

PAWEŁ KAWALEC

WYJAŚNIĆ TO PODAĆ MODEL PRZYCZYNOWY

Dlaczego cień tego pnia ma taką właśnie długość? Dlaczego najlepiej jest nie doprawiać obiadu solą? Dlaczego naciskając przycisk, kierujemy pilota w kierunku panelu odbiornika telewizyjnego? Znamy odpowiedzi. Czy jednak są one zadowolające?

Kwestia ta powróci jeszcze, choć na marginesie poniższych rozważań, których głównym tematem będzie jej odwrotność: Czy odpowiedzi na powyższe pytania spełniają kryteria zadowolającego wyjaśniania? Śledząc najważniejsze z dotychczas zgłoszonych propozycji, staram się wykazać, że nie. Zadowolające wyjaśnianie musi bowiem łączyć dwie cechy: uwzględniać zależności przyczynowe oraz pozostawać w zgodzie z doświadczeniem¹. Te dwa warunki są nie tylko konieczne², lecz także – jak uzasadniam w tekście – wystarczające dla podania zadowolającego wyjaśnienia.

1. PRZYCZYNOWOŚĆ A WSPÓŁCZESNE KONCEPCJE WYJAŚNIANIA

O wyraźnie wyartykułowanych koncepcjach wyjaśniania można mówić dopiero w XX w. Jednakże długi cień rzuca na nie, sięgająca starożytności, dedukcjonistyczna idea, którą zwięźle streszcza formuła:

Dr PAWEŁ KAWALEC – Wydział Filozofii KUL, Katedra Metodologii Nauk, adres do korespondencji: 20-950 Lublin, Al. Raławickie 14, e-mail: pawel.kawalec@kul.lublin.pl

¹ Mimo radykalnie różnych stanowisk co do istnienia praw w naukach społecznych, John T. Roberts (2004) i Harold Kincaid (2004) konkludują, że wyjaśnianie opiera się w nich na wiedzy przyczynowej.

² Nie są też wzajemnie wykluczające, co sugerowałaby krytyka D. Hume'a.

(D) Wyjaśnić to wydedukować z pryncypiów.

W dotychczasowej filozofii nauki bezwzględnie dominuje ona nad inną formułą – równie długowieczną – przyczynową:

(P) Wyjaśnić to podać przyczynę.

Obie te formuły współtworzą podwaliny Arystotelesa koncepcji wyjaśniania (Hankinson 1998, s. 161-167; Losee 1993, s. 9-12), której mniej istotnym założeniem było, aby przesłanki były prawdziwe, niedowodliwe oraz lepiej znane od wniosku (*Analityki wtóre*, 71 b 20 – 72 a 5).

Za Wesleyem Salmonem, najwybitniejszym współczesnym teoretykiem wyjaśniania, powtórzmy:

Poszukiwanie naukowego zrozumienia świata ma korzenie sięgające czasów starożytnych znacznie poprzedzających Arystotelesa. Mimo że dalekie od zakończenia, to *naukowe* poszukiwanie okazało się bardzo owocne w wielu dziedzinach badań. Filozofowie i naukowcy, co najmniej od czasów Arystotelesa, starali się też określić, na czym takie zrozumienie polega. To *filozoficzne* poszukiwanie, aż do całkiem niedawna, było znacznie mniej udane. Twierdzą jednak, że w okresie ostatnich czterech dekad dokonał się znaczący postęp filozoficzny (1990, s. 1; tłum. P. K.).

Początek wyróżnionym przez Salmona dekadom daje przełomowy artykuł Carla G. Hempła i Paula Oppenheima *Studies in the logic of explanation* (1948/1965). Wszystko, co od tego czasu napisano o wyjaśnianiu, nawiązuje do zaproponowanej tam koncepcji. Jej też poświęcimy więcej uwagi.

W tej koncepcji³ wyartykułowano przede wszystkim formułę dedukcyjno-nomologiczną (DN), jako tzw. model dedukcyjno-nomologiczny (DN). To, co ma być wyjaśnione, czyli *eksplanandum*, jest opisem (jednostkowego) zdarzenia. *Eksplanans*, czyli to, co wyjaśnia, zawiera zasadniczo dwa elementy: prawo⁴ (gr. *nomos*, stąd nomologiczny) oraz opis warunków początkowych, w których zaistniało wyjaśniane zdarzenie. Wyjaśnianie uznaje się za zadowalające przede wszystkim wtedy, gdy eksplanandum wynika logicznie z tak skonstruowanego eksplanansu (stąd dedukcyjny). Oprócz tego żąda się, by podane prawo zawierało treść empiryczną (było confirmowalne przez doświadczenie),

³ Uwzględniam tu także późniejsze prace C. G. Hempła, zwł. 1965.

⁴ Współcześnie sygnalizuje się tu słabszy wymóg, jakim jest zdanie prawdopodobne (ang. *lawlike*), a więc niekoniecznie spełniające wszystkie warunki nakładane na prawa naukowe. Pod takie określenie podpadają w szczególności hipotezy naukowe.

a eksplanans był prawdziwy. Tylko ten ostatni warunek, jaki mają spełniać zadowalające wyjaśnienia, Hempel uznawał za „empiryczny”, czyli pozalogiczny oraz niesyntaktyczny. Wynikanie logiczne, prawo naukowe i konfirmowalność traktował jako pojęcia logiczne (1965, s. 247), które można zdefiniować na podstawie samej formy logicznej zdań, czyli czysto syntaktycznie. Jeśli wyjaśnienie spełnia tylko trzy pierwsze, ale nie czwarty warunek, to jest „wyjaśnieniem potencjalnym”, a nie „faktycznym”.

Wzorując się na sukcesach logiki w klaryfikacji podstaw matematyki, autorzy dedukcyjno-nomologicznej koncepcji wyjaśniania uznali logikę za zasadnicze narzędzie w analizie pojęć naukowych (Psillos 2002, s. 220). Powstała w ten sposób – krótko opisana wyżej – charakterystyka wyjaśniania, która była w zasadzie strukturalno-syntaktyczna. Miało to ugruntować obiektywność wyjaśniania naukowego poprzez jednoznaczne ujawnienie jego formy logicznej oraz wykazanie, że może być sensownie aplikowana. Wyjaśnianie, w tym ujęciu, jest szczególnego rodzaju prawomocnym wnioskowaniem dedukcyjnym, a wyróżnia je występowanie w przesłankach praw naukowych⁵.

W zaproponowanym przez Hempla ujęciu formuła dedukcjonistyczna:

(D*) Wyjaśnić to wydedukować z prawa naukowego,

jako szczególny przypadek, implikuje (P). Nie wszystkie bowiem wyjaśnienia dedukcyjno-nomologiczne są przyczynowe, ale „wyjaśnianie przyczynowe jest szczególnego rodzaju wyjaśnianiem dedukcyjno-nomologicznym” (Hempel 1965, s. 300). Takie ujęcie wyjaśniania przyczynowego ma u podstaw specyficzne rozumienie przyczynowości⁶. To, że jedno zdarzenie jest przyczyną drugiego, należy rozumieć – według Hempla – jako stwierdzenie, że istnieją prawa, ze względu na które wystąpienie zdarzenia wcześniejszego jest nomologicznie wystarczające dla wystąpienia zdarzenia późniejszego.

Podstawowe zarzuty przeciw modelowi DN, kwestionujące jego adekwatność jako modelu wyjaśniania, można uznać za konsekwencję pomniejszenia

⁵ Wielu autorów podkreśla, że to pojęcie wyjaśniania jest przede wszystkim *e p i s t e - m i c z n e* – wyjaśnić jakieś zdarzenie to wykazać, że należało się go spodziewać, wzięwszy pod uwagę prawa rządzące zachodzeniem tego rodzaju zdarzeń oraz warunki wstępne, w których ono faktycznie zaistniało.

⁶ Uwagi poczynione przez K. Poppera (1934/2002, s. 53-55) m o ż n a odczytać jako tendencję przeciwstawną w tym względzie koncepcji Hempla i bliższą rozważanej w niniejszym artykule. Por. też Hajduk 1970, s. 95-96.

roli przyczynowości w wyjaśnianiu (Psillos 2002, s. 224)⁷, innymi słowy, potraktowania (P) jako logicznej konsekwencji (D*).

Pierwsza grupa zarzutów, wykazujących, że model DN nie stanowi warunku wystarczającego wyjaśniania, może być zilustrowana przykładem formułowanym mniej więcej następująco. Zaczniemy od wyjaśnienia zgodnego z DN: wysokość pnia i kąt względem Słońca, jako warunek wstępny, oraz prawa optyki geometrycznej (wraz z elementarną trygonometrią) składają się na eksplanans, z którego wynika logicznie długość cienia. Sylvain Bromberger (1966) zaproponował odwrócenie tego przykładu: przy długości cienia i określonym kącie względem Słońca, jako warunku wstępnym, oraz tych samych prawach można logicznie wydedukować wysokość pnia. Zgodnie z DN należałoby zatem uznać, że nie tylko długość pnia wyjaśnia długość jego cienia, ale też na odwrót. To ostatnie, choć zgodne z DN, razi swoją niezgodnością z intuicją, gdyż to nie cień powoduje, że pień ma taką lub inną wysokość.

