

Rozdział II. O pojęciu informacji w perspektywie teoretyczno-prawnej

Wojciech Załuski

§ 1. Pojęcie informacji w języku potocznym

Termin „informacja” nie posiada w języku potocznym precyzyjnego znaczenia. Chcąc wydobyć z tego języka jakąś jego definicję, adekwatną dla rozmaitych kontekstów jej występowania¹, nie zdołamy wyjść poza bardzo ogólne stwierdzenia typu „informacja jest pewną porcją znaczącej, tj. w jakimś sensie ważnej, wiedzy”, „informacja jest znaczącym faktem czy pewnym zbiorem takich faktów”, „informacja jest czymś znaczącym, co zawarte jest w mowie lub piśmie”. Podobna sytuacja zachodzi w przypadku języka prawniczego (co oczywiście nie jest faktem zaskakującym, zważywszy, iż język ten stanowi rejestr języka potocznego). Powiada się w nim, że np. „prawo powinno chronić informację, jaką są dane osobowe”, „prawo powinno ułatwiać przepływ informacji między podmiotami gospodarczymi” itd. I znów jedyną definicją informacji adekwatną dla tych różnych kontekstów wydaje się być ogólna konstatacja, iż informacja jest pewną porcją znaczącej wiedzy. Ta niestaranność w posługiwaniu się pojęciem informacji nie powinna jednak niepokoić. Nazwa „informacja” nie wydaje się być bowiem wyjątkiem na tle innych pojęć języka potocznego, które przecież mogą spełniać swoje główne zadanie, tj. zapewniać efektywną komunikację, nawet wtedy, gdy nie są wzorcowo precyzyjne. A właśnie powyższa definicja informacji – mimo swej ogólnikowości – wydaje się być wystarczająca dla realizacji tak ujętego zadania zarówno języka potocznego, jak i tego rejestru, który nas szczególnie tu interesuje, tj. języka prawni-

¹ Te konteksty to np. „informacja o wybuchu epidemii przejęła wszystkich zgrozą”, „czerwone światło przekazuje informację o nakazie zatrzymania się”, „televizja nie zawsze jest źródłem rzetelnej informacji”.

czego. Pomimo tego chciałbym w niniejszej pracy zwrócić uwagę na fakt, że istnieją bardzo precyzyjne definicje informacji, eksplikujące w ciekawy sposób jej potoczne znaczenie. Celem moich rozważań będzie prezentacja dwóch najbardziej znanych spośród nich i próba wykazania, że mogą być one przydatne dla rozważań prawniczych. Definicje, o których mowa, zostały sformułowane na gruncie 2 matematycznych dyscyplin – teorii informacji (definicja informacji statystycznej) i teorii decyzji (definicja informacji pragmatycznej) – i doczekały się wszechstronnego filozoficznego komentarza.

§ 2. Eksplikacje potocznego pojęcia informacji

Zanim przejdę do omówienia obu definicji – eksplikacji potocznego rozumienia informacji – sformułuję pewien ogólny warunek, który, jak sądzę, musi być spełniony przez każdą taką eksplikację, jeśli ma być ona przydatna z punktu widzenia użytkownika języka prawniczego. Warunek ten będzie więc jednym z kryteriów oceny relewantności tych definicji dla rozważań prawniczych. Warunek ten można ująć tak: **każda eksplikacja pojęcia informacji musi respektować intuicję znaczeniową zawartą w jego potocznym użyciu, tj. intuicję, że informacja jest porcją wiedzy, i to wiedzy znaczącej.** Jak zobaczymy, obie definicje spełniają ten warunek. Należy ponadto podkreślić, że nie są one wzajemnie sprzeczne. Różnica między nimi polega na tym, że odmiennie dookreślają pojęcie ważności wiedzy stanowiące element potocznego rozumienia „informacji”: definicja statystyczna precyzuje to pojęcie w kategoriach czysto epistemicznych, a więc w oderwaniu od kontekstu decyzyjnego, natomiast definicja pragmatyczna uzależnia „ważność” wiedzy od roli, jaką pełni ona w procesach decyzyjnych. Poniższa prezentacja obu definicji ma na celu wskazanie podstawowych intuicji, jakie za nimi stoją. Tak ujęty cel usprawiedliwia pominięcie przez wielu subtelności matematycznych związanych z tym definicjami i ograniczenie do niezbędnego minimum symboliki matematycznej².

² W opracowaniu tego ujęcia szczególnie pomocne były dla mnie prace *Klemensa Szaniawskiego*: hasło „Informacja”, w: *Filozofia a nauka. Zarys Encyklopedyczny*, Warszawa 1987, s. 244–251, oraz artykuły zebrane w dziele *O nauce, rozumowaniu i wartościach. Pisma wybrane*, Warszawa 1994: „Nauka jako proces poszukiwania informacji”, s. 110–119, „Informacja a decyzja”, s. 253–266, „Pragmatyczna wartość informacji”, s. 305–325, „O pojęciu informacji”, s. 323–331, „Informacja i decyzja jako narzędzia filozofii nauki”, s. 345–356, „Dwa pojęcia informacji”, s. 367–378, „O wartości informacji w grach”, s. 379–389, „Typy informacji i ich rola w metodologii nauk”

I. Pojęcie informacji w świetle teorii informacji – informacja statystyczna

Teoria informacji powstała w latach 20-tych zeszłego stulecia. Dojrzałą postać nadał jej C. Shannon, pracując nad problemem przesyłania i najefektywniejszego kodowania informacji³. W ramach tej teorii funkcjonuje rozróżnienie między informacją i jej materialnym nośnikiem, zwanym sygnałem⁴. Rozróżnienie to nie jest ściśle przestrzegane w języku potocznym, w którym mówi się, na przykład, że „białe pasy na jezdni są informacją dla pieszych o możliwości przejścia na drugą stronę”, podczas gdy precyzyjnie powinno się powiedzieć, że „białe pasy na jezdni dostarczają pieszym informacji o możliwości przejścia na drugą stronę”. Jak wynika z tego rozróżnienia, informacja jest cechą przypisaną sygnałowi, czy, ujmując rzecz ściślej, funkcją sygnału jest przekazywanie jakiejś informacji. Warto przy tym podkreślić, że w ramach teorii informacji sygnał rozumiany jest bardzo szeroko: po pierwsze, sygnałem może być praktycznie rzecz biorąc wszystko – na przykład, słowo pisane, dźwięk, zapach, obraz, gest – a więc każdy przedmiot i każde zdarzenie; po drugie, sygnał może nieść informację w sposób naturalny (dym świadczy – a więc niesie informację – o ogniu) lub na mocy konwencji (przykład z pasami na jezdni). W toku dalszych rozważań będę mówił jednak tylko o sygnałach zdaniowych. Nie pozbawia to owych rozważań ogólności, gdyż wystąpienie każdego sygnału da się opisać w zdaniu, które stwierdza ten fakt (co pokazuje, że każdy sygnał można zinterpretować jako sygnał zdaniowy). Niekiedy informację zawartą w zdaniu określa się mianem „semantycznej”; tak rozumiana informacja semantyczna jest więc odmianą informacji statystycznej i nie stanowi kolejnego typu informacji⁵. Główne idee teorii informacji zrekonstruuję w postaci 3 tez.

