Paula Thagarda koncepcja rozumowania abdukcyjnego*

Mariusz Urbański
Uniwersytet im. Adama Mickiewicza w Poznaniu

Streszczenie. *Na przykładzie koncepcji rozumowania abdukcyjnego, autorstwa Paula Thagarda, w artykule niniejszym omawia się sposób pojmowania abdukcji charakterystyczny dla eksplanacyjno-koherencyjnego modelu tego typu rozumowań, który w chwili obecnej oferuje najbardziej satysfakcjonujące połączenie psychologicznej adekwatności oraz efektywności obliczeniowej definiowanych w jego ramach procedur generowania i oceny hipotez abdukcyjnych.*

Słowa kluczowe: *rozumowania abdukcyjne, eksplanacyjno-koherencyjny model abdukcji, teoria koherencji eksplanacyjnej*

1. Abdukcja

Prawdopodobnie najbardziej zwięzły opis struktury rozumowania abdukcyjnego zaproponowany został przez Charlesa Sandersa Peirce'a (1931–1958, 5.189)[1]:

Obserwujemy zaskakujące zjawisko C.
Gdyby A było prawdziwe, zachodzenie C byłoby oczywistością.
Mamy zatem podstawy, by podejrzewać, że A jest prawdziwe.

W myśl tej charakterystyki rozumowanie abdukcyjne prowadzi do wniosków, za pomocą których próbujemy nadawać sens zadziwiającym zjawiskom, osiągać zrozumienie zaskakujących wydarzeń bądź tłumaczyć sobie budzące zaciekawienie informacje. Abdukcja, jako jedna z form rozumowania uprawdopodabniającego, jest rozumowaniem zawodnym: ponieważ wniosek dostarcza więcej informacji, niż zawierają

*Artykuł niniejszy jest zmodyfikowaną wersją podrozdziału 4.2 książki „Rozumowania abdukcyjne" (Urbański 2009a) o tym samym tytule.

[1] Jeśli w bibliografii nie zaznaczono inaczej, tłumaczenia cytatów pochodzą od autora.

Studia z Kognitywistyki i Filozofii Umysłu, **6(1)**: 97–120, 2012

przesłanki, prawdziwość przesłanek rozumowania nie gwarantuje prawdziwości abdukcyjnego wniosku. W ujęciu Peirce'a abdukcja ma charakter twórczy: „wnioskowanie hipotetyczne [czyli abdukcyjne] postuluje istnienie czegoś odmiennego od dotychczas bezpośrednio obserwowanego, i to często czegoś, czego nie można by zaobserwować bezpośrednio" (Peirce 1931–1958, 2.640). Mimo że jego natura jest dość tajemnicza (Peirce mówi nawet o „instynkcie" (Peirce 1931–1958, 5.172) i „tajemniczej mocy zgadywania" (Peirce 1931–1958, 6.530), to jednak nie ma charakteru irracjonalnej iluminacji:

> [...] choć w niewielkim stopniu ograniczana logicznymi regułami, jest jednak [abdukcja] logicznym wnioskowaniem, w którym wniosek stwierdzany jest co prawda jedynie problematycznie bądź w trybie przypuszczającym, tym niemniej posiadającym dobrze określoną logiczną strukturę (Peirce 1931–1958, 5.188).

W literaturze przedmiotu znaleźć można kilka różnych i nierównoważnych koncepcji, dotyczących logicznej struktury abdukcji. Podzielić je można na trzy grupy, kryterium podziału konstruując na podstawie odpowiedzi na dwa pytania: po pierwsze, czy przed hipotezą abdukcyjną stawia się zadanie wyjaśniania zjawisk, którym za pomocą abdukcji próbujemy „nadawać sens", po drugie zaś, czy pomiędzy hipotezą a owymi zjawiskami (czy raczej, ściślej rzecz ujmując, reprezentacjami zjawisk) powinny zachodzić związki wynikania logicznego. Różne kombinacje odpowiedzi na te pytania prowadzą do wyróżnienia trzech modeli abdukcji: modelu eksplanacyjno-dedukcyjnego, eksplanacyjno-koherencyjnego i apagogicznego (Urbański 2009a, 2009b).

Pojmowanie abdukcji w duchu modelu eksplanacyjno-dedukcyjnego jest charakterystyczne dla badań nad abdukcją prowadzonych z perspektywy, którą moglibyśmy nazwać obliczeniową. Ich głównym celem jest zdefiniowanie efektywnej procedury generowania możliwie najlepszych hipotez abdukcyjnych. Pierwszorzędne znaczenie ma tu więc uchwycenie precyzyjnie definiowalnego związku między hipotezą a zjawiskiem, którego ma dotyczyć (czy raczej znów: jego reprezentacją). Taki charakter ma związek dedukcyjny: w modelu eksplanacyjno-dedukcyjnym H traktuje się jako hipotezę abdukcyjną dla A na gruncie teorii X gdy spełnione sa dwa warunki:

1. A nie wynika (logicznie) z X oraz
2. A wynika (logicznie) z X i H łącznie.

Zakłada się przy tym zwykle, że zarówno A jak i H są zdaniami.

Eksplanacyjny charakter abdukcji jest tutaj w pewnym sensie efektem ubocznym, który swe źródło ma w spostrzeżeniu, że najbardziej oczywistym sposobem reprezentowania oczywistego w Peirce'owskim

schemacie wymogu związków treściowych między hipotezą abdukcyjną a zjawiskiem jest w tym kontekście postulowanie, by związek między nimi opisywać za pomocą dedukcyjno-nomologicznego modelu wyjaśniania. Trzeba jednak pamiętać, że aby sensownie posługiwać się pojęciem wyjaśniania dedukcyjno-nomologicznego w konstrukcji stosownej procedury abdukcyjnej należałoby posłużyć się odpowiednio bogatym językiem. Procedury abdukcyjne, definiowane w ramach modelu eksplanacyjno-dedukcyjnego, często tego warunku nie spełniają (por. Aliseda 2006), albo spełniają tylko pozornie (por. Bylander i in. 1995, Josephson i Josephson 1994).

W koncepcjach abdukcji, tworzonych w ramach modelu eksplanacyjno-koherencyjnego (Leake 1995, 1993; Thagard 2007a), zwraca się uwagę na istotne słabości modelu eksplanacyjno-dedukcyjnego, podkreślając jednocześnie, że głównym zadaniem abdukcji jest wyjaśnianie zaskakujących zjawisk. Zasadnicze różnice między dwoma modelami eksplanacyjnymi można zwięźle ująć w kilku punktach (Urbański 2009a, s. 60–67):

1. Co prawda traktowanie zdań jako wyłącznych składników rozumowań (przesłanek i wniosku) jest wygodne, ale nietrafne. Po pierwsze, posługujemy się często rozumowaniami, w których rolę przesłanek i wniosku mogą odgrywać inne wyrażenia (np. pytania; por. Wiśniewski 2011). Po drugie, składnikami rozumowań mogą być także reprezentacje niewerbalne (Nelsen 1997, Magnani 2009).
2. Wyjaśnianie dedukcyjno-nomologiczne traktowane jest jako tylko jednym z możliwych związków eksplanacyjnych między hipotezą a wyjaśnianym zjawiskiem i to związkiem wcale nie najbardziej typowym.
3. W adekwatnym modelu rozumowań, których celem jest wyjaśnianie zaskakujących zjawisk, nie sposób pominąć pytań o to, co decyduje, że postanawiamy szukać wyjaśnień tych a nie innych zjawisk i kiedy w ogóle warto uruchamiać procedurę poszukiwań hipotez abdukcyjnych. Innymi słowy, termin „zaskakujące" z Peirce'owskiego schematu abdukcji zasługuje na poważne potraktowanie.
4. Model eksplanacyjno-dedukcyjny nie pozwala na uchwycenie istotnego elementu przynajmniej niektórych rozumowań abdukcyjnych, a mianowicie ich kreatywności.
5. W modelu eksplanacyjno-dedukcyjnym postuluje się takie kryteria oceny hipotez abdukcyjnych, które są nierealistyczne z punktu

widzenia rzeczywistych rozumowań (niesprzeczność, teoriomnogościowo rozumiana prostota, minimalność logiczna).