Uogólniając, model DN dopuszcza symetryczność między zdaniem opisującym przyczynę a opisującym skutek – jedno może wyjaśniać drugie i na odwrót⁸. Związek przyczynowy jednak jest asymetryczny i dlatego DN nieadekwatnie ujmuje wyjaśnianie przyczynowe, które także jest asymetryczne – tylko przyczyna wyjaśnia skutek, a nie na odwrót.

Odpowiedź, jaką na powyższą trudność zaproponował Hempel (1965, s. 352), można streścić następująco. Odróżniając prawa następstwa (uwzględniające zmianę w czasie) od praw współwystępowania (synchroniczna zależ-

⁷ Zestawiając koncepcję Arystotelesa ze współczesnymi, Hankinson komentuje: „To, w jakim stopniu potrafimy określić właściwy kierunek wyjaśniania, będzie zależało od tego, czy uda nam się pokazać, że jedna własność rzeczywiście zależy przyczynowo od drugiej. A to z kolei można odczytać z tego, w jaki sposób nasza interwencja zmieniłaby daną sytuację. [...] Współczesne koncepcje odrzucające pojęcie przyczynowości jako anachronizm, który należałoby zastąpić przez bezkierunkowe zależności funkcyjne (klasyczny wyraz daje temu Russell 1912), nie mogą się odnieść do tych faktów. Podobnie Hempla nomologiczno-dedukcyjny model wyjaśniania, gdzie konkretny przypadek wyjaśnia się przez subsumpcję pod prawo, nie potrafi zróżnicować takich przypadków (1998, s. 166-167; tłum. P. K.).

⁸ Nie podejmuję tu systematycznej dyskusji na temat innej tezy o symetrii, mianowicie między wyjaśnianiem a przewidywaniem. Dotyczyć jej będą jedynie pewne uwagi w punkcie 2. o kontrfaktycznym wnioskowaniu w bayesowskich modelach przyczynowych. Szczegółową analizę tej symetrii podaje Jan Woleński (1979/1996). Podsumowuje ją następująco: „symetria wyjaśniania i przewidywania jest istotną cechą metodologiczną systemów teoretycznych, a być może także kryterium odróżniającym systemy teoretyczne od wiedzy nieteoretycznej” (1979/1996, s. 265). Taka symetria – zdaniem S. Nowaka – przesądza o praktycznej doniosłości wyjaśnień przyczynowych (1985, s. 354).

ność funkcyjna między zmiennymi), stwierdza się wyjaśnienia przyczynowe tylko w przypadku tych pierwszych. Zależność między zmiennymi jest symetryczna tylko w przypadku praw współwystępowania, a te, zdaniem Hempla, nie są prawami przyczynowymi. Model DN jest symetryczny tylko w przypadku praw współwystępowania, ale ponieważ prawa te nie są przyczynowe, więc i oparte na nich wyjaśnienia nie są przyczynowe. Symetrii modelu DN nie można więc uznać za przejaw nieadekwatności do zależności przyczynowej, gdyż taka, w przypadku praw współwystępowania, nie zachodzi.

Jedno z praw dla gazów idealnych określa zależność ciśnienia i objętości niezależnie od czasu, jest więc prawem współwystępowania. Wiadomo jednak, że to kompresja gazu spowodowała wzrost jego ciśnienia. Georg H. von Wright (1973), w obronie modelu DN, próbował uchylić tę trudność. Symetrię tego modelu można złamać dla konkretnego zastosowania funkcyjnej zależności określonej w prawie współwystępowania, gdy jedna ze zmiennych (np. objętość gazu czy wysokość pnia) jest poddana manipulacji. Skoro w danym przypadku dokonano manipulacji na jednej ze zmiennych, to tylko ona może wystąpić w opisie warunków wstępnych w eksplanansie, a pozostałe, zależne od niej funkcyjnie, będą występować w eksplanandum. Uzupełnienie modelu DN o manipulowalność pozwoli uniknąć trudności z nieadekwatnością względem asymetrii przyczynowej:

(D*M) Wyjaśnić to wydedukować z prawa naukowego zastosowanego do zmiennej manipulowanej.

Formuła ta nie brzmi ładnie – i nie bez powodu. Pierwotne zamierzenie Hempla, jak odnotowaliśmy wcześniej, związane było z dedukowaniem (P) z (D*). Wzbogacona wersja (D*M) zawiera pojęcie „manipulowalności”, co oznacza definitywną klęskę pierwotnego zamierzenia Hempla, by (P) było elementem wtórnym, wyprowadzalnym z bardziej „obiektywnych”, a w zasadzie z formy logicznej wyjaśnienia.

Zarzuty z drugiej grupy wykazują, że DN nie stanowi warunku koniecznego wyjaśniania. Także w ostatecznej konkluzji prowadzą one do stwierdzenia, że pominięto w tym modelu istotne rozważania przyczynowe.

Tę grupę zarzutów zainicjował przykład Michaela Scrivena, ilustrujący fakt, że – przynajmniej niektóre – zadowalające wyjaśnienia nie muszą się odwoływać do praw naukowych, gdyż wystarczy, że określą przyczynę danego zdarzenia. Na przykład moje przypadkowe uderzenie kolanem w biurko jest wyjaśnieniem tego, skąd się wzięła abstrakcyjnie rozbiegana plama

atramentu na dywanie, mimo że nie podaję odnośnych praw Newtona czy innych (Scriven 1962, s. 90).

Oczywiście nie mamy tu do czynienia z sytuacją, w jakiej nie potrafiąc podać pełnego i zadowalającego wyjaśnienia, proponujemy tylko wyjaśnienie, określane przez Hempla jako „szkic wyjaśnienia”. Podane wyżej wyjaśnienie uznajemy za pełne (Scriven 1962, s. 94), twierdząc tym samym, że nie wszystkie wyjaśnienia muszą być nomologiczne, oparte na prawach naukowych. Stanowisko Scrivena, że istnieją wyjaśnienia konkretnych zdarzeń nie odwołujące się do praw naukowych, ma istotną presupozycję. Zależec będzie bowiem od metafizycznej tezy o istnieniu przyczynowości jednostkowej, a mianowicie o istnieniu zależności przyczynowych między konkretnymi zdarzeniami, których nie da się podporządkować zależnościom między odpowiadającymi im typami zdarzeń. Podany wyżej przykład musielibyśmy uzupełnić o (uzasadnioną na gruncie metafizyki przyczynowości) tezę, że moje uderzenie kolanem i strącenie kałamarza są do siebie w relacji, która nie jest szczególnym przypadkiem np. relacji, jakie zachodzą między uderzeniami powodującymi wstrząs biurka a spadaniem z niego przedmiotów.

Koncepcję, która uwzględnia wniosek z kontrprzykładu Scrivena, ale nie jego założenia metafizyczne, zaproponował ostatnio James Woodward; bardziej szczegółowo omówimy ją poniżej.

Jako uzupełnienie modelu DN Hempel zaproponował model dedukcyjno-statystyczny (DS). Jeśli wśród przesłanek znajduje się nomologiczne twierdzenie dotyczące regularności statystycznych (np. każdy atom C^{14} ma prawdopodobieństwo rozpadu w okresie 5730 lat wynoszące 0,5), to otrzymujemy szczególny przypadek wnioskowania o schemacie DN, w którym wniosek-ekspanandum jest pewną regularnością statystyczną. Jak zaobserwował Salmon (1990, s. 53), co nadal jest potwierdzane (Psillos 2002, s. 242), model DS, który skupiał szczególną uwagę Hempla w (1965), jest szczególnym rodzajem modelu DN, dlatego też dziedziczy omówione wyżej trudności.

Kolejne modele wyjaśniania projektowano, biorąc pod uwagę trudności modelu DN⁹, zwłaszcza z adekwatnością do zależności przyczynowych. Model SR (statystycznej relewancji) Salmona (1971) jest najważniejszą z pro-

⁹ P. Kitcher i W. C. Salmon tak podsumowują swoją analizę pragmatycznego modelu wyjaśniania B. van Fraassena (1977; 1980): „van Fraassen zaproponował piękną koncepcję pragmatyki wyjaśniania, która powinna być traktowana jako uzupełnienie raczej niż alternatywa w stosunku do tradycyjnych stanowisk w kwestii wyjaśniania” (1998, s. 188). Z tego też względu pragmatyczny model wyjaśniania van Fraassena nie będzie tu osobno rozpatrywany.

pozycji zgłoszonych po koncepcji Hempła. Podanie wyjaśnienia polega tu zasadniczo na określeniu, jakiego rodzaju cechy są statystycznie skojarzone z eksplanandum, a jakie nie. Podajmy uproszczony przykład wyjaśnienia w modelu SR. Pytamy, dlaczego osoba O , będąca nastolatkiem (N), jest przestępcą (P). Załóżmy, że jedyne cechy statystycznie relewantne dla P wśród N to płeć (K lub M) oraz miejsce zamieszkania (U – miasto, W – wieś). Otrzymujemy zatem partycję N na cztery możliwe klasy: NKU , NKW , NMU , NMW . Na wyjaśnienie składać się będą: (1) określenie prawdopodobieństwa bycia przestępcą dla całej klasy nastolatków (2) oraz dla każdej z czterech klas wyróżnionych; (3) określenie, do której z nich należy osoba O . W ten sposób uzyskamy informację, jaki zestaw cech okazał się statystycznie istotny dla wyjaśnianego faktu¹⁰.