s. 411–421, „O definiowaniu informacji”, s. 445–451. Korzystałem także z: J. Hintikka, *The Varieties of Information and Scientific Explanation*, w: B. van Rotselaar, J.F. Staal (red.), *Logic, Methodology and Philosophy of Science*, Dordrecht, Amsterdam, s. 151–171, A.M. Jagłom, *I.M. Jagłom*, *Prawdopodobieństwo i informacja*, Warszawa 1963. Nie podaję dalej konkretnych odniesień do tych prac, gdyż ostateczny układ prezentacji (wraz z tezami 1–3, gdy idzie o informację statystyczną) oraz w dużej części notacja są propozycją własną.

³ Por. C. Shannon, W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana–Illinois 1962.

⁴ W języku potocznym częściej używa się nazwy „znak”.

⁵ Istnieje jednak także inne rozumienie informacji semantycznej, które koncentruje się na samej treści zdania i zmierza do uchwycenia różnic w doniosłości zdań mających identyczną wartość (w sensie informacji statystycznej).

Będę stosował następującą notację⁶. F oznacza pytanie, na które zdania s_1, s_2, \dots, s_n są odpowiedziami **dopuszczonymi jako możliwe** przez daną osobę – odbiorcę sygnału. Zdania te tworzą zbiór $S = (s_1, s_2, \dots, s_n)$. Jeśli dana osoba jest przekonana, że wie, które zdanie jest prawdziwe, zbiór ten jest wówczas zbiorem jednoelementowym.

Teza 1: **Informacja jest cechą przypisaną zdaniu s_i nie w sposób absolutny, lecz w sposób relatywny względem stanu wiedzy odbiorcy, do którego zdanie to jest skierowane**⁷. Innymi słowy, to samo zdanie s_i może dostarczać różnych informacji w zależności od tego, jaki system przekonań posiada odbiorca tego sygnału zdaniowego. Prosty przykład zilustruje tę tezę. Zdanie „Warszawa jest największym miastem w Polsce” nie dostarcza żadnej informacji temu, kto wie, że Warszawa jest największym miastem w Polsce, tj. wie, że to zdanie jest jedyną poprawną odpowiedzią na pytanie o to, jakie jest największe miasto w Polsce. Dostarcza już jednak pewnej informacji temu, kto nie jest pewien, czy takim miastem jest Warszawa czy Kraków, gdyż eliminuje jedną z dwóch możliwych odpowiedzi, jaką brał pod uwagę. I wreszcie, dostarcza bardzo dużej informacji temu, kto, dysponując skromną wiedzą na temat geografii Polski, waha się, czy wskazać na Warszawę, Kraków, Wrocław, Gdańsk, czy Bielsko-Białą⁸. Ten przykład prowadzi nas do bardziej rozbudowanej wersji tezy 1.

⁶ Notacja i tezy 1–3 są propozycją własną, mającą na celu możliwe prostą prezentację głównych idei teorii informacji.

⁷ Stwierdzając, iż „informacja jest cechą przypisaną zdaniu s_i ”, posługuję się pewnym skrótem myślowym; precyzyjnie należałoby powiedzieć „informacja jest cechą, która byłaby przypisana zdaniu s_i , gdyby okazało się prawdziwe”.

⁸ Można by próbować kwestionować tezę, iż informacja jest **zawsze** zrelatywizowana do systemu przekonań odbiorcy, argumentując, że np. zdanie „Ziemia kręci się wokół Słońca” nie podlega takiej relatywizacji, gdyż **dla każdego** odbiorcy niesie tę samą niską ilość informacji (jako że jest oczywiste i nie usuwa żadnych wątpliwości). Na ten argument można łatwo odpowiedzieć. Zdanie „Ziemia kręci się wokół Słońca” jest oczywiste dla współcześnie żyjących – **oczywiste w świetle powszechnie akceptowanego przez nich – heliocentrycznego – systemu wiedzy astronomicznej**. Jednak, na przykład, dla *Arystotelesa, Ptolemeusza, czy św. Tomasza z Akwinu*, zdanie „Ziemia faktycznie kręci się wokół Słońca” niesie ogromną ilość informacji, gdyż jest skrajnie nieprawdopodobne na tle ich – geocentrycznego – systemu wiedzy, tj. sprzeczne z tym systemem (por. teza 2). Stosując zaproponowaną wyżej notację, można opisać sytuację *Arystotelesa, Ptolemeusza, czy św. Tomasza z Akwinu* tak; na pytanie F : „czy poprawny jest system geocentryczny czy heliocentryczny?” dopuszczali oni tylko jedną odpowiedź s „System geocentryczny opisuje strukturę Wszechświata”. Ich zbiór wyglądał zatem tak: $S = (s)$. Okazało się jednak, że s jest fałszywym, zaś prawdziwe jest zdanie t „System heliocentryczny opisuje strukturę Wszechświata”, w ogóle nieuwzględniane w S ; dlatego zdanie t pociągało za sobą całkowitą przebudowę systemu wiedzy