Niezależnie od różnic, jakie dzielą eksplanacyjno-dedukcyjny i eksplanacyjno-koherencyjny model abdukcji, ich wspólny eksplanacyjny mianownik pozwala uznać, że w obu tych modelach abdukcja utożsamiana jest z jakąś formą (mniej lub bardziej wyrafinowaną) wnioskowania do najlepszego wyjaśnienia (IBE, *Inference to the Best Explanation* (Harman 1965, Lipton 1991).

Z kolei model apagogiczny, podobnie jak model eksplanacyjno-dedukcyjny, wyrasta z dążenia do konstrukcji efektywnej obliczeniowo procedury abdukcyjnej i, podobnie jak model eksplanacyjno-koherencyjny, zmierza w kierunku możliwie adekwatnego opisu takich rozumowań. Zwraca się tutaj jednak uwagę, że związek między abdukcją a wyjaśnianiem nie jest tak ścisły, jak głoszą oba modele eksplanacyjne (Hintikka 2007): hipotezy abdukcyjne mogą służyć celom eksplanacyjnym równie dobrze co predykcyjnym, mogą mieć na celu unifikację pozornie niepowiązanych praw, eliminację hipotez konkurencyjnych itp. Ponadto, nawiązując do Arystotelesowskiego pojęcia redukcji prostej, podkreśla się tutaj, że dla charakterystyki rozumowania abdukcyjnego, bardziej istotne są związki epistemiczne między przesłankami a wnioskiem niż związki prawdziwościowe. Tak jak dedukcja jest wnioskowaniem zachowującym prawdziwość, tak abdukcja jest wnioskowaniem zachowującym niewiedzę: wnioski rozumowań abdukcyjnych, niezależnie od ich wartości logicznej, nie zaspokajają epistemicznych potrzeb rozumującego podmiotu, są zawsze prowizoryczne i tymczasowe (Gabbay i Wodds 2005).

Spośród trzech wymienionych modeli abdukcji póki co tylko model eksplanacyjno-koherencyjny oferuje satysfakcjonujące połączenie psychologicznej adekwatności i (potencjalnej) efektywności obliczeniowej definiowanych w jego ramach procedur abdukcyjnych. Przyjrzymy się bliżej najbardziej dla tego modelu charakterystycznej koncepcji abdukcji, zaproponowanej przez Paula Thagarda. Interesować nas będą szczególnie te jej aspekty, które związane są ze sposobem pojmowania natury rozumowań, z zakładanym modelem wyjaśniania oraz z kryteriami oceny hipotez abdukcyjnych.

Wedle Thagarda abdukcja jest rozumowaniem, na które składa się pięć etapów (Thagard 2007a, Thagard i Litt 2008). Abdukcja rozpoczyna się od emocjonalnej reakcji zdziwienia (etap pierwszy), wywołanej identyfikacją zjawiska zarazem zaskakującego i interesującego na tyle, by stać się celem wyjaśniania, którego zadaniem jest maksymalizowanie koherencji zbioru reprezentacji mentalnych. Jeśli poszukiwanie hipotez

wyjaśniających (etap drugi) wśród hipotez uprzednio stosowanych lub w każdym razie znanych kończy się niepowodzeniem, próbujemy konstruować nowe hipotezy (etap trzeci), a następnie poddajemy je porównawczej ocenie (etap czwarty), bazującej na, powiązanych z oceną koherencji, kryteriach konsiliencji i podobieństwa, a także prostoty. Przy czym granice między generowaniem a oceną hipotez nie są ostre, procesy te przenikają się nawzajem i mogą przebiegać równolegle. Wreszcie, akceptacja wybranej hipotezy powiązana jest z pojawieniem się poznawczego usatysfakcjonowania[2] (etap piąty): proces abdukcji tak jak zaczyna się, tak i kończy reakcją emocjonalną[3].

2. Rozumowanie

Według Thagarda (1988, s. 51–53; 1992, s. 53–54) abdukcja to rozumowanie, za pomocą którego nadajemy sens zjawiskom zaskakującym. Takie pojmowanie roli abdukcji nie odbiega w gruncie rzeczy od funkcji przypisywanej tego typu rozumowaniom na gruncie modelu eksplanacyjno-dedukcyjnego. Zasadnicza różnica pomiędzy Thagarda modelem abdukcji a modelem dedukcyjnym leży gdzie indziej, a mianowicie w odpowiedzi na pytanie następujące: czym są składniki rozumowań?

W poglądach na naturę rozumowań możemy wyróżnić dwa zasadnicze stanowiska (Adler i Rips 2008). Wedle pierwszego z nich rozumowania mają charakter językowy, a w każdym razie są reprezentowalne za pomocą (mniej lub bardziej skomplikowanych) struktur językowych (Rips 1994). Wedle drugiego, rozumowania są modelami, reprezentującymi stany rzeczy (Johnson-Laird 1983). W pierwszym przypadku reguły przekształcania reprezentacji, będących elementami rozumowań, to reguły wnioskowania, charakteryzujące (syntaktyczne lub semantyczne) związki między sądami (bądź zdaniami): przesłankami a wnioskiem. W drugim przypadku są to reguły przekształcania i integrowania modeli[4]. W modelu eksplanacyjno-dedukcyjnym konsekwentnie preferuje się ten pierwszy sposób pojmowania rozumowań. W modelu eksplanacyjno-koherencyjnym bywa różnie, ale akurat Thagard w swojej koncepcji rozumowań rezygnuje z pojęcia sądu na rzecz „biologicznie realistycznych pojęć, związanych ze strukturami neuronalnymi"; rozu-

[2] Co nie znaczy, rzecz jasna, że każde rozumowanie abdukcyjne dociera do etapu poznawczego usatysfakcjonowania.

[3] Thagard (2000, s. 193–194) nazywa takie emocje *metakoherencyjnymi*.

[4] Nb. argumenty na rzecz każdego z tych stanowisk bazują przede wszystkim na badaniach o charakterze behawioralnym, podczas gdy wyniki badań neuroobrazowych zdają się sugerować możliwą perspektywę ich integracji (Goel 2007, Reverberi i in. 2007).

mowania są, w jego ujęciu, przekształceniami struktur neuronalnych (a rozumowania wyrażane za pomocą zdań są jedynie szczególnym przypadkiem takich struktur; Thagard 2007a, s. 19)[5], są procesami „w znacznej mierze nieświadomymi, w których wiele informacji, przetwarzanych równolegle, wiązanych jest w spójną całość" (Thagard 2000, s. 3).

Takie podejście daje liczne korzyści (Thagard 2007a, s. 18–20), których zamierzonym wspólnym mianownikiem są, jak wolno sądzić, prostota i realizm. Po pierwsze, nie musimy zakładać, że składników rozumowań jest nieskończenie wiele i że są wśród nich takie, które nigdy nie zostaną przez nikogo pomyślane bądź wypowiedziane (jak to ma miejsce w przypadku sądów bądź zdań). Nie musimy zatem badać abstrakcyjnych obiektów, których sposób istnienia jest niezwykle tajemniczy. Po drugie, nie musimy kłopotać się równie tajemniczą relacją między abstrakcyjnymi sądami a sądzącymi podmiotami, istotami z krwi i kości. Po trzecie, unikamy raf i mielizn rozmaitych paradoksów, w które obfituje filozoficzna analiza nastawień sądzeniowych[6]. Nie znaczy to oczywiście, że wskutek takiego posłużenia się brzytwą Ockhama nastawienia sądzeniowe w ogóle znikają z pola zainteresowań analizy rozumowań, ale że ulubione stany mentalne filozofów (przekonanie, wątpienie itp.) należy traktować jako struktury i procesy neuronalne. Po czwarte, za składniki rozumowań możemy więc uznać niewerbalne reprezentacje, pochodzące z rozmaitych modalności percepcyjnych i, po piąte, potraktować rozumowania jako procesy holistyczne, w których informacje zawarte w przesłankach czerpane są z wielu źródeł naraz (por. Thagard 1992).

Neuronalna koncepcja rozumowań nie jest rzecz jasna wolna od problemów. Czy teoria, wiążąca rozumowania ze specyficznymi strukturami – mózgami o takiej a nie innej budowie – nie jest zbyt wąska[7]?