Zwróćmy uwagę zwłaszcza na dwa podstawowe założenia modelu SR. Jedno z nich wyrażane jest przez (P), a drugie przez następującą formułę:

(S) Relewancja statystyczna odzwierciedla zależności przyczynowe.

Jeśli (S) rozumieć w ten sposób, że na podstawie relewancji statystycznej można w każdym przypadku jednoznacznie określić zależności przyczynowe, to (S) jest zdaniem fałszywym, co obszernie ilustruje wiele współczesnych prac (Pearl 2000; Spirtes, Glymour i Scheines 2000). Posłużmy się przykładem Salmona. Ciśnienie atmosferyczne A powoduje występowanie burz B oraz położenie strzałki S barometru. Oczywiście położenie strzałki nie jest przyczyną burzy, a więc S jest statystycznie nieistotne dla B , to znaczy $P(B | AS) = P(B | A)$. Natomiast A jest relewantne dla B , gdyż $P(B | AS) \neq P(B | S)$. Tym samym Salmon uznaje A za istotne dla wyjaśnienia B , a S za nieistotne. Jednakże z tymi zależnościami statystycznymi zgodnych jest więcej struktur przyczynowych niż ta opisana przez Salmona.

¹⁰ Kanoniczną postać modelu SR można scharakteryzować następująco. Pytanie o wyjaśnienie ma postać: Dlaczego x , będący członkiem klasy wyznaczonej przez cechę A , ma także cechę B ? Odpowiedź rozpoczyna się od wprowadzenia homogenicznej partycji A , czyli wyróżnienia klasy podzbiorów C_i zbioru A , które są wzajemnie rozłączne, a łącznie wyczerpujące. Partycja jest tak skonstruowana, że nie można uzyskać innej statystycznie relewantnej partycji żadnego z podzbiorów AC_i względem B , a więc nie ma innych cech D_k w A takich, że $P(B | AC_i) \neq P(B | AC_i D_k)$. Na właściwą część wyjaśnienia składają się: (1) określenie prawdopodobieństwa B w A , czyli $P(B | A) = p$; (2) określenie prawdopodobieństwa B dla każdego podzbioru A w homogenicznej partycji, czyli $P(B | AC_i) = p_i$ oraz (3) określenie, do której partycji należy x (Salmon 1971, s. 76-77).

Mianowicie odwrócenie zależności przyczynowej między A i S daje dokładnie takie same zależności statystyczne, jak w przykładzie Salmona, ale S byłoby wówczas istotne w wyjaśnieniu B, przynajmniej jako przyczyna pośrednia.

Ta prosta modyfikacja przykładu Salmona pokazuje, że znajomość relewancji statystycznej nie zawsze pozwala na jednoznaczne określenie kierunku zależności przyczynowej, a tym samym na określenie, co mogłoby pełnić funkcję eksplanansu – czy ciśnienie atmosferyczne, czy położenie strzałki barometru. Okazuje się więc, że model SR dziedziczy podobne do DN trudności z nieadekwatnością do asymetrii przyczynowej.

Salmon, uznając tę i inne trudności modelu SR, zaproponował inny model (1984). Model CM (przyczynowo-mechanicystyczny) zmierza do podania bogatszej charakterystyki zależności przyczynowych niż statystyczne. Zasadnicze pojęcie, które zostało tu wprowadzone, to *process przyczynowy*, rozumiany jako proces fizyczny, który może przenosić swoją strukturę w sposób ciągły, np. piłka (ale nie jej cień). *Interakcja przyczynowa* następuje wówczas, gdy jeden proces przyczynowy po zetknięciu z drugim zmienia jego strukturę, np. zderzenie cząstek, które zmienia ich kierunek oraz energię kinetyczną. Wyjaśnienie danego zdarzenia polega w modelu CM przede wszystkim na określeniu (przynajmniej części) procesów przyczynowych i interakcji prowadzących do niego oraz opisanu procesów i interakcji, które je tworzą. Na wyjaśnienie ruchu czerwonej i białej bili składa się to, że są one procesami przyczynowymi, podobnie jak kij bilardowy oraz podanie interakcji, jakie między nimi zaszły, a więc uderzenia kija w białą bilę, a tej w czerwoną. Cień zaś rzucany przez jedną z bil na drugą nie stanowi składowej wyjaśnienia, gdyż nie jest procesem przyczynowym.

Zasadniczy problem z tą propozycją Salmona polega na tym, że kategorie procesu i interakcji przyczynowej są zbyt ogólne, by wyodrębnić te własności – w tym przypadku kul bilardowych – które rzeczywiście są przyczynowo i eksplanacyjnie istotne, zwłaszcza masy i prędkości przed zderzeniem. Gdyby np. zdarzyło się tak, że niewielka ilość talku z kija pozostała na białej bili, a następnie przy uderzeniu część została na powierzchni bili czerwonej, wówczas – zgodnie z podanymi przez Salmona kryteriami procesu przyczynowego – talk ten należałoby uznać za jeden z elementów wyjaśnienia w modelu CM.

Z trudności obu modeli Salmona wynika pewien istotny wniosek w odniesieniu do formuły (P) – w większości przypadków nie uda nam się jednoznacznie wskazać właściwego modelu przyczynowego. Ogólna zatem charak-

terystyka wyjaśniania powinna być wolna od tego założenia. Ten wniosek rozwijam w punkcie 3.

Niekiedy próbuje się przywrócić model DN Hempla, poprzez uzupełnienie go jakimś dodatkowym elementem, uznając, że podaje on adekwatne warunki konieczne. Takie uzupełnienie ma na celu uchylenie zarzutów, że ten model nie podaje warunku wystarczającego wyjaśniania.

Najważniejszą z takich propozycji jest model unifikacjonistyczny¹¹:

(D*U) Wyjaśnić to wydedukować z unifikującego prawa naukowego.

Wyprowadzenie dedukcyjne z prawa naukowego – a szerzej, z uogólnienia o szerokim zakresie – Philip Kitcher uznaje za warunek konieczny, lecz niewystarczający wyjaśnienia. To wyjaśniające uogólnienie musi ponadto unifikować bogaty zestaw zjawisk. Warunek ten spełniają wzorcowe przykłady, jak dokonana przez Newtona unifikacja teorii ruchu ciał niebieskich i ziemskich czy Maxwella unifikacja zjawisk elektryczności i magnetyzmu.

Kitcher nie podał ścisłych kryteriów zadowalającego wyjaśnienia, jedynie w przybliżeniu określił, na czym ma polegać unifikacja. Im mniej różnych schematów wyjaśnienia używa się przy wyprowadzaniu możliwie największej liczby różnorodnych zjawisk oraz im ściślejsze ograniczenia nakładają one na to, co można z nich wyprowadzić, tym bardziej zunifikowane jest proponowane wyjaśnienie. Pomimo że nie doprecyzowuje on liczby wniosków i wzorców wyjaśnień ani ograniczeń, jak postąpić, gdyby te żądania prowadziły w odmiennych kierunkach, to jednak stwierdza, że w konkretnych przypadkach jest jasne, który z kandydatów na wyjaśnienie najlepiej spełnia te wymogi.

Czy długość cienia jest wyjaśnieniem wysokości pnia? Powszechnie przyjęte wzorce wyjaśniania w przypadku rozmiaru przedmiotów – np. gór, planet, organizmów czy artefaktów – odwołują się do „warunków, w których taki obiekt powstał, oraz modyfikacji, którym następnie był poddany” (Kitcher 1989, s. 485). Dodanie do tych wzorców kolejnego, który pozwalałby wyjaśniać rozmiary przedmiotów na podstawie cienia, jaki rzucają, jest niezgodne, jak twierdzi Kitcher, z wymogami modelu unifikacjonistycznego. Po pierwsze, byłyby to nowy wzorec dodany do obecnie istniejących, ale nie

¹¹ Podstawowy zarys tej koncepcji podał Michael Friedman (1974), systematycznie jednak tę koncepcję zaprezentował Philip Kitcher (1989); dlatego też model unifikacjonistyczny wiąże się z jego nazwiskiem.

pozwalaby na wyprowadzenie żadnych nowych wniosków. Po drugie, gdyby wśród dotychczas przyjętych wzorców wyjaśniania zastąpić wzorzec odwołujący się do powstania i późniejszych modyfikacji przez wzorzec odwołujący się do cienia, wówczas znacznie zmniejszyłaby się liczba przypadków, które można byłoby wyjaśnić. Nie można więc wzorca odwołującego się do cienia przedmiotu ani dodać do dotychczas przyjętych, ani wprowadzić go na miejsce wzorca odwołującego się do powstania i modyfikacji. W obu przypadkach okazuje się, że wzorzec odwołujący się do cienia jest mniej unifikujący niż dotychczas przyjęty. Cień pnia nie wyjaśnia jego wysokości.

Przy ocenie tego modelu zwróćmy uwagę na to, jaką przypisuje on rolę przyczynowości. Kitcher (1989, s. 477) twierdzi, że zależność przyczynowa jest wtórna względem zależności eksplanacyjnej. Niezależny od wyjaśniania porządek przyczynowy nie istnieje. Unifikacja eksplanacyjna nie zakłada żadnych pojęć przyczynowych, a określenie zależności przyczynowych opiera się na tej właśnie unifikacji.

