochwalić żadnymi eksperymentami oraz przyczynowymi wyjaśnieniami. Wciąż jest to program „w budowie” i jak dotąd nie doprowadził do żadnej rewolucji, którą wieścili niektórzy z jego zwolenników. Z tych powodów twierdzę, że entuzjazm Kirchhoffa, Froesego i innych jest przedwczesny.

ARGUMENTY NA RZECZ TEZY KOGNITYWISTYCZNEJ

Przypomnijmy: podejście kognitywistyczne w odniesieniu do problemu ciągłości zachodzącej pomiędzy życiem a umysłem opiera się na przekonaniu, według którego takiej ciągłości nie ma (Hohwy, 2015; Hohwy, 2017). Umysłu nie da się wyjaśnić tylko w kategoriach biologicznej samoorganizacji i autonomii, jak twierdzą niektórzy zwolennicy nonkognitywizmu i teorii autopojezy. Powodem tego ma być, przypomnijmy, przekonanie, że wyjaśnienie umysłu i procesów

Wywód ten opiera się na wielu definicjach projektujących („Posiadać umysł to posiadać reprezentację”; „Posiadać umysł to być systemem autopojetycznym i adaptacyjnym”). Wystarczy zanegować to, co znajduje się w definiensie, aby zauważyć jawną fałszywość tych wyrażań (przykładowo: „Posiadać umysł to nie posiadać reprezentacji” — zdanie to odniesione do wszystkich systemów poznawczych jest jawnie fałszywe, ponieważ ludzie oraz część zwierząt realizują swoje procesy poznawcze [przynajmniej ich część] na podstawie reprezentacji: pojęciowych, symbolicznych, wyobrażeniowych itd.). Można oczywiście bronić tego ujęcia, twierdząc, że wyznacza ono warunki minimalne, tzn. konieczne, ale niewystarczające do posiadania umysłu, niemniej wciąż problematyczny jest status tego rodzaju definicji oraz wyjaśnień budowanych na ich podstawie. Trudno byłoby bowiem bronić tezy, że są to definicje sprawozdawcze, chociażby z powodu ciągłych sporów wśród badaczy, dotyczących tego, jak w ogóle rozumieć życie. Maher zaznacza, że w swoich analizach opiera się na sformułowanej przez Thompsona tezie *minds in life*, która brzmi „wszystkie żyjące rzeczy (*things*) posiadają umysł” (Maher, 2007: 121). Teza ta jednak znowu jest zawołowaną definicją projektującą o postaci: „Żywe jest to, co posiada umysł”.

poznawczych wymaga odwołania się do jakiejś koncepcji informacji semantycznej. Jak wiemy, zdaniem Kirchhoffa i Froese go do wykazania przejścia pomiędzy życiem a prostą umysłowością wystarczy pojęcie informacji syntaktycznej. Istnieją jednak argumenty, które pozwalają podważyć przekonanie tych badaczy i podkreślają konieczność odwołania się do informacji semantycznej¹⁵.

ARGUMENT Z NIEWYSTARCZALNOŚCI WYJAŚNIEŃ OPARTYCH NA INFORMACJI SYNTAKTYCZNEJ

Kirchhoff i Froese bronią swojego stanowiska, wykorzystując tezę Godfrey-Smitha i Sterelno, że informację biologiczną należy wiązać z informacją Shannona (Godfrey-Smith & Sterelny, 2016). Charakterystyczne dla tych autorów jest „redukcyjne” podejście do analizy życia, które powiązałem z ujęciami monoatrybutywnymi. Ma to określone konsekwencje dla ich rozumienia informacji biologicznej. W artykule *Information in biology* (Godfrey-Smith, 2007) Godfrey-Smith stwierdza, że pojęcie informacji w biologii jest kluczowe dla zrozumienia bardzo wielu zjawisk i mechanizmów. Posługuje się nim genetyka, biologia rozwojowa czy teoria ewolucji. Charakterystyczna dla pojęcia informacji jest rola, która wiąże je z opisem relacji pomiędzy genami a różnymi procesami życiowymi. Zdaniem wielu biologów przyczynową rolę genów należy rozumieć w sensie „przenoszenia informacji”. Informację biologiczną można tutaj interpretować na dwa sposoby: albo minimalistycznie jako informację syntaktyczną, albo w sposób mocniejszy i bardziej kontrowersyjny, czyli jako informację semantyczną. Z informacją shannonowską ma się do czynienia zawsze, gdy zachodzi jakaś korelacja czy zdarzenie warunkowe. Można zatem powiedzieć, że geny zawierają informację o wytwarzanych przez siebie białkach albo że zawierają one informację o fenotypie całego organizmu. Niemniej nie mówi się w ten sposób więcej niż wtedy, kiedy się mówi, że dym zawiera informację o ogniu albo pierścienie drzewa zawierają informację o jego wieku. Pojawia się jednak pewna wątpliwość dotycząca tego, czy używając takiego minimalnego pojęcia informacji, jesteśmy w stanie odróżnić biologię od innych nauk, przykładowo geologii.