[5]Pogląd ten wpisuje się w głoszony przez Thagarda (2000), *kognitywny naturalizm*.

[6]Jeśli Lewis Carroll i Charles Dodgson to jeden i ten sam człowiek, to jak mogę być przekonany, że Carroll napisał „Alicję w Krainie Czarów", nie będąc jednocześnie przekonanym, że uczynił to Dodgson, skoro przekonania te najwyraźniej mają tę samą treść? (por. Hintikka 1992, s. 305–308; Anderson i Owens 1990).

[7]Przypomnijmy znany cytat z A. M. Turinga:

> „Najważniejszą sprawą jest, aby spróbować wytyczyć linię oddzielającą właściwości mózgu człowieka, o których chcemy dyskutować od tych, które nas nie interesują. Weźmy skrajny przypadek, nie interesuje mnie to, że mózg ma konsystencję zimnej owsianki. Nie powiemy przecież: «Ta maszyna jest całkiem twarda, czyli nie jest mózgiem, a więc nie może

Jakie grupy neuronów biorą udział w procesach, będących składnikami rozumowań? Odpowiedź Thagarda, udzielana co prawda raczej *implicite*, jest prosta: nawet jeśli jest to teoria skrojona tylko na miarę mózgów ziemskich organicznych istot rozumnych, to ma przynajmniej tę zaletę, że dokładnie określa, co powinno stać się przedmiotem badania i że przedmiot ten jest namacalny. Tego ostatniego nie musimy co prawda traktować jako wartości, dla której nie zaszkodzi poświęcić ogólności analizy. Trzeba jednak przyznać, że jest to perspektywa dość atrakcyjna z praktycznego punktu widzenia, jeśli chcemy poszukiwać odpowiedzi na pytania jak przebiegają rzeczywiste procesy rozumowań u ludzi, jakim podlegają ograniczeniom, a także skąd się biorą i na czym polegają błędy w rozumowaniach.

Tym sposobem np. przekształcanie reprezentacji wizualnych w geometrycznych „dowodach bez słów" (Nelsen 1997) możemy potraktować nie jako postrzeżeniową bazę, ale jako rozumowanie, w którym generujemy hipotezy wyjaśniające jakie przekształcenia łączą poszczególne obrazy. Reprezentacje nie są myślowym półproduktem, przeznacznym do dalszej obróbki. Pełnią takie same role, jak zdania w rozumowaniach werbalizowanych: są przesłankami i wnioskami, uznawanymi bądź tylko pomyślanymi. Wolno zapewne przyjąć, że w takim razie relacje między nimi mają taki sam charakter, jak relacje między zdaniami i że moglibyśmy mówić co najmniej o związkach wynikania między takimi reprezentacjami, a tym samym o ich wartościach logicznych. Łatwo zauważyć, że wkroczylibyśmy w ten sposób na dość śliski grunt: aparat pojęciowy logiki formalnej wcale nie musi szczególnie dobrze nadawać się do analizy związków między reprezentacjami bazującymi na modalnościach zmysłowych. Nawet jeśli nie pójdziemy dalej w tym kierunku, pozostaje do rozwiązania istotny problem. Otóż eksplanacyjny charakter abdukcji wymaga wprowadzenia w tym kontekście neuronalnie ugruntowanego pojęcia wyjaśniania.

3. Wyjaśnianie

Thagard (1992, s. 118–130) wyróżnia sześć sposobów interpretowania pojęcia wyjaśnienia na gruncie filozofii, nauk kognitywnych i sztucznej inteligencji. Wyjaśnianie jego zdaniem bywa interpretowane jako: rozumowanie dedukcyjne, zależność statystyczna, zastosowanie schematu, porównanie bazujące na analogii, relacja przyczynowa, akt językowy.

myśleć»." (Newman i in., s. 3–4; tłum. P. Łupkowski; por. Łupkowski 2010).

Wspólnym mianownikiem tych sześciu sposobów interpretowania pojęcia wyjaśnienia jest, wedle Thagarda, przyczynowość: hipoteza wyjaśnia eksplanandum, jeśli jest w stanie przedstawić je jako ostatni element ciągu przyczynowo-skutkowego. Odwołanie do związków przyczynowo-skutkowych ma też być pomocne w neutralizowaniu znanych paradoksów wyjaśniania (Brody 1972, Hausman 1982).

Rzecz jasna natychmiast powstaje pytanie, co musimy wiedzieć o związkach przyczynowo-skutkowych, by wyjaśnianie przyczynowe[8] móc uczynić fundamentem abdukcji; co najmniej od czasów Hume'a wiadomo, że nie jest to pytanie łatwe. Jak odróżniać związki, polegające jedynie na współwystępowaniu zjawisk od związków rzeczywiście przyczynowych? Jak od związków rzeczywiście przyczynowych odróżniać statystyczne zależności między zdarzeniami?

Istotne problemy są dwa (Bunge 1979, s. xviii–xix): ontologiczny (czym jest przyczynowość?) i metodologiczny (na podstawie jakich kryteriów rozpoznajemy związki przyczynowe?). Problem ontologiczny Thagarda nie interesuje, natomiast jego rozwiązanie problemu metodologicznego jest dość szczególnej natury. Otóż, powiada Thagard, ludzie mają intuicyjną zdolność (zmysł?) rozróżniania między związkami przyczynowymi a zależnościami czysto statystycznymi. Jej źródłem ma być po części rozumienie mechanizmów powiązań przyczynowych, po części – co bardziej istotne – naturalna inklinacja do traktowania pewnych zjawisk jako przyczynowo powiązanych, inklinacja bazująca na podstawowych własnościach naszego systemu percepcyjnego (Thagard 2007a, s. 15). śr Na poparcie tej tezy Thagard przytacza argumenty, pochodzące z badań kognitywnych, neurologicznych i z zakresu psychologii rozwojowej (por. Hilton 2007). Płyną one z eksperymentów, sugerujących że zarówno kilkumiesięczne dzieci (Leslie i Keeble 1987, Baillargeon i in. 1995, Mandler 2004) jak i osoby dorosłe (Michotte 1963, Kadaba i in. 2007) traktują pewne zjawiska jako przyczynowo powiązane oraz że za identyfikowanie związków przyczynowych odpowiedzialne są określone obszary mózgu (prawy środkowy zakręt czołowy i prawy dolny płacik ciemieniowy (Fugelsang i in. 2005).

4. Sztuczny model neuronalny

Nawet jeśli takie ominięcie metodologicznego problemu przyczynowości jest poznawczo mało satysfakcjonujące, to jednak świadczące na jego rzecz dane empiryczne są interesujące w stopniu wystarczającym przynajmniej do tego, żeby, korzystając z płynących z nich wniosków,

[8] Na temat wyjaśniania przyczynowego por. Kawalec 2006.

spróbować rekonstrukcji takiego modelu abdukcji za pomocą sztucznych sieci neuronowych (SSN).

Trudno oczywiście oczekiwać, żeby SSN umożliwiały adekwatne reprezentowanie procesów mózgowych. Możemy tu mówić co najwyżej o mniej lub bardziej udanej emulacji. Tym niemniej wykorzystanie odpowiednio wyrafinowanych sieci do modelowania rozumowań prowadzić może do interesujących wniosków (McClelland 2009, Thomas i McClelland 2008). Odpowiednio wyrafinowanych, bo np. prosty, lokalistyczny model porównawczej oceny konkurencyjnych hipotez, zdefiniowany przez Thagarda (1989), w którym przez neurony reprezentowane są sądy, a przez pobudzające bądź opóźniające wagi, przyporządkowane poszczególnym sygnałom – związki między sądami, nie jest szczególnie pouczający.

Bardziej interesujący jest model, opisany zwięźle przez Thagarda i Litta (2008), bazujący na *Neural Engineering Framework* (NEF) Eliasmitha i Andersona (2003)[9]. Dzięki założeniom, dotyczącym definiowania i przekształcania reprezentacji neuronalnych, a także sposobu charakteryzowania dynamiki neuronalnej (Eliasmith i Anderson 2003, s. 15), NEF oferować ma, zdaniem swoich autorów, możliwość stosunkowo realistycznego modelowania procesów poznawczych i symulowania aktywności różnych obszarów mózgu (Eliasmith 2005). Z kolei jako narzędzie do modelowania dystrybucji informacji między neuronami służą Thagardowi i Littowi zredukowane reprezentacje holograficzne (*holographic reduced representations*, HRR (Plate 1995 2003), za pomocą których złożone relacje strukturalne (jak np. gramatyczna struktura języka) mogą być kodowane w SSN, systemach z definicji rozproszonych.