Teza 1': **Informacja jest cechą przypisaną zdaniu s_i w sposób relatywny względem stanu wiedzy odbiorcy, gdyż ów stan determinuje kształt jego zbioru S , od liczności którego zależy ilość informacji zawartej w s_i . Zależność tę można ująć tak: im liczniejszy jest zbiór S , tym większą informację niesie zdanie s_i (przy założeniu równego prawdopodobieństwa należących do zbioru S zdań).** Wynika stąd, że im większą niepewność zdanie usuwa, tym bardziej jest „informatywne”. Zaś niepewność jest tym większa, im liczniejszy jest S , tj. zbiór możliwych odpowiedzi na dane pytanie. W skrajnym przypadku, gdy zbiór S jest jednoelementowy, a zawarte w nim zdanie jest prawdziwe (przypadek, gdy ktoś wie, że Warszawa jest stolicą Polski i słyszy zdanie s_i konstatujące ten fakt), zdanie s_i nie niesie żadnej informacji, gdyż nie ma żadnej niepewności, którą mogłoby usuwać. Teza 1' nie mówi jednak, w jaki sposób należy mierzyć zawartość informacji w danym zdaniu s_i . Ponadto, dotyczy tylko zbioru zdań, które są równo-prawdopodobne: tylko w takim przypadku zawartość informacji w zdaniach należących do zbioru S zależy od liczności tego zbioru. Dopuszczane przez odbiorcę zdania – potencjalne odpowiedzi na dane pytanie – mają jednak często różne wartości prawdopodobieństw. Wracając do naszego przykładu: ktoś, na przykład, może przypisywać zdaniu „Kraków jest największym miastem w Polsce” prawdopodobieństwo równe 0.2, zaś zdaniu „Warszawa jest największym miastem w Polsce” prawdopodobieństwo równe 0.8. W takim przypadku zdanie „Warszawa jest największym miastem Polski” niesie dla niego mniejszą informację, niż niosłoby zdanie „Kraków jest największym miastem Polski”, gdyby było prawdziwe. Jest to intuicyjnie zrozumiałe: im bowiem wyższe prawdopodobieństwo ktoś przypisuje danemu zdaniu⁹, tym mniejszą informację owo zdanie niesie, **ponieważ tym mniejszą niepewność usuwa.** W ten sposób doszliśmy do trzeciej – fundamentalnej – tezy teorii informacji, wskazującej na istnienie ścisłego związku między pojęciami informacji i prawdopodobieństwa.

Teza 2: **Informacja zawarta w zdaniu s_i jest odwrotnością, czyli malejącą funkcją prawdopodobieństwa zdania s_i ; innymi słowy, zdanie s_i niesie tym większą informację dla odbiorcy, im mniejsze prawdopodobieństwo ów odbiorca wcześniej temu zdaniu przypisywał.** Warto dodać, że z uwagi na psychologiczną tezę, mówiącą, że im bardziej prawdopodobne jest dane zdanie,

astronomicznej, niosło ogromną ilość informacji. Ten przykład jasno dowodzi tezy mówiącej o relatywizacji informacji do systemu wiedzy.

⁹ To znaczy, im bardziej jego sytuacja przypomina sytuację, w której zdanie to stanowi jedyny element zbioru zdań branych przez niego pod uwagę.

tym mniejsze zaskoczenie wywoła u odbiorcy, gdy okaże się prawdziwe, można powiedzieć, że wielkość informacji zawartej w zdaniu mierzy się stopniem zaskoczenia, jakie jego prawdziwość wywołuje u odbiorcy. Teza 2 mówi, że informacja jest malejącą funkcją prawdopodobieństwa zdania, w związku z czym należy w tym miejscu zadać pytanie o kształt tej funkcji, czyli o **konkretną miarę** zawartości informacji w zdaniu. Najprostsze rozwiązanie zaproponowane w literaturze jest takie: należy utożsamić informację niesioną przez zdanie s_i z prawdopodobieństwem nieziszczenia się tego zdania, czyli z prawdopodobieństwem sytuacji, gdy prawdziwe okazuje się zdanie $\sim s_i$ – w ten sposób powstaje funkcja c^{10} :

$$\text{(Funkcja } c\text{): } c(s) = p(\sim s) = 1 - p(s)$$

Częściej używa się funkcji logarytmicznej (o podstawie dwójkowej) ze znakiem minus, który ją przekształca z rosnącej w malejącą¹¹; jest to tzw. funkcja i :

$$\text{(Funkcja } i\text{): } i(s) = -\log_2 p(s)$$

Warto przedstawić kilka uwag technicznych na temat obu funkcji. Obie są określone dla liczb ułamkowych (gdyż prawdopodobieństwo zmienia się w przedziale od zera do jedności). Obie maleją do zera wraz ze wzrostem prawdopodobieństwa (gdy $p = 1$, obie, jak łatwo sprawdzić, przyjmują wartość 0). Różnica między nimi ujawnia się dla zdań, których prawdopodobieństwo jest bliskie zera: zdaniom o prawdopodobieństwie zerowym funkcja c przypisuje informację równą jedności, zaś funkcja i nie jest dla nich określona, gdyż jej wartość rośnie do nieskończoności, w miarę jak prawdopodobieństwo zdania maleje do zera. Nie wdając się w szczegółową analizę wad i zalet każdej z tych funkcji, chcę tylko podkreślić, iż niekwestionowaną zaletą funkcji i jest to, że – w przeciwieństwie do funkcji c – spełnia ona pewien bardzo intuicyjny postulat, zgodnie z którym informacja zawarta w 2 sygnałach łącznie jest sumą informacji każdego z nich osobna wtedy i tylko wtedy, gdy sygnały te są od siebie probabilistycznie niezależne¹².

¹⁰ Funkcja ta została wprowadzona przez *Carnapa* i *Bar-Hillela*. Por. *R. Carnap, Y. Bar-Hillel, An Outline of a Theory of Semantic Information, Technical Report 24, Harvard University Press, Cambridge, Mas., London, England 1952.*

¹¹ Prawdopodobieństwo określa liczba z przedziału $[0, 1]$, a logarytm takiej liczby nie jest nigdy dodatni. Po to, aby otrzymać liczbę dodatnią, dodaje się znak minus do logarytmu. Opuszczając ten znak, otrzymuje się wielkość ujemną, którą można interpretować jako miarę niedoboru informacji.