Część autorów podkreśla, że informacja Shannona jest zbyt uboga, aby mogła wyjaśnić wiele zjawisk badanych w biologii. Należy do nich między innymi problem potencjalnych błędów. Mianowicie: informacja Shannona nie pozwala wyjaśnić, w jaki sposób dana zmienna może nieść fałszywą informację o innej zmiennej. Niektórzy twierdzą (Maynard-Smith, 2000; Sterelny, Smith, & Dickison, 1996), że aby wyjaśnić możliwość błędnej reprezentacji, należy

¹⁵ Ograniczam się tutaj wyłącznie do tych dyskusji z zakresu filozofii biologii, które wiążą się z problemem wyjaśnienia zjawiska życia i odpowiedzią na pytanie o to, czy informacja syntaktyczna jest satysfakcjonująca eksplanacyjnie dla tego zagadnienia. Nie dyskutuję tutaj zatem sformułowanego przez Hutto i Myina (Hutto & Myin, 2013) trudnego problemu treści i jego rozwiązania.

wprowadzić naturalistyczne pojęcie funkcji jako efektu procesów naturalnej selekcji (Bielecka, 2018). Mówienie o funkcjach pozwala wyjaśnić nieprawidłowości w działaniu określonych mechanizmów czy struktur. Przykładowo: funkcją serca jest pompowanie krwi, a nie wydawanie efektów akustycznych. Dzieje się tak, ponieważ jest to funkcja, która doprowadziła do tego, że serce jest faworyzowane przez dobór naturalny (Wright, 1973). Bez pojęcia funkcji nie dałoby się wyjaśnić tego, że serce może nie działać. Do wyjaśnienia błędnej reprezentacji lub dysfunkcji potrzebny jest zatem jakiś słownik normatywny. Samo pojęcie informacji syntaktycznej nie pozwala jednak na jego sformułowanie. Część badaczy uważa, że podobna strategia „teleofunkcjonalna” może pomóc zrozumieć semantyczne własności genów, a także struktur biologicznych o właściwościach semantycznych (Maynard-Smith, 2000; Sterelny & Griffiths, 1999). Można oczywiście utrzymywać, że posiadanie danej funkcji wcale nie implikuje posiadania treści. Funkcja nóg do chodzenia wcale nie reprezentuje chodu (Godfrey-Smith, 2007: 108). Jednocześnie Sterelny, Smith i Dickison (Sterelny, Smith, & Dickison, 1996) twierdzą, że istnieje związek pomiędzy funkcjami ewolucyjnymi i semantycznymi w odniesieniu do niektórych struktur biologicznych: geny dzięki swoim funkcjom reprezentują wyniki, które mają wytwarzać. Dodają również, że inne, także niezależne, czynniki pełniące podobne funkcje rozwojowe mają również właściwości semantyczne¹⁶. Zatem nie tylko geny, ale także czynniki pozagenetyczne powinno się scharakteryzować w kategoriach informacji semantycznej.

Podsumowując powyższy wątek, należy stwierdzić, że trwa spór dotyczący tego, czy i w jaki sposób należy wykorzystać pojęcie informacji semantycznej w biologii. Badacze są zgodni co do tego, że pojęcie informacji Shannona nakłada pewne warunki minimalne dla zastosowania informacji w wyjaśnianiu zjawisk, niemniej to dopiero informacja semantyczna może pozwolić na znaczny postęp teoretyczny. Nie jest zatem jasne, dlaczego biolodzy czy filozofowie biologii mieliby rezygnować z pełniejszego wyjaśnienia badanych przez siebie zjawisk. Może być również tak, że część zjawisk da się wyjaśnić przy użyciu wyłącznie informacji Shannona, zaś wyjaśnienie innych zjawisk zakłada wykorzystanie informacji semantycznej. Przedwczesne jest jednak uznanie samej informacji syntaktycznej za w pełni satysfakcjonującą w odniesieniu do wymogu wyjaśnienia życia i zjawisk z nim powiązanych¹⁷.

¹⁶ „Informacja intencjonalna wydaje się lepszym kandydatem na zrozumienie sposobu, w jaki geny przenoszą informacje rozwojowe i nic innego nie robią” (Sterelny & Griffiths, 1999: 104, przekład własny).

¹⁷ Alistair Isaac (Isaac, 2019) powiada, że w informacji syntaktycznej są już *implicite* zawarte pewne intuicje dotyczące semantycznej wartości komunikatu. Shannon wprost twierdził, że zaproponowana przez niego teoria informacji nic nie mówi na temat znaczenia czy treści informacji, ponieważ nie jest to problem istotny z punktu widzenia inżyniera. Nie oznacza to jednak, że informacja w komunikacie nie ma treści.

