LOGIKA WNIOSKOWAŃ EMPIRYCZNYCH

Krystyna Misiuna
Uniwersytet Warszawski

I. WNIOSKOWANIA EMPIRYCZNE JAKO WNIOSKOWANIA SUPRAKLASYCZNE

Działania, jakie podejmujemy w życiu codziennym, poprzedzone są z reguły wnioskowaniami opartymi na niezupełnej informacji: zbiór sądów I jest niezupełny, jeśli nie potrafimy rozstrzygnąć czy konkretny sąd $i \in I$, ani też nie potrafimy rozstrzygnąć czy $\neg i \in I$. Zdając sobie sprawę, że informacja, jaka doprowadziła nas do danego wniosku, jest niezupełna, dopuszczamy możliwość wycofania się z tego wniosku lub dodania nowych konkluzji w sytuacji, kiedy uzyskamy nowe informacje, przy zachowaniu dotychczasowych przesłanek. Jednak liczne przykłady codziennych wnioskowań wskazują na to, że takie wnioskowania są podstawą naszych decyzji. Wszystkim tym wnioskowaniom towarzyszy przyjmowane milcząco założenie, że w normalnych warunkach taki a taki wniosek wynika z takiej a takiej przesłanki, jak na przykład: *Jeśli przekręcę kluczyk w stacyjce samochodu, to silnik samochodu zacznie pracować.* To ukryte tu założenie o normalności, pozwalające na wykluczenie wyjątkowych okoliczności, nie daje się sformalizować na gruncie logiki klasycznej, gdyż może to być jeden z dwu przypadków:

1. $F(x) \rightarrow S(x)$ jest w normalnych warunkach prawdziwe;
2. $[F(x) \wedge G(x)] \rightarrow S(x)$ nie jest w normalnych warunkach prawdziwe.

Mamy więc do czynienia z wnioskowaniami, w których wniosek nie wynika z konieczności z przesłanek, jak w przypadku wnioskowań czysto dedukcyjnych, a relacja konsekwencji, na jakiej opierają się takie wnioskowania, jest relacją niemonotoniczną. Niemonotoniczne relacje konsekwencji są w pewnym sensie silniejsze niż klasyczna relacja konsekwencji, gdyż są nadzbiorami klasycznej relacji konsekwencji, i dlatego nazywane są relacjami supraklasycznymi (Makinson 2005). Interesujące jest to, że istnieją również supraklasyczne relacje konsekwencji będące relacjami monotonicznymi. Jak zatem można pogodzić ich istnienie z faktem maksymalności klasycznej relacji konsekwencji? Okazuje się, że jest to możliwe tylko wtedy, kiedy te supraklasyczne relacje konsekwencji, zarówno monotoniczne jak też niemonotoniczne, nie są domknięte na podstawianie, czyli nie są strukturalne. Nie istnieje bowiem supraklasyczna relacja konsekwencji w tym samym języku, co język klasycznej relacji konsekwencji, która jest domknięta na podstawianie, z wyjątkiem samej klasycznej relacji konsekwencji i relacji pełnej. W przypadku wnioskowań, w których relacje konsekwencji są relacjami supraklasycznymi może być tak, że konkluzje są fałszywe przy prawdziwości wszystkich przesłanek.

I. 1. Supraklasyczne relacje konsekwencji we wnioskowaniach entymematycznych

Przykładem supraklasycznych relacji konsekwencji, które są zarazem monotoniczne, jak też przechodnie i zwrotne, są relacje konsekwencji, z jakimi mamy do czynienia we wnioskowaniach entymematycznych . Niech L będzie językiem rachunku zdań, natomiast $K \subseteq L$ niech będzie ustalonym zbiorem formuł, który pełni rolę dodatkowych założeń. Niech A będzie zbiorem formuł, a x indywidualną formułą. Zdefiniujemy relację konsekwencji ze względu na zbiór założeń K w następujący sposób:

Definicja: $A \Rightarrow_K x$ wtedy i tylko wtedy, gdy nie istnieje taka interpretacja v, że $v(K \cup A) = 1$, a $v(x) = 0$.

Dla każdego zbioru K relacja konsekwencji ze względu na zbiór K jest nadzbiorem klasycznej relacji konsekwencji: $\Rightarrow \subseteq \Rightarrow_K$, lecz nie jest to relacja konsekwencji domknięta na podstawianie. Makinson (2005, 25) podaje następujący przykład, pokazujący na czym polega w tym przypadku to, że mamy do czynienia z relacją konsekwencji, która nie jest domknięta na podstawianie.

Przykład: Niech $K = \{p\}$, gdzie p jest zmienną zdaniową, $A = \{q\}$, $x = p \wedge q$. Niech σ będzie funkcją podstawiania, która przyporządkowuje zmiennej zdaniowej q tę samą zmienną zdaniową, czyli $\sigma(q) = q$, natomiast zmiennej p przyporządkowuje r, czyli $\sigma(p) = r$. Wtedy otrzymujemy, że $\sigma(x)$ nie jest konsekwencją $\sigma(A)$ ze względu na zbiór K, ponieważ $K \cup \{\sigma(q)\} = K \cup \{q\} = \{p, q\}$. Jest tak dlatego, że podstawienie stosujemy do zbioru A i konkluzji x, a nie do zbioru K, który pozostaje stały. Zatem $\sigma(x) = r \wedge q$ nie jest konsekwencją klasyczną zbioru $\{p, q\}$.

Przykładem eliptycznego argumentu, który staje się wnioskowaniem ze względu na jednoelementowy zbiór K jest następujący argument: „Sokrates jest człowiekiem, zatem jest on śmiertelny”, gdzie K = { Każdy człowiek jest śmiertelny}. Dodatkowa przesłanka czyni ten argument formalnie i materialnie poprawnym. Konkluzja „Sokrates jest śmiertelny” jest klasyczną konsekwencją zbioru przesłanek: { Każdy człowiek jest śmiertelny. Sokrates jest człowiekiem}.

Relacja supraklasyczna zdefiniowana w powyższy sposób jest jednak monotoniczna. Niech $A \Rightarrow_K x$ i $A \subseteq B$. Pokażemy, że $B \Rightarrow_K x$. Na podstawie definicji supraklasycznej relacji konsekwencji ze względu na zbiór K oraz na podstawie tego, że $A \Rightarrow_K x$ dostajemy $A \cup K \Rightarrow x$. Ponieważ klasyczna relacja konsekwencji jest monotoniczna i ponieważ $A \subseteq B$ dostajemy $B \cup K \Rightarrow x$, czyli $B \Rightarrow_K x$.

I. 2. Supraklasyczne relacje konsekwencji niemonotoniczne we wnioskowaniach entymematycznych

Przejdziemy od monotonicznych relacji konsekwencji, jakimi są relacje ze względu na zbiór K, do niemonotonicznych relacji konsekwencji, jeśli weźmiemy pod uwagę maksymalne podzbiory K' zbioru K niesprzeczne z A. Podzbiór K' zbioru K jest niesprzeczny z A, jeśli istnieje klasyczna interpretacja v, taka że $v(K' \cup A) = 1$, natomiast K' jest maksymalnie niesprzeczny z A, jeśli jest niesprzeczny z A i nie jest podzbiorem właściwym żadnego podzbioru K, który jest niesprzeczny z A. Entymematyczną niemonotoniczną relację konsekwencji ze względu na zbiór K definiujemy na tej podstawie w następujący sposób:

Definicja: $A \rightarrow_K x$ wtedy i tylko wtedy, gdy $K' \cup A \Rightarrow x$ dla każdego podzbioru $K' \subseteq K$ maksymalnie niesprzecznego z A.

Przykład (Makinson 2005, 31): Niech $K = \{p \rightarrow q, q \rightarrow r, r \rightarrow \neg p\}$, $A = \{p\}$. K jest sprzeczny z A, więc musimy znaleźć maksymalne podzbiory K niesprzeczne z A. Będą to:

$K' = \{p \rightarrow q, q \rightarrow r\}$;

$K'' = \{ p \rightarrow q, r \rightarrow \neg p\}$;

$K''' = \{ q \rightarrow r, r \rightarrow \neg p\}$.

Zgodnie z definicją $A \rightarrow_K x$ jeśli $K_i \cup \{p\} \Rightarrow x$ dla i = 1,2,3. W związku z tym dostajemy:

$Cn(K' \cup \{p\}) = Cn (\{p, q, r\})$;

$Cn(K'' \cup \{p\}) = Cn(\{p, q, \neg r\})$;

$Cn(K''' \cup \{p\}) = Cn(\{p, \neg q, \neg r\})$.

Ponieważ $A \Rightarrow x$ wtedy i tylko wtedy, gdy $x \in Cn(A)$, gdzie $Cn(A) = \{x: A \Rightarrow x\}$, zatem x jest niemonotoniczną konsekwencją ze względu na zbiór K, jeśli x należy do wyżej wymienionych trzech zbiorów, czyli jeśli x jest klasyczną konsekwencją alternatywy:

$x \in Cn((p \wedge q \wedge r) \vee (p \wedge q \wedge \neg r) \vee (p \wedge \neg q \wedge \neg r))$.

