© Witold Marciszewski • Zakład Metodologii Nauk Społecznych
Wyższa Szkoła Administracji Publicznej im. Stanisława Staszica w Białymstoku
Wykład z Metodologii Nauk – 2007/2008

1. Sprawdzalność empiryczna hipotezy a jej wartość informacyjna

Od czego zależy sprawdzalność empiryczna hipotezy? Mówimy o hipotezie, że jest sprawdzalna empirycznie, gdy o jej wartości logicznej (prawda lub fałsz) da się się rozstrzygać na podstawie obserwowalnych zmysłowo faktów. Zdania rejestrujące takie fakty określamy mianem *spostrzeżeniowych* lub *obserwacyjnych*. Dotykamy tu fundamentalnego w filozofii problemu – roli poznawczej doświadczenia zmysłowego. Opozycyjne w tej kwestii stanowiska zajmują racjonalizm (Kartezjusz, Leibniz etc.) i empiryzm (Hume, Comte etc.).

Według *racjonalizmu* ludzkie poznanie świata rozwija się w interakcji doświadczenia zmysłowego z tą zdolnością rozumu, która dostarcza poznania prawd abstrakcyjnych. Takich, jak prawa logiki i matematyki, prawdy moralne, zasady metafizyczne (przyczynowość, celowość, czasoprzestrzenność świata, powszechna zmienność, złożoność i ewolucja itp.).[1]

W opozycji do racjonalizmu *empiryzm* głosi, że nie istnieje nic takiego jak poznanie rozumowe, nie istnieje też nic, co miałoby być jego przedmiotem, jak zbiory, liczby, struktury, algorytmy, programy komputerowe, wartości moralne czy estetyczne. Całe zatem poznanie sprowadzałoby się do doświadczenia zmysłowego. Sądy zaś abstrakcyjne nie dotyczyłyby rzeczywistości, lecz były jedynie regułami przekształcania sądów spostrzeżeniowych w hipotezy i prawa nauki. Skrajną postać empiryzmu głosiło Koło Wiedeńskie w latach 20-tych i 30-tych ubiegłego wieku.

Empiryzm w stylu Koła Wiedeńskiego został przezwyciężony przez namysł nad realnym funkcjonowaniem nauki. Prowadzi ów namysł do wniosku, że nie istnieją takie czyste dane doświadczenia zmysłowego, na których dałoby się budować teorię, a które same żadnej by teorii nie zakładały. Znaczy to, że już na etapie obserwacji ingerują w nasze poznanie pojęcia brane z jakiejś teorii. Np. badacz drobnoustrojów, *postrzegając* przez mikroskop obiekt, którego długość oszacowałby na jeden milimetr, nie powie, że ma w polu widzenia obiekt tej długości, ale że ma on długość 1/1000 mm. Wie bowiem na podstawie *teorii* optycznej, że jego mikroskop daje powiększenie tysiąckrotne i odpowiednio do tego interpretuje swe spostrzeżenie. U samych zaś podstaw teorii empirycznych, w szczególności przyrodniczych, znajdują się przesłanki brane z teorii matematycznych, jak geometria (niezbędna m.in. do budowy mikroskopu), arytmetyka etc.

Innego rodzaju przykładów dostarczają zdania obserwacyjne stwierdzające przynalezność obiektu do pewnej klasy. Np. gdy zauważam, że nadlatuje bocian, niezbędne jest do tego zauważenia, bym już posiadał pojęcie ogólne wyrażane słowem „bocian”; za jego pomocą interpretuję swe spostrzeżenie. A pojęcia ogólne nie są zapisem czystych i pierwotnych danych doświadczenia zmysłowego, lecz wynikiem pewnego procesu abstrakcji.

Teza o o nieistnieniu czystych danych zmysłowych, które wystarczyłyby za wszystkie podstawowe przesłanki teorii empirycznej, same zaś żadnej teorii nie zakładały, należy do istoty postempirystycznej metodologii nauk. Jej pionierem jest Karl Popper (1902-1994), a z pozycji pragmatyzmu krytykę poglądu empirystycznego dopełnia Willard van Orman Quine (1908-2000). Rdzeniem koncepcji Poppera jest postulat *falsyfikowalności* teorii oraz postulat, żeby teoria przyczyniała się do *wzrostu wiedzy*. Oba są nadbudowane nad pojęciem informacji, toteż trzeba zacząć od wyjaśnieniu tego pojęcia.