W odniesieniu do omawianego wyżej przykładu można jednak zakwestionować adekwatność wniosku Kitchera dotyczącego przyczynowości. Przygodny fakt, że nie wszystkie przedmioty rzucają cień, nie może być – jak w modelu Kitchera – jedną z podstaw modelu wyjaśniania. Co byłoby bowiem, gdyby wszystkie przedmioty rzucały wystarczająco dużo cieni, aby określić wszystkie ich wymiary? W takim przypadku drugi z podanych wyżej argumentów Kitchera upada. Nie byłoby bowiem powodu, żeby nie zastąpić wzorca odwołującego się przy wyjaśnianiu rozmiaru przedmiotów do ich powstania i modyfikacji wzorcem odwołującym się do ich cieni. Ten ostatni zresztą byłby prostszy.

Ten problem modelu zaproponowanego przez Kitchera można ogólnie scharakteryzować następująco (Woodward 2002, s. 49). Pojęcie unifikacji, które ma wzbogacić Hempla model wyjaśniania (D^*), jest w ostateczności pojęciem opisowym (unifikacja jako ekonomia opisu czy kompresja informacji ma prowadzić do wywnioskowania jak najwięcej z jak najmniejszej liczby założeń lub przy jak najmniejszej liczbie schematów wyjaśniania) i nie uchroni (D^*) przed zarzutami nieadekwatności do asymetrii zależności przyczynowej. Zunifikowany opis, który nie zawiera informacji o zależnościach przyczynowych, występuje w licznych przykładach w nauce, np. w klasyfikacjach biologicznych, geologicznych i astronomicznych. Jeśli wiem, że dany organizm należy do określonej kategorii klasyfikacyjnej, np. jest ssakiem, to mogę na tej podstawie wiele wywnioskować na jego temat: że ma kręgosłup, serce, że jest żyworodny itd. Tego typu schematy zostaną jednak

uznane przez naukowców za opisowe¹², gdyż nie stanowią wyjaśnienia, np. nie podają przyczyn wyjaśniających, dlaczego ten organizm ma serce.

Innym, oprócz klasyfikacji, przykładem zunifikowanego ujęcia informacji opisowych w sposób ekonomiczny są różnorodne procedury statystyczne, np. analiza czynnikowa czy techniki skalowania wielowymiarowego. Na przykład, znając „obciążenie” każdego z n testów na inteligencję, można wyprowadzić $n \cdot (n-1)/2$ wniosków o wzajemnych korelacjach tych testów (Woodward 2002, s. 49). Taka unifikacja jednak nic nie mówi o zależnościach przyczynowych.

Współcześnie wśród metodologów i filozofów nauki dość powszechne jest przekonanie, że główny problem dotychczasowych modeli wyjaśniania tkwi w ich niedakwatości do zależności przyczynowych. Bardziej szczegółowy powód to – zdaniem Woodwarda – fakt, że dotychczasowe modele wyjaśniania nie „uwzględniają schematu kontrfaktycznej zależności między eksplanansem a eksplanandum” (2002, s. 50). Ponadto „zwykliśmy sądzić, że wyjaśnianie ma coś wspólnego z określeniem zależności przyczynowych, a – jak kontynuuje Woodward – nie da się zaprzeczyć, że między nimi a zależnościami kontrfaktycznymi zachodzi ścisły związek. Jeśli wyjaśnienia odwołują się do przyczyn, to wydaje się bardzo prawdopodobne, że jakaś postać kontrfaktycznej teorii wyjaśniania musi być poprawna” (2000, s. 210)

Tę ideę zwięźle wyraża kolejna formuła:

(K) Wyjaśnić to podać zależność kontrfaktyczną¹³.

Zmiana wysokości pnia (przy pozostawieniu pozostałych czynników przyczynowych bez zmian) wydłuży lub skróci długość jego cienia. Manipulacje zaś długością cienia, nie naruszające bezpośrednio samego pnia, nie doprowadzą do zmiany jego wysokości. Długość cienia jest więc kontrfaktycznie zależna od wysokości pnia, a tym samym jest też, zdaniem Woodwarda, przez nią wyjaśniona. Oczywiście, odwrotna zależność nie zachodzi, gdyż z dwóch możliwych tu zdań, wyrażających zależność kontrfaktyczną, prawdziwe jest tylko jedno: „Gdyby zmienić wysokość pnia, zmieniłaby się długość jego

¹² W odniesieniu do nauk społecznych stwierdzenie to szczegółowo uzasadnia Marek Styczeń (1977, s. 141-143).

¹³ Kontrfaktyczne teorie przyczynowości rozpowszechniły się we współczesnej filozofii, a także w metodologii nauk, pod wpływem Davida Lewisa (1973), który zainicjował także opracowanie formalnej semantyki zdań kontrfaktycznych.

cienia”. Stąd wynika także asymetria wyjaśniania, odzwierciedlająca asymetrię kontrfaktycznej zależności przyczynowej¹⁴.

Mimo zbieżności z konkluzją wcześniej omawianego kontrprzykładu Scrivena, model zaproponowany przez Woodwarda (1997; 2000; 2003) opiera się na kategoriach znacznie mniej metafizycznie angażujących. W zasadzie ta ostatnia propozycja polega na zastąpieniu praw naukowych w wyjaśnianiu przez relacje niezmiennicze¹⁵. Istotną cechą tego typu relacji jest to, że nie tylko zdają sprawę z zaistniałych regularności (tego, co się zdarza lub zdarzyło), lecz także pozwalają określić, co by się zdarzyło, gdyby podjąć określone interwencje. W przypadku wspomnianego wcześniej prawa dla gazu idealnego można powiedzieć, że odgrywa ono istotną rolę w wyjaśnianiu nie dlatego, że wzrost objętości gazu traktuje się jako nomicznie spodziewany, ale dlatego, że pozwala określić, jak zmieniłoby się jego ciśnienie, gdyby zmieniły się warunki. Jak podkreśla Woodward, ważnym elementem takiego wyjaśniania jest więc umieszczenie eksplanandum „w przestrzeni możliwych alternatyw” (2000, s. 209).

Model zaproponowany przez Woodwarda¹⁶ prowadzi do przemyślenia na nowo roli wyprowadzania dedukcyjnego w modelu wyjaśniania. Hempel w swoim modelu DN eksponował wyprowadzenie z prawa naukowego, gdyż to wyprowadzenie miało pokazać, dlaczego należało się spodziewać zdarzenia, o którym mowa w eksplanandum. Według Woodwarda wyprowadzenie z prawa naukowego jest – jak podkreśla, niekiedy tylko – skutecznym sposobem przekazania informacji o zależności kontrfaktycznej. Wyprowadzenie wysokości pnia z długości cienia na podstawie praw optyki geometrycznej jest przykładem na to, że nie wszystkie wyprowadzenia zgodne ze schematem wprowadzonym przez Hempla w modelu DN podają zależność kontrfaktycz-

¹⁴ Występujący tu implicite warunek nakładany na zdania kontrfaktyczne występujące w wyjaśnieniach Woodward wyraża następująco: „zdania kontrfaktyczne istotne dla wyjaśniania mają poprzedniki, które stają się prawdziwe dzięki specjalnemu rodzajowi egzogenicznego procesu przyczynowego, który nazywam interwencją. [...] interwencje można traktować jak manipulacje, które człowiek mógłby przeprowadzić w wyidealizowanych warunkach eksperymentalnych” (1997, s. 29). To, czy jest możliwa taka ludzka interwencja, jest faktem przygodnym i nie przesądza o związku przyczynowości z wyjaśnianiem. Por. też Nowak 1985, s. 327-339; Brzeziński 1996, s. 280-286.

¹⁵ Zaletą tego podejścia jest to, że stosuje się do tych nauk szczegółowych, w których można wskazać zależności inwariantne, ale nie można przywołać bezwyjątkowych regularności (praw).

¹⁶ Zasadnicze elementy tej koncepcji wyjaśniania można też odczytać w pracach S. Nowaka (zwł. 1985, s. 351-381).

ną. Schemat wyprowadzenia eksplanandum, określony w tym modelu, nie jest więc warunkiem koniecznym zależności kontrfaktycznej.

Nie jest także jej warunkiem wystarczającym, co podkreśla Woodward. Są bowiem inne sposoby ukazania zależności kontrfaktycznej niż wyprowadzenie z prawa naukowego. I tak moje uderzenie kolanem w biurko jest wyjaśnieniem strącenia kałammarza, gdyż zachodzi w tym przypadku (asymetryczna) zależność kontrfaktyczna. Zaznaczymy za Woodwardem, że:

Inne sposoby reprezentacji, takie jak diagramy oraz grafy, w podobny sposób przekazują informacje o zależności przyczynowej bez wyraźnego wprowadzania wyprowadzeń (2002, s. 51).

Zastanawiając się nad perspektywami badań w teorii wyjaśniania, Woodward dodaje o tych środkach reprezentacji:

[...] są zarówno bogatsze niż sposoby reprezentacji standardowo wykorzystywane przez filozofów (logika, teoria prawdopodobieństwa bez żadnych dodatków), jak i bliższe technikom wykorzystywanym w samej nauce. Ani logika, ani teoria prawdopodobieństwa nie może ująć modalnych i kontrfaktycznych elementów, które są zasadnicze dla wyjaśniania (2002, s. 51).

Przejdźmy zatem do bliższego studium sposobów reprezentacji, które Woodward – podobnie jak wielu innych współczesnych teoretyków wyjaśniania i przyczynowości¹⁷ – uznaje za tak obiecujące.