Alternatywa ta jest klasycznie równoważna z $p \wedge (q \vee \neg r)$, a zatem konsekwencjami niemonotonicznymi ze względu na zbiór K są konsekwencje klasyczne powyższej formuły.

Widzimy więc, że zbiór tych konsekwencji jest większy niż zbiór klasycznych konsekwencji zbioru A. Ponieważ nasza relacja konsekwencji ze względu na zbiór K jest teraz niemonotoniczna, więc jest tak, że $A \rightarrow_K x$, lecz nie zachodzi $A \cup B \rightarrow_K x$, dla zbiorów formuł A, B. Natomiast dla indywidualnych formuł możemy mieć $a \rightarrow_K x$, a przy tym niespełniony warunek: $a \wedge b \rightarrow_K x$.

Przykład ilustrujący brak monotoniczności relacji $\rightarrow_K$ (Makinson 2005, 32): Niech $K = \{p \rightarrow q, q \rightarrow r\}$, gdzie p, q, r są zmiennymi zdaniowymi. Wtedy $p \rightarrow_K r$, ponieważ przesłanka p jest niesprzeczna z całym zbiorem K i $\{p\} \cup K \Rightarrow r$. Lecz nie zachodzi $\{p, \neg q\} \rightarrow_K r$, ponieważ zbiór przesłanek teraz nie jest niesprzeczny z całym zbiorem K, lecz tylko z podzbiorem $K' = \{ q \rightarrow r\}$. Jednak nie zachodzi relacja klasycznej konsekwencji $\{p, \neg q\} \cup K' \Rightarrow r$, gdyż istnieje taka interpretacja v, która jest kontr-modelem, a mianowicie $v(p) = 1$, $v(q) = 0$ i $v(r) = 0$. Chociaż uzyskaliśmy tu dodatkową informację w postaci dodatkowej przesłanki, to utraciliśmy jedno z założeń ze zbioru K. To, że wycofujemy się z wcześniejszego wniosku r możemy interpretować w ten sposób, że informacja, jakiej dostarcza nowa przesłanka jest bardziej wiarygodna w stosunku do wcześniejszych założeń, które doprowadziły do uznania r. We wnioskowaniach niemonotonicznych uznanie lub odrzucenie wniosku zależy nie tylko od struktury logicznej przesłanek i ewentualnie dodatkowych założeń, jak to ma miejsce w przypadku wnioskowań czysto dedukcyjnych, lecz również od treści tych przesłanek i założeń. Zależność tę dobrze ilustruje następujący prosty przykład. Z przesłanki „To stworzenie jest ptakiem" wyprowadzamy niemonotonicznie wniosek: „To stworzenie lata". Jeśli jednak uzyskamy dodatkową informację: „To stworzenie jest pingwinem", to przestajemy uznawać wcześniejszy wniosek „To stworzenie lata".

Wnioskowania, które nie są monotoniczne w zwykłym sensie mogą spełniać warunek ostrożnej monotoniczności, tak jak jest w przypadku niemonotonicznych relacji konsekwencji ze względu na dodatkowy zbiór założeń. Jest to słabsza postać monotoniczności w tym sensie, że warunek zwykłej monotoniczności implikuje warunek ostrożnej monotoniczności, lecz nie na odwrót. W najprostszej postaci warunek ostrożnej monotoniczności przybiera następującą postać:

(OM) Jeśli $A \rightarrow_K x$ i $A \rightarrow_K y$, to $A \cup \{x\} \rightarrow_K y$.

W przypadku niemonotonicznej relacji konsekwencji modulo zbiór założeń K mamy do czynienia z pragmatycznym dylematem. Polega on na tym, że jeśli K nie jest domknięty na klasyczną konsekwencję, czyli gdy $K \neq Cn(K)$, to niemonotoniczna relacja konsekwencji modulo K może być zależna od sposobu sformułowania K, a nie od jego zawartości;

natomiast jeśli K jest domknięty na klasyczną konsekwencję, czyli gdy $K = Cn(K)$, to ta relacja konsekwencji jest identyczna z klasyczną relacją konsekwencji w przypadku sprzeczności zbioru A z K. Poniższy przykład pokazuje, że otrzymamy różne niemonotoniczne relacje konsekwencji, odpowiednio modulo K i K', dla klasycznie równoważnych zbiorów K i K'. Niech $K = \{p \to (q \wedge r), r \to \neg p\}$, a $K' = \{p \to q, p \to r, r \to \neg p\}$, które są klasycznie równoważne. Niech $A = \{p\}$. Zatem A jest sprzeczne z K, a podzbiorami K, które są maksymalne niesprzeczne z A są:

$\{p \to (q \wedge r)\}$ i $\{r \to \neg p\}$.

Ponieważ q nie jest konsekwencją klasyczną zbioru $\{p, r \to \neg p\}$ więc tym samym nie zachodzi relacja $A \to_K q$.

Podzbiorami K', które są maksymalnie niesprzeczne z A są:

$\{p \to q, p \to r\}$ i $\{p \to q, r \to \neg p\}$.

Tym samym q jest klasyczną konsekwencją obu tych podzbiorów K wraz ze zbiorem $A = \{p\}$, a zatem w tym przypadku dostajemy: $A \to_{K'} q$.

I. 3. Supraklasyczne relacje konsekwencji z ograniczonym zbiorem interpretacji

Relacje konsekwencji, które są nadzbiorami klasycznej relacji konsekwencji możemy uzyskać również wtedy, kiedy ograniczymy zbiór interpretacji, czyli gdy weźmiemy pod uwagę podzbiór W całego zbioru interpretacji V i zdefiniujemy relację konsekwencji nie ze względu na V, lecz ze względu na W:

Definicja: $A \Rightarrow_W x$ wtedy i tylko wtedy, gdy nie istnieje interpretacja v należąca do W taka, że $v(A) = 1$, a $v(x) = 0$.

Otrzymamy w ten sposób supraklasyczne relacje konsekwencji, które nie są domknięte na podstawianie. Niech warunek supraklasyczności wyraża następująca implikacja:

$$\text{Jeśli } a \Rightarrow x, \text{ to } a \Rightarrow_W x.$$

Załóżmy, że $a \Rightarrow x$. Zatem $v(x) = 1$ dla każdej interpretacji v, przy której $v(a) = 1$. Tym samym $v(x) = 1$ dla każdej interpretacji $v \in W$, przy której $v(a) = 1$, a zatem $a \Rightarrow_W x$, co dowodzi zachodzenie warunku supraklasyczności relacji konsekwencji ze względu na zbiór interpretacji W.

Jeśli weźmiemy pod uwagę skończony język, czyli język generowany przez Boolowskie spójniki ze skończonego zbioru zmiennych zdaniowych, to dla takiego języka supraklasyczne monotoniczne relacje konsekwencji ze względu na zbiór założeń K są tymi samymi relacjami konsekwencji co supraklasyczne monotoniczne relacje konsekwencji ze względu na zbiór interpretacji W. Główna różnica między tymi dwoma rodzajami supraklasycznych relacji konsekwencji polega na tym, że relacje konsekwencji z ograniczonym zbiorem interpretacji nie zawsze spełniają warunek zwartości, lecz informatyk, który pracuje na skończonym języku, może uważać te dwa rodzaje konsekwencji za równoważne, gdyż na skończonym języku spełniają one warunek zwartości.

I. 4. Preferencyjne relacje konsekwencji

Przejście od supraklasycznych monotonicznych relacji konsekwencji do supraklasycznych niemonotonicznych odbywa się za pośrednictwem modeli preferencyjnych. Modelem takim jest para (W, <), gdzie W jest, tak jak poprzednio, zbiorem interpretacji, a < jest prrciwzwrotną i przechodnią relacją na W. Dysponując pojęciem modelu preferencyjnego możemy zdefiniować preferencyjną relację konsekwencji w następujący sposób:

Definicja: $A \rightarrow_{<} x$ wtedy i tylko wtedy, gdy dla każdej interpretacji $v \in W$, która jest minimalna ze względu na < i która spełnia A, v(x) = 1. Inaczej mówiąc, gdy każdy minimalny ze względu na < model A jest modelem x.

Niech **A** będzie zbiorem wszystkich interpretacji w zbiorze W, które spełniają A, czyli $\mathbf{A} = \{v \in W: v(A) = 1\}$ jest zbiorem modeli formuł zawartych w A. Niech $\min_{<} \mathbf{A}$ będzie zbiorem wszystkich minimalnych elementów **A** czyli zbiorem modeli minimalnych formuł zawartych w A. Przy takiej notacji możemy naszą definicję sformułować w następujący sposób: $A \rightarrow_{<} x$ wtw, gdy $v \in \mathbf{x}$ jeśli $v \in \min_{<} \mathbf{A}$, co możemy sprowadzić do następującej równoważności:

$$A \rightarrow_{<} x \quad \text{wtw, gdy} \quad \min_{<} \mathbf{A} \subseteq \mathbf{x}.$$

Korzystając z powyższej notacji możemy zdefiniować warunek zastopowania modeli preferencyjnych:

$$\text{Jeśli } v \in \mathbf{A}, \text{ to albo } v \in \min_{<} \mathbf{A}, \text{ albo istnieje } u < v, \text{ gdzie } u \in \min_{<} \mathbf{A}.$$

Poprzednio zdefiniowaną relację konsekwencji $A \Rightarrow_{W} x$ możemy uważać za szczególny przypadek preferencyjnej relacji konsekwencji, gdy < jest relacją pustą. Preferencyjne relacje konsekwencji są niemonotoniczne.