*

Komunikaty, które sobie wzajem nadajemy różnią się cechą, na którą istnieje trafne polskie określenie: *treściwość*. Jej przeciwieństwo cechuje taki tekst, o którym powiada się kolokwialnie, że „leje wodę”, to znaczy, jest beztreściwy czy mało treściwy. A mówiąc bardziej technicznie – dostarcza małej ilości informacji; jeszcze inaczej: ma niewielką wartość (lub zawartość) informacyjną.

Potrafimy nieraz oszacować porównawczo, że jakiś tekst zawiera więcej informacji od innego, czyli jest bardziej treściwy. Porównajmy pod tym względem takie oto obietnice składane przez majętnego dżentelmena jakiejś damie.

[1] Najwybitniejszym przedstawicielem nowoczesnego racjonalizmu był w 20. wieku genialny matematyk i wybitny filozof Kurt Gödel. Zob. tekst M. K. Solomona *On Kurt Gödel's Philosophy of Mathematics*, do którego odsyłacz znajduje się na stronie www.calculemus.org/lect/index.html w spisie wykładów 2006/07.

1. Kupię ci perły i brylanty.
2. Kupię ci perły.
3. Kupię ci brylanty.
4. Kupię ci perły lub brylanty.

Zawartość informacyjna (ilość informacji) zdania 1, w skrócie $I(1)$, ma się do pozostałych, jak następuje.

$$I(1) > I(2), \quad I(1) > I(3), \quad I(2) > I(4), \quad I(3) > I(4); \quad I(1) > I(4).$$

Zdanie 1, jako obietnica, jest treściwsze czyli bardziej informacyjne od 2 i od 3, skoro więcej obiecuje. A z kolei 2 (jak i 3) niesie więcej informacji niż 4 (choć liczy mniej słów), ponieważ trudniej jest dotrzymać 2 niż 4. Kto złoży obietnicę 4, ma otwarte dwie możliwości, a więc łatwiej mu się wywiązać (np. kupi to, co tańsze, jeśliby w tym momencie miał problem z gotówką); obietnice 2 i 3 są bardziej od 4 konkretne i w tym sensie treściwsze, a tym samym trudniejsze do dotrzymania.

Niedotrzymanie obietnicy czyni zdanie zdanie ją wyrażające wypowiedzią fałszywą. Mamy wtedy do czynienia z faktem falsyfikacji. Falsyfikacja (niedotrzymanie) tym bardziej zagraża, im obietnica jest konkretniejsza czyli im więcej zawiera informacji. I tak, łatwiej o falsyfikację zdania 1 niż zdania 4, czyli 1 jest wypowiedzią mającą wyższy stopień falsyfikowalności.

Bardziej precyzyjne, to jest, ilościowe oszacowanie nie zawsze jest możliwe. Warto jednak zapoznać się z przypadkami, kiedy taka możliwość zachodzi. Pozwalają one lepiej rozumieć, na czym treściwość polega. Gdy nasza intuicja wyćwiczy się w obliczaniu ilości informacji, trafniej i łatwiej dokona oszacowań także tam, gdzie informacja jest mało podatna na określenie ilościowe.

W opowiastce tak prostej, żeby naszą metodę dało się w niej przedstawić jak najprzejrzyściej, rozpatrzmy przypadek poszukiwań drobnego obiektu na szachownicy; niech będzie nim mrówka. Wyobraźmy sobie, że za wiadomość o miejscu pobytu mrówki wyznaczona jest nagroda, której wielkość zależy od ilości dostarczonej informacji, mianowicie dziesięć dolarów za każdą jednostkę informacji (jakie to jednostki, wyjaśni się niebawem). Wypłata zaś nastąpi wtedy, gdy wiadomość okaże się prawdą.