ARGUMENT Z BŁĘDÓW POZNAWCZYCH

Artemy Kolchinsky oraz David Wolpert (Kolchinsky & Wolpert, 2018) twierdzą, że klasyczne shannonowskie pojęcie informacji syntaktycznej nie pozwala dobrze wyjaśnić tego, w jaki sposób i dlaczego otoczenie jest przyczynowo konieczne dla utrzymania istnienia danego systemu (żywego albo sztucznego), rozumianego tutaj jako jego dążenie do stanu niskiej entropii. Chodzi zatem o wyjaśnienie sposobu, w jaki system, unikając równowagi termodynamicznej, jest jednocześnie podatny na różne kontryfaktyczne interwencje ze strony środowiska. Kolchinsky i Wolpert twierdzą, że dobre wyjaśnienie korelacji system–środowisko jest możliwe, o ile skorzysta się z pojęcia informacji semantycznej. Tę ostatnią rozumieją jako taką informację, jaką system fizyczny musi mieć o swoim środowisku, aby utrzymać się przy istnieniu. Dodają przy tym, że należy jednak skonstruować teorię, która będzie odporna na zarzut głoszący, że informacja semantyczna ma znaczenie jedynie dla obserwatora czy badacza danego systemu, nie zaś dla tego systemu. Aby uniknąć tego zarzutu, należy mieć takie pojęcie informacji semantycznej, jakie jest oparte wyłącznie na wewnętrznej dynamice systemu, niezależnie od jego uwarunkowań ewolucyjnych oraz pochodzenia.

Punkt wyjścia analiz Kolchinsky’ego i Wolperta jest podobny do punktu wyjścia teorii autopojezy. Mianowicie: systemom autonomicznym (takim jak żywe organizmy, ale także niektóre roboty), które utrzymują swoje istnienie w obrębie danego środowiska, nie można przypisać funkcji samoorganizacji z zewnątrz i autonomii. Opis tych ostatnich jest niezależny zarówno od obserwatora, jak i historii ewolucyjnej. Zdaniem zwolenników autopojezy nie wymaga on korzystania z pojęć semantycznych. Wystarczy odnieść się do dynamicznego sprzężenia ze środowiskiem oraz bezpośredniej (pozbawionej treści) intencjonalności. Inaczej sprawę stawiają Kolchinsky i Wolpert. Twierdzą oni, że system autonomiczny, aby utrzymać się przy życiu, musi nieustannie aktualizować swoją wiedzę na temat otoczenia, czyli uzyskiwać określone, a nie dowolne, informacje na podstawie ciągłej obserwacji. W tym sensie prosta bakteria realizuje ten sam mechanizm przetwarzania informacji co makler giełdowy: bakteria chemotaktyczna wyczuwa stężenie chemiczne w swoim środowisku i następnie podąża w jego kierunku, tym samym lokalizując źródło pożywienia i energii. Podobnie makler: on również musi nieustannie zbierać określone informacje, aby szybko reagować na dynamiczne zmiany na giełdzie. Takie autonomiczne zachowanie odróżnia tego rodzaju bakterię od „pasywnych struktur” w stylu komórek Bénarda, które powstają jedynie w wyniku oddziaływań zewnętrznych (w tym przypadku podgrzewania danej cieczy od dołu).

To, co odróżnia zachowania systemów autonomicznych od pasywnych, to możliwość popełnienia błędu czy pomylenia się (Bickhard, 2014). Owa możliwość pojawia się zawsze wtedy, gdy system wykorzystuje informację, która

może „utrudnić” mu utrzymanie własnego istnienia. Oznacza to, że relacja semantyczna pomiędzy systemem a środowiskiem jest asymetryczna, w przeciwieństwie do informacji syntaktycznej, która jest symetryczna jako informacja wzajemna (Kirchhoff & Froese, 2017). Przykładowo zakłada się, że bakteria chemotaktyczna ma informacje o swoim środowisku, ale nie oczekuje się, iż środowisko naturalne będzie mieć jakieś informacje semantyczne o bakterii.

Należy zatem stwierdzić, że pełne wyjaśnienie samoorganizacji i autonomii systemów żywych oraz sztucznych wymaga zastosowania informacji semantycznej, która pozwala wykazać asymetryczność relacji zachodzącej pomiędzy tym systemem a jego środowiskiem. Utrzymanie przez system niskiej entropii zakłada zatem możliwość popętnienia lub uniknięcia błędu, co odpowiednio zwiększa lub zmniejsza stan równowagi termodynamicznej systemu z jego otoczeniem.