Przykład niemonotonicznej preferencyjnej relacji konsekwencji (Makinson 2005, 69): Niech język L zawiera trzy zmienne zdaniowe p, q, r. Niech W będzie zbiorem dwu interpretacji v' i v'' zdefiniowanych w następujący sposób: v'(p) = v''(p) = 1, v'(q) = 0, v''(q) = 1, v'(r) = 1, v''(r) = 0. Interpretacje te uporządkowane są tak, że v' < v''. W takim modelu preferencyjnym p $\rightarrow_{<}$ r, ponieważ najmniejszą interpretacją, przy której p jest prawdziwe jest v’ i r jest też prawdziwe przy v’. Lecz w modelu tym nie zachodzi preferencyjna relacja konsekwencji p $\wedge$ q $\rightarrow_{<}$ r, ponieważ najmniejszą interpretacją, przy której p $\wedge$ q jest prawdziwe jest v’’, a r jest fałszywe przy v’’. Przykład ten pokazuje zarazem, że niemonotoniczna preferencyjna relacja konsekwencji może nie być przechodnia, gdyż mamy: p $\wedge$ q $\rightarrow_{<}$ p i p $\rightarrow_{<}$ r, lecz nie zachodzi p $\wedge$ q $\rightarrow_{<}$ r.

Łatwo można wykazać supraklasyczność preferencyjnych relacji konsekwencji. Wyrazimy supraklasyczność przy pomocy warunku: Jeśli A$\Rightarrow$x, to A$\rightarrow_{<}$ x. Niech preferencyjna relacja konsekwencji będzie określona przez model preferencyjny (W, <). Dla dowodu załóżmy, że nie zachodzi relacja A$\rightarrow_{<}$ x. Pokażemy, że nie zachodzi też wtedy A$\Rightarrow$x. Jeśli nie zachodzi relacja preferencyjnej konsekwencji, to istnieje minimalny model zbioru A: v, który nie jest modelem x, a zatem v(x) = 0. Ponieważ jednak v jest minimalnym modelem zbioru formuł A, więc v(A) = 1. Zatem nie zachodzi klasyczna relacja konsekwencji A$\Rightarrow$x.

Preferencyjne relacje konsekwencji nie zawsze spełniają warunek ostrożnej monotoniczności, który w przypadku indywidualnych formuł ma postać:

Jeśli a$\rightarrow$b i a$\rightarrow$x, to a $\wedge$b $\rightarrow$x.

Chociaż własność ta czasami nie przysługuje preferencyjnej relacji konsekwencji, to jednak zawsze występuje w przypadku skończonych modeli preferencyjnych. Ogólnie mówiąc zachodzi ona wtedy, gdy model spełnia warunek zastopowania gwarantujący istnienie minimalnych modeli preferencyjnych. Preferencyjne relacje konsekwencji mogą być również pozbawione własności zachowania niesprzeczności. Jest to własność opisana przez warunek:

Jeśli A$\rightarrow$f, to A$\Rightarrow$f,

Gdzie f jest klasyczną sprzecznością. Własność ta zachodzi dla supraklasycznej relacji konsekwencji modulo K, lecz może nie zachodzić dla niemonotonicznej preferencyjnej relacji konsekwencji z tego powodu, że niektóre interpretacje są pominięte. Gdy interpretacja, przy której a jest prawdziwe nie należy do W, to dostaniemy v(a) = 0 dla wszystkich v należących do W, pomimo tego, że a nie jest klasyczną sprzecznością, a zbiór minimalnych modeli preferencyjnych sądu a będzie pusty i tym samym niemonotoniczną preferencyjną konsekwencją a będzie f. Zachowanie niesprzeczności obowiązuje dla preferencyjnych modeli zastopowanych w następującej postaci: Jeśli A $\rightarrow_{<}$ f, to A$\Rightarrow_{W}$ f.

I. 5. Konsekwencja klasyczna definiowana przez funkcję prawdopodobieństwa

Z niemonotonicznością mamy też do czynienia w przypadku wnioskowań statystycznych. Prawdopodobieństwo warunkowe sądu x ze względu na sąd a może spadać, jeśli dołączymy do a dodatkową informację b: $p(x/a \wedge b) < p(x/a)$, gdzie p jest funkcją prawdopodobieństwa. Aksjomatyka Kołmogorowa definiuje pojęcie funkcji prawdopodobieństwa. Dziedziną tej funkcji może być zbiór wszystkich formuł Boolowskich języka zdaniowego, natomiast przeciwdziedziną przedział domknięty liczb rzeczywistych [0, 1]. Jeden z tych aksjomatów mówi, że

$$p(x) \leq p(y), \text{ jeśli } x \Rightarrow y.$$

Funkcja prawdopodobieństwa $p: L \to [0, 1]$ jest bogatsza niż funkcja interpretacji $v: L \to \{0, 1\}$, a tym samym teoria prawdopodobieństwa jest znacznie bardziej złożona niż logika funkcji prawdziwościowych. To prowadzi do tego, że nie istnieje jeden sposób opisywania relacji klasycznej konsekwencji w terminach probabilistycznych. Jednym z takich sposobów, odwołującym się do wspomnianego aksjomatu Kołmogorowa, jest następująca definicja:

Definicja: Dla dowolnych formuł Boolowskich a, x: $a \Rightarrow x$ wtedy i tylko wtedy, gdy $p(a) \leq p(x)$, dla każdej funkcji $p: L \to [0, 1]$.

Zachodzenie warunku: jeśli $a \Rightarrow x$, to $p(a) \leq p(x)$ zagwarantowane jest wyżej wymienionym aksjomatem Kołmogorowa. Łatwo zauważyć, że zachodzi też warunek w odwrotną stronę: jeśli $p(a) \leq p(x)$, to $a \Rightarrow x$. Dla dowodu załóżmy, że x nie jest konsekwencją klasyczną a. Zatem istnieje taka funkcja interpretacji $v: L \to \{0, 1\}$, że $v(a) = 1$; $v(x) = 0$. Ponieważ funkcja v jest szczególnego rodzaju funkcją prawdopodobieństwa, gdyż spełnia wszystkie aksjomaty Kołmogorowa, więc otrzymujemy $v(a) > v(x)$, co kończy dowód.

Inne aksjomaty Kołmogorowa definiujące funkcję prawdopodobieństwa:

1. $0 \leq p(x) \leq 1$
2. $p(x) = 1$ dla pewnej formuły x
3. $p(x \vee y) = p(x) + p(y)$, jeśli x jest klasycznie sprzeczne z y.

Funkcja prawdopodobieństwa jest rozumiana tu jako dowolna funkcja ze zbioru formuł języka domkniętego na operacje Boolowskie w zbiór liczb rzeczywistych spełniająca cztery aksjomaty Kołmogorowa. Chociaż funkcja ta ma tę samą dziedzinę, jaką posiada funkcja interpretacji w klasycznym rachunku zdań, to zdecydowanie różni się od tej ostatniej. W klasycznym rachunku zdań wartościowanie zmiennych zdaniowych może być jednoznacznie rozszerzone na zbiór wszystkich formuł, lecz funkcja prawdopodobieństwa nie. Będą istniały wartościowania zmiennych zdaniowych przypisujące im liczby rzeczywiste z przedziału [0, 1], które mogą być rozszerzone na więcej niż jeden sposób do funkcji p na zbiorze wszystkich formuł zdaniowych spełniających aksjomaty Kołmogorowa.

Przykład (Makinson, 2005, 115): Rozważmy język z dwiema zmiennymi zdaniowymi: q , r . Niech funkcja wartościowania f przyporządkowuje im f(q) = f(r) = 0,5. Istnieje wiele sposobów, w jakie możemy rozszerzyć ją do funkcji prawdopodobieństwa p: L $\rightarrow$ [0, 1]. Przy jednym z nich mamy $p(q \wedge r) = p(\neg q \wedge \neg r) = 0{,}5$ i $p(q \wedge \neg r) = p(\neg q \wedge r) = 0$. Natomiast przy drugiej funkcji prawdopodobieństwa możemy mieć: $p'(q \wedge r) = p'(q \wedge \neg r) = p'(\neg q \wedge r) = p'(\neg q \wedge \neg r) = 0{,}25$. Rozważmy teraz klasyczny język zdaniowy L ze skończoną liczbą zmiennych zdaniowych n. Przez opis stanu tego języka rozumiemy koniunkcję n zmiennych, z których każda jest albo zwykłą zmienną, albo jej negacją. Niech f będzie funkcją na zbiorze 2^n opisów stanu naszego języka w zbiór [0, 1] taką, że jej wartości sumują się do 1. Funkcję taką nazywamy *rozkładem prawdopodobieństwa.* Wtedy f może być rozszerzona w jednoznaczny sposób do funkcji prawdopodobieństwa p: L $\rightarrow$ [0, 1] spełniającej aksjomaty Kołmogorowa.