Do zalet neuronalnego modelu rozumowań abdukcyjnych Thagard i Litt zaliczają, po pierwsze, możliwość analizowania za jego pomocą multimodalnych aspektów tego typu przetwarzania danych, możliwość, która wymyka się ujęciom tradycyjnym, postulującym wyłacznie językowy sposób reprezentowania składników rozumowań. Po drugie, naturalne odzwierciedlenie, za pomocą odpowiednich populacji neuronów, znajdują tu informacje o emocjonalnymwymiarze rozumowania abdukcyjnego, rozpiętego między poważnie potraktowanym zaskoczeniem, uruchamiającym procedurę poszukiwania hipotez, a poznawczym usatysfakcjonowaniem[10].

[9]Implementujący NEF program symulacyjny NESim (*Neural Engineering Simulator*) i jego następca Nengo (*Neural Engineering Objects*), oba działające w środowisku MATLAB, wraz z dokumentacją dostępne są pod adresem: `http://compneuro.uwaterloo.ca/codelibrary/codelibrary.html`.

[10]Co do możliwości modelowania emocji za pomocą narzędzi symbolicznych

Z grubsza rzecz ujmując, model działa następująco. Sieci neuronowe uczone są najpierw reguł przyczynowych oraz związków eksplanacyjnych. Wykrycie potencjalnego eksplanandum, którym może być zjawisko, zdanie bądź jakikolwiek inny rodzaj danych, możliwych do reprezentowania za pomocą SSN, uruchamia moduł odpowiedzialny za reakcję emocjonalną, inicjującą dalszą aktywność. Działania kolejnych populacji neuronów (reprezentujących zapamiętane reguły przyczynowe oraz zasady dokonywania operacji na HRR) zmierzają do odnalezienia poprzednika reguły, której zjawisko wyjaśniane jest następnikiem, generując w ten sposób wyjaśnienia przyczynowe. Sieci zdolne są także do dokonywania prostej oceny hipotez, bazującej na sprawdzeniu, czy wartość, uzyskiwana przez wybrane wyjaśnienie na wyjściu, jest wystarczająco wyższa od wartości hipotez konkurencyjnych (co wymaga dobrania odpowiednich wartości progowych). Jeśli generowanie hipotezy wyjaśniającej kończy się sukcesem, sygnał emocjonalny przełączany jest z zaskoczenia na usatysfakcjonowanie, co z kolei prowadzi do akceptacji wybranej przyczyny jako adekwatnego wyjaśnienia eksplanandum i powrotu sygnału emocjonalnego do wyjściowego, neutralnego poziomu.

Ponieważ epatowanie Czytelnika technicznymi szczegółami bez możliwości zaprezentowania działania modelu nie wydaje się celowe, poprzestaniemy na dalszych wskazówkach bibliograficznych. Ogólny opis modelu podają cytowani Thagard i Litt (2008), natomiast szczegółowy opis operacji konstruowania HRR – Eliasmith i Thagard (2001). Dokładną charakterystykę, bazującego na podobnych założeniach, afektywnego modelu podejmowania decyzji ANDREA (*Affective Neuroscience of Decision through Reward-based Evaluation of Alternatives*) przedstawiają Litt, Eliasmith i Thagard (2008).

W tak prostym modelu trudno oczywiście oczekiwać istotnej reprezentacji dla rzeczywistej kreatywności: sieć najpierw uczy się reguł przyczynowych, aby później móc wśród wyuczonych związków odnajdywać najbardziej adekwatne hipotezy wyjaśniające. Mając jednak na uwadze jego potencjał warto zadać pytanie, czy za pomocą takich niedeterministycznych procedur można skonstruować narzędzie generowania i oceny hipotez abdukcyjnych charakteryzujące się mechaniczną efektywnością?

Thagard (1988, s. 64) wyróżnia dwa znaczenia terminu „mechaniczna efektywność". W pierwszym, mocniejszym znaczeniu, mechaniczna jest choćby metoda tabel zerojedynkowych w klasycznym rachunku zdań: jest to algorytm, który dla dowolnej formuły A pozwala w skończonym

– a dokładniej szczególnego rodzaju logik modalnych – por. Adam i in. 2009.

czasie rozstrzygnąć, czy jest ona tautologią, czy nie. W drugim znaczeniu mechaniczność procedury oznacza możliwość jej implementacji na przykład w działającym programie komputerowym. Taki charakter ma program PI (*Processes of Induction*), podstawowe narzędzie modelowania organizacji i rozwoju wiedzy w obliczeniowej filozofii nauki (*Computational Philosophy of Science*, Thagard 1988) , taki charakter ma też neuronalny model abdukcji Thagarda i Litta. Algorytmy, gwarantujące zautomatyzowane rozwiązywanie zadań eksplanacyjnych, prawdopodobnie nie istnieją. Tym niemniej, wedle Thagarda, można sensownie utrzymywać, że proces odkrywania jest przynajmniej w przybliżeniu algorytmiczny i że składa się nań wiele procedur, które co prawda nie gwarantują osiągania rozwiązań, ale w sposób istotny ich osiąganie wspomagają.

5. Kryteria oceny hipotez

Jeśli, jak Thagard, utożsamimy abdukcję z wnioskowaniem do najlepszego wyjaśnienia, to za kryteria oceny hipotez, generowanych przez procedurę abdukcyjną, przyjdzie nam uznać te same kryteria, na podstawie których jedno wyjaśnienie uznalibyśmy za lepsze od innych. Jednak określenie co to właściwie znaczy, że jedno wyjaśnienie jest lepsze od innych, nie jest sprawą oczywistą (Grobler 2006, s. 103).

Thagard (1988, s. 75–99) proponuje trzy kryteria: konsiliencję, prostotę i podobieństwo (analogię). Ich źródłem mają być studia przypadków rozumowań i argumentów, pochodzących od Darwina, Lavoisiera, Huygensa i Harveya. Czwartym kryterium oceny hipotez jest wedle Thagarda, jak wolno sądzić, koherencja (Thagard 2000, s. 15–16). Przy czym pojęcia kryterium nie należy rozumieć tu w kategoriach warunków koniecznych i wystarczających, a w kategoriach jednego z możliwych standardów oceny, konfrontowanego z innymi kryteriami w procesie oceny hipotez (Thagard 1988, s. 78).

5.1. Konsiliencja

Termin „konsiliencja" w ciągu ostatnich dziesięciu lat zrobił sporą karierę, głównie dzięki Edwardowi O. Wilsonowi (1998). Według Wilsona konsiliencja wiąże się przede wszystkim z jednością wiedzy (zgodnie zresztą z podtytułem jego książki). Wilson odwołuje się do Williama Whewella, który jako pierwszy miał użyć tego terminu w kontekście filozoficznym: „Whewell pisał o konsiliencji jako dosłownie «zbieganiu się» wiedzy dzięki łączeniu faktów i opartych na nich teorii empirycz-

nych z różnych dziedzin w jedną wspólną teorię wyjaśniającą" (Wilson 1998, s. 8).

„Jońskie oczarowanie"[11], z którego wyrasta wilsonowska koncepcja konsiliencji, ma uniwersalistyczny charakter, wyrażający się w przekonaniu o zasadniczej jedności nauki. Thagard, który również odwołuje się do Whewella, interpretuje konsiliencję w znacznie skromniejszy, rzec można lokalistyczny sposób. Teoria posiada własność konsiliencji w tym większym stopniu, im bardziej unifikuje i systematyzuje wiedzę i z im bardziej różnych dziedzin przedmiotowych pochodzą fakty, które zdolna jest wyjaśniać. Jako że wraz z rozwojem teorii może zmieniać się zarówno poziom unifikowania i systematyzowania, jak i jej obszar zastosowań, konsiliencja teorii jest własnością dynamiczną i może zmieniać się w czasie. Rzecz jasna, to co istotne dla oceny teorii czy hipotezy to nie prosta liczba wyjaśnianych przez nie faktów, a ich różnorodność i względna istotność (Thagard 1988, s. 80–81). Wskutek tego, w ocenie konsiliencji pojawić się może perspektywa podmiotowa: ocena stopnia różnorodności wyjaśnianych faktów zakłada odpowiednio bogatą wiedzę na temat dziedzin, z których pochodzą, a względna istotność hipotez zależy od epistemicznego celu podmiotu.