¹² Jeśli więc mamy dwa zdania „27 jest podzielne przez 9” i „27 jest podzielne przez 3”, to informacja zawarta w koniunkcji „27 jest podzielne przez 9 i 27 jest podzielne przez 3” nie jest sumą informacji zawartej w każdym zdaniu z osobna, gdyż drugie zdanie wynika z pierwszego (wiadomo, że każda liczba podzielna przez 9 jest także podzielna przez 3), nie wnosi więc żadnej

Jak wspomniano wyżej, do pomiaru informacji najczęściej stosuje się logarytm o podstawie dwójkowej. Jednostką informacji jest wtedy tzw. bit (*Binary digiT* – cyfra w systemie dwójkowym). Zdanie o $p = 1/2$, na przykład, zdanie mówiące, że wypadnie reszka po rzucie rzetelną monetą, zawiera jeden bit informacji, gdyż $-\log_2(1/2) = 1$. Inaczej mówiąc, jeden bit jest miarą informacji otrzymanej w wyniku uzyskania odpowiedzi na elementarne pytanie, tj. na pytanie, na które odpowiedź może brzmieć tylko „tak” lub „nie”¹³. Zilustrujemy pojęcie informacji statystycznej kilkoma przykładami:

a) Mamy kostkę o 32 bokach. Pytamy o prawdopodobieństwo wypadnięcia, powiedzmy, ścianki oznaczonej cyfrą 7. S w tym przypadku składa się z 32 zdań (np. s_1 to „wypadnie ścianka 1”). Zatem prawdopodobieństwo, że wypadnie ścianka 7, wynosi $1/32$. Informacja zawarta w zdaniu stwierdzającym zajście tego faktu równa się $-\log_2(1/32) = 5$ bit. Jeśli ograniczymy zbiór S do 16, rzucając kostką o 16 bokach, to zdanie „wypadła ścianka 7” niesie informację $-\log_2(1/8) = 3$ bit.

b) Prawdopodobieństwo alternatywy „wypadnie ścianka numer 7 lub ścianka numer 8” w sytuacji, gdy rzucamy kostką o 32 ściankach, wynosi $1/32 + 1/32 = 1/16$. Zatem informacja w nim zawarta wynosi $-\log_2(1/16) = 4$ bity. Zauważmy, że przykład ten sugeruje prosty sposób zdefiniowania tego, czym jest banał, czyli tautologia. Otóż informacja zawarta w banalnej – tautologicznej – wiadomości „wypadnie ścianka 1 lub ścianka 2 lub ścianka 3... lub ścianka 32” wynosi zero bitów, gdyż prawdopodobieństwo zdarzenia opisanego w tej wiadomości wynosi 1 (wiadomo, że wypadnie któraś z 32 ścianek), zaś $\log_2(1) = 0$. Inaczej mówiąc, zdanie to jest banalne, gdy nie eliminuje żadnej możliwości, dzięki czemu okaże się prawdziwe niezależnie od tego, co się wydarzy. Analogicznie, banalna, bo nic nie mówiąca o świecie, jest informacja „jutro będzie świecić słońce lub nie będzie świecić słońce”¹⁴.

c) Z tezy 2, mówiącej o zachodzeniu odwrotnej zależności między prawdopodobieństwem zdania i informacją w nim zawartą, wynika, że ilość informacji zawarta w danym zdaniu zwiększy się, jeśli prawdziwe zdanie było mniej prawdopodobne. Jednym ze sposobów tworzenia zdań mniej prawdopodob-

dodatkowej informacji. Można wykazać w ścisły sposób, że tylko funkcja logarytmiczna spełnia ten bardzo intuicyjny postulat.

¹³ Bit jest bardzo małą jednostką, dlatego używa się zwykle jednostki większej – bajtu, który jest równy 8 bitom. Bajt z kolei występuje w jeszcze większych porcjach – kilobajtach, megabajtach, gigabajtach, tj. jednostkach utworzonych z potęg liczby 2¹⁰. Np. kilobajt to 2¹⁰ bajtów = 1024 bajty, a 1 megabajt to 2²⁰ = 1048575 bajty.

¹⁴ Por. W. Marciszewski, *Sztuczna Inteligencja*, Kraków 1998, s. 111.

nych – a więc bardziej „nasyconych” informacją – jest łączenie ich za pomocą funktora koniunkcji. Na przykład, prawdopodobieństwo koniunkcji „w pierwszym rzucie wypadnie ścianka 7, a w drugim – ścianka 8” jest mniejsze niż każdego z jej składników; jej prawdziwość spowoduje więc większy przyrost informacji (10 bitów) niż prawdziwość któregoś z jej składników (5 bitów)¹⁵.

Warto dodać, że informację statystyczną można przypisywać nie tylko pojedynczemu zdaniu s_i , lecz także zbiorowi wszystkich możliwych zdań w danej sytuacji, tj. S jako całości. Znaczy to, że można mówić o informacji dostarczonej przez odpowiedź na dane pytanie bez precyzacji, o którą odpowiedź – czyli o które konkretnie zdanie s_i – chodzi. Miarą jest wtedy średnia ilość informacji dostarczonej przez możliwe odpowiedzi na to pytanie; średnią tę liczy się, mnożąc ilość informacji każdej możliwej odpowiedzi na dane pytanie, tj. każdego s_i , przez jej prawdopodobieństwo, i następnie sumując uzyskane w ten sposób iloczyny. Przekładając ten słowny opis na matematyczną formułę, dochodzimy do słynnego wzoru *Shannona*, pozwalającego obliczyć średnią ilość H informacji zawartej w zbiorze S możliwych odpowiedzi (s_1, s_2, \dots, s_n) na pytanie F :

Wzór *Shannona*: $H(S) = - \sum_s p(s) \log_2 p(s)$. Zatem średnia wartość informacji H zmiennej S jest wartością oczekiwaną logarytmu jej funkcji rozkładu prawdopodobieństwa (wziętą ze znakiem minus, aby wartość była nieujemna).

Dysponując już pojęciem średniej ilości informacji, możemy sformułować tezę 3.