Jeśli p jest funkcją prawdopodobieństwa przyporządkowującą sądowi a pewną wartość ze zbioru [0, 1] i dowiemy się, że a jest prawdziwe, to jak zmodyfikujemy p na podstawie tej informacji? W takiej sytuacji powinniśmy przejść od funkcji p do funkcji $p(x/a) = p(x \wedge a)/p(a)$. Jest to funkcja prawdopodobieństwa warunkowego określona na tej samej dziedzinie, na której określona jest funkcja p. Rozważmy jeszcze inne sposoby definiowania klasycznej relacji konsekwencji w terminach prawdopodobieństwa.

Definicja: Niech t będzie dowolną liczbą rzeczywistą z przedziału [0, 1] taką, że t jest różne od 0; a, x niech reprezentują dowolne Boolowskie formuły, wtedy $a \Rightarrow x$ wtw, gdy dla każdej funkcji prawdopodobieństwa p jeśli $p(a) \geq t$, to $p(x) \geq t$.

Liczba t nazywana jest wielkością progową tego warunku. Inna definicja ten sam warunek wyraża równoważnie w następujący sposób:

Definicja: $a \Rightarrow x$ wtedy i tylko wtedy, gdy $p(x/a) \geq t$ lub $p(a) = 0$, dla każdej funkcji prawdopodobieństwa p.

Definicja ta odwołuje się do prawdopodobieństwa warunkowego. Jeszcze inny warunek definiujący klasyczną relację konsekwencji w terminach prawdopodobieństwa to:

Definicja: $a \Rightarrow x$ wtedy i tylko wtedy, gdy $p(a \rightarrow x) \geq t$; lub równoważnie $p(\neg a \vee x) \geq t$, dla każdej funkcji prawdopodobieństwa p.

I. 6. Supraklasyczne relacje konsekwencji definiowane w terminach prawdopodobieństwa

Mając zdefiniowaną klasyczną relację konsekwencji w terminach prawdopodobieństwa możemy zastanawiać się nad tym, jak można otrzymać supraklasyczne relacje konsekwencji i ewentualnie niemonotoniczne. Jak się okazuje do relacji supraklasycznych prowadzi ograniczenie zbioru funkcji prawdopodobieństwa do pewnego niepustego podzbioru Q zbioru P wszystkich funkcji prawdopodobieństwa dla danego języka. Na przykład w przypadku warunku $a \Rightarrow x$ wtw, gdy $p(a) \leq p(x)$, jeśli jest taka formuła Boolowska, która nie jest tautologią, lecz taka że $p(x) = 1$ dla każdej funkcji p ze zbioru Q, to chociaż x nie będzie klasyczną konsekwencją dowolnej formuły, to będzie konsekwencją supraklasyczną. Relacja ta jednak będzie monotoniczna. Załóżmy, że spełniony jest warunek $p(a) \leq p(x)$ dla każdej funkcji prawdopodobieństwa w zbiorze Q. Wtedy $p(a \wedge b) \leq p(a) \leq p(x)$ dla wszystkich $p \in Q$. Natomiast w przypadku warunku odwołującego się do prawdopodobieństwa warunkowego: $p(x/a) \geq t$ lub $p(a) = 0$, gdy ponadto zbiór Q nie jest domknięty na prawdopodobieństwo warunkowe, relacja konsekwencji może nie być monotoniczna. W szczególnym przypadku jest tak wtedy, gdy Q jest zbiorem {p}. Przykład (Makinson 2005, 128): Musimy znaleźć odpowiednie formuły Boolowskie a, b, x i odpowiednią funkcję p taką, że $p(x/a) \geq t$, podczas gdy $p(x/a \wedge b) < t$. Załóżmy, że mamy w języku tylko q i r, a rozkład prawdopodobieństwa dla wszystkich czterech opisów stanu ma wartość 0,25, wartość $t = 0,5$, przy czym

$a = q \vee \neg q$

$x = q \vee r$

$b = \neg q \wedge \neg r$.

Wtedy dostajemy $p(a) = 1$, $p(x/a) = p(x \wedge a)/p(a) = p(x) = p(q \vee r) = 0,75$.

Natomiast $p(x/a \wedge b) = p(x \wedge a \wedge b)/p(a \wedge b) = p[(q \vee r) \wedge (q \vee \neg q) \wedge (\neg q \wedge \neg r)]/ p[(q \vee \neg q) \wedge (\neg q \wedge \neg r) = 0/0,25 = 0$. Zatem relacja konsekwencji między konkluzją x i przesłanką a zachodzi, natomiast nie zachodzi, gdy do przesłanki a dodamy przesłankę b, a zatem relacja konsekwencji definiowana przez warunek: $p(x/a) \geq t$ nie jest monotoniczna. W ten sposób pokazaliśmy tym samym, że możemy zdefiniować następującą niemonotoniczną relację konsekwencji: Dla dowolnej liczby rzeczywistej $t \in [0, 1]$ i takiej, że $t \neq 0$ oraz dowolnej funkcji prawdopodobieństwa p: $a \rightarrow_{tp} x$ wtedy i tylko wtedy, gdy $p(x/a) \geq t$ lub $p(a) = 0$. Definicja ta wyraża pojęcie wystarczającego prawdopodobieństwa x na podstawie danego a.

Makinson (2005, 130) podaje również przykład niemonotonicznej relacji konsekwencji zdefiniowanej w następujący sposób: a $\to^{+}_{tp}$ x wtedy i tylko wtedy, gdy $p(x/a) \geq p(x)$, pod warunkiem, że $p(a) \neq 0$. W tym przypadku warunek a prowadzi do prawdopodobieństwa x, które jest wyższe niż mogłoby być bez tego warunku.

Supraklasyczne relacje konsekwencji zdefiniowane w terminach funkcji prawdopodobieństwa, w przeciwieństwie do innych relacji supraklasycznych, nie posiadają własności koniunkcji w konkluzji, jaka przysługuje również klasycznej relacji konsekwencji, która sprowadza się do tego, że:

$$\text{Jeśli } A \Rightarrow x \text{ i } A \Rightarrow y, \text{ to } A \Rightarrow x \wedge y.$$

Przykład: Rozważmy relację konsekwencji z poprzedniego przykładu $\to_{tp}$, gdzie wyraża ją warunek: $p(x/a) \geq t$ lub $p(a) = 0$, oraz ten sam język i ten sam rozkład prawdopodobieństwa jak w poprzednim przykładzie. Niech a będzie tautologią, $x = q$, $y = r$, $t = 0{,}5$. Wtedy dostajemy $p(x/a) = p(x) = 0{,}5 \geq t$ oraz $p(y/a) = p(y) = 0{,}5$. Zatem x wynika supraklasycznie z a i y wynika supraklasycznie z a. Natomiast $p(x \wedge y/a) = p(x \wedge y) = 0{,}25 < t$, więc z a nie wynika supraklasycznie $x \wedge y$.

Własność koniunkcji w konkluzji może być postrzegana jako nieintuicyjna, gdy w założeniu ma być konkluzją dużej liczby formuł, a nie tylko dwu. Mówi o tym tzw. paradoks loterii. Jeśli loteria ma dużą liczbę losów n, wtedy dla każdego losu jest wysoce prawdopodobne, że nie będzie losem wygrywającym i tym samym racjonalnie jest mieć przekonanie, że tak jest. Jednak jest też pewne, że pewien los wśród tych n, wygra, a zatem racjonalnie jest też mieć takie przekonanie. Mamy więc sytuację, w której z jednej strony racjonalnie jest być przekonanym o każdym elemencie zbioru n + 1 sądów, lecz z drugiej strony nie jest racjonalne być przekonanym o koniunkcji ich wszystkich, ponieważ jest to sąd logicznie sprzeczny. Tym samym nie jest racjonalne w tym przypadku respektowanie reguły koniunkcji w konkluzji. Supraklasyczne relacje konsekwencji definiowane przez funkcję prawdopodobieństwa, w przeciwieństwie do innych relacji suraklasycznych, nie spełniają też reguły (prawego) osłabienia:

Jeśli x wynika supraklasycznie z A i $x \Rightarrow y$, to y wynika supraklasycznie z A.

I. 7. Supraklasyczne relacje konsekwencji definiowane w terminach funkcji możliwości

Funkcje możliwości definiowane są tak jak funkcje prawdopodobieństwa z tą różnicą, że aksjomat Kołmogorowa mówiący o dodawaniu prawdopodobieństw zastępujemy następującym warunkiem:

$$p(x \vee y) = \max(p(x), p(y)).$$

Ta zmiana pociąga to, że mamy teraz $p(x) = 1$ lub $p(\neg x) = 1$ dla dowolnej formuły x, podczas gdy dla prawdopodobieństwa zachodzi równość: $p(\neg x) = 1 - p(x)$. W terminach funkcji możliwości możemy zdefiniować supraklasyczną relację konsekwencji.