2. MODELOWANIE PRZYCZYNOWOŚCI W SIECIACH BAYESOWSKICH

Bardziej szczegółowe rozważanie formuły (K), symbolizującej model wyjaśniania kontrfaktycznego, poprzedzimy analizą technik modelowania przyczynowości, które okazują się mieć daleko idące implikacje dla filozofii i metodologii nauk¹⁸.

¹⁷ Do tego grona zaliczają się przede wszystkim: polemizująca z Woodwardem Nancy Cartwright (2002), jego bliski współpracownik, Daniel M. Hausman (1997; 1998), oraz kolega z California Institut of Technology, Christopher Hitchcock (2003; 2004).

¹⁸ Przede wszystkim dla teorii przyczynowości, ale w konsekwencji także – jak staram się tu wykazać – dla koncepcji wyjaśniania oraz innej podstawowej kategorii metodologii nauk, a mianowicie prawa naukowego. Por. Hitchcock 2004; Psillos 2002; Woodward 2003.

Zacznijmy od klasycznego przykładu eksperymentu (Wainer 1989), w którym fumiganty glebowe (F) wykorzystuje się do kontrolowania populacji węgorka (W), co ma wpłynąć na wzrost plonów z uprawy owsa (P). Fumiganty mogą wpływać (zarówno pozytywnie, jak negatywnie) na plony upraw owsa bezpośrednio, niezależnie od zmian populacji węgorka. Zasadniczy problem dotyczy w tym przypadku określenia całkowitego wpływu fumigantów na plony upraw (a więc bezpośrednio i pośrednio – przez regulację populacji węgorka), przy uwzględnieniu komplikacji, jakie są spowodowane występowaniem innych czynników.

Najprostszy sposób polegałby na przeprowadzeniu kontrolowanego eksperymentu. To jednak często w warunkach naturalnych jest niewykonalne, gdyż rolnicy sami chcą decydować, które pola zostaną poddane fumigacji. Ponadto decyzję o ilości zastosowanych fumigantów podejmują, biorąc pod uwagę wielkość populacji węgorka w roku poprzednim (F_0), a to jest nie tylko wielkością nieznaną, lecz także mocno skorelowaną z populacją w roku bieżącym.

Można jednak wykorzystać fakt, że analiza laboratoryjna próbek gleby pozwala ustalić populację węgorka przed (W_1) i po (W_2) zastosowaniu fumigantów. Ponadto, ponieważ fumiganty są aktywne tylko przez krótki okres, więc można bezpiecznie założyć, że nie mają wpływu na wzrost tej populacji węgorka (W_3), która przetrwa fumigację. Ten wzrost zależy więc od populacji ptaków i innych drapieżników (D), co jest skorelowane z zeszłoroczną populacją węgorka, a więc i samą fumigacją.

Na podstawie tych rozważań można, posługując się procedurą opisaną przez Judea Pearla (2000, s. 49-54) oraz Petera Spirtesa, Clarka Glymoura i Roberta Scheinesa (2000, s. 73-155), skonstruować diagram przyczynowy¹⁹, który podano na rysunku poniżej (por. Pearl 1995, s. 669-670).

¹⁹ Szczegółowe omówienie konstrukcji różnych, także nieeksperymentalnych, typów modeli podaje Z. Brzeziński (1996, s. 343-430).

Wielkości mierzalne, takie jak populacja węgorka przed (W_1) i po fumigacji (W_2) oraz po zakończeniu sezonu (W_3), a także plony upraw owsa (P) i ilość fumigantów (F), reprezentowane są przez kropki. Wielkości nieznane (lub niemierzalne), takie jak W_0 oraz populacja ptaków (i innych drapieżników) D , są reprezentowane przez małe okręgi.

Zmienne na diagramie połączone dwoma rodzajami krawędzi: liniami ciągłymi w przypadku zmiennych mierzalnych i liniami przerywanymi w przypadku zmiennych nieznanymi (lub niemierzalnymi). W obu przypadkach strzałka sygnalizuje kierunek zależności przyczynowej. Każda z takich strzałek reprezentuje autonomiczny, tj. całkowicie niezależny od pozostałych, mechanizm przyczynowy. Brak połączeń między zmiennymi oznacza negatywne założenie przyczynowe o nieistnieniu między nimi związku przyczynowego. Na przykład brak krawędzi między W_1 a P oznacza, że populacja węgorka nie wpływa na plony z upraw przed fumigacją, a cały wpływ jest zapośredniczony przez tę populację po fumigacji, W_2 oraz W_3 .

Wykorzystując wypracowane przez Pearl'a reguły (2000, s. 85-89), można określić całkowity wpływ fumigantów na plony upraw:

$$P(p|f) = \sum_{w_1} \sum_{w_2} \sum_{w_3} P(p|w_2, w_3, f) P(w_2|w_1, f) \sum_f P(w_1, f), \quad (1)$$

gdzie $P(p|f)$ oznacza prawdopodobieństwo osiągnięcia plonów na poziomie $P = p$, pod warunkiem, że fumigacja jest ustalona (przez niezależną interwencję) na poziomie $F = f$.

Uogólniając powyższy przykład, przechodzę do podania ogólnej postaci modelu przyczynowego²⁰. *S t r u k t u r a p r z y c z y n o w a* $D = \langle V, L \rangle$ składa się ze zbioru zmiennych V oraz zbioru zależności funkcyjnych, jakie zachodzą między nimi, L o postaci $X_i = f_i(PA_i, U_i)$ – gdzie $PA_i \subset V$ jest zbiorem rodziców X_i (Pearl 2000, s. 14), a U_i jest losowym zaburzeniem danej zależności funkcyjnej²¹. Można to zaburzenie opisać także jako podsumowanie informacji o tych czynnikach przyczynowych wpły-

²⁰ Systematyczną dyskusję różnych pojęć modelu występujących w naukach empirycznych podaje Ryszard Wójcicki (1979). Z wyróżnionych pojęć modelu przyjęte w tej pracy najbliższe jest kategorii „obiektów matematycznych jako modeli operacyjnych” (Wójcicki 1974, s. 290-294). Por. też Mejbaum 1995, s. 146-149.

²¹ Za J. Pearl'em przyjmuję konwencję stosowania ujednoczonego zapisu dla zbiorów zmiennych i zmiennych, jeśli w danym kontekście nie powoduje to błędnej interpretacji (por. Pearl 2000, s. 9).

wających na daną zmienną, które nie zostały uwzględnione jako zmienne modelu (por. Hausman i Woodward 2004, s. 149).

(Bayesowski) model przyczynowy – za Pearlem (2000, s. 44) – definiuje następująco:

(M) *M o d e l e m p r z y c z y n o w y m* jest para $M = \langle D, \Theta_D \rangle$, którą tworzą: struktura przyczynowa D oraz zbiór parametrów Θ_D zgodnych z D . Parametry Θ_D każdej zmiennej $X_i \in V$ przypisują funkcję $X_i = f_i(\text{PA}_i, U_i)$ oraz każdej zmiennej $U_i \in V$ przypisują miarę prawdopodobieństwa $P(U_i)$, przy założeniu, że każda ze zmiennych U_i ma rozkład niezależny od pozostałych.

Utworzenie takiego modelu pociąga za sobą określenie łącznego rozkładu $P(M)$ na zmiennych w nim występujących. Ten rozkład odzwierciedla niektóre własności struktury przyczynowej, przede wszystkim zerowe korelacje cząstkowe między zmiennymi, czego graficznym odpowiednikiem będzie brak strzałek na grafie. I tak na rysunku 1 brak strzałki między F a W_3 oznacza, że – zgodnie z danym rozkładem – F jest niezależne probabilistycznie od W_3 , wzięwszy pod uwagę W_2 . Mówiąc inaczej, fumigacja oddziałuje na populację węgora tylko w krótkim okresie po jej zastosowaniu, ale pozostaje bez wpływu na jej dalszy rozwój.

Jednym z założeń konstruowania tak rozumianych modeli przyczynowych jest to, że zaobserwowane rozkłady traktuje się jako *s t a b i l n e* (Pearl 2000, 48). Niech $I(P)$ oznacza zbiór wszystkich relacji niezależności probabilistycznej, które są implikowane przez P . $I(P(\langle D, \Theta_D \rangle))$ jest *s t a b i l n e* w tw $I(P(\langle D, \Theta_D \rangle)) \subseteq I(P(\langle D, \Theta'_D \rangle))$.

Innymi słowy, stabilność oznacza, że zmiana parametrów z Θ na Θ' nie prowadzi do zmiany zbioru relacji niezależności probabilistycznych. Dzięki takiej własności rozkładu można wyodrębnić te niezależności, które są strukturalne, a więc nie zależą od takich, a nie – nawet nieco – innych wartości funkcji i rozkładów²².

²² Ilustracją może być następujący model. Załóżmy, że $z = f_1(x, z_1)$ oraz $y = f_2(x, u_2)$. W tym przypadku zmienne Z i Y są niezależne ze względu na X dla wszystkich funkcji f_1 oraz f_2 . Gdyby do tego modelu dodać zależność $y = f_3(z, u_3)$ oraz przyjąć liniową postać funkcji: $z = \gamma x + u_1$, $y = \alpha x + \beta z + u_2$, gdzie ze względu na wartości poszczególnych parametrów akurat jest tak, że $\alpha = -\beta\gamma$, wówczas Y i X byłyby niezależne. Taka jednak niezależność jest niestabilna, gdyż znika, gdy tylko parametry w równaniu $\alpha = -\beta\gamma$ zmienią nieco swoje wartości (por. Pearl 2000, s. 49).