Dla Wilsona ostatecznym ideałem konsiliencji jest jednolita teoria wszystkiego, z której wyprowadzalne byłyby prawa poszczególnych nauk[12]. Zdaniem Thagarda przykłady teorii oskarżanych o wyjaśnianie zbyt szerokiej klasy zjawisk (jak teoria flogistonowa albo psychoanaliza) skłaniają do rozważenia możliwości nałożenia pewnych ograniczeń na dopuszczalny poziom konsiliencji teorii czy hipotezy. Jednak aby uniknąć nakładania takich ograniczeń *ad hoc*, należy je powiązać z innymi kryteriami oceny, a w szczególności z prostotą.

5.2. Prostota

To, że prostota jest istotnym kryterium oceny teorii i hipotez, jest raczej oczywiste. Odpowiedź na pytanie, co to znaczy, że jedna teoria czy hipoteza jest prostsza od innej, już oczywista nie jest. Jej mało skomplikowana wersja brzmi: chodzi o prostotę syntaktyczną. Im mniej środ-

[11] Termin „jońskie oczarowanie" bądź, wedle tłumacza „Konsiliencji" na język polski, „zauroczenie" (*Ionian enchantment*), oznacza „przekonanie – o wiele głębsze niż jedynie robocza hipoteza – że świat jest uporządkowany i że można go wyjaśnić za pomocą niewielkiej liczby praw przyrody". Termin ten nawiązuje do poglądów jońskich filozofów przyrody, w myśl których istnieje wspólna podstawa całej materialnej rzeczywistości.

[12] Nb. na tej podstawie Fodor (1998) zarzuca Wilsonowi promowanie fizykalistycznego redukcjonizmu.

ków językowych potrzeba do wyrażenia teorii czy hipotezy, tym lepiej. Prostota rozumiana jako zwięzłość ma niewątpliwe zalety (i w pewnych okolicznościach może stanowić cenne dodatkowe kryterium oceny hipotez), ale takie jej pojmowanie wydaje się jednak nadmiernie uproszczone. Inna możliwość to utożsamienie prostoty z minimalnością logiczną, która również ma swoje zalety, ale potrafi sprawić niespodzianki. Gabbay i Woods (2006, s. 20), cytując Goddu (2002, s. 15) i Hitchcocka (2002, s. 158), poddają pod rozwagę argument następujący:

(P1) Wszystkie małpy są naczelnymi.

(W) Zatem wszystkie małpy są ssakami.

Najprostsza entymematyczna przesłanka tego wnioskowania (którą nb. możemy interpretować jako hipotezę abdukcyjną) brzmi:

(P2) Wszystkie naczelne są ssakami.

Ale logicznie słabszą od niej jest przesłanka następująca:

(P2') Nie wszystkie małpy są naczelnymi lub wszystkie małpy są ssakami.

A przecież trudno oczekiwać, żebyśmy odrzucili przesłankę (P2) na rzecz (P2').

Prostota, która interesuje Thagarda, ma charakter pragmatyczny i w przypadku abdukcji nierozerwalnie powiązana jest z wyjaśnianiem. Jest funkcją rozmiaru i stopnia skomplikowania zbioru hipotez, a także relacji między samymi hipotezami, potrzebnymi na gruncie danej teorii do wyjaśnienia zaskakujących zjawisk. Dlatego właśnie Lavoisiera tlenowa teoria spalania ciał jest prostsza od teorii flogistonowej. Teoria flogistonowa wymaga np. akceptacji hipotez, głoszących, że flogiston istnieje i że w pewnych okolicznościach posiada masę dodatnią, a w innych ujemną. Teoria Lavoisiera tłumaczy fakty, nie odwołując się do tak karkołomnych konstrukcji a ponadto jest oszczędniejsza ontologicznie[13]. To właśnie tak rozumiana prostota nakłada pewne ograniczenia na konsiliencję: teoria zarazem konsilientna i prosta nie tylko wyjaśnia szeroką gamę zjawisk, ale w wyjaśnianiu tym nie odwołuje się do nadmiernie wielkiej liczby hipotez *ad hoc*, o wąskim zakresie zastosowania (Thagard 1988, s. 83)[14].

Załóżmy zatem, że mamy dwie teorie, wyjaśniające te same zjawiska za pomocą dwóch zbiorów hipotez abdukcyjnych. Które hipotezy

[13] Choć akurat zasadą brzytwy Ockhama, zakazującą namnażania bytów ponad potrzebę, w ocenie konkurencyjnych hipotez należy posługiwać się ostrożnie (Thagard 1988, s. 86).

[14] Sformułowanie „nadmiernie wielka liczba" nie jest szczególnie precyzyjne, ale trudno tu o jakąś precyzyjną miarę. Pamiętać też należy, że nie każdą hipotezę *ad hoc* trzeba traktować jak zło wcielone.

są prostsze? Jeśli jeden ze zbiorów zawierałby się w drugim, odpowiedź byłaby łatwa. Ale jeśli nie pozostają one w żadnych interesujących relacjach teoriomnogościowych? Gdyby któryś z nich był zbiorem sprzecznym albo miał inne nieporządane własności, byłby to argument na rzecz wyboru konkurencyjnego zbioru hipotez (zakładając oczywiście, że niesprzeczność jest wartością, na której nam zależy). Jeśli nie, to pozostaje nam pieczołowite dokonywanie ich drobiazgowych jakościowych porównań w kontekście konkretnych, eksplanacyjnych zastosowań. „Pojęcie prostoty jest bardzo złożone", powiada Thagard (1988, s. 84)[15].

5.3. Podobieństwo

Argumenty z podobieństwa są istotne w ocenie abdukcji, ponieważ podobieństwo ma istotny wpływ na jakość hipotez eksplanacyjnych (Thagard 1988, s. 94). Mowa tu nie tylko o podobieństwie zjawisk wyjaśnianych do już wyjaśnionych, ale także o strukturalnym podobieństwie rozumowań, reguł i rozwiązań problemów, stosowanych w wyjaśnianiu różnych zjawisk. Przykładem odwołania się do podobieństwa wyjaśnianych zjawisk byłoby, zastosowane przez Huygensa (1690, s. 4), porównanie niektórych własności rozchodzenia się dźwięku i światła jako argumentu na rzecz tezy, że światło również ma naturę falową. Przykładem użycia podobnych narzędzi do różnych dziedzin byłoby zastosowanie teorii doboru naturalnego do organizmów społecznych (Wilson 1975, Hamilton 1964)[16].

Jednak to nie sama obecność analogii sprawia, że podobieństwo jest istotnym kryterium oceny hipotez abdukcyjnych. Huygens (1690, s. 10–12) na przykład wskazuje nie tylko na podobieństwa w zachowaniu dźwięku i światła, ale także na istotne różnice między nimi, traktując mimo to analogie w zachowaniach dźwięku i światła jako ważki argument na rzecz swojej tezy. Rzecz w tym, że stosując argument z podobieństwa w ocenie hipotez odwołujemy się (albo odwoływać powinniśmy) nie do prostego współwystępowania zjawisk, ale do współwyjaśniania. W szkolnym schemacie wnioskowania przez analogię z faktu, że obiekt A posiada własności P, Q, R, S i z faktu, że obiekt B posiada własności P, Q, R wyprowadzany jest wniosek, że B ma własność

[15] Thagard przedstawia co prawda sposób implementacji jakościowych kryteriów konsiliencji i prostoty w programie PI ale, jak sam przyznaje, jako model procesów wyjaśniania PI ma liczne braki (Thagard 1988, s. 92).

