

## 1. Sprawdzalność empiryczna hipotezy a jej wartość informacyjna

Główny temat niniejszych szkiców z metodologii nauk rodzi się z pytania, od czego zależy sprawdzalność empiryczna hipotezy. To znaczy, taka właściwość, że o tym, czy hipoteza zasługuje na akceptację daje się rozstrzygnąć na podstawie obserwowalnych zmysłowo faktów. Jest to dla studium filozofii temat doniosły nie tylko dlatego, że stanowi ciąg dalszy zaczynającego się kursem logiki treningu w sprawności rozumowania. Także dlatego, że pytanie to nawiązuje do fundamentalnego w nauce o poznaniu problemu – roli doświadczenia. Dalece odmienne stanowiska zajmują w tej sprawie racjonalizm i empiryzm.

Według *racjonalizmu* ludzkie poznanie świata rozwija się w interakcji doświadczenia zmysłowego z tą aktywnością intelektu, która owocuje poznaniem prawd abstrakcyjnych, jak prawa logiki, zasady teorii mnogości i arytmetyki, zasady metafizyczne (przyczynowość, celowość, czasoprzestrzenność świata, powszechna zmienność i ewolucja itp.).<sup>1</sup>

W opozycji do racjonalizmu *empiryzm* głosi, że nie istnieje nic takiego jak poznanie intelektualne, gdyż nie istnieje nic, co miałyby być poznania takiego przedmiotem, mianowicie zbiory, liczby, struktury, algorytmy, programy komputerowe itd. Całe zatem poznanie sprowadza się do doświadczenia zmysłowego. Zasady zaś abstrakcyjne nie dotyczą realnego świata lecz są jedynie środkami przekształcania danych zmysłowych. W skrajnej postaci tego rodzaju empiryzm był głoszony przez Koło Wiedeńskie.

Empiryzm taki został przewyżniony przez namysł nad realnym funkcjonowaniem nauki posługujący się analizą logiczną; jest to znaczący wkład w ów wiekowy spór epistemologiczny. Namysł taki prowadzi do wniosku, że nie istnieją czyste dane doświadczenia zmysłowego, na których dałoby się budować teorię, a któreby same żadnej teorii nie zakładały; ten negatywny wynik prowadzi na trop przypuszczenia, że już na etapie obserwacji ingerują jakieś pojęcia czy zasady abstrakcyjne (co na swój sposób antycypował Kant; tego jednak tematu nie będzie się w obecnych szkicach poruszać).

Wniosek o nieistnieniu czystych danych zmysłowych, które same wystarczyłyby za przesłanki teorii, stanowi istotny składnik postempirystycznej metodologii nauk empirycznych. Tworzy ją postulat falsyfikowalności teorii oraz postulat, żeby teoria jak najobficiej przysparzała wiedzy. Oba są nadbudowane nad pojęciem informacji, stąd poświęca się temu pojęciu pierwszy z czterech szkiców.

\*

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 majątnego dżentelmena jakiejś damie.

<sup>1</sup> Najwybitniejszym przedstawicielem nowoczesnego racjonalizmu był w 20. wieku genialny matematyk i wybitnie oryginalny filozof Kurt Gödel. Zob. tekst M. K. Solomona *On Kurt Gödel's Philosophy of Mathematics*, do którego odsyłać znajduje się na stronie [www.calculumus.org/lect/index.html](http://www.calculumus.org/lect/index.html).

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, jeśli 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 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 opowieści 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 się 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, jeśli 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. *binary digit*).

Metodę podaje następujący wzór, którym  $P(z)$  oznacza prawdopodobieństwo (logiczne) zdania  $z$ , zaś  $I(z)$  oznacza ilość informacji (wiedzy) dostarczanej przez zdanie  $z$ .<sup>2</sup>

<sup>2</sup> Wzór ten pochodzi od twórcy teorii informacji, utworzonej dla potrzeb inżynierii telekomunikacyjnej, 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 nadaje się on także do opisywania wielkości i przyrostu informacji pojmowanej jako treść zdania.

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

Informatorzy, o których tu mowa nie dysponują żadnymi danymi, które byłyby zaczerpnięte z obserwacji błąkającej się mrówki. Przynieszone 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.

- A.  $P(z) = 1 \Rightarrow I(z) = \log_2 1 = 0$ , bo  $2^0 = 1$ .
- B.  $P(z) = 1/2 \Rightarrow I(z) = \log_2 2 = 1$ , bo  $2^1 = 2$ .
- C.  $P(z) = 1/4 \Rightarrow I(z) = \log_2 4 = 2$ , bo  $2^2 = 4$ .
- D.  $P(z) = 1/8 \Rightarrow I(z) = \log_2 8 = 3$ , bo  $2^3 = 8$ .
- E.  $P(z) = 1/16 \Rightarrow I(z) = \log_2 16 = 4$ , bo  $2^4 = 16$ .
- F.  $P(z) = 1/32 \Rightarrow I(z) = \log_2 32 = 5$ , bo  $2^5 = 32$ .
- G.  $P(z) = 1/64 \Rightarrow I(z) = \log_2 64 = 6$ , 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, że 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  $I(z)$  przy danym  $P(z)$  wyraża intuicję, że wielkość informacji jest odwrotna do wielkości prawdopodobieństwa. Odwrotność można też ująć za pomocą innej funkcji, mianowicie odejmowania:

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

Dla zaznaczenia, że jest to odmienne, bo inaczej zdefiniowane, pojęcie informacji, zastosowano na jej oznaczenie inny symbol:  $Inf$ . 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 wewnątrznie 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. Należy to tak zinterpretować, że zawartość informacyjna zdania sprzecznego jest nieskończona, co jednak nie jest zaletą, lecz kolosalną wadą. Znaczą bowiem, że do konsekwencji zdania sprzecznego należą wszystkie możliwe zdania, a więc i wszystkie zdania fałszywe.<sup>3</sup>

Wedle drugiej z naszych miar informacji, zdanie spreczne ma też informację maksymalną, co z logicznego punktu widzenia implikuje również zawieranie się w niej wszystkich zdań fałszywych, ale to maksimum jest w tym przypadku dobrze określone, pokrywając się z liczbą jeden. W takiej, czy innej interpretacji, sprzeczność jawi się jako największe dla poznania świata zagrożenie (por. rozdz. 4, ostatnie pytanie egzaminacyjne).

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. ♠

---

<sup>3</sup> O logarytmie zera mówi się też, że nie jest on wielkością zdefiniowaną, co wprowadza pewien niuans do interpretacji zdań sprzecznych z informacyjnego punktu widzenia, ale szczegół ten pominiemy w obecnym uproszczonym przedstawieniu sprawy.