Bayesowskie modele przyczynowości pozwalają na dwojakiemu rodzaju wnioskowania. Pierwsze dotyczą sytuacji nieeksperymentalnych, gdy dokonuje się przewidywania na podstawie związków przyczynowych, które są zawarte w modelu. Takie wnioskowanie w zasadzie w niczym nie odbiega od wnioskowania w sieciach bayesowskich, gdzie zasadniczym twierdzeniem, pozwalającym uaktualniać informacje na podstawie danych doświadczenia, jest twierdzenie Bayesa (por. Kawalec 2003).

Drugie dotyczą sytuacji, gdy interwencja na zmienną X modelu wprowadza modyfikację w związki przyczynowe. Przeprowadzenie wnioskowania wymaga wykorzystania odpowiednich reguł, których zastosowanie z kolei zależne jest od zależności probabilistycznych zachodzących w odpowiednich modyfikacjach danego modelu. Odwołując się do modelu w postaci równań, można powiedzieć, że usunięto z nich wszystkie, w których X jest zmienną zależną, a w pozostałych uwzględniono konkretną wartość, jaką zmienna ta uzyskała poprzez interwencję.

Przyjmując, że przed i po interwencji pozostałe związki przyczynowe nie uległy zmianie oraz że w obu przypadkach mamy te same czynniki zakłócające, otrzymujemy podstawę wnioskowań kontrfaktycznych, dzięki którym można np. przewidzieć efekty określonego działania (Pearl 2000, s. 219). I tak, w powyżej dyskutowanym przykładzie, aby określić całkowity wpływ fumigacji na wydajność upraw owsa przy ustaleniu fumigacji na poziomie $F = f$ przez – niekoniecznie ludzką – interwencję, pomija się zależność funkcyjną F od W_0 , a w pozostałych zależnościach zastępuje się F przez $F = f$. Rezultatem tego kontrfaktycznego wnioskowania jest równanie (1), które określa interesujący nas wpływ fumigacji na uprawy owsa w sposób zależny tylko od danych doświadczenia.

3. WYJAŚNIANIE A KLASY MODELI

Skoro u podstaw modeli bayesowskich leży idea reprezentacji wnioskowań o zależnościach kontrfaktycznych, to w trywialny sposób spełniają one postulat (K) Woodwarda. Wyraża to następująca formuła:

(MP) Wyjaśnić to podać model przyczynowy.

Z punktu widzenia badań empirycznych istnieje zasadnicza trudność z takim postulatem. Otóż naukowiec dysponuje jedynie próbkami zdarzeń, na

podstawie których dokonuje oszacowania rozkładu $P_{[O]}$ dla zmiennych zaobserwowanych $O \subseteq V$. Dopiero w ten sposób przystępuje się do określenia zależności przyczynowych i ostatecznego sformułowania modelu. W zasadzie jednak zbiór zmiennych V jest nieznany, a istnieje wiele modeli, które byłyby zgodne z zaobserwowanym rozkładem. Różnice między nimi sprowadzałyby się do odmiennych zależności przyczynowych oraz występowania innych zmiennych ukrytych (nieobserwowalnych). Jak podkreśla Pearl: „bez żadnych ograniczeń na rodzaj modeli branych pod uwagę naukowiec nie będzie w stanie określić struktury badanych zjawisk” (2000, s. 45).

Formuła (MP) prowadziłyby zatem do sytuacji, w której rywalizują ze sobą różne modele i każdy z nich określany byłby – zgodnie z (MP) – jako wyjaśnienie. Wydaje się to zasadniczo niezgodne z intuicyjnym rozumieniem wyjaśniania. Czy można jednak podać lepszą wersję (MP)?

W wielu omówionych w punkcie 1. koncepcjach wyjaśniania przyjmowano implicite założenie, że można je d n o z n a c z n i e zidentyfikować poprawne wyjaśnienie. Trudności, jakie rozważałem w odniesieniu do modeli zaproponowanych przez Salmona (część z nich uznał on sam), wyraźnie jednak wskazują – jak skonkludował Woodward – że jest to zbyt wygórowane oczekiwanie i należałoby je raczej zastąpić odniesieniem do pewnej klasy alternatywnych modeli²³.

W swojej krytyce algorytmów, które Peter Spirtes, Clark Glymour i Robert Scheines zaproponowali jako mechanizm generujący bayesowskie modele przyczynowe z danych nieeksperymentalnych, David Freedman podkreśla, że „te algorytmy prowadzą do estymacji nie modeli liniowych, lecz klas równoważnych modeli liniowych” (1997b, s. 180-181). Różnica jest zasadnicza, gdyż o tych drugich nie można jednoznacznie twierdzić – na co zwróciłem wcześniej uwagę za Pearlem – że w każdym przypadku poprawnie została zidentyfikowana zależność przyczynowa (Freedman 1997a, s. 144-145).

Spśród różnych analiz Freedmana, wykazujących niejednoznaczność wniosków przyczynowych na podstawie danych doświadczenia, najbardziej przekonujące – jak sądzę – jest studium związku przyczynowego między spożyciem soli a ciśnieniem krwi. Freedman oraz Diana Petitti analizują „hi-

²³ S. Kamińskiego studium wyjaśniania, zwłaszcza przyczynowego, potwierdza takie rozwiązanie: „Trudno [...] byłoby to zrobić [tj. wskazać jednoznacznie właściwą przyczynę] w wyjaśnianiu rzeczywistości wziętej w aspekcie kwalitatywnym lub kwantytatywnym, chyba że w wyjątkowo prostej sytuacji. Zawsze bowiem zachodzi ogromne bogactwo (wielość i różnorodność) czynników sprawczych i warunków” (1989, s. 175).

potęgę dotyczącą soli”, którą zwięźle wyrażają: „Powszechnie sądzi się, że spożycie soli prowadzi do podwyższonego ciśnienia krwi oraz większego ryzyka ataku serca czy udaru” (2002, s. 1). Jej konsekwencją jest wniosek, że spożycie soli powinno być drastycznie ograniczone. Szczegółowo analizując podstawowe rodzaje dowodów, które mają potwierdzać „hipotezę dotyczącą soli”, a mianowicie eksperymenty na zwierzętach i ludziach oraz badania obserwacyjne na ludziach (tzw. program Intersalt przeprowadzony w 52 krajach), wykazują, że nie dają wystarczającego potwierdzenia tej hipotezie. Mianowicie w przypadku osób młodych (tj. do 20 roku życia) korelacja wydaje się odwrotna (choć nie w 4 egzotycznych spośród 52 krajów), a efekt dla osób o normalnym ciśnieniu krwi – niezauważalny. Jeśli zaś chodzi o wpływ soli na zdrowie, to daje się on zauważyć zwłaszcza u osób otyłych, ale jest różnokierunkowy.

Ogólny wniosek zatem brzmiałby, że istnieje wiele modeli, które są obserwacyjnie zgodne z danymi, jakich dostarczają dotychczasowe badania tej hipotezy. Wśród nich znalazłby się także model zgodny z „hipotezą dotyczącą soli”, ale tylko po odpowiednim przekształceniu, tj. po zawężeniu wyjaśnianego obszaru zjawisk.

Wziąwszy pod uwagę powyższe rozważania, dwa wyróżnione na początku tego artykułu podstawowe warunki zadowalającego wyjaśnienia, a mianowicie uwzględnienie zależności przyczynowych oraz zgodność z doświadczeniem, wyrazimy formułą:

(W) Wyjaśnić to podać klasę równoważnych obserwacyjnie modeli przyczynowych.

Aby wyeksplikować „równoważność obserwacyjną” w terminologii, w której sformułowaliśmy model przyczynowy, wprowadzę kilka pojęć pomocniczych. S t r u k t u r ą u k r y t ą SU nazywamy parę $\langle D, O \rangle$ (gdzie $O \subseteq V$ jest zbiorem zaobserwowanych zmiennych). Struktura ukryta $SU = \langle D, O \rangle$ jest p r e f e r o w a n a w stosunku do innej struktury ukrytej $SU' = \langle D', O \rangle$ wtw dla dowolnego zbioru parametrów Θ_D zgodnych z D istnieje $\Theta'_{D'}$ takie, że $P_{[O]}(\langle D', \Theta'_{D'} \rangle) = P_{[O]}(\langle D, \Theta_D \rangle)$.

Wybór prostoty, który jest nakładany w powyższym określeniu, dotyczy siły ekspresyjnej danej struktury, a nie jej opisu syntaktycznego. Może być bowiem tak, że jedna struktura ukryta wymaga znacznie większej liczby parametrów niż inna, a jednak to właśnie ona będzie preferowana, gdyż może się dostosować do bogatszego zbioru rozkładów prawdopodobieństwa na

zmiennych obserwowalnych. Jednym z powodów, dla których naukowcy wybierają prostsze teorie, jest to, że takie teorie są bardziej ograniczające, a tym samym bardziej podatne na falsyfikację. Dają bowiem naukowcom mniej okazji do zbyt dokładnego dopasowania do danych *ex post*, a przez to większą wiarygodność zyskują takie, dla których dopasowanie zostanie rzeczywiście stwierdzone (Popper 1934/2002)²⁴.