ARGUMENT Z KOMUNIKACJI

Nicholas Shea, Peter Godfrey-Smith oraz Rosa Cao (Shea, Godfrey-Smith, & Cao, 2018) zwracają uwagę na fakt, że wielu badaczy twierdzi, iż informacja syntaktyczna jest wystarczająca do wyjaśnienia interakcji i komunikacji prostych systemów¹⁸, a mimo tego domyślnie korzystają z koncepcji, która jest bogatsza niż ta zaproponowana przez Shannona. Tendencja owa jest szczególnie dostrzegalna w kontekstach, w których liczy się nieprawdziwość informacji. Wspomniani autorzy proponują taką teorię informacji, która w sposób bardziej adekwatny zda sprawę z wymiaru informacyjnego w tych prostych modelach. W tym celu wprowadzają pojęcie treści funkcjonalnej, które oddaje stopień, w jakim różne stany świata są zaangażowane w wymianę informacyjną pomiędzy nadawcami i odbiorcami. Dany komunikat ma treść funkcjonalną wtedy, gdy określone procesy selekcji utrzymują równowagę systemu. Oznacza to, że dla określonego sygnału istnieje takie zachowanie jego odbiorcy, jakie nie pojawia się w przypadku reakcji na inny sygnał. Oczywiście odbiorca może reagować w taki sam sposób również na inne sygnały, ale wykluczona jest możliwość, aby reagował on w ten sam sposób na wszystkie możliwe sygnały w danym środowisku (Shea, Godfrey-Smith, & Cao, 2018: 1015–1016). Shea i współautorzy powołują się tutaj na pogląd Briana Skyrmsa, który głosi, że zawartość informacyjna sygnału to wektor zmian dokonanych przez sygnał w rozkładzie prawdopodobieństwa po stanach świata (Skyrms, 2010). W tym ujęciu sygnały

¹⁸ Tendencję tę dobrze wyrażają Carl T. Bergstrom i Martin Rosvall, twierdząc: „Kiedy teoretycy informacji mówią o kodowaniu, to nie myślą oni o własnościach semantycznych. Wszelkie własności semantyczne są wpychane w książkę kodów, czyli interfejs pomiędzy strukturą źródłową i strukturą kanału, który dla teoretyków informacji jest tak samo interesujący, jak książka telefoniczna dla socjologów” (Bergstrom & Rosvall, 2011: 171, przekład własny).

nie tylko zawierają treści na temat stanów świata (w odniesieniu do dyspozycji nadawcy), ale także niosą informacje o dokonanych czynnościach odbiorcy (w odniesieniu do jego dyspozycji). Pogląd Skyrmsa, na co zwraca uwagę Godfrey-Smith, prowadzi do ważnego pytania dotyczącego tego, co odbiorca sygnału faktycznie otrzymuje od jego nadawcy. Na to pytanie można podać dwie możliwe odpowiedzi: (1) jeżeli odbiorca dowiaduje się, jaki świat prawdopodobnie jest, to jednocześnie może się dowiedzieć, jak zmieniły się prawdopodobieństwa; albo (2) jeżeli odbiorca nie dowie się na podstawie sygnału, jaki świat prawdopodobnie jest, to w reakcji na brak dobrej informacji może podjąć pewne działania (Godfrey-Smith, 2011). Można zatem stwierdzić, że analiza informacji dotyczy nie tylko prawdziwości treści, którą informacja ta przekazuje, ale także jej nieprawdziwości czy niepełności. Zdaniem Skyrmsa błędna informacja pojawi się wtedy, kiedy sygnał zmniejsza prawdopodobieństwo zajścia danego stanu rzeczy, albo wtedy, gdy to prawdopodobieństwo zwiększa.

Shea, Godfrey-Smith i Cao uważają, że pojęcie treści funkcjonalnej jest użytecznym narzędziem w analizie komunikacji w systemach sygnalizujących. Swoją tezę popierają kilkoma przykładami. Omówię najprostszy z nich (Shea, Godfrey-Smith, & Cao, 2018: 1018–1019). Wyobraźmy sobie, że istnieje system sygnalizacji, który opiera się na dwóch równie prawdopodobnych stanach rzeczy, dwóch prostych sygnałach i dwóch działaniach. Istnieją różne możliwe strategie nadawcze i odpowiadające im strategie odbiorcze. Są to kombinacje zachowań nadawcy i odbiorcy, w których obie strony uzyskują tyle samo w każdym procesie. Dzieje się tak, ponieważ sygnały są wykorzystywane do korelacji działań odbiorcy z określonym stanem świata. W tym prostym przypadku nadawca wysyła sygnał M1, odpowiadając na stan świata S1. Następnie odbiorca wytwarza sygnał A1, odpowiadając na informację o S1 zawartą w M1 itd. Pomiędzy nadawcą i odbiorcą sygnału jest utrzymana równowaga na podstawie określonego stanu świata. Treść funkcjonalną sygnału (M1) określa się poprzez zbadanie zachowania odbiorcy, które jest specyficzne dla tego, a nie innego sygnału (A1) w odniesieniu do określonego stanu rzeczy (S1), który jest warunkiem utrzymania stanu równowagi. Stabilizacja systemu jest oparta na wymianie informacyjnej pomiędzy nadawcą a odbiorcą w odniesieniu do określonych (probabilistycznych) stanów świata.