Intuicja podpowiada, że zachodzi zależność między wielkością nagrody a ryzykiem błędu. Najmniej ryzykowne będzie doniesienie, że mrówka jest gdzieś na szachownicy, bez najmniejszego sprecyzowania miejsca. Jest to wiadomość (na gruncie wiedzy poszukiwaczy) całkowicie pewna, czyli zabezpieczona w stu procentach przed falsyfikacją, gdyż pokrywa się z już posiadaną wiedzą; taką pewność najwyższą oznaczamy liczbą 1. Ale z tego samego względu byłaby to wiadomość bezwartościowa poznawczo, tak że informator musiałby zadowolić się zerem dolarów. Co bierze się stąd, że ilość dostarczonej informacji równa sie zeru. Tak zaczyna się rysować interesująca nas prawidłowość: w przypadku pewności, czyli prawdopodobieństwa maksymalnego (1), mamy minimalną ilość informacji (0).

Skoro nie opłaca się być takim asekurantem, inny informator podejmuje ryzyko błędu i precyzuje informację z dokładnością do połówki pola szachownicy; powiada np., że mrówka znajduje się na górnej połowie. Z powodu takiej dokładności, większej od zera, naraża się na falsyfikację w przypadku, jeśliby odnaleziono mrówkę nie w górnej lecz w dolnej połowie pola. O ile jego doniesienie się potwierdzi, czyli nie zostanie obalone, to jaka wtedy należy się mu nagroda? Żeby na to odpowiedzieć, musimy mieć miarę ilości informacji i metodę jej obliczania. Stosowana w takim pomiarze jednostka informacji nazywa się *bit* (skrót od ang. *bi*nary digi*t*).

Metodę podaje następujący wzór, którym $P(z)$ oznacza prawdopodobieństwo zdania z, zaś $inf(z)$ oznacza ilość informacji (wiedzy) dostarczanej przez zdanie z.[2]

$$inf(z) = -log_2 P(z) \text{ czyli } inf(z) = log_2 1/P(z).$$

Informatorzy, o których tu mowa nie dysponują żadnymi danymi, które byłyby zaczerpnięte z obserwacji błąkającej się mrówki. Przynoszone przez nich wiadomości są wynikiem czystego zgadywania. Można, oczywiście, rozpatrywać przypadki, gdy odgadywanie jest wspomagane obserwacjami, ale jako bardziej skomplikowane mniej by się one nadawały do uchwycenia tej zależności między informacją i prawdopodobieństwem, którą wyraża powyższe równanie. Obliczmy według niego liczbę bitów w kolejnych, coraz to dokładniejszych, a więc coraz mniej prawdopodobnych, lokalizacjach mrówki.

[2] Wzór ten pochodzi od twórcy teorii informacji, utworzonej dla potrzeb inżynierii telekomunikayjnej, Claude Shannona (1916-2001). Shannon już w swej pracy magisterskiej zauważył uniwersalność kodu binarnego, twierdząc, że ciągami zer i jedynek da się opisać tekst, obraz i dźwięk. Gdy prawdopodobieństwo we wzorze Shannona pojąć jako prawdopodobieństwo logiczne – to jest, będące cechą zdań (a nie zdarzeń, jak w innym rodzaju prawdopodobieństwa), określaną na podstawie ich formy logicznej (jak koniunkcja, alternatywa etc.), to wzór ten nadaje się także do opisywania wielkości i przyrostu informacji pojmowanej jako treść zdania.

A. $P(z) = 1 \quad \Rightarrow inf(z) = log_2 1 = 0$, bo $2^0 = 1$.

B. $P(z) = 1/2 \Rightarrow inf(z) = log_2 2 = 1$, bo $2^1 = 2$.

C. $P(z) = 1/4 \Rightarrow inf(z) = log_2 4 = 2$, bo $2^2 = 4$.

D. $P(z) = 1/8 \Rightarrow inf(z) = log_2 8 = 3$, bo $2^3 = 8$.

E. $P(z) = 1/16 \Rightarrow inf(z) = log_2 16 = 4$, bo $2^4 = 16$.

F. $P(z) = 1/32 \Rightarrow inf(z) = log_2 32 = 5$, bo $2^5 = 32$.

G. $P(z) = 1/64 \Rightarrow inf(z) = log_2 64 = 6$, bo $2^6 = 64$.

W tym skrajnie uproszczonym przypadku, gdy ułamek wyrażający prawdopodobieństwo ma w liczniku jedność, a w mianowniku jakąś potęgę liczby dwa, stosunek wielkości informacji do wielkości prawdopodobieństwa danego zdania da się określić bardzo prosto: *zdanie ma tyle bitów informacji, ile wynosi wykładnik potęgi, do której trzeba podnieść liczbę dwa, żeby otrzymać liczbę obecną w mianowniku* (można więc obejść się tu bez logarytmowania, ale formuła z logarytmem jest należycie ogólna, obejmująca wszystkie przypadki).