[16] Wnikliwą analizę argumentów z podobieństwa, wnioskowań przez analogię i związków między podobieństwem a analogią przedstawia Szymanek 2008; por. także Bartha 2010, Abrantes 1999, Gentner i in. 2001).

S (Ziembiński 2002, s. 191)[17]. Thagard (1988, s. 93) proponuje schemat, w którym z faktu, że obiekty A i B posiadają własności P, Q, R i z faktu, że posiadanie przez A własności S wyjaśnia, dlaczego A jest P, Q, R wyprowadzany jest wniosek, że ewentualne posiadanie przez B własności S byłoby obiecującym wyjaśnieniem posiadania przez B własności P, Q, R. Wniosek, że B posiada własność S byłby być może zbyt daleko idący, ale podobieństwa między A i B podnoszą wartość wyjaśnienia posiadania własności P, Q, R za pomocą S. Przedmiotem analogii są zatem nie tylko własności obiektów, ale także, a nawet przede wszystkim, relacje między obiektami i ich własnościami (Gentner 1983). Dotyczy to zwłaszcza relacji przyczynowych (Thagard i Shelley 2001), do których to wedle Thagarda, jak pamiętamy, przede wszystkim odwoływać ma się wyjaśnianie. Tym sposobem argumenty z podobieństwa wspierać mogą nie tylko nowe hipotezy, ale także stanowić dodatkowe uzasadnienie dla hipotez już uprzednio postulowanych.

Nie znaczy to oczywiście, że generowanie hipotez abdukcyjnych sprowadzać się ma do szukania podobieństw między tym, co nowe, a tym, co znane. W historii nauki znajdziemy mnóstwo przykładów kreatywnych rozumowań abdukcyjnych, w których postulowane jest istnienie nowych obiektów i które formułowane są za pomocą nowych pojęć. Jak pisze Thagard (1988, s. 95), „stosowanie znanych modeli nie stanowi istoty wyjaśniania, ale jest pomocne", *ceteris paribus*.

Poszukiwanie najlepszego wyjaśnienia za pomocą kryteriów konsiliencji, prostoty i podobieństwa, samo w sobie niełatwe, komplikowane jest jeszcze przez fakt, że kryteria te sprzężone są ze sobą w sposób uniemożliwiający prostą maksymalizację wszystkich trzech naraz (Thagard 1988, s. 98). Odwołując się do terminologii koneksjonistycznej moglibyśmy powiedzieć, że kryteria konsiliencji, prostoty i podobieństwa nakładają na siebie nawzajem pewne ograniczenia, a ostateczna ocena hipotez jest efektem ich wzajemnych interakcji (por. Rumelhart 1989). Wspominaliśmy już, że prostota hamuje konsilientne tendencje do generowania hipotez *ad hoc*. Z kolei konsiliencja, niekiedy wbrew prostocie, nakłania do akceptowania dodatkowych hipotez, jeśli wyjaśniają one dodatkowe zjawiska. Podobieństwo może kłócić się z pozostałymi dwoma kryteriami, jeśli prosta i konsilientna hipoteza ma radykalnie nowy charakter. Skoro jednak rzeczywisty proces oceny hipotez abdukcyjnych jest skomplikowany i wielowymiarowy, taka interpretacja jego komplikację i wielowymiarowość dobrze oddaje.

[17] Przy czym wnioskowaniu takiemu towarzyszyć mogą dodatkowe zastrzeżenia na temat jego racjonalności, dotyczące charakteru związku między własnościami P, Q, R, S.

5.4. Koherencja

Swoją teorię koherencji Thagard (2000, rozdz. 6; por. Thagard i Shelley 2001, s. 337) streszcza za pomocą kilku tez. Po pierwsze, dążenie do maksymalizowania koherencji jest istotą każdego rozumowania. Reguły takie jak *modus ponendo ponens* same w sobie nie charakteryzują rzeczywistych wnioskowań, ponieważ ich konkluzje mogą być sprzeczne z uprzednio zaakceptowanymi informacjami. Jedyna reguła wnioskowania brzmi: zaakceptuj wniosek, jeśli maksymalizuje on koherencję. Po drugie, koherencja jest istotna nie tylko ze względu na kwestię akceptowania bądź odrzucania wniosków rozumowań, ale może być również powiązana z przypisywaniem pozytywnej bądź negatywnej charakterystyki emocjonalnej interesującym nas reprezentacjom (zdaniom, pojęciom, obiektom itd). Po trzecie, istotą procesu maksymalizowania koherencji jest porządkowanie przestrzeni reprezentacji za pomocą sieci wzajemnie powiązanych zależności; proces ten można scharakteryzować za pomocą określonych algorytmów.

Thagard i Verbeurgt (1998) definiują kilka takich algorytmów (w widoczny sposób skłaniając się ku algorytmowi koneksjonistycznemu[18]). W przypadku każdego z nich chodzi o to, aby dokonać podziału zadanego zbioru reprezentacji na zbiór elementów zaakceptowanych i zbiór elementów odrzuconych w sposób, który gwarantuje spełnienie jak największej liczby zależności między elementami[19]. Jeśli dwa elementy są ze sobą koherentne, istnieje między nimi zależność pozytywna, jeśli nie – zależność negatywna. Zależność pozytywna może zostać spełniona albo poprzez zaakceptowanie obu elementów, albo poprzez odrzucenie obu, natomiast zależność negatywna jedynie poprzez odrzucenie jednego z elementów i zaakceptowanie drugiego.

Spośród sześciu wymienianych przez Thagarda rodzajów koherencji: eksplanacyjnej, bazującej na podobieństwie, pojęciowej, dedukcyjnej, percepcyjnej i deliberatywnej (Thagard i in. 2002) interesuje nas przede wszystkim ten pierwszy, charakteryzowany przez teorię koherencji eksplanacyjnej (*Theory of Explanatory Coherence*, TKE) za pomocą następujących zasad (Thagard 1992; por. 2007b, b.d.):

E1 (zasada symetrii): koherencja eksplanacyjna jest relacją symetryczną;

[18] Implementowanemu w środowisku Java za pomocą programu ECHO: `http://cogsci.uwaterloo.ca/JavaECHO/jecho.html`.

[19] Za pomocą terminu „spełnianie zależności" tłumaczę angielski termin *constraint satisfaction*. Zwyczajowe tłumaczenie tego terminu jako „spełnianie ograniczeń" wydaje sie tu nie na miejscu, biorąc pod uwagę, że owe zależności mogą być tak negatywne, jak i pozytywne.

E2 (zasada wyjaśniania): (a) hipoteza jest koherentna z wyjaśnianymi przez siebie obserwacjami lub innymi hipotezami; (b) hipotezy, które wspólnie wyjaśniają to samo eksplanandum są ze sobą koherentne; (c) im więcej hipotez potrzeba, by wyjaśnić jedno eksplanandum, tym niższy poziom ich wzajemnej koherencji;

E3 (zasada podobieństwa): podobne hipotezy wyjaśniające podobne zjawiska są ze sobą koherentne;

E4 (zasada priorytetu danych): poziom akceptacji elementów, charakteryzujących wyniki obserwacji, nie musi zależeć od związków koherencji;

E5 (zasada sprzeczności): elementy sprzeczne nie są ze sobą koherentne;

E6 (zasada konkurencji): jeśli elementy P i Q wyjaśniają to samo eksplanandum, a nie są eksplanacyjnie powiązane ze sobą, to nie są ze sobą koherentne[20];

E7 (zasada akceptacji): akceptacja elementu zależy od jego koherencji z reprezentacjami, do systemu których należy.