Powyższa analiza wzmacnia powszechne przekonanie głoszące, że mózg ludzki jest dobrym przykładem systemu sygnalizacyjnego, który ma własności zarówno semantyczne, jak i reprezentacyjne. Neurony lub całe ich populacje są nadawcami i odbiorcami informacji przenoszonych w potencjale czynnościowym. Cao (Cao, 2012) zauważa jednak, że dany system, w tym przypadku mózg, może przetwarzać faktyczną informację semantyczną, jeżeli spełni dwa wymagania: (1) odbiorca informacji musi być w stanie czerpać korzyści ze swojego otoczenia na podstawie sygnałów, które otrzymuje; oraz (2) musi mieć on pewną elastyczność wyboru reakcji w stosunku do odbieranego sygnału. Zdaniem Cao,

jeżeli odniesie się te wymogi do sygnalizacji na poziomie pojedynczego neuronu lub nawet całej ich populacji, to okazuje się, że w najlepszym wypadku mamy tutaj do czynienia jedynie z bardzo słabą sygnalizacją w sensie semantycznym, o ile w ogóle się ona pojawi. Raczej uaktywnia się bogata ilościowo informacja w sensie korelacji. Taki wniosek z pewnością spełnia oczekiwania i wspiera argumentację niektórych badaczy, na przykład Kirchhoffa i Froese'go czy Hutto i Myina. Niemniej, na co stanowczo zwraca uwagę Cao, prawdziwa reprezentacja świata i aktywność istniejącego w nim systemu opierają się dopiero na odbiorze bogatej w treść informacji, która angażuje cały organizm, do którego konkretna populacja neuronów wnosi bardzo niewielki wkład.

Omówione przeze mnie trzy argumenty pozwalają podać w wątpliwość przekonanie nonkognitywistów o tym, że sama informacja syntaktyczna w sposób adekwatny i niekontrowersyjny pozwala wyjaśnić ciągłość zachodzącą pomiędzy życiem a umysłem. Argumenty te wzmacniają wiele zarzutów, które sformułowaliśmy podczas omawiania propozycji Kirchhoffa i Froese'go.

ZAKOŃCZENIE

Zdaniem nonkognitywistów minimalizacja energii swobodnej w sensie fristonowskim odnosi się jedynie do informacji syntaktycznej. Informacja Shannona jest konieczna i wystarczająca do wyjaśnienia organizacji i funkcjonowania organizmów żywych oraz podstawowych form umysłowości. Założenie o eksplanacyjnej wystarczalności informacji syntaktycznej zakłada odrzucenie przekonania, zgodnie z którym informacja semantyczna oraz powiązane z nią treść semantyczna oraz reprezentacje pełnią jakąś istotną funkcję eksplanacyjną. Można to rozumieć na dwa sposoby: (1) albo wyjaśnienia oparte na pojęciu treści lub reprezentacji nie wnoszą nic nowego do rozumienia życia i poznania; (2) albo pojęcia takie jak treść oraz reprezentacja nie mają żadnych realnych desygnatów i mogą pełnić w najlepszym wypadku funkcje instrumentalne (Egan, 2014), co w efekcie może także prowadzić do przyjęcia podejścia (1). Ujęcia te można, za Chemero, wiązać kolejno z tak zwaną hipotezą dotyczącą wiedzy oraz hipotezą dotyczącą natury (Chemero, 2014)¹⁹. Dla potrzeb niniejszych analiz określe je jako „antyreprezentacjonistyczną hipotezę dotyczącą wiedzy” oraz „antyreprezentacjonistyczną hipotezę dotyczącą natury”.

¹⁹ Pierwsza z nich głosi, że „najlepsze modele i/lub wyjaśnienia naturalnych systemów poznawczych nie będą odwoływać się do jawnych reprezentacji. Tzn. nasze najlepsze modele i/lub wyjaśnienia systemów poznawczych nie będą zawierać nic, co stanowi reprezentacje zgodnie z naszą teleologiczną teorią reprezentacji”; druga zaś, że „naturalne systemy poznawcze nie korzystają z reprezentacji. Tzn. wewnątrz systemu poznawczego nie istnieje nic, co spełnia warunki określane przez teleologiczną teorię reprezentacji” (Chemero, 2014: 84). Rozszerzam w tym miejscu hipotezy sformułowane przez Chemero o wymiar biologicznego życia.