Definicja: Dla dowolnej funkcji możliwości p definiujemy relację $\rightarrow_p$ przy pomocy reguły: $a \rightarrow_p x$ wtw, gdy albo $p(a \wedge \neg x) < p(a \wedge x)$, albo $\neg a$ jest tautologią.

Można wykazać, że każda taka relacja $\rightarrow_p$ jest supraklasyczna, a ponadto dla języków skończonych, spełnia warunek koniunkcji w konkluzji, kumulacyjną przechodniość: jeśli $a \rightarrow x$ i $a \wedge x \rightarrow y$, to $a \rightarrow y$, ostrożną monotoniczność: jeśli $a \rightarrow x$ i $a \rightarrow y$, to $a \wedge x \rightarrow y$ i racjonalną monotoniczność. Racjonalna monotoniczność definiowana jest za pomocą warunku:

(RM) Jeśli $a \rightarrow x$ i nie zachodzi relacja $a \rightarrow \neg b$, to $a \wedge b \rightarrow x$.

I. 8. Supraklasyczne relacje konsekwencji w teoriach zmiany przekonań

Wnioskowania niemonotoniczne mają wiele wspólnego z logiką zmiany przekonań, która opisuje to, w jaki sposób istniejące przekonania mogą być modyfikowane. Koncentruje się ona na takich trzech operacjach jak: rozszerzenie, kontrakcja i rewizja.

Definicja: Niech K będzie zbiorem formuł Boolowskich reprezentujących zbiór przekonań, a niech symbolizuje formułę. Przez rozszerzenie K przez a , K + a, rozumiemy zbiór wszystkich konsekwencji K i a: $Cn(K \cup \{a\})$, gdzie Cn jest klasyczną konsekwencją. Gdy K jest sprzeczne z a, K + a = Fm (zbiór wszystkich formuł języka).

Kontrakcja jest usunięciem sądu a ze zbioru przekonań K w taki sposób, że zbiór K – a nie pociąga a, o ile a nie jest tautologią. Rewizja jest procesem dodania sądu do zbioru przekonań K w taki sposób, że otrzymany w ten sposób zbiór K * a jest niesprzeczny, o ile a jest niesprzeczne. W przeciwieństwie do rozszerzenia te dwa ostatnie pojęcia nie są jednoznacznie wyznaczone. Nie istnieje jedyna operacja kontrakcji lub rewizji, która byłaby „właściwa" dla wszystkich sytuacji. Z konsekwencją niemonotoniczną najbardziej związana jest operacja rewizji. Jest to operacja dwuargumentowa, podczas gdy operacja konsekwencji jest jednoargumentowa. Możemy ją łatwo przekształcić w operację jednoargumentową: *a(K) rewizji K modulo a. Operacje te nie spełniają inkluzji: $K \subseteq *a(K)$, ponieważ możemy nie mieć $K \subseteq K * a$. Taka inkluzja nie może zachodzić jeśli a jest niesprzeczne, a jednocześnie sprzeczne z K, ponieważ coś będzie musiało być usunięte z K dla odzyskania niesprzeczności. Trudno byłoby więc myśleć o tej operacji jako o rodzaju inferencji. Natomiast jednoargumentowa operacja $*K(a)$: operacja rewizji przy pomocy a modulo K może być odczytana jako operacja konsekwencji. Również odwrotnie, operacje konsekwencji spełniające odpowiednie warunki mogą być postrzegane jako funkcje rewizji.

Proces kontrakcji i rewizji nie są od siebie niezależne. Rewizja może być rozumiana jako złożony proces kontrakcji a następnie rozszerzenia. Najpierw usuwamy z K to, co jest sprzeczne z a, a potem włączamy a. Stąd definicja: $K * a = Cn((K - \neg a) \cup \{a\})$.

Definicja: Rozważmy zbiór przekonań K jako zbiór formuł Boolowskich (domkniętych lub nie na konsekwencję klasyczną). Dla każdej formuły a, niech K_a będzie rodziną wszystkich podzbiorów $K' \subseteq K$, które są maksymalnie niesprzeczne z a. Niech δ będzie funkcją selekcji podrodziny $\delta(K_a) \subseteq K_a$. Wtedy operację rewizji zdefiniujemy (Makinson, 146):

$$K * a = Cn(\cap\{(K': K' \in \delta(K_a)\} \cup \{a\}) .$$

Jest zwyczaj odróżniania rewizji przekonań od ich *aktualizacji*. Rewizja jest procesem zmiany przekonań na podstawie zmiany w naszym sposobie myślenia lub na podstawie dodatkowej informacji. Aktualizacja przekonań jest procesem ich zmiany na podstawie zmian, jakie odbywają się w świecie. Niech S będzie zbiorem, którego elementami są światy (stany). Dla każdego takiego elementu s rozważamy rodzinę relacji $<_s$. Każda taka relacja rozumiana jest jako relacja odległości między światami od świata s. Relacje takie muszą być przeciwzwrotne, przechodnie i zastopowane oraz spełniające warunek, który mówi, że żaden świat nie jest tak blisko s jak s.

Definicja: Niech K będzie zbiorem formuł Boolowskich rozumianym jako zbiór przekonań, który ma być uaktualniony, niech a będzie formułą, która uaktualnia. Rezultat aktualizacji K przez a, K#a, jest definiowany jako zbiór wszystkich formuł x, które są prawdziwe w każdym stanie s', który spełnia a i jest tak blisko jak to tylko możliwe pewnego stanu s spełniającego K (s' jest minimalnym stanem ze względu na relację $<_s$ w zbiorze wszystkich a-stanów), czyli:

$$K\# a = \{x: \min_{<s} \mathbf{a} \subseteq \mathbf{x} \text{ dla każdego } s \in \mathbf{K}\}.$$

Jeśli $K \subseteq K'$, to $K\#a \subseteq K'\#a$. Warunek ten wyraża monotoniczność argumentu K. Jednak warunek monotoniczności nie musi zachodzić dla uaktualniającej formuły a, a zatem nie zawsze jest tak, że: $K\#a \subseteq K\#(a \wedge b)$. W teorii rewizji przekonań zachodzi warunek: $K * a =$

$Cn(K \cup \{a\})$, jeśli K jest niesprzeczne z a, natomiast w teorii aktualizacji nie mamy podobnego warunku. Tak jak w przypadku rewizji, aktualizacja może być zdefiniowana jako relacja konsekwencji modulo K. Każdy wybór K określa osobną relację $\Uparrow_K$ zdefiniowaną jako:

$$A \Uparrow_K x \text{ wtw, gdy } x \in K\#A$$

Czyli dla każdego K-stanu s, x jest prawdziwe w każdym stanie t który jest minimalny ze względu na relację $<_s$ w zbiorze wszystkich A-stanów.

II. INFORMACJA

Informacja, w jednym ze swych ujęć, bywa utożsamiana z naturalnym, w przeciwieństwie do językowego, znaczeniem, w jakim na przykład dym znaczy informacyjnie ogień. Przy takim ujęciu to, jaką informację dostarczają rzeczy jest sprawą obiektywną, gdyż informacja nie wymaga do swego istnienia, tak jak tego wymaga wiedza, istnienia świadomych istot. Informacja może być też definiowana bardzo ogólnie jako *struktura realizowana w fizycznym świecie poddająca się interpretowaniu lub wykorzystywaniu w rozsądny sposób przez odbiorcę* (Rott, 2008, 458). Matematyczna teoria komunikacji stawia pytanie, czy zdarzenia, jakie występują u odbiorcy zmieniają prawdopodobieństwo tego, co wystąpiło w źródle informacji. Taka teoria zakłada istnienie kanału komunikacyjnego między źródłem a odbiorcą, jak ilustruje to następujący przykład (Dretske, 2008, 33). Jest ośmiu pracowników a jeden z nich musi wykonać jakieś nieprzyjemne zadanie. Ich kierownik wydaje zlecenie, aby wybrali spośród siebie tego, kto wykona to zadanie i poinformowali go o tym. Posyłają wiadomość „H. został wybrany”. Teoria komunikacji identyfikuje wielkość informacji związaną z wystąpieniem danego zdarzenia z redukcją niepewności, z eliminacją możliwości reprezentowaną przez to zdarzenie. Na początku było 8 uprawnionych osób do wykonania zadania, następnie te 8 możliwości zostały zredukowane do jednej przez wybór H. Nie ma więc dalej niepewności, kto wykona zadanie. Gdy liczba możliwości zostaje zredukowana w ten sposób, wielkość informacji jest funkcją tego, jak wiele możliwości było na początku (w tym przypadku 8) i odpowiednich prawdopodobieństw (0, 125 dla każdej). Tym samym wielkość informacji, mierzona w bitach, generowanej przez wystąpienie jednej z n możliwości jest logarytmem o podstawie 2 z n (potęga, do której 2 należy podnieść, aby otrzymać n):

$I = - \log_2 p(n) = -\log_2 1/n = \log_2 n$, czyli

$$I = \log_2 n$$

W naszym przykładzie $I = \log_2 8 = 3$ bity. Gdybyśmy mieli 16 pracowników, to selekcja H generowałaby 4 bity informacji. Większa liczba informacji związana jest z większą redukcją niepewności. Powstaje pytanie, jaka wielkość tej informacji dociera do odbiorcy. Czy wiadomość, jaka dociera na kartce papieru do kierownika, niesie informację, że H. został wybrany? Podobne pytanie zadajemy wtedy, gdy pytamy o to, jak dużo informacji, o tym kto ukradł i zjadł ciasto, niosą okruchy ciasta na ustach Jasia. Teoria komunikacji pozwala na obliczenie przekazywanej informacji, na podanie miary wiarygodności kanału informacyjnego łączącego źródło i odbiorcę. Istnieje 8 możliwości, jakie mogą wystąpić w źródle informacji i 8 wyników, jakie może dotrzeć do odbiorcy. Istnieje więc 64 prawdopodobieństw warunkowych między tymi zdarzeniami tego rodzaju jak:

p(H. został wybrany/ Imię „H" pojawia się na kartce).