Teza 3: Średnia ilość informacji H zawarta w możliwych zdaniach tworzących zbiór S zwana jest entropią informacyjną S lub miarą nieokreśloności pytania F . Oto stosowny przykład. Na pytanie F , jakie są możliwe wyniki rzutu rzetelną monetą, możemy odpowiedzieć: „orzeł” lub „reszka”. Zbiór S – zbiór możliwych odpowiedzi na F – jest zatem dwuelementowy. Prawdopodobieństwo każdego z tych dwóch zdań wynosi $1/2$, i każde z nich zawiera informację równą – $\log_2(1/2) = 1$ bit. Zatem średnia ilość informacji zawartej w odpowiedziach na powyższe pytanie, tj. entropia zbioru S wyniesie: $H(S) = - [1/2 \log_2(1/2) + 1/2 \log_2(1/2)] = 1$ bit. Entropia informacyjna – miara nieokreśloności pytania – przyjmuje największą wartość wtedy, gdy wszystkie odpowiedzi na dane pytanie są jednakowo prawdopodobne, gdyż wtedy wła-

¹⁵ Zważywszy, że zdania „wypadnie 7” i „wypadnie 8” są probabilistycznie niezależne, prawdopodobieństwo ich koniunkcji obliczamy, mnożąc prawdopodobieństwa każdego z nich osobno, tj. $1/32 \cdot 1/32 = 1/1024$. Stosując definicję informacji binarnej, otrzymujemy $-\log_2(1/1024) = 10$ bitów. Definicja informacji binarnej miałaby nieco bardziej złożoną formę, gdybyśmy chcieli ją zastosować do zdań zależnych.

śnie istnieje najmniejsza szansa odgadnięcia, która z nich jest prawdziwa¹⁶, zaś uzyskana informacja usuwa największą niepewność (w takim przypadku entropia jest równa logarytmowi liczby tych odpowiedzi). Entropia jest zaś najmniejsza – wynosi zero – w odniesieniu do takiego pytania, na które odpowiedź jest znana. **Pojęcie entropii informacyjnej jest niezbędne dla obliczenia przyrostu informacji w sytuacji, gdy zdanie, o którym dowiadujemy się, że jest prawdziwe, nie jest jednym ze zdań tworzących S, lecz jest powiązane z nimi w sposób probabilistyczny.** Nawiązując do wcześniejszego przykładu, takim zdaniem może być np. zdanie E, „Warszawa jest stolicą Polski”, które wydaje się powiązane probabilistycznie ze zdaniem „Warszawa jest największym miastem w Polsce”, tj. podnosi jego prawdopodobieństwo, lecz nie przesądza o jego prawdziwości. Ażeby obliczyć ilość informacji zawartej w E względem S, tj. $I(E, S)$, musimy odwołać się do pojęcia entropii informacyjnej. Miarą tej zawartości będzie średnia redukcja entropii zbioru S, tj. redukcja niepewności co do zdań tworzących S, uzyskana dzięki E:

$$I(E, S) = H(S) - H(S/E)$$

Zarysowana w niniejszym punkcie definicja informacji posiada liczne zastosowania w informatyce, telekomunikacji, językoznawstwie, cybernetyce, metodologii nauk (wykorzystano ją tu m.in. w badaniach nad kwestią prawdopodobieństwa konkurencyjnych hipotez, tj. stopnia ich potwierdzenia przez dane eksperymentalne). W punkcie 3 przedstawię możliwe sposoby jej wykorzystania w analizach prawniczych.

II. Pojęcie informacji w świetle teorii decyzji – informacja pragmatyczna

Jest oczywiste, że informacja stanowi zasadniczy element każdego procesu decyzyjnego – nie dysponując pełną informacją, nie jesteśmy w stanie podjąć optymalnej decyzji. Co to jednak znaczy, że informacja jest niezbędna dla podjęcia dobrej decyzji? W jaki sposób można oceniać informację ze względu na jej przydatność w podejmowaniu decyzji? Czy przydatność informacji dla decydenta jest prostą funkcją ilości informacji statystycznej? Definicja informacji statystycznej przedstawiona w poprzednim punkcie nie mówi nic o jej potencjalnych zastosowaniach w kontekstach decyzyjnych. W rzeczy samej, wartościując informację z punktu widzenia jej praktycznej doniosłości, tj. relatywizując ją do określonego problemu decyzyjnego, wprowadzamy do niej

¹⁶ Zbiór możliwych odpowiedzi jest wtedy niejako najbardziej chaotyczny.

nowy element – element pragmatyczny. Dlatego informacja, o której mowa w związku z problemem podejmowania decyzji, nosi nazwę **pragmatycznej**.

Pojęcie informacji pragmatycznej najłatwiej jest zrozumieć na prostym przykładzie. Wyobraźmy sobie sytuację, w której osoba X ma do wyboru dwa działania: a_1 i a_2 . Skutki tych działań zależą od okoliczności – „możliwych stanów rzeczy” – s_1 i s_2 , na które X nie ma wpływu. X potrafi jednak przypisać im prawdopodobieństwo; załóżmy, że prawdopodobieństwo wystąpienia s_1 wynosi 0.2, zaś s_2 – 0.8. Poniższa tabelka ilustruje tę sytuację:

	$s_1 (p = 0.2)$	$s_2 (p = 0.8)$
a_1	0	8
a_2	5	5

Liczby w tabelce reprezentują wartości, jakie decydent przypisuje możliwym skutkom swoich działań (miarą tych wartości jest „użyteczność”). Skutki te są wyznaczone nie tylko przez dane działanie, lecz także przez możliwe stany rzeczy, niezależne od decydenta (np. skutkiem działania a_1 i możliwego stanu rzeczy s_2 jest jakieś zdarzenie oceniane przez X-a na 8 jednostek użyteczności). Warto podać egzemplifikację abstrakcyjnej sytuacji decyzyjnej opisanej w tabelce. Niech a_1 będzie, na przykład, decyzją „udzielić pożyczki Y-owi”, a_2 – „nie udzielić pożyczki Y-owi”, s_1 – „Y nie zwróci pożyczki”, zaś s_2 – „Y zwróci pożyczkę”¹⁷. Dla pełnej analizy tej sytuacji będziemy potrzebować jeszcze jednego pojęcia – oczekiwanej (średniej) użyteczności (*expected utility* – EU) działania. Otóż EU działania a_i jest sumą użyteczności możliwych skutków tego działania pomnożonych przez prawdopodobieństwa ich wystąpienia. Zatem:

$$EU(a_1) = 0.2 \cdot 0 + 0.8 \cdot 8 = 6.4$$

$$EU(a_2) = 0.2 \cdot 5 + 0.8 \cdot 5 = 1 + 4 = 5$$

Zgodnie z kryterium racjonalnego wyboru, nakazującym wybierać to działanie spośród dostępnych w danej sytuacji, którego oczekiwana użyteczność jest największa, decydent powinien wybrać działanie a_1 , czyli udzielić pożyczki Y-owi.