Hipoteza dotycząca natury ma charakter ontyczny, czyli orzeka coś na temat natury życia i (lub) umysłu, zaś hipoteza dotycząca wiedzy jest epistemiczna, ponieważ dotyczy sposobu, w jaki powinno się uprawiać naukę czy, w interesującym nas kontekście, kognitywistykę. Trudno powiedzieć, pod którą z hipotez podpisują się zwolennicy nonkognitywizmu. Ich deklaracje w tym aspekcie nie są jednoznaczne. Nie głoszą oni bowiem globalnego antyrepresentacjonizmu: wyższe funkcje poznawcze mogą w ich ujęciu mieć charakter reprezentacyjny. Kognitywistyczna alternatywa zaproponowana przez Jakoba Hohwy'ego odrzuca obie te hipotezy. Jego zdaniem bowiem wyjaśnienia reprezentacyjne pełnią istotną funkcję ekplanacyjną w kognitywistyce i neuronaukach, ponieważ systemy poznawcze są reprezentacyjne *per se* (Kiefer & Hohwy, 2018). Kognitywizm zatem nie tyle odrzuca obie te hipotezy, ile przeformułowuje je na modłę reprezentacyjną.

W tym miejscu dochodzimy do drugiej różnicy dzielącej zwolenników kognitywizmu i nonkognitywizmu, czyli różnicy w podejściu do natury wyjaśnień. Kirchhoff i Froese podkreślają (Kirchhoff & Froese, 2017), że ich stanowisko odróżnia od podejścia kognitywistycznego przekonanie, iż pomiędzy życiem a umysłowością istnieje bezpośrednie przejście. Sama ta teza nie jest do końca jasna, bowiem można ją rozumieć: albo jako (1) twierdzenie, że życie jest konstytutywne dla pojawienia się umysłu; albo jako (2) twierdzenie, że zjawiska życia i umysłowości można wyjaśnić, korzystając z tego samego modelu wyjaśnień. Niemniej wydaje się, że omawiany spór ostatecznie dotyczy nie tyle tych dwóch kwestii, ile różnic wynikających z przyjęcia innych założeń dotyczących hipotez o wiedzy i naturze. Z tego też powodu trudno jednoznacznie stwierdzić, na ile opcje kognitywistyczna i nonkognitywistyczna są zasadne. Twierdzą jednak, że biorąc pod uwagę przedstawioną przeze mnie analizę oraz argumenty, jakie można sformułować przeciwko tezie o ekplanacyjnej wystarczalności pojęcia informacji syntaktycznej, należy opowiedzieć się za stanowiskiem kognitywistycznym w kwestii natury relacji pomiędzy życiem i umysłem. Stanowisko to nie jest wolne od słabości (van Es & Myin, 2020), niemniej wydaje się, że wciąż oferuje ono wartościowe i nietrywialne wyjaśnienie wielu zjawisk poznawczych i biologicznych.

BIBLIOGRAFIA

- Allen, M. (2018). The foundation: Mechanism, prediction, and falsification in Bayesian enactivism. *Physics of Life Reviews*, 24, 17–20. <https://doi.org/10.1016/j.plrev.2018.01.007>.
- Allen, M. & Friston, K.J. (2018). From cognitivism to autopoiesis: Towards a computational framework for the embodied mind. *Synthese*, 195(6), 2459–2482. <https://doi.org/10.1007/s11229-016-1288-5>.
- Andrews, M. (2021). The math is not the territory: navigating the free energy principle. *Biology and Philosophy*, 36, 30. <https://doi.org/10.1007/s10539-021-09807-0>.