W ten sposób obliczamy wielkość zawartej w zdaniu informacji mierzoną w bitach. Liczba bitów stanowi odwrotność prawdopodobieństwa danego zdania. Gdy jest ono maksymalne, jak w wierszu A, zdanie ma 0 bitów informacji. Istotnie, gdy szachownica reprezentuje cały świat, informacja, że mrówka jest gdzieś w świecie jest minimalna (zerowa), jako że prawdopodobieństwo jest maksymalne. Gdy prawdopodobieństwo wynosi 1/2 (wiersz B), doniesienie, ze mrówka jest na tej oto, a nie na drugiej połówce, ma już wartość informacyjną, mniejszą od każdej innej niezerowej, a większą od zera; jest to więc jeden bit. Gdy wskaże się połówkę połówki rozważanej powierzchni, podwoi to poprzednią ilość informacji, będzie więc dwa bity. I tak dalej. Odpowiednio, za wiadomość, która jest żadna czyli zerowa (A) należy się zero pieniędzy. Za wiadomość wskazującą na określoną połowę należy się (przy przeliczniku dziesięć) dolarów 10, za wiadomość z wiersza C – 20 dolarów. Itd.

Wzór Shannona podający sposób obliczenia $inf(z)$ przy danym $P(z)$ wyraża intuicję, że wielkość informacji jest odwrotna do wielkości prawdopodobieństwa. Odwrotność można też ująć za pomocą innej funkcji, rozważanej przez Karla Poppera.[3]

$$cont(z) = 1 - P(z)$$

Dla zaznaczenia, że jest to odmienne, bo inaczej zdefiniowane, pojęcie informacji, zastosowano na jej oznaczenie symbol $cont$. Podczas, gdy wielkość informacji obliczana według wzoru Shannona może się wyrazić dowolnie wielką liczbą dodatnią, liczba użyta za miarę informacji według obecnego wzoru jest liczbą rzeczywistą dodatnią z przedziału między zero i jeden.

Obie miary zgodne są w tym, że przypisują tautologiom informację zerową. Różnią się natomiast, gdy idzie o obliczanie ilości informacji w zdaniu wewnętrznie sprzecznym (np. w formie $p \wedge \neg p$) czyli mającym prawdopodobieństwo zerowe. Gdy $P = 0$, należy (wg Shannona) obliczyć $log_2 0$, z którym sprawa ma się tak, że gdy ciąg ułamków, wyrażających prawdopodobieństwa, zmierza do zera jako do granicy, wartość logarytmu zmierza do minus nieskończoności.

Wedle drugiej z naszych miar informacji, zdanie sprzeczne ma informację maksymalną, co z logicznego punktu widzenia implikuje zawieranie się w niej również wszystkich zdań fałszywych. Sprzeczność jawi się jako największe dla poznania świata zagrożenie.

Podsumujmy. Wśród hipotez różniących się między sobą ilością informacji najwyższy stopień empirycznej sprawdzalności cechuje tę, która zawiera informacji najwięcej. Im więcej bowiem informacji, tym więcej rozpoznajemy obserwowalnych faktów, o których wiadomo, czy byłyby z daną hipotezą zgodne. Im więcej zaś takich faktów, tym bardziej jest ona sprawdzalna empirycznie. ♠

[3] Zob. Karl Popper, *Logika odkrycia naukowego*, przełożyła Urszula Niklas, PWN, Warszawa 1977, s. 302. Oryginał: *The Logic of Scientific Discovery*, Basic Books, Inc., New York 1959, s. 374. Zob. też Klemens Szaniawski, *Prawdopodbieństwo* w: Witold Marciszewski (red.), *Logika formalna. Zarys encyklopedyczny z zastosowaniem do informatyki i lingwistyki*, PWN, Warszawa 1987, s. 415; z tego tekstu wzięte są oznaczenie *inf* i *cont* (od ang. *content*).