¹⁷ Bardziej adekwatnego opisu sytuacji udzielania pożyczki dostarcza teoria gier. Opis taki jednak byłby bardziej skomplikowany. Dlatego posługuję się prostszymi narzędziami, jakim jest teoria decyzji.

Jaką rolę odgrywa w tej – jak i każdej innej – sytuacji decyzyjnej informacja **usuwająca niepewność**?¹⁸ Otóż racjonalny decydent uzależnia wybór działania od uzyskanej informacji, wykorzystując ją w taki sposób, by zmaksymalizować użyteczność skutku¹⁹. Jeśli więc dowie się, że zachodzi s_1 , to wybierze działanie a_2 , gdyż w sytuacji, w której prawdopodobieństwo s_1 równe jest jedności, a_2 jest działaniem lepszym; jeśli zaś dowie się, że zachodzi s_2 , to wybierze a_1 (czyli to działanie, które wybierał w sytuacji niepewności, tj. w sytuacji, w której nie wiedział, który z możliwych stanów rzeczy zachodzi). W obu przypadkach decydent uzyskuje maksimum tego, co w ogóle może uzyskać (czyli 5 i 8) przy danym stanie świata. Te spostrzeżenia można precyzyjnie ująć, odwołując się do pojęcia informacji pragmatycznej. Prezentuję je w wersji zaproponowanej przez *Klemensa Szaniawskiego*. Otóż zauważmy, że uzyskanie informacji likwidującej niepewność jest równoważne rozszerzeniu zbioru możliwych zachowań decydena o nowe działanie (nazwijmy je działaniem a_0), które można scharakteryzować w taki sposób²⁰:

	$s_1(p = 0.2)$	$s_2(p = 0.8)$
a_0	5	8

Możemy teraz ustalić wartość informacji zdania mówiącego, że zachodzi określony stan świata, ze względu na rozważany problem decyzyjny. Oczekiwana użyteczność działania, które stało się dostępne dla decydena dzięki zdobyciu informacji likwidującej niepewność, obliczamy w taki sam sposób jak użyteczność dla a_1 i a_2 .

$$EU(a_0) = 0.2 \cdot 5 + 0.8 \cdot 8 = 1 + 6.4 = 7.4$$

¹⁸ Ściśle rzecz biorąc, powinniśmy mówić o likwidowaniu ryzyka, a nie niepewności, ponieważ w teorii decyzji różni się sytuacje ryzyka, gdzie rozkład prawdopodobieństwa na możliwych stanach świata jest znany (a tak jest właśnie w naszej sytuacji), oraz sytuacje niepewności, gdzie rozkład ten nie jest znany. Myślę jednak, że ta nieścisłość terminologiczna, podyktowana chęcią „odciążenia” tego przeglądowego artykułu od zbyt wielu technicznych pojęć, w które i tak już obfituje, nie doprowadzi do żadnych nieporozumień.

¹⁹ Por. *H. Schneeweiss*, *Probability and Utility. Dual Concepts in Decision Theory*, w: *G. Menges* (red.), *Information, Inference and Decision*, Reidel Publishing Company, Dordrecht, Boston 1974, s. 113–145.

²⁰ To działanie nie stało się dostępne w sensie fizycznym, gdyż decydent ma nadal do wyboru dwa działania a_1 i a_2 . Taka charakterystyka stanowi jednak bardzo przejrzysty **opis abstrakcyjnej sytuacji uwzględniającej to, że decydent może uzyskać jedną z dwóch** (w naszym przykładzie) **informacji likwidujących niepewność**; w rzeczywistości decydent otrzyma oczywiście jedną – konkretną – informację.

Oczekiwana użyteczność a_0 jest zatem większa o 1 (gdyż $7.4 - 6.4 = 1$) od najwyższej oczekiwanej użyteczności, jaką możemy uzyskać w sytuacji niepewności. Ta różnica to wartość informacji w rozważanym problemie decyzji, czyli **informacja pragmatyczna** ze względu na problem decyzji. **Informacja pragmatyczna jest więc różnicą między oczekiwaną użytecznością, która jest osiągalna wtedy, kiedy dysponuje się informacją usuwającą niepewność, a oczekiwaną użytecznością osiągalną bez tej wiedzy**²¹. Rozważany kazus był jednak bardzo prosty. W bardziej złożonych sytuacjach, np. wtedy, gdy dostępnych sposobów działania i możliwych stanów świata jest bardzo wiele, jak i wtedy, gdy informacja ma bardziej złożony charakter, pojęcie informacji pragmatycznej przybiera postać nieco bardziej skomplikowaną. Załóżmy, przykładowo, że w omawianym kazusie osoba X dowiadyuje się od znajomej pracownicy banku, że stan konta Y-a jest pusty. Stan konta jest pewnym nośnikiem informacji na temat potencjalnej wypłacalności Y-a. Ale związek między tym nośnikiem a informacją, że Y nie będzie w stanie spłacić X-owi pożyczki, jest tylko probabilistyczny, gdyż Y może dysponować innymi środkami niż pieniężne, lub może np. trzymać pieniądze w domu, co umożliwi mu spłatę długu zaciągniętego u X-a. W takim przypadku nie mamy pewności, jaki stan rzeczy – s_1 , czy s_2 – ma faktycznie miejsce. Tę nową sytuację, powstałą po uzyskaniu przez X-a informacji zawartej w zdaniu e o stanie konta Y-a można opisać w następujący sposób: zamiast absolutnych prawdopodobieństw przypisanych zdaniom o możliwych stanach rzeczy, tj. $p(s)$, mamy prawdopodobieństwa względne $p(s/e)$, tj. prawdopodobieństwa zdań o możliwych stanach rzeczy ze względu na jakieś zdanie e (np. zamiast prawdopodobieństwa absolutnego zdania, że Y nie spłaci długu, mamy prawdopodobieństwo względne, że Y nie spłaci długu, przy założeniu, że stan jego konta jest pusty; znajomość e zmniejsza tu niepewność co do s). Wynika stąd, że redukcja niepewności jest tylko w granicznym przypadku tożsama z jej całkowitą likwidacją – ze wskazaniem, jaki stan rzeczy rzeczywiście zachodzi. Natomiast w sytuacji, gdy związek między informacją o określonym fakcie a stanem świata jest probabilistyczny, wartość pragmatyczna tej informacji jest nadwyżką średniej użyteczności osiągalnej dzięki znajomości owego faktu pozostającego w probabilistycznym związku ze stanem świata współ-wyznaczającym (wraz z dzia-

²¹ Łatwo zauważyć, że informacja pragmatyczna wynosi 0, jeśli istnieje taki sposób postępowania, który przy wszystkich stanach rzeczy przynosi maksimum użyteczności – nie ma wtedy problemu decyzyjnego, i w konsekwencji informacja, która miałaby pomóc go ewentualnie rozstrzygnąć, jest zbędna.

laniem decydenta) skutki oceniane przez decydenta (tę nadwyżkę mierzy się oczywiście względem sytuacji, gdy decydent nie miał tej informacji, tj. gdy jego niepewność co do s była wyższa).