- Bergstrom, C.T. & Rosvall, M. (2011). The transmission sense of information. *Biology and Philosophy*, 26, 159–176.
- Bickhard, M.H. (2014). Model interaktywistyczny. (Przeł. M. Miłkowski). *Przegląd Filozoficzno-Literacki*, 2(39), 109–165.
- Bielecka, K. (2018). *Błądź, więc myślę. Co to jest błędna reprezentacja?*. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.
- Boden, M.A. (2000). Autopoiesis and life. *Cognitive Science Quarterly*, 1, 117–145.
- Brooks, R.A. (1991). Intelligence without representation. *Artificial Intelligence*, 47, 139–159.
- Bruineberg, J., Dołęga, K., Dewhurst, J., & Baltieri, M. (2020). *The emperor's new Markov blankets*. [Preprint].
- Cao, R. (2012). A teleosemantic approach to information in the brain. *Biology and Philosophy*, 27, 49–71. <https://doi.org/10.1007/s10539-011-9292-0>.
- Caticha, A. (2015). The basics of information geometry. *AIP Conference Proceedings*, 1641, 15–26.
- Chemero, A. (2014). Antyrepresentacjonizm i nastawienie dynamiczne. (Przeł. P. Gładziejewski). *Przegląd Filozoficzno-Literacki*, 2(39), 79–107.
- Clark, A. (2017). How to knit your own Markov Blanket: Resisting the second law with metamorphic minds (s. 1–19). W: T. Metzinger & W. Wiese (Red.). *Philosophy and predictive processing*. Frankfurt am Main: MIND Group. <https://doi.org/10.15502/9783958573031>.
- Di Paolo, E. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and Cognitive Sciences*, 4(4), 429–452.
- Dretske, F. (2004). *Naturalizowanie umysłu*. (Przeł. B. Świątczak). Warszawa: Wydawnictwo IFIS PAN.
- Egan, F. (2014). How to think about mental content. *Philosophical Studies*, 170, 115–135. <https://doi.org/10.1007/s11098-013-0172-0>.
- Feynman, R. (1980). *Wykłady z mechaniki statystycznej*. (Przeł. K. Makowski). Warszawa: Państwowe Wydawnictwo Naukowe.
- Friston, K.J. (2009). The free-energy principle: A rough guide to the brain?. *Trends in Cognitive Sciences*, 13(7), 293–301.
- Friston, K.J. (2010). The free-energy principle: A unified brain theory?. *Nature Reviews Neuroscience*, 11, 127–138.
- Friston, K.J. (2013). Life as we know it. *Journal of The Royal Society Interface*, 10, [b.p.]. <https://doi.org/10.1098/rsif.2013.0475>.
- Friston, K. J. (2019). *A free energy principle for a particular physics*. [Preprint]. arXiv:1906.10184v1 (dostęp 03.05.2021).
- Friston, K.J., Adams, R.A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, 3, 1–20.
- Friston, K.J. & Stephan, K.E. (2007). Free energy and the brain. *Synthese*, 159, 417–458.
- Friston, K.J., Wiese, W., & Hobson, J.A. (2020). Sentience and the origins of consciousness: From cartesian duality to Markovian monism. *Entropy*, 22, 516–516. <https://doi.org/10.3390/E22050516>.
- Froese, T. (2014). Review of radicalizing enactivism: Basic minds without content by D.D. Hutto, E. Myin. *The Journal of Mind and Behavior*, 35, 1–2.
- Godfrey-Smith, P. (2007). Information in biology (s. 103–119). W: D. Hull, M. Ruse (Red.). *The Cambridge companion to the philosophy of biology*. Cambridge: Cambridge University Press.
- Godfrey-Smith, P. (2011). Review of signals, by Brian Skyrms. *Mind*, 120, 1288–1297.
- Godfrey-Smith, P. & Sterelny, K. (2016). Biological information. W: E.N. Zalta (Red.). *The Stanford encyclopedia of philosophy* (Summer 2016 Edition). Dostęp: <https://plato.stanford.edu/archives/sum2016/entries/information-biological/> (10.03.2021).