Kolejne dwa pojęcia, które należy tu wprowadzić, to: struktura ukryta minimalna oraz zgodna z rozkładem. SU jest minimalna w stosunku do klasy struktur ukrytych w tym żaden członek tej klasy nie jest preferowany w stosunku do SU.

Ponadto $SU = \langle D, O \rangle$ jest zgodna z rozkładem P^* na zmiennych O , jeśli istnieje taka parametryzacja Θ_D , że $P_{|\Theta|}(\langle D, O \rangle) = P^*$.

Dwa modele są więc obserwowalne równoważne ze względu na zaobserwowany rozkład P w tym składające się na nie (ukryte) struktury przyczynowe są minimalne oraz zgodne z P^* ²⁵. Podana wyżej formuła (W) odwołuje się do klasy zawierającej wszystkie takie modele ze względu na dane doświadczenia, jakimi dysponujemy. Zmniejszenie liczby modeli należących do tej klasy, a więc zgodnych z danymi doświadczenia, może się dokonać jedynie poprzez uzyskanie dodatkowych informacji o uprzedniości czasowej jednych zmiennych względem innych lub dzięki przeprowadzeniu odpowiedniej interwencji.

Nie ma oczywiście gwarancji, że prawdziwy model przyczynowy należy do klasy modeli równoważnych obserwacyjnie. Każdy model, poza tą klasą, będzie jednak wymagał bardziej wymyślnych korektur parametrów – często wprowadzanych *ex post* – aby pasował do danych.

Na ogół klasa modeli zgodnych z (zaobserwowanym) rozkładem będzie bogatsza niż jeden tylko prawdziwy model przyczynowy. Dotyczy to przede wszystkim badań nieeksperymentalnych, w których nieeksperymentalny roz-

²⁴ Zbiór niezależności probabilistycznych, które wynikają z danego modelu, nakłada ograniczenia na jego siłę ekspresji. Niekiedy więc test preferencji i równoważności obserwacyjnej sprowadza się do sprawdzenia implikowanych niezależności probabilistycznych, co z kolei można określić na podstawie topologii grafu reprezentującego ten model. W przypadku jednak modeli ze zmiennymi ukrytymi prawdziwość ta nie zachodzi, gdyż – jak wykazali Verma i Pearl (1990) – niektóre struktury ukryte nakładają na zaobserwowany rozkład ograniczenia numeryczne, a nie niezależnościowe. W takich przypadkach ocena preferencji jest zadaniem trudniejszym.

²⁵ Modele te ponadto są identyfikowalne, tzn. wpływ przyczynowy jednej zmiennej na drugą daje się jednoznacznie określić, jeśli wziąć pod uwagę strukturę tego modelu oraz dane doświadczenia. Precyzyjne definicje podaje Pearl (2000, s. 77).

kład może nie zdeterminować kierunku zależności przyczynowej między zmiennymi oraz może prowadzić do błędnego oszacowania współczynników takiej zależności lub pozostawić dodatkowe, zbędne zmienne w modelu (tzw. problem wspólnej przyczyny).

Oprócz dopuszczenia alternatywnych modeli, (W) w zasadzie zachowuje pierwotne intuicje związane z (P) oraz (D) jako podstawami teorii wyjaśniania, którą zarysował Arystoteles. Jest jednak ważny wyjątek, na który chcę zwrócić uwagę²⁶.

Wprowadzony przez Arystotelesa sposób rozumienia (P) nieuchronnie prowadzi do przyjęcia naturalistycznych tez metafizycznych²⁷. Jak podkreśla Hankinson:

Przyczyny celowe [które są podstawowym spośród czterech rodzajów wyróżnionych przez Arystotelesa] są więc częścią rzeczywistości w tym sensie, że dążenie do formy, które reprezentują, jest bezpośrednio wpisane w strukturę rzeczy. Nie są one jakimiś nieuchwytnymi jeszcze-nie-zrealizowanymi przedmiotami, które w jakiś tajemniczy sposób obdarzone byłyby mocą przyczynową; są one raczej nastawionymi na przyszły rozwój elementami zaczątkowej struktury organizmów – struktury, której rzeczywiste istnienie pozwala Arystotelesowi na odrzucenie poglądu, że Wszechświat jest kontrolowany przez opatrnościową rękę dobrodusznego bóstwa, a jednocześnie nieredukowanie go do tego, co – jak przynajmniej on uznawał – jest absurdalną przypadkowością czystego mechanizmu (1998, s. 146).

Wprowadzony model wyjaśniania (W) nie pozwoliłby na tego typu konkluzje jako wykraczające poza zgodność z danymi doświadczenia, na podstawie których można co najwyżej porównywać modele empirycznie równoważne. Podobnie poza kryteriami zadowalającego wyjaśnienia pozostają wszelkie opowieści sięgające do tego, co miałyby być ukryte „głębiej” pod danymi doświadczenia.

Strukturalny charakter zależności przyczynowych w bayesowskich modelach przyczynowości gwarantuje ich niezmienniczość na manipulacje. Ta niezmienniczość jednak nie jest cechą epistemologiczną tych modeli, nie jest jedynie testem zachodzenia związku przyczynowego między dwiema zmiennymi. Nie ma bowiem, na co zwrócił uwagę Woodward (2000, s. 204-205), niebanalnego alternatywnego sposobu określenia warunków prawdziwości dla

²⁶ Inna różnica to dopuszczalność wyjaśnień przyczynowych dla zdarzeń jednostkowych w bayesowskich modelach przyczynowości (Pearl 2000), zgodnie z postulatem D. Davidsona (1967/2001, s. 161-162). Por. też Mellor 1995, s. 130-131.

²⁷ Bardziej konsekwentny od Arystotelesa w wyprowadzaniu naturalistycznych implikacji teorii wyjaśniania był jego wybitny uczeń, Teofrast (por. Hankinson 1998, s. 191-192).

zdań wyrażających zależności przyczynowe. W związku z tym, jeśli chcemy w charakterystyce metafizycznej mówić o tym, *czym jest* przyczynowość, to – pozostając w zgodzie z doświadczeniem – musimy się odwołać jedynie do pojęć interwencji i niezmienniczości.

W tekście *Różnorodność przyczyn a antynaturalizm metodologiczny* starałem się wykazać, że przy zastosowaniu (W), wśród dopuszczalnych metodologicznie wyjaśnień naciśnięcia przycisku i skierowania pilota w kierunku panelu odbiornika telewizyjnego, znajdują się takie, które zawierają zmienne reprezentujące stany mentalne²⁸. Jeśli nawet zignorować techniki psychometrii w zastosowaniu do pomiaru psychopatologii (Farmer, McGuffin i Williams 2002), gdzie stany mentalne traktuje się jak zmienne mierzalne, to i tak można na podstawie (W) podać ogólną konkluzję, że do klasy modeli przyczynowych, którą uznamy za metodologicznie satysfakcjonujące wyjaśnienie, należeć będą takie modele, które zawierają zmienne ukryte – w tym przypadku stany mentalne – a przy tym są identyfikowalne oraz zgodne z danymi doświadczenia²⁹.

Analogicznie można argumentować, że (W) nie implikuje „metafizycznego pluralizmu nomologicznego”, który Nancy Cartwright (1999, s. 31) rozumie jako „doktrynę, że naturą rządzą różne systemy praw w różnych dziedzinach, niekoniecznie ze sobą powiązane w jakiś systematyczny czy jednorodny sposób; składanka praw (*patchwork of laws*)”.

Na zakończenie chciałbym podkreślić pewną własność koncepcji (W), która może się stać źródłem nieporozumień. Niezależnie od tego, czy wyjaśnienia traktuje się jak argumenty, czy też nie (Salmon 1998, s. 102-104), intuicyjne rozumienie tego pojęcia wiąże się z pewnego rodzaju wnioskowaniem. To z kolei może prowadzić do pozornej trudności: że podanie modelu nie jest związane z wnioskowaniem, a więc nie może być wyjaśnieniem adekwatnym. Przypomnijmy więc to, co zostało wyraźnie przedstawione w p. 2 – bayesowskie modele przyczynowości są *j e d n o c z e ś n i e* mechanizmem inferencyjnym, który pozwala przeprowadzać wnioski przyczyno-

²⁸ Dokładniej, podważając argument wykluczający moc przyczynową stanów mentalnych, bronię tezy antynaturalizmu eksplanacyjnego w postaci następującej: „modeli intencjonalnych nie da się zastąpić modelami neuronalno-fizjologicznymi bez uchylenia zgodności z danymi doświadczenia (adekwatności empirycznej)” (Kawalec 2004, s. 4).

²⁹ Funkcjonowanie tego typu modeli przyczynowych w naukach społecznych szczegółowo omawia S. Nowak (1985, s. 373-381).

we, a w przypadku modeli identyfikowalnych także sprawdzać zgodność tych wniosków z danymi doświadczenia³⁰.

Jak Arystoteles, wyjaśniamy to, co dane w doświadczeniu, przez przyczyny, ale – wbrew niemu – niezwiązani naturą.