Po pierwsze zatem, przedmiotem zainteresowania TKE jest akceptacja bądź odrzucenie elementów przestrzeni interesujących nas reprezentacji, bazujące na ich wzajemnej koherencji (E1). W przestrzeni tej istnieją elementy wyróżnione, charakteryzujące wyniki obserwacji, które to elementy mogą być akceptowane niezależnie od powiązań z innymi elementami (E4). Poziom akceptacji każdego spośród pozostałych elementów co najmniej pośrednio zależy od jego związków z elementami wyróżnionymi, bezpośrednio natomiast od poziomu akceptacji wszystkich elementów, z którymi jest koherentny (E7). Wyjaśnianie i koherencja powiązane są ze sobą poprzez zależności bazujące na konsiliencji, podobieństwie i logicznych związkach między elementami. Poziom koherencji hipotezy wprost proporcjonalnie zależy od jej mocy wyjaśniającej (E2c) i obszaru zastosowań (E2a), a więc konsiliencji. Również podobieństwo pozostaje w pozytywnej zależności w stosunku do koherencji (E3). Koherencja dwóch hipotez maksymalizowana jest, gdy hipotezy wyjaśniają zjawiska podobne (E3), gdy wspólnie wyjaśniają te same zjawiska lub gdy jedna z nich wyjaśnia drugą (E2b, E6). Koherencja pozostaje w negatywnej zależności w stosunku do sprzeczności[21] (E5) i do braku związków eksplanacyjnych między hipotezami (E6).

[20] P i Q są eksplanacyjnie powiązane jeśli jedno z nich wyjaśnia drugie lub jeśli wspólnie wyjaśniają to samo eksplanandum.

[21] Ponieważ Thagard posługuje się terminem *contradiction* należy mniemać, że ma na myśli sprzeczność logiczną. Dwa zdania są logicznie sprzeczne wtedy i tylko wtedy, gdy nie mogą one być łącznie prawdziwe ani łącznie

Koherencja pełni więc w ocenie hipotez wyjaśniających rolę metakryterium, sprzężonego co najmniej z konsiliencją i podobieństwem. Abdukcyjne nadawanie sensu zjawiskom zaskakującym polega w gruncie rzeczy na takiej ich interpretacji, która pasuje do dostępnych danych lepiej, niż interpretacje alternatywne. Z kolei najlepszą interpretacją jest taka, która oferuje najbardziej spójny, koherentny obraz tego, co próbujemy zrozumieć (Thagard i Verbeurgt 1998, s. 3). Dedukcyjne związki między hipotezą a eksplanandum są tylko jednym z możliwych sposobów pojmowania koherencji, i nie mają uprzywilejowanego charakteru w porównaniu ze związkami, bazującymi choćby na podobieństwie czy zależnościach percepcyjnych.

Zaznaczmy jeszcze, że TKE nie ma wiele wspólnego z koherencyjną teorią prawdy (której zresztą Thagard (2007b, s. 29–30) zarzuca, że prowadzi do epistemologicznego idealizmu). Nie jest też typową koherencyjną teorią uzasadniania, z uwagi na postulowanie istnienia elementów wyróżnionych, akceptowanych bądź odrzucanych „fundamentalistycznie". Najbliżej jej zapewne do poglądów Susan Haack (1993), dla których ukuła ona termin *foundherentism*.

Model eksplanacyjno-koherencyjny bazuje na tym samym, pochodzącym od Peirce'a schemacie abdukcji, co model eksplanacyjno-dedukcyjny. Jednak w odróżnieniu od tego drugiego próbuje charakteryzować rozumowania abdukcyjne w sposób możliwie bliski ich rzeczywistym zastosowaniom. Dzieje się tak głównie z uwagi na fakt, że przedmiotem zainteresowania w modelu eksplanacyjno-koherencyjnym jest abdukcja traktowana jako proces, który nie tylko posiada określoną strukturę i prowadzi do określonych rezultatów, ale także jako rozumowanie, które w rzeczywistości przeprowadzane jest zawsze przez konkretne podmioty. W konsekwencji, na uruchomienie tego procesu, jego przebieg i zakończenie istotny wpływ mają czynniki podmiotowe: potrzeby epistemiczne, zasoby informacyjne, na bazie których generowane i oceniane są hipotezy abdukcyjne, a wreszcie, co jest szczególnie istotne w modelu proponowanym przez Paula Thagarda, emocje, zarówno te wynikające z wyjściowej niewiedzy i końcowego osiągnięcia pożądanej informacji, jak i te, które towarzyszyć mogą reprezentacjom, stanowiącym elementy rozumowania abdukcyjnego. W konsekwencji, pojęcie konkluzywności rozumowania abdukcyjnego jest w modelu koherencyjnym znacznie trudniejsze do zdefiniowania, niż w modelu dedukcyjnym, a zatem i modelowanie abdukcji jest w pierwszym z nich bardziej kłopotliwe, niż w drugim. Za podsumowanie korzyści, mimo tych trudności płynących jednak z podejścia eksplanacyjno-koherencyjnego, niech posłużą słowa

---

fałszywe. *Mutatis mutandis*, zapewne podobnie należałoby rozumieć sprzeczność między innymi rodzajami reprezentacji.

Thagarda (2007a, s. 5): „Abdukcja, zamiast przypominać nieco głupkowatego kuzyna dedukcyjnej reguły *modus ponendo ponens*, jest tak naprawdę znacznie bogatszą i bardziej produktywną (*more powerful*) formą myślenia."

Literatura

Abrantes, P. (1999). Analogical Reasoning and Modeling in the Sciences. *Foundations of Science*, 4(3):237–270.

Adam, C., Herzig, A., Longin, D. (2009). A logical formalization of the OCC theory of emotions. *Synthese*, 168(2):201–248.

Adler, J. E., Rips, L. R. (2008). *Reasoning. Studies of Human Inference and Its Foundations*. Cambridge University Press, Cambridge.

Aliseda, A. (2006). *Abductive Reasoning. Logical Investigations into Discovery and Explanation*. Springer, Dordrecht.

Anderson, C. A., Owens, J. (red.) (1990). *Propositional Attitudes: The Role of Content in Logic, Language, and Mind*. CSLI Publications, Stanford University.

Baillargeon, R., Kotovsky, L., Needham, A. (1995). The aqusition of physical knowledge in infancy. W: D. Sperber, D. Premack, A. J. Premack (red.), *Causal cognition: A multidisciplinary debate*, 79–116. Clarendon Press, Oxford.

Bartha, P. (2010). *By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments*. Oxford University Press, Oxford.

Brody, B. (1972). Towards an Aristotelian theory of scientific explanation. *Philosophy of Science*, 39:20–31.

Bunge, M. (1979). *Causality and Modern Science*. Dover Publications, New York, 3 wyd.

Bylander, T., Allemang, D., Tanner, M. C., Josephson, J. R. (1995). The Computational Complexity of Abduction. Rap. tech., Department of Computer and Information Science, The Ohio State University, Columbus.

Eliasmith, C. (2005). Cognition with neurons: a large-scale, biologically realistic model of the Wason task. W: L. B. B. Bara, M. Bucciarelli (red.), *Proceedings of the XXVII Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates, Mahwah, NJ.

Eliasmith, C., Anderson, C. H. (2003). *Neural engineering: Computation, representation and dynamics in neurobiological systems*. MIT Press, Cambridge, MA.

Eliasmith, C., Thagard, P. (2001). Integrating structure and meaning: A distributed model of analogical mapping. *Cognitive Science*, 25:245–286.

Fodor, J. (1998). Look! (recenzja *Consilience: The Unity of Knowledge Edwarda O. Wilsona)*. London Review of Books, 29. października 1998. `http://www.lrb.co.uk/v20/n21/fodo01_.html`. Uzyskano: 15.06.2009.

Fugelsang, J. A., Roser, M. E., Corballis, P. M., Gazzaniga, M. S., Dunbar, K. N. (2005). Brain mechanisms underlying perceptual causality. *Cognitive Brain Research*, 24:41–47.

Gabbay, D. M., Woods, J. (2005). *The Reach of Abduction. Insight and Trial*. Elsevier.

Gabbay, D. M., Woods, J. (2006). Advice on Abductive Logic. *Logic Jour nal of the IGPL*, 14(2):189–219.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7:155–170.

Gentner, D., Holyoak, K. J., Kokinov, B. K. (red.) (2001). *The Analogical Mind: Perspectives from Cognitive Science*. MIT Press, Cambridge, MA.

Goddu, G. C. (2002). The most important and fundamental distinction in logic. *Informal Logic*, 22:1–17.

Goel, V. (2007). Anatomy of deductive reasoning. *Trends in Cognitive Science*, 11(10):435–441.

Grobler, A. (2006). Metodologia nauk. Wydawnictwo Aureus, Wydawnictwo Znak, Kraków.