- Heller, M., Lubański, M., & Ślaga, S. (1997). *Zagadnienia filozoficzne współczesnej nauki*. Warszawa: Wydawnictwo ATK.
- Hohwy, J. (2014). The self-evidencing brain. *Noûs*, 50(2), 259–285.
- Hohwy, J. (2015). The neural organ explains the mind (s. 1–22). W: T. Metzinger & J.M. Windt (Red.). *Open MIND*. Frankfurt am Main: MIND Group. <https://doi.org/10.15502/9783958570016>.
- Hohwy, J. (2017). How to entrain your evil demon (s. 1–15). W: T. Metzinger & W. Wiese (Red.). *Philosophy and predictive processing*. Frankfurt am Main: MIND Group. <https://doi.org/10.15502/9783958573048>.
- Hutto, D. (2006). Against passive intellectualism: Reply to Crane (s. 121–149). W: R. Menary (Red.). *Radical enactivism*. London: John Benjamins Publishing Company.
- Hutto, D. & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content*. Cambridge: MIT Press.
- Ieropoulos, I., Melhuish, C., & Greenman, J. (2003). Artificial metabolism: Towards true energetic autonomy in artificial life (s. 792–799). W: W. Banzhaf, I. Ziegler, T. Christaller, P. Dittrich, & J.T. Kim (Red.). *Advances in artificial life. ECAL 2003. Lecture notes in computer science* (t. 2801). Berlin–Heidelberg: Springer.
- Isaac, A. (2019). The semantics latent in Shannon information. *The British Journal for the Philosophy of Science*, 1(70), 103–125. <https://doi.org/10.1093/bjps/axx029>.
- Kiefer, A. & Hohwy, J. (2018). Content and misrepresentation in hierarchical generative models. *Synthese*, 195, 2387–2415.
- Kirchhoff, M. (2018). Autopoiesis, free energy, and the life–mind continuity thesis. *Synthese*, 195, 2519–2540.
- Kirchhoff, M. & Froese, T. (2017). Where there is life there is mind: In support of a strong life–mind continuity thesis. *Entropy*, 19(4), 1–18. <https://doi.org/10.3390/e19040169>.
- Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: Autonomy, active inference and the free energy principle. *Journal of the Royal Society Interface*, 15, 1–11. <https://doi.org/10.1098/rsif.2017.0792>.
- Kolchinsky, A. & Wolpert, D.H. (2018). Semantic information, autonomous agency and non-equilibrium statistical physics. *Interface Focus*. [Preprint]. <https://doi.org/10.1098/rsfs.2018.0041>. arXiv:1806.08053v3.
- Korbak, T. (2021). Computational enactivism under the free energy principle. *Synthese*, 198, 2743–2763. DOI: 10.1007/s11229-019-02243-4.
- Latawiec, A. (1992). Wokół pojęcia zjawiska biologicznego. *Studia Philosophiae Christianae*, 2(28), 241–254.
- Maher, Ch. (2017). *Plant minds: A philosophical defense*. Oxford: Routledge.
- Maynard-Smith, J. (2000). The concept of information in biology. *Philosophy of Science*, 67(2), 689–699. <https://doi.org/10.1086/392768>.
- Mayr, E. (2002). *To jest biologia*. (Przeł. J. Szacki). Warszawa: Prószyński i S-ka.
- McGregor, S. (2017). The Bayesian stance: Equations for ‘as-if’ sensorimotor agency. *Adaptive Behavior*, 2(25), 72–82. <https://doi.org/10.1177/1059712317700501>.
- Millikan, R.G. (1984). *Language, thought, and other biological categories: New foundations for realism*. Cambridge: MIT Press.
- Miłkowski, M. (2015). The hard problem of content: Solved (long ago). *Studies in Logic, Grammar and Rhetoric*, 41(1), 73–88. <https://doi.org/10.1515/slgr-2015-0021>.
- Miłkowski, M. (2018). Autopoiesis nie wywołała rewolucji (s. 217–227). W: G. Króliczak, K. Łastowski, Ł. Przybylski, P. Przybysz, & M. Urbański (Red.). *Filozof w krainie umysłów. Profesorowi Andrzejowi Klawiterowi w darze*. Poznań: Wydawnictwo Naukowe Wydziału Nauk Społecznych UAM.

- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco: Morgan Kaufmann Publishers.
- Piekarski, M. (2020). *Mechanizmy predykcyjne i ich normatywność*. Warszawa: Liberi Libri.
- Poczobut, R. (2011). Od rozszerzonego życia do rozszerzonego umysłu (poznania). *Przegląd Filozoficzno-Literacki*, 2–3(31), 217–238.
- Ramsey, W.M. (2007). *Representation reconsidered*. Cambridge: Cambridge University Press.
- Ramstead, M.J.D., Badcock, P.B., & Friston, K.J. (2017). Answering Schrödinger's question: A free-energy formulation. *Physics of Life Reviews*, 24, 1–16. <https://doi.org/10.1016/j.plrev.2017.09.001>.
- Shannon, E.A. (1948). Mathematical theory of communication. *Bell Labs Technical Journal*, 27, 379–423.
- Shea, H., Godfrey-Smith, P., & Cao, R. (2018). Content in simple signaling systems. *British Journal for the Philosophy of Science*, 69(4), 1009–1035. doi: 10.1093/bjps/axw036.
- Skyrms, B. (2010). *Signals: Evolution, learning, & information*. Oxford–New York: Oxford University Press.
- Sterelny, K. & Griffiths, P.E. (1999). *Sex and death: An introduction to the philosophy of biology*. Chicago: University of Chicago Press.
- Sterelny, K., Smith, K.C., & Dickison, M. (1996). The extended replicator. *Biology and Philosophy*, 11(3), 377–403. <https://doi.org/10.1007/BF00128788>.
- Thompson, E. (2011). Précis of mind in life. *Journal of Consciousness Studies*, 18, 10–22.
- Wiese, W. & Friston, K.J. (2021). Examining the continuity between life and mind: Is there a continuity between autopoietic intentionality and representationality?. *Philosophies*, 6, 1–17. <https://doi.org/10.3390/philosophies6010018>.
- Wright, L. (1973). Functions. *The Philosophical Review*, 2(82), 139–168.
- van Es, T. & Myin, E. (2020) Predictive processing and representation: How less can be more (s. 7-24). W: D. Mendonça, M. Curado, & S.S. Gouveia (Red.). *The philosophy and science of predictive processing*. London: Bloomsbury.
- Villalobos, M. (2013). Enactive cognitive science: Revisionism or revolution?. *Adaptive Behavior*, 3(21), 159–167. <https://doi.org/10.1177/1059712313482953>.
- Villalobos, M. & Dewhurst, J. (2018). Enactive autonomy in computational systems. *Synthese*, 195(5), 1891–1908. <https://doi.org/10.1007/s11229-017-1386-z>.