Kończąc prezentację obu ujęć informacji, należy raz jeszcze podkreślić, że nie są one wzajemnie sprzeczne; dostarczają one po prostu innych miar wartości informacji: w ramach definicji statystycznej miarą tą jest spadek niepewności, natomiast w ramach definicji pragmatycznej jest nią użyteczność.

§ 3. O przydatności pojęć informacji statystycznej i pragmatycznej dla rozważań prawniczych

Nie ulega wątpliwości, że obie definicje wnoszą jasność do naszego aparatu pojęciowego. Pojęcie informacji statystycznej, dostarczając precyzyjnej charakterystyki relacji między informacją i prawdopodobieństwem, wyostrza naszą intuicję, iż sygnał niesie tym więcej informacji, im większą niepewność usuwa. Z kolei pojęcie informacji pragmatycznej jest dobrą eksplikacją mglistej intuicji, że informacja jest tym bardziej przydatna dla decydenta, im bardziej zwiększa użyteczność sposobu działania decydenta w porównaniu z sytuacją, w której nie posiadał on tej informacji. Wydaje się więc, że obie definicje są zgodne z warunkiem, który postawiłem na wstępie względem każdej eksplikacji potocznego pojęcia informacji (przypomnijmy: mówi on, że z takiej eksplikacji powinno wynikać, że informacja jest porcją wiedzy, i to wiedzy znaczącej). Czy można jednak stąd wyprowadzić wniosek, że obie definicje wnoszą coś istotnego do rozważań prawniczych w porównaniu z potocznym rozumieniem tego pojęcia? Łatwiej chyba udzielić odpowiedzi na to pytanie, gdy idzie o pojęcie informacji pragmatycznej. Dlatego od niego rozpocznę moje rozważania.

Jak wynika z analizy pojęcia informacji pragmatycznej, informacja tego rodzaju pozwala decydentowi podjąć bardziej korzystną dla niego decyzję, gdyż eliminuje negatywne skutki niepewności co do możliwych stanów rzeczy współ-wyznaczających skutki jego możliwych działań. Zatem dzięki uzyskanej informacji decydent może zwiększyć użyteczność swoich działań, tj. dzięki niej przysporzą mu one więcej korzyści. A zważywszy, że maksymalizacja spo-

łecznego dobrobytu, bogactwa, użyteczności²² (ang. *social welfare, wealth, utility*) jest jednym z najważniejszych celów prawa²³, należy uznać, że informacja pragmatyczna – jako warunek konieczny osiągnięcia możliwie wysokiej użyteczności – również powinna być jednym z celów prawa. To proste rozumowaniu prowadzi nas do sformułowania postulatu wobec prawa, który wynika z uznania maksymalizacji społecznej użyteczności za jeden z naczelných celów prawa: **prawo powinno się przyczyniać do dostarczenia decydom informacji niezbędnej do podejmowania optymalnych decyzji, tj. powinno się przyczyniać do zwiększenia informacji pragmatycznej.** O tym, że maksymalizacja informacji pragmatycznej w istocie jest jednym z najważniejszych celów prawa, świadczy bardzo wiele przepisów prawnych (np. przepisy dotyczące zawierania umowy sprzedaży, które nakładają na sprzedawcę obowiązek wyjawienia potencjalnemu nabywcy towaru jego wad i przewidują określone sankcje za niedopełnienie tego obowiązku). Warto zauważyć, że prawo zwykle nie realizuje celu, jakim jest zwiększanie informacji pragmatycznej, w bezpośredni sposób, lecz w sposób pośredni – wpływając na motywację podmiotów (np. za pomocą sankcji), które są w posiadaniu takich informacji.

Relacje między prawem i informacją pragmatyczną są jednak dużo bardziej złożone, niż wynika to z powyższych rozważań. Otóż tezę, iż prawo powinno przyczyniać się do zwiększania informacji decydom, która jest niezbędna dla podjęcia przez nich decyzji maksymalizujących ich użyteczność, należy opatrzyć co najmniej dwoma zastrzeżeniami. Po pierwsze, przyczynianie się do zwiększania ilości informacji pragmatycznej wśród decydom może uchozić za dość oczywiste zadanie prawa tylko w przypadku tych jego gałęzi, które regulują dobrowolne, a więc wzajemnie korzystne, umowy między decydentami. Jak bowiem pisałem wyżej, warunkiem maksymalizacji użyteczności decydom – osiągalnej między innymi na drodze dobrowolnych umów – jest posiadanie wystarczającej informacji przez decydom, tj. potencjalne strony umowy; brak takich informacji powstrzymuje strony od zawierania wzajemnie korzystnych umów, i w efekcie obniża społeczny dobrobyt, bogactwo, użyteczność. Relacje między prawem i informacją w innych gałęziach prawa są bardziej skomplikowane i trudne do ujęcia w postaci prostego postulatu, podobnego do przedstawionego wyżej (np. przepisy procedury karnej, przyznające

²² Między tymi pojęciami zachodzą subtelne różnice znaczeniowe, którymi jednak, biorąc pod uwagę cel pracy, nie będę się dalej zajmował.