Haack, S. (1993). *Evidence and Inquiry: Towards Reconstruction in Epistemology*. Basil Blackwell, Oxford.

Hamilton, W. D. (1964). The Genetical Evolution of Social Behaviour, I–II. *Journal of Theoretical Biology*, 7:1–16, 17–52.

Harman, G. (1965). Inference to the Best Explanation. *Philosophical Review*, 74(1):88–95.

Hausman, D. (1982). Constructive empiricism contested. *Pacific Philosophical Quarterly*, 61:21–28.

Hilton, D. (2007). Causal Explanation. From Social Perception to Knowledge-Based Attribution. W: A. Kruglanski, E. T. Higgins (red.), *Social Psychology. Handbook of Basic Principles*, 232–253. The Guilford Press, New York–London, 2 wyd.

Hintikka, J. (1992). W stron¦ ogólnej teorii indywiduacji i identyfikacji. W: *Eseje logiczno-filozoficzne*. Taum. A. Grobler. WN PWN, Warszawa.

Hintikka, J. (2007). Abduction – Inference, Conjecture, or an Answer to a Question? W: *Socratic Epistemology. Explorations of Knowledge-Seeking by Questioning*, 38–60. Cambridge University Press.

Hitchcock, D. (2002). A note on implicit premisses. *Informal Logic*, 22:158–159.

Huygens, C. (1690). *Traité de la lumière*. Leiden. `http://www.gutenberg.org/etext/14725`. Uzyskano: 15.06.2009. Angielski przekład dostępny online: *Treatise on Light*, tłum. Silvanus P. Thompson.

Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge University Press, Cambridge, MA.

Josephson, J. R., Josephson, S. G. (red.) (1994). *Abductive Inference: Computation, Philosophy, Technology*. Cambridge University Press, Cambridge.

Kadaba, N., Irani, P. P., Leboe, J. (2007). Visualizing Causal Semantics Using Animations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1254–1261.

Kawalec, P. (2006). *Przyczyna i wyja±nianie. Studium z filozofii i metodologii nauk*. Wydawnictwo KUL, Lublin.

Leake, D. B. (1993). Focusing construction and selection of abductive hypotheses. W: *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, 24–29. IJCAI.

Leake, D. B. (1995). Abduction, Experience and Goals: A Model of Everyday Abductive Explanation. *The Journal of Experimental and Theoretical Artificial Intelligence*, 7:407–428.

Leslie, A. M., Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, 25:265–288.

Lipton, P. (2004). *Inference to the Best Explanation*. Routledge, Londyn.

Litt, A., Eliasmith, C., Thagard, P. (2008). Neural affective decision theory: Choices, brains, and emotions. *Cognitive Systems Research*, 9:252–273.

Łupkowski, P. (2010). *Test Turinga. Perspektywa sędziego*. Wydawnictwo Naukowe UAM, Poznań.

Magnani, L. (2009). *Abductive Cognition. The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning*. Springer-Verlag, Berlin–Heidelberg.

Mandler, J. M. (2004). *The foundations of mind: Origins of conceptual thought*. Oxford University Press, Oxford.

McClelland, J. L. (2009). The Place of Modeling in Cognitive Science. *Topics in Cognitive Science*, 1:11–38.

Michotte, A. (1946). *La Perception de la Causalité*. Institut Supérieur de Philosophie, Louvain. Angielski przekład: *The perception of causality*, tłum. T. Miles i E. Miles, Basic Books, 1963.

Nelsen, R. B. (1997). *Proofs without Words: Exercises in Visual Thinking*. Mathematical Association of America.

Newman, A. H., Turing, A. M., Jefferson, G., Braithwaite, R. B. (1952). Can automatic calculating machines be said to think?, Broadcast discussion transmited on BBC (14 and 23 Jan. 1952). The Turing Digital Archive (www.turingarchive.org), Contents of AMT/B/6.

Peirce, C. S. (1931 – 1958). Collected Works. Harvard University Press, Cambridge, MA.

Plate, T. A. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks*, 6:623–641.

Plate, T. A. (2003). *Holographic Reduced Representation: Distributed Representation for Cognitive Structures*. CSLI Publications, Stanford, CA.

Reverberi, C., Cherubini, P., Rapisarda, A., Rigamonti, E., Caltagirone, C., Frackowiak, R. S., Macaluso, E., Paulesua, E. (2007). *Neural basis of generation of conclusions in elementary deduction*. NeuroImage, 38:752–762.

Rips, L. J. (1994). *The Psychology of Proof: Deductive Reasoning in Human Thinking*. MIT Press, Cambridge, MA.

Rumelhart, D. E. (1989). The Architecture of Mind: A Connectionist Approach. W: M. I. Posner (red.), *Foundations of Cognitive Science*, 133–159. MIT Press, Cambridge, MA.

Sun, R. (red.) (2008). *The Cambridge Handbook of Computational Psychology*. Cambridge University Press, Cambridge.

Szymanek, K. (2008). *Argument z podobieństwa*. Wydawnictwo Uniwersytetu Śląskiego, Katowice.

Thagard, P. (1988). *Computational Philosophy of Science*. MIT Press, Cambridge, MA.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, 12:435–467.

Thagard, P. (1992). *Conceptual Revolutions*. Princeton University Press, Princeton, NJ.

Thagard, P. (2000). *Coherence in Thought and Action*. MIT Press, Cambridge, MA.

Thagard, P. (2007a). Abductive inference: From philosophical analysis to neural mechanisms. W: A. Feeney, E. Heit (red.), *Inductive reasoning: Cognitive, mathematical, and neuroscientific approaches*, 226–247. Cambridge University Press, Cambridge.

Thagard, P. (2007b). Coherence, Truth, and the Development of Scientific Knowledge. *Philosophy of Science*, 74:28–47.

Thagard, P. (b.d.). Coherence: The Price is Right. Uzyskano: 15.06.2009. `http://cogsci.uwaterloo.ca/Articles/Pages/coh.price.html`.

Thagard, P., Eliasmith, C., Rusnock, P., Shelley, C. P. (2002). Knowledge and coherence. W: R. Elio (red.), *Common sense, reasoning, and rationality*, 104–131. Oxford University Press, New York.

Thagard, P., Litt, A. (2008). Models of scientific explanation. W: Sun (2008), 549–564.

Thagard, P., Shelley, C. P. (2001). Emotional Analogies and Analogical Inference. W: Gentner et al. (2001), 335–362.

Thagard, P., Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science*, 22:1–24.

Thomas, M. S. C., McClelland, J. L. (2008). Connectionist models of cognition. W: Sun (2008), 23–58.

Urbański, M. (2009a). *Rozumowania abdukcyjne*. Wydawnictwo Naukowe UAM, Poznań.

Urbański, M. (2009b). Trzy modele rozumowań abdukcyjnych. W: M. Urbański, P. Przybysz (red.), *Funkcje umysłu*, tom 27 z serii Poznańskie Studia z Filozofii Humanistyki, 303–315. Zysk i S-ka, Poznań.

Wilson, E. O. (1975). *Sociobiology: The New Synthesis*. Harvard University Press, Cambridge, MA. Polski przekaad (wersja skrócona): *Socjobiologia*, tłum. M. Siemiński, Zysk i S-ka, Poznań 2001.

Wilson, E. O. (1998). *Consilience: The Unity of Knowledge*. Alfred A. Knopf Publisher, New York. Polski przekład: *Konsiliencja. Jedność wiedzy*, tłum. J. Mikos, Zysk i S-ka, Poznań 2002.

Wiśniewski, A. (2011). Answering by Means of Questions in View of Inferential Erotetic Logic. *Logique et Analyse*, w druku.

Ziembiński, Z. (2006). *Logika praktyczna*. PWN, Warszawa.

Paul Thagard's model of abductive reasoning

Mariusz Urbański
Adam Mickiewicz University in Poznań

Abstract. *In this paper an explanatory-coherentist model of abductive reasoning is described on the example of the model of abduction proposed by Paul Thagard. Currently, it offers the most satisfactory combination of psychological adequacy and computational effectivenes of the modelled processes of abductive hypotheses generation and evaluation.*

Keywords: *abductive reasoning, explanatory-coherentist model of abduction, theory of explanatory coherence*