Informacja, jaka dociera do odbiorcy – informacja przeniesiona – identyfikowana jest z funkcją tych 64 prawdopodobieństw i obliczana według wzoru:

$$I_t = I – E,$$

gdzie E jest miarą statystycznej niezależności między zdarzeniami występującymi w źródle i u odbiorcy. E nazywane bywa entropią warunkową lub *ekwiwokacją*. Załóżmy, że na kartce zawsze napisane jest imię tego, kto został wybrany. Wtedy prawdopodobieństwa warunkowe mają wartość 1 lub 0, na przykład:

p(H. został wybrany/Imię „H" pojawia się na kartce) = 1

p(B. została wybrana/Imię „H" pojawia się na kartce) = 0

p(B. została wybrana/Imię „B" pojawia się na kartce) = 1.

Mamy w takiej sytuacji wiarygodny kanał informacyjny, E = 0, a cała informacja wygenerowana przez wybór dociera do swego celu.

Prawdopodobieństwa warunkowe mogą ulec zmianie, jeśli ten, kto przenosił kartkę papieru z imieniem zgubił ją i napisał imię przypadkowo wybrane, mimo tego, że jest to imię „H". Wtedy bowiem dla każdego z 64 prawdopodobieństw warunkowych dostajemy tę samą wartość, co w następującym przypadku:

p(H. został wybrany/ Imię „H" pojawia się na kartce) = 1/8.

Statystyczna funkcja definiująca E przybiera teraz maksymalną wartość 3 bity, a wielkość przeniesionej informacji równa jest 0. W przeciwieństwie do poprzedniego przypadku, gdzie mieliśmy maksymalną komunikację, mamy tutaj zerową komunikację.

W teorii Dretske'go ekwiwokację obliczamy według wzoru:

$$E = \Sigma p(r_i)\, E(r),$$

gdzie E(r) jest entropią warunkową, która mierzy naszą niepewność dotyczącą s jeśli wiemy, że r wystąpiło.

$$E(r) = - \Sigma p(s_i/r_i)\log_2 p(s_i/r_i).$$

W naszym przykładzie E(r) = -8 x 1/8 x log(1/8) = 3 bity. Tak więc ekwiwokacja E, która mierzy naszą niepewność dotyczącą s w sytuacji, gdy nie wiemy, jaką wartość przyjęło r, otrzyma wartość 1/8 x 3 x 8 = 3 bity. Krytykę tego ujęcia ekwiwokacji znajdujemy w artykule Lombardi (2004), które podaje zarazem wzór na ekwiwokację w sensie Shannona.

Możemy też podać przykład pośredni, gdzie informacja przeniesiona równa jest 2,75 bita, gdyż E = 0,25. Teoria komunikacji daje odpowiedź kiedy ktoś wie, że H został wybrany, a mianowicie tylko wtedy, gdy wielkość informacji przeniesionej jest taka jak wielkość informacji wygenerowanej, lecz nie wtedy, gdy wielkość informacji przeniesionej jest mniejsza niż 3 bity.

Teoria komunikacji korzysta z wartości średnich i w niektórych przypadkach nie jest w stanie uchwycić potocznego sensu informacji. Można też argumentować, że kanał komunikacyjny, który czasami jest niewiarygodny nie jest wystarczająco dobry, aby wiedzieć w tych przypadkach, kiedy informacja przesłana jest zgodna z wygenerowaną, na przykład taki jest kanał, w którym E = 0,25, gdyż przenosi czasami mylące informacje, a w konsekwencji nie może być kanałem, do którego można mieć zaufanie. Przekonania, jakie formułujemy na jego podstawie nie mogą mieć pewności charakterystycznej dla wiedzy.

Jeśli to, co generuje informację posiada inne możliwości, to możemy mówić o wygenerowanej informacji przez wybór, lecz jeśli mamy do czynienia z koniecznym stanem rzeczy, to żadna informacja nie jest generowana. Konieczny stan rzeczy generuje zerową informację. Fakty, które generują informację są faktami przygodnymi. Dretske ogranicza w związku z tym teorię wiedzy do wiedzy empirycznej.

Jeśli kanał komunikacyjny nie przenosi całej wygenerowanej informacji, to wystarcza on do wytworzenia prawdziwego przekonania, lecz nie do wytworzenia wiedzy. Aby wiedzieć, co wydarzyło się w źródle informacji musimy otrzymać całą wygenerowaną informację.

II. 1. Informacja jako miara niepewności

Czasami mówi się, że „informacja” ma w teorii komunikacji inne znaczenie niż znaczenie w języku potocznym, gdzie uważamy stwierdzenie za posiadające większą informację, jeśli jego treść dotyczy jakiejś rzeczy mało prawdopodobnej. Stwierdzenie o zdarzeniu, które ma wysokie prawdopodobieństwo dostarcza mało informacji, natomiast stwierdzenie o zdarzeniu, które ma małe prawdopodobieństwo dostarcza dużo informacji. Słowo „informacja” używane w życiu codziennym ma dwa aspekty: aspekt niespodzianki i znaczenia. Matematyczna teoria informacji jest wyłącznie związana z aspektem niespodzianki. Teoria ta ma swoje początki w inżynierii komunikacji, gdzie ważny był tylko czynnik niespodzianki. Poza tym znaczenie okazało się zbyt trudnym pojęciem do badania matematycznego. Stąd informacja w matematycznej teorii informacji ma ograniczone znaczenie jako miara niespodzianki, a stwierdzenia potocznie uważane za bezsensowne mogą mieć wysoką zawartość informacyjną.

Treść informacyjna zdarzenia, I(z), powinna być funkcją malejącą jego prawdopodobieństwa, p(z), a więc jeśli $p(z) \leq p(z')$, to $I(z) \geq I(z')$.

Przykład (Applebaum, 1996, 106): Załóżmy, że wybieramy losowo kartę z talii 52 kart. Rozważmy następujące zdarzenia:

E’ = karta jest jedną z kart karo;

E’’ = karta jest siódemką;

$E' \cap E''$ = karta jest siódemką karo. Zdarzenia E’ i E’’ są niezależne.

Zdarzenia te mają następujące prawdopodobieństwa:

$p(E') = 13/52 = 1/4$

$p(E'') = 4/52 = 1/13$,

$p(E' \cap E'') = 1/52$.

Z naszej dyskusji odniesionej do tego przykładu otrzymujemy:

$I(E' \cap E'') \geq I(E'') \geq I(E')$;

$I(E' \cap E'') = I(E') + I(E'')$;

$I(E) \geq 0$ dla każdego zdarzenia E.

Funkcja, która spełnia 3 powyższe warunki musi mieć postać:

$$I(E) = -K \log_a(p(E)),$$

gdzie a i K są stałymi. Jeśli zdarzenie E jest pewne, to nie niesie ono informacji, gdyż log(1) = 0. Przyjmuje się standardowo, że K = 1, a = 2, a zatem definicja informacji zdarzenia E ma następującą postać:

$$I(E) = - \log_2(p(E)).$$

Jednostkami informacji są bity. Dostaniemy 1 bit informacji gdy wybierzemy jedną z dwu równie prawdopodobnych możliwości. Jaką informację niosą zdarzenia z poprzedniego przykładu? Jak wygenerowana informacja zależy od prawdopodobieństwa naszego wyboru?

$I(E') = - \log(1/4) = 2$ bity, gdyż $2^2 = 4$; ponieważ zachodzi równość: $-\log 1/x = \log x$.

$I(E'') = -\log(1/13) = 3,70$ bity

$I(E' \cap E'') = I(E') + I(E'') = 5, 70$ bitów.

Im większe jest prawdopodobieństwo naszego wyboru, tym mniejsza jest informacja, jaką niesie nasz wybór.

Definicja: Średnia informacja, miara niepewności, entropia używane są na oznaczenie tego samego i obliczane według wzoru:

$$H = -\Sigma p_i \log(p_i)$$

gdzie suma uogólniona obliczana jest po wszystkich wartościach od 1 do n.