²³ Zdaniem zwolenników nurtu w filozofii prawa zwanego *Law and Economics* jest to cel najważniejszy.

dziennikarzowi prawo do odmowy składania zeznań co do okoliczności objętych tajemnicą dziennikarską podpadają pod zupełnie inną regułę dotyczącą informacji – mianowicie nakaz ochrony dostępu do informacji w imię realizacji innej wartości, jaką jest swoboda poszukiwań dziennikarskich). Po drugie, nawet w odniesieniu do gałęzi prawa regulujących zawieranie wzajemnie korzystnych umów teza, iż prawo powinno przyczyniać się do zwiększania informacji pragmatycznej, jest zbyt dużym uproszczeniem. Można bowiem zapytać, jak na przykład miałyby wyglądać w jej świetle uzasadnienie przepisów prawa pracy ograniczających uprawnienia pracodawcy do zdobywania informacji o pewnych szczegółach stanu zdrowia potencjalnego pracownika. Taki przepis z pewnością obniża użyteczność pracodawcy, gdyż odcina go od informacji, które mogą być niezbędne dla podjęcia optymalnej, tj. maksymalizującej jego użyteczność, decyzji dotyczącej zatrudnienia pracownika. Obecność tego przepisu w systemie prawa uzasadnia jednak troska pracodawcy o pewne wartości, które mogą wchodzić w konflikt z maksymalizacją użyteczności. Taką wartością może być np. godność osoby. Wartość ta byłaby naruszona, gdyby wymagano od pracownika ujawnienia szczegółów dotyczących jego stanu zdrowia (być może wstydliwych i mogących rzutować na decyzję pracodawcy o jego zatrudnieniu). Oba zastrzeżenia odsłaniają złożoność zadania, przed jakim często staje pracodawca, regulując kwestie związane z informacją: musi on mianowicie dokonać „ważenia” wartości (takich jak godność, ochrona prywatności, swoboda uprawiania określonego zawodu itd.) dla określonych typów sytuacji, a następnie zdecydować, czy wartości uwikłane w dany typ uzasadniają „zawieszenie” czy ograniczenie postulatu maksymalizacji użyteczności, i związanego z nim na zasadzie środków – cel wymogu przyczyniania się do zwiększania ilości pragmatycznej wśród decydentów. To ograniczenie wymogu, by prawo przyczyniało się do zwiększania ilości pragmatycznej wśród decydentów (w tym także sędziów), nie zmniejsza rzecz jasna znaczenia, jakie dla rozważań prawniczych ma pojęcie informacji pragmatycznej. Łatwo bowiem zauważyć, że aby porównać „ważone” wartości relewantne dla danej sytuacji (np. wartość godności pracownika, która byłaby naruszona, gdyby pracownik musiał ujawnić szczegóły dotyczące jego stanu zdrowia) z użytecznością, jaką traci dany decydent na skutek zatajenia przed nim ważnej dlań informacji (np. pracodawca traci użyteczność, zatrudniając pracownika w sytuacji niepełnej informacji na temat jego stanu zdrowia), trzeba dysponować jakąś miarą wartości informacji, a takiej miary dostarcza właśnie pojęcie informacji pragmatycznej (*vide*: kasus z pożyczką z punktu 2.2 ilustrujący najprostsza metodę jej obliczania).

Podsumujmy. Informacja pragmatyczna wydaje się przydatna w kontekście rozważań prawniczych, gdyż pozwala nadać precyzję i – w wielu przypadkach – dokonać aproksymatywnej kwantyfikacji tych informacji, które mają znaczenie dla podejmowania decyzji pozostających w obrębie zainteresowań pracodawcy. Mówiąc bardziej ogólnie, odwoływanie się do takich precyzyjnie zdefiniowanych pojęć jak informacja pragmatyczna jest charakterystyczne dla takiego, jeszcze wciąż słabo rozpowszechnionego wśród prawników, podejścia do prawa, które czyni szeroki użytek z metod matematycznych, takich jak teoria decyzji, teoria gier i teoria wyboru społecznego. Warto dodać, że w ramach tego nurtu wyróżniono szereg podtypów informacji pragmatycznej, niezbędnych dla przeprowadzenia precyzyjnych analiz już bardziej konkretnych zagadnień prawnych²⁴.

Z pewnością trudniej będzie uzasadnić tezę o korzyściach płynących dla prawa z odwołania się do definicji statystycznej. Jaki bowiem związek ma ujęcie informacji w postaci funkcji logarytmicznej, czy spór o to, jak powinna się „zachowywać” intuicyjnie trafna funkcja informacji, gdy prawdopodobieństwo zdania zbliża się do zera, z kwestiami, które interesują prawnika? Na pierwszy rzut oka wydaje się, że żaden. Sądzę jednak, że taka odpowiedź z co najmniej dwóch powodów nie jest właściwa. Po pierwsze, pojęcie informacji statystycznej może być przydatne dla tego nurtu badań nad prawem, który w literaturze anglosaskiej jest określany mianem *Artificial Intelligence and the Law*. W ramach tego nurtu bada się takie problemy jak, np. czy system prawa jest Maszyną *Turinga*, tj. czy wszystkie zagadnienia prawne są rozstrzygalne, lub czy proces podejmowania decyzji przez sędziego może być modelowany za pomocą maszyn cyfrowych, i jakie konsekwencje płyną stąd dla klasycznego pytania stawianego w ramach filozofii prawa o status logiczny rozumowań prawniczych. Dla analizy tych problemów przydatny może być opis norm prawnych korzystający ze statystycznej definicji informacji. Po drugie, pojęcie informacji statystycznej może się okazać przydatne także w badaniach nad językiem prawniczym, w szczególności nad takimi kwestiami, jak jego nieostrość (ang. *vagueness*), rozmytość (ang. *fuzziness*) czy otwarta tekstowość (ang. *open texture*) (choć trzeba przyznać, że w dotychczasowej refleksji nad tymi zagadnieniami pojęcie informacji statystycznej pojawia się niezmiernie rzadko). Te 2 racje wydają się być wystarczające dla uzasadnienia tezy o przydatności pojęcia in-

²⁴ Egzemplifikacją takiego podejścia do prawa jest praca *D.G. Baird, R.G. Gertner, R.C. Picker, Game Theory and the Law*, Harvard University Press, Cambridge, Mas., London, Englad, 2000. Por. także *W. Zaluski, Game Theory in Jurisprudence*, Copernicus Center Press, Krakow 2013.

formacji statystycznej dla rozważań prawniczych. Wypada jednak się zgodzić, że o ile definicja pragmatyczna informacji jest przydatna zarówno dla teoretyka, jak i dla dogmatyka prawa, o tyle definicja statystyczna, z uwagi na dość wąski i hermetyczny krąg problemów, z którymi się wiąże, może okazać się interesująca raczej tylko dla teoretyka prawa.