BIBLIOGRAFIA

- B r o m b e r g e r S. (1966), Why-questions, w: *Mind and cosmos*, red. R. G. Colodny, Pittsburgh: University of Pittsburgh Press, s. 86-111.
- B r z e z i ń s k i Z. (1996), *Metodologia badań psychologicznych*, Warszawa: PWN.
- C a r t w r i g h t N. (1999), *The dappled world. A study of the boundaries of science*, Cambridge: Cambridge University Press.
- C a r t w r i g h t N. (2002), Against modularity, the Causal Markov Condition and any link between the two: comments on Hausman and Woodward, „*The British Journal for the Philosophy of Science*”, nr 53, s. 411-53.
- D a v i d s o n D. (1967/2001²), Causal relations, w: *t e n ż e*, *Essays on actions and events*, Oxford: Clarendon Press, s. 149-162.
- F a r m e r A., M c G u f f i n P., W i l l i a m s J. (2002), *Measuring psychopathology*, Oxford: Oxford University Press.
- F r e e d m a n D. (1997a), From association to causation via regression, w: *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences*, red. V. R. McKim, S. P. Turner, Notre Dame: University of Notre Dame Press, s. 111-161.
- F r e e d m a n D. (1997b), Rejoinder to Spirtes and Scheines, w: *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences*, red. V. R. McKim, S. P. Turner, Notre Dame: University of Notre Dame Press, s. 176-182.
- F r e e d m a n D. A., P e t i t t i D. B. (2002), *Salt and blood pressure: conventional wisdom reconsidered*. Technical Report 573. Department of Statistics, Berkeley: University of California.
- F r i e d m a n M. (1974), Explanation and scientific understanding, „*Journal of Philosophy*”, nr 71, s. 5-19.

³⁰ Podstawy takiego nieporozumienia w przypadku ogólnym usuwa R. Wójcicki: „Istnieje wyraźne pokrewieństwo między pojęciem modelu oraz pojęciem prawa. Każde prawo jest w istocie rzeczą modelem zjawiska, którego dotyczy, z uwagi na ten typ prawidłowości, pod który podpada prawidłowość opisywana prawem” (1974, s. 300).

- H a j d u k Z. (1970), Wyjaśnianie dedukcyjne, „Roczniki Filozoficzne”, 18, z. 3, s. 69-99.
- H a n k i n s o n R. J. (1998), Cause and explanation in Ancient Greek thought, Oxford: Clarendon Press.
- H a u s m a n D. M. (1997), Causation, agency, and independence, „Philosophy of Science”, 64, Supplement, s. 15-25.
- H a u s m a n D. M. (1998), Causal asymmetries, Cambridge: Cambridge University Press.
- H a u s m a n D. M., W o o d w a r d J. (2004), Modularity and the causal Markov condition: a restatement, „The British Journal for the Philosophy of Science”, nr 55, s. 147-161.
- H e m p e l C. G. (1965), Aspects of scientific explanation, New York: The Free Press.
- H e m p e l C. G., O p p e n h e i m P. (1965), Studies in the Logic of Explanation, w: C. G. H e m p e l, Aspects of scientific explanation, New York: The Free Press, s. 245-296.
- H i t c h c o c k C. (2003), Of Humean bondage, „The British Journal for the Philosophy of Science”, nr 54, s. 1-25.
- H i t c h c o c k C. (2004), Prediction versus accommodation and the risk of overfitting, „The British Journal for the Philosophy of Science”, nr 55, s. 1-34.
- K a m i ń s k i S. (1989), Wyjaśnianie w metafizyce, w: t e n ż e, Jak filozofować?, Lublin: TN KUL, s. 151-176.
- K a w a l e c P. (2003), Zagadnienia metodologiczne w bayesowskiej teorii konfirmacji, „Roczniki Filozoficzne”, 51, z. 1, s. 113-142.
- K a w a l e c P. (2004), Różnorodność przyczyn a antynaturalizm metodologiczny, Manuskrypt.
- K i n c a i d H. (2004), There are laws in the social sciences, w: Contemporary debates in philosophy of science, red. C. Hitchcock, Oxford: Blackwell Publishing, s. 168-185.
- K i t c h e r P., S a l m o n W. C. (1998), Van Fraassen on explanation, w: W. C. S a l m o n, Causality and explanation, New York: Oxford University Press, s. 178-190.
- L e w i s D. (1973), Counterfactuals, Cambridge, MA: Harvard University Press.
- L o s e e J. (1993³), A historical introduction to the philosophy of science, Oxford: Oxford University Press.
- M e j b a u m W. (1995), Wyjaśnianie i wyjaśnienie. Zarys teorii eksplanacji, Szczecin: FRNF.
- M e l l o r D. H. (1995), The facts of causation. Londyn: Routledge.
- P e a r l J. (2000), Causality: models, reasoning, and inference, Cambridge: Cambridge University Press.
- P e a r l J. (1995), Causal diagrams for empirical research, „Biometrika”, nr 82, s. 669-710.
- P o p p e r K. (1934/2002), Logika odkrycia naukowego, tłum. U. Niklas, Warszawa: PWN.

- P s i l l o s S. (2002), Causation and explanation, Montreal: McGill-Queen's University Press.
- R o b e r t s J. T. (2004), There are no laws of the social sciences, w: Contemporary debates in philosophy of science, red. C. Hitchcock, Oxford: Blackwell Publishing, s. 151-166.
- R u s s e l l B. (1912-13), On the notion of cause, „Proceedings of the Aristotelian Society”, nr 13, s. 1-26.
- S a l m o n W. C. (1990), Four decades of scientific explanation, Minneapolis: University of Minnesota Press.
- S a l m o n W. C. (1971), Statistical explanation and statistical relevance. With contributions by J. G. Greeno and R. C. Jeffrey, Pittsburgh: University of Pittsburgh Press.
- S a l m o n W. C. (1998), Causality and explanation, New York: Oxford University Press.
- S c r i v e n M. (1962), Explanations, predictions, and laws, w: Scientific Explanation, Space, and Time. H. Feigl, G. Maxwell, „Minnesota Studies in the Philosophy of Science”, 3, Minneapolis: University of Minnesota Press, s. 170-230.
- S p i r t e s P., G l y m o u r C., S c h e i n e s R. (2000²), Causation, prediction, and search, Cambridge, MA: The MIT Press.
- S t y c z e Ń M. (1977), Two methods of taxonomy, w: Problems of formalization in the social sciences, red. K. Szaniawski, Wrocław-Warszawa: Zakład Narodowy im. Ossolińskich – Wydawnictwo PAN, s. 139-162.
- V a n F r a a s s e n B. (1977), The pragmatics of explanation, „American Philosophical Quarterly”, nr 14, s. 143-150.
- V a n F r a a s s e n B. (1980), The scientific image, Oxford: Clarendon Press.
- V o n W r i g h t G. H. (1993), On the logic of the causal relations, w: Causation, red. E. Sosa, M. Tooley, Oxford: Oxford University Press, s. 105-124.
- W a i n e r H. (1989), Eelworms, bullet holes, and Geraldine Ferraro: some problems with statistical adjustment and some solutions, „Journal of Educational Statistics”, nr 14, s. 121-140.
- W o l e Ń s k i J. (1979/1996), Wyjaśnianie a przewidywanie, w: t e n ż e, W stronę logiki, Kraków: Aureus, s. 251-65.
- W o o d w a r d J. (1997), Explanation, invariance and intervention, „Philosophy of Science”, 64, Supplement, s. S26-S41.
- W o o d w a r d J. (2002), Explanation, w: The Blackwell guide to the philosophy of science, red. P. Machamer, M. Silberstein, Oxford: Blackwell Publishers, s. 37-54.
- W o o d w a r d J. (2000), Explanation and invariance in the special sciences, „The British Journal for the Philosophy of Science”, nr 51, s. 197-254.
- W o o d w a r d J. (2001), Law and explanation in biology: invariance is the kind of stability that matters, „Philosophy of Science”, nr 68, s. 1-20.
- W o o d w a r d J. (2003), Making things happen: a theory of causal explanation, New York: Oxford University Press.

- W ó j c i c k i R. (1979), Models and theories of empirical phenomena, w: t e n - ż e, Topics in the formal methodology of empirical sciences, Dordrecht–Wrocław: Reidel–Ossolineum, s. 157-163.
- W ó j c i c k i R. (1974), Metodologia formalna nauk empirycznych, Wrocław–Warszawa: Zakład Narodowy im. Ossolińskich–Wydawnictwo PAN.

EXPLANATION AND CAUSAL MODELS

S u m m a r y

A survey of the modern theories of explanation leads us to conclude that it is J. Woodward's counterfactual approach that seems to set out a promising agenda for setting up a satisfactory model of explanation. Such a model would need to come to terms with two stipulations, i.e. to reflect causal asymmetry and be empirically adequate. To meet the latter I modify the counterfactual model of explanation to require a class of observationally equivalent (Bayesian) causal models instead of the true causal model. For in most cases empirical adequacy of causal models, i.e. their identifiability and consistency with observed distributions, would neither grant the stronger conclusion Woodward embeds in his counterfactual account of explanation nor license metaphysical naturalism.

Słowa kluczowe: model DN, model przyczynowy, naturalizm, wyjaśnianie, C. G. Hempel, W. C. Salmon, P. Kitcher, S. Nowak, J. Woodward.

Key words: model DN, causal model, naturalism, explanation, C. G. Hempel, W. C. Salmon, P. Kitcher, S. Nowak, J. Woodward.