Przykład (Applebaum, 109): Załóżmy, że mamy monetę, która nie jest rzetelna tak, że prawdopodobieństwo wyrzucenia reszki jest odpowiednio: 0,95, 0,60 i 0,5 w przypadku monety rzetelnej. Jaka jest entropia w każdym przypadku? Obliczymy entropię w tym przypadku według wzoru:

$$H(p) = -p \log(p) – (1 – p)\log(1 – p).$$

Otrzymamy odpowiednio następujące wyniki:

0, 286 bitów

0,971 bitów

1 bit.

Tam gdzie entropia jest niska, wynik ten możemy interpretować tak, że osoba, która używa tej monety jest prawie pewna wygranej. Jeśli entropia jest wysoka, jak w drugim przypadku, to osoba jest znacznie mniej pewna wygranej. Natomiast jest w stanie maksymalnej niepewności, gdy entropia przyjmuje maksymalną wartość, jak w trzecim przypadku.

II. 2. Informacja jako miara treści

W przeciwieństwie do statystycznej teorii komunikacji, teoria informacji rozwijana przez Carnapa i Bar-Hillela nazywana jest teorią semantycznej informacji. Informację pojmują oni jako eliminację niepewności. Im więcej logicznych możliwości stwierdzenie s wyklucza, tym mniejszą pozostawia niepewność, tym bardziej jest informujące. Definiują informację jako *miarę treści* stwierdzenia s w terminach prawdopodobieństwa s :

$$\mathrm{cont}(s) = 1 – p(s)$$

Jest to miara absolutnej liczby możliwości, jaką informacja umożliwia nam pominąć. Inaczej mówiąc, informacja semantyczna lub *treść* zdania deklaratywnego jest zbiorem światów możliwych, w których to zdanie jest fałszywe, czyli zbiorem światów, które to zdanie wyklucza:

$$\mathbf{cont}(s) = \{x \in W: \neg s \text{ jest prawdziwa w } x\}$$

Dla dowolnego zdania prawdziwego logicznie t mamy **cont**(t) = ∅. Teoria semantycznej informacji nie tylko zainteresowana jest samym pojęciem semantycznej informacji, lecz również jej miarą. Podstawowa intuicja jest taka, że informacja zdania s jest odwrotnie proporcjonalna do prawdopodobieństwa stanu rzeczy, które ono opisuje, tak jak to wyraża wcześniej podana definicja zaczerpnięta z pracy Shannona. Bar-Hillel i Carnap wprowadzają również pojęcie *miary informacji* inf dla zapewnienia addytywności w tym sensie, że

$$\inf(s \wedge s') = \inf(s) + \inf(s'),$$

co nie zachodzi dla cont, gdyż s i s' mogą mieć wspólną treść. Stąd też definicja inf postaci:

$$\inf(s) = -\log_2 p(s)$$

lub równoważna z definicją powyższą definicja następująca:

$$\inf(s) = \log_2 1/1 - \mathrm{cont}(s).$$

Zdanie logicznie prawdziwe ma minimalną miarę informacji, ponieważ

$$\inf(t) = \log 1/1 - 0 = 0.$$

Zdania logicznie równoważne będą miały tę samą miarę informacji, inf(s) = inf(s') a również ich semantyczna informacja jest identyczna: **cont**(s) = **cont**(s'). Bar-Hillel i Carnap zwracają uwagę, że może istnieć sens, w jakim prawdy logiczne nie posiadają zerowej informacji, jak na przykład w sensie *psychologicznej* informacji dla danej osoby. Jednak w klasycznej semantyce światów możliwych zdania logicznie prawdziwe są we wszystkich światach możliwych prawdziwe, a zdania logicznie równoważne prawdziwe są w identycznych zbiorach światów możliwych.

Pojęcie informacji odwołujące się do niespodzianki lub nieprzewidywalności definiowane jest jako:

$$\inf(s) = -\log p(s) \quad \text{lub} \quad \inf(s) = \log[1/p(s)]$$

Relatywne pojęcia odpowiednio nieprzewidywalności i treści:

$$\inf(s/t) = \inf(s \wedge t) - \inf(t)$$

$$\mathrm{cont}(s/t) = \mathrm{cont}(s \wedge t) - \mathrm{cont}(t).$$

Definicje te mówią jak dużo informacji dodaje s do informacji danej nam przez t. Hintikka (1967) definiuje odpowiednio te pojęcia w następujący sposób:

$$\inf(s/t) = -\log p(s/t)$$

$$\mathrm{cont}(s/t) = \mathrm{cont}(t \to s).$$

Relatywna nieprzewidywalność jest związana z relatywnym (względnym) prawdopodobieństwem. Natomiast miara treści s relatywnie do t jest miarą treści stwierdzenia, że jeśli t jest prawdziwe, to prawdziwe jest s. Informacja, jaką s dodaje do informacji t musi być informacją najsłabszego stwierdzenia, które gdy dodane do t będzie implikowało s, a stwierdzeniem tym jest $t \rightarrow s$.

Możemy również mówić o relatywnej informacji i relatywnej treści w innym sensie. Zamiast pytać o to, jaką informację lub treść s dodaje do t, zakładamy, że t jest prawdziwe, i pytamy o to, jak informacja lub treść mogłyby być opisywane pod takim warunkiem. W rezultacie otrzymalibyśmy *warunkową* informację i *warunkową* treść:

$$\mathrm{inf}(s/t) = -\log p(s/t)$$

$$\mathrm{cont}(s/t) = 1 - p(s/t).$$

Relatywna i warunkowa informacja definiowane są w ten sam sposób, różnica dotyczy definiowania treści.

Możemy w związku z pojęciem informacji mówić o redukcji niepewności dotyczącej tego, co mówi dane stwierdzenie lub o redukcji niepewności dotyczącej czegoś innego, co opisywane jest przez inne stwierdzenie. Tym samym możemy mówić o informacji, jaką h dostarcza o g w następujący sposób:

$\mathrm{Inf}(g) - \mathrm{inf}(g/h) = \log[p(g/h)/p(g)] = \log[p(g \wedge h)/p(g)p(h)]$ lub

$\mathrm{Cont}(g) - \mathrm{cont}(g/h) = 1 - p(g \vee h)$ (informacja relatywna)

$\mathrm{Cont}(g) - \mathrm{cont}(g/h) = p(g/h) - p(g)$ (informacja warunkowa).

Ten pierwszy warunek nazywa Hintikka (1967, 317): transinf(h/g). Inf(g) reprezentuje tutaj niepewność, której pozbywamy się, gdy dowiadujemy się, że g, natomiast inf(g/h) jest informacją, jaką g dodaje do informacji h, czyli niepewnością dotyczącą g, która pozostaje po tym, kiedy dowiemy się, że h jest prawdziwe. Zatem różnica mierzy redukcję naszej niepewności dotyczącą g, która ma miejsce, gdy wiemy, że h. Nasze zaskoczenie, że zachodzi g może być większe, gdy wcześniej dowiedzieliśmy się, że h jest prawdziwe. Jeśli g i h są niezależnymi zdarzeniami, to wartość tej różnicy jest zerowa.

Przykład: Niech s będzie stwierdzeniem „Dzisiaj w Warszawie jest zimno”. Mogę nie być zainteresowana inf(s) lub cont(s), lecz mogą mnie interesować warunki pogodowe na jutro (stwierdzenie t), a zatem mogę być zainteresowana informacją, jaką s dostarcza o t . W statystycznej teorii informacji Shannona mamy do czynienia z pojęciem informacji przeniesionej, wtedy h stwierdza otrzymanie wiadomości, a g stwierdza jej wysłanie. Różnica między inf(g) a informacją przeniesioną transinf (h/g), która jest równa inf(g/h), nazywana jest w statystycznych teoriach informacji ekwiwokacją .

Mogę też być zainteresowana redukcją niepewności co do tego, czy będzie zimno czy nie. Niech e = dzisiaj jest zimno, h = będzie zimno jutro, to informacja, jakiej dostarcza e o stanie rzeczy h lub ¬h jest następująca (Hintikka 1967, 318):

$$p(h/e)transinf(e/h) + p(\neg h/e)transinf(e/\neg h).$$

Z teoriodowodowym i semantycznym ujęciem informacji mamy też do czynienia w logice intuicjonistycznej. W przeciwieństwie do logiki epistemicznej jest to rachunek wiedzy implicite, gdzie znaczenia standardowych stałych logicznych powiązane są z byciem znanym lub dowodliwym. Logika intuicjonistyczna posiada swój system dedukcji naturalnej dzięki swej Brouwerowsko-Heytingowsko- Kołmogorowskiej interpretacji, lecz posiada też semantykę w stylu modalnym, gdzie światy są rozumiane jako poziomy informacji w procesie badawczym. Inferencyjna i semantyczna informacja pozostają więc bardzo blisko. Lecz logika intuicjonistyczna ma specyficzne ujęcie konsekwencji, wykraczające poza to, co chcielibyśmy mieć w teorii informacji. Stąd też podjęto usiłowania wyabstrahowania ogólniejszego ujęcia informacji. Tak na przykład van Benthem proponuje modalną teorię informacji nad modelami rozumianymi jako poziomy informacji uporządkowane przez inkluzję. Istnieją próby łączenia logiki epistemicznej z logikami substrukturalnymi i logikami adaptywnymi. Nie ma jednak zgody co do integracji inferencyjnego i semantycznego ujęcia informacji. Najważniejszym problemem jest ciągle formalna teoria informacji. Tacy filozofowie jak Dretske badają pojęciową naturę informacji, natomiast tzw. dynamiczny zwrot w logice ogranicza informację do jej komunikacyjnych i epistemicznych własności.

II. 3. Informacja jako zbiór epistemicznych możliwości

Mamy dwa pojęcia informacji: deklaratywna informacja (statyczna) i dynamiczna. Bycie poinformowanym, że p, w tym sensie, że podmiot posiada informację, że p odróżniamy od stawania się poinformowanym, że p, co jest dynamicznym procesem. W tym pierwszym przypadku mamy stan informacyjny podmiotu w określonym czasie. Jedno z kluczowych ujęć informacji jest takie, gdzie informację podmiotu traktuje się jako zbiór wszystkich relewantnych możliwości, do jakich podmiot ma kognitywny dostęp. Jeśli podmiot posiada informację, że p lub q, to taka informacja odsłania 3 możliwości: światy, w których p jest prawdziwe, lecz q fałszywe; światy, w których q jest prawdziwe, lecz q fałszywe; światy w których p i q jest prawdziwe. Jeśli teraz podmiot otrzyma informację, że p jest fałszywe, spowoduje to zmniejszenie liczby możliwości, którymi kognitywnie rozporządza podmiot, do jednej, światów w których zachodzi negacja p i q. Możemy powiedzieć, że informacja, jaką posiada podmiot nie odróżnia pewnych światów, tak jak to ma miejsce w przypadku informacji p lub q. Powiemy, że światy pozostają do siebie w relacji epistemicznej nieodróżnialności dla podmiotu a:

$$w \sim_a w',$$

gdy informacja, jaką rozporządza a nie odróżnia w od w' (Jago, 2006). Jeśli w różni się od w' tym, że r jest prawdziwe w pierwszym, lecz nie jest w drugim, to stan informacyjny podmiotu nie zawiera informacji, że r lub, że $\neg$r. Stan informacyjny jest modelowany przez klasę światów, których podmiot nie może odróżnić. Intuicyjnie, im większa klasa nieodróżnialnych światów, tym mniej informacji. Informacja służy do odróżnienia świata aktualnego od światów możliwych. Informacyjna treść stwierdzenia, że p rozumiana jest jako zmiana w relacji nieodróżnialności podmiotu polegająca na tym, że podmiot po takiej zmianie może odróżnić te światy, w których p jest prawdziwe, od tych w których nie jest. Światy rozumiane są tutaj jako epistemicznie możliwe. Epistemiczne możliwości nie są metafizycznymi światami możliwymi. Mówimy, że sąd S posiada treść informacyjną dla podmiotu a, gdy S mógłby spowodować zmiany w stanie informacyjnym a. Załóżmy, że a jest poinformowany, że $p \rightarrow q$ i że p. Czy zatem q jest informujące? Podmiot a po takiej informacji wyklucza światy, w których $p \wedge \neg q$ jest prawdziwe, a następnie wyklucza światy, w których $\neg p$ jest prawdziwe. Pozostają więc światy, w których q jest prawdziwe, a tym samym stawanie się poinformowanym, że q nie spowoduje zmiany. Załóżmy teraz, że $A \rightarrow B$ jest logicznie prawdziwe oraz że a jest poinformowany, że A. Czy B może posiadać treść informacyjną? Ponieważ mamy do czynienia z prawdą logiczną, więc jest to sąd prawdziwy we wszystkich światach. Dodanie B nie spowoduje zmiany w światach, które a uważa za możliwe, a B nie zawiera treści informacyjnej ponad tę, jaka jest w A. Jest to problem, który nosi nazwę informacyjnego przeładowania. Jeśli podmiot jest poinformowany, że p, jest on również poinformowany o nieskończonej liczbie zdań, które wynikają logicznie z p. Tak więc konsekwencje zbioru sądów S zawierają treść informacyjną, którą zawierają sądy w S. Z podobnym problemem mamy do czynienia w przypadku wiedzy, gdzie nosi on nazwę logicznej wszechwiedzy. Zgodnie z tym poglądem tautologie nie są informujące, gdyż tautologie mają prawdopodobieństwo 1, a log(1) =0. Nie potrafimy go uniknąć, jeśli odwołujemy się do światów możliwych. Pojawia się on już w najsłabszej logice informacji K.

Wyróżnia się 3 główne logiczne ujęcia informacji (van Benthem, Martinez, 2008):

1. logika epistemiczna i doksastyczna jako logika informacji: informacja jest zawsze o czymś, dla kogoś i związana jest z dynamicznym procesem jej aktualizacji; informacja jest tu postrzegana jako zasięg epistemicznych możliwości (opcji).
2. teoria sytuacji jako logika informacji eksplikuje ten aspekt informacji, że jest ona o czymś, jej intencjonalność. Aspekt ten nie znajduje swego wyrazu w algorytmicznej teorii informacji.
3. teoria dedukcji jako logika informacji ujmuje ten aspekt informacji, który polega na jej byciu zawartej implicite (zakodowanej) w czymś, na przykład, w danych, z których dedukcja wyłania informację. Do tego nurtu możemy włączyć teorię Kołmogorowa, gdzie wielkość informacji w bitach dla danego ciągu definiowana jest jako długość (mierzona w bitach) najkrótszego programu, który produkuje ten ciąg i kończy pracę.

Jak powinniśmy mierzyć wielkość informacji o danym zjawisku, które dane jest nam za pomocą obserwacji? Teoria informacji Shannona i algorytmiczna teoria informacji opierają się na idei, że wielkość ta może być mierzona przez minimalną liczbę bitów potrzebnych do opisania danej sytuacji. Różnica między tymi teoriami polega na tym, że teoria Shannona opis ten pojmuje w terminach rozkładu prawdopodobieństwa, podczas gdy algorytmiczna teoria Kołmogorowa zajmuje stanowisko nieprobabilistyczne i odwołuje się do wielkości najkrótszego programu, który produkuje (drukuje) ciąg reprezentujący tę obserwację. Na przykład 10.000 wydrukowanych jedynek zawiera informację log10.000 bitów, ponieważ tej wielkości program produkuje taki ciąg:

dla i: = 1 do 10.000; drukuj 1.

Podobna definicja może być podana dla ciągów nieskończonych, lecz w takim przypadku program produkuje element po elemencie bez zatrzymania się.

II. 4. Przykład wnioskowania supraklasycznego

Rozważmy przykład pokazujący związek, w jakim pozostają trzy rozważane w tym artykule pojęcia: pojęcie supraklasycznej relacji konsekwencji, prawdopodobieństwa i informacji.
Załóżmy, że wnioskujemy supraklasycznie w następujący sposób:
Przesłanka a: *Jan miał bezpośredni kontakt z osobą chorą na grypę przed dwoma dniami*.
Wniosek x: *Jan dzisiaj zachorował na grypę*.
Załóżmy, że prawdopodobieństwo warunkowe p(x/a):

p(Jan dzisiaj zachorował na grypę/Jan miał bezpośredni kontakt z osobą chorą na grypę przed dwoma dniami) $\geq$ t,

gdzie t jest wielkością progową z przedziału [0, 1] i $t \neq 0$.

Powstaje pytanie, jak zmieni się powyższe wnioskowanie, gdy dostaniemy informację **b**: *Jan zaszczepił się przeciwko grypie przed dwoma tygodniami*. Należy wtedy oszacować prawdopodobieństwo warunkowe uwzględniające tę nową informację, czyli: $p(x/\, a \wedge \mathbf{b})$:

p(Jan dzisiaj zachorował na grypę/ Jan miał bezpośredni kontakt z osobą chorą na grypę przed dwoma dniami i Jan zaszczepił się przeciwko grypie przed dwoma tygodniami):

$$p(x/a \wedge \mathbf{b}) = p(x \wedge a \wedge \mathbf{b})/p(a \wedge \mathbf{b}).$$

Jeśli przyjęlibyśmy że $p(\mathbf{b}) = 1$, to tym samym informacja, jakiej dostarcza **b** byłaby inf $\mathbf{b} = -\log p(\mathbf{b}) = \log 1 = 0$, natomiast $p(x/a \wedge b) = p(x/a)$.

Jeśli $p(\mathbf{b}) < 1$ i ponadto $p(x/a \wedge \mathbf{b}) < t$, to będziemy musieli wycofać się z uznanego wcześniej wniosku x: *Jan dzisiaj zachorował na grypę*, ponieważ nie jest wtedy spełniony następujący warunek: $a \wedge \mathbf{b} \rightarrow_{tp} x$ wtedy i tylko wtedy, gdy $p(x/\, a \wedge \mathbf{b}) \geq t$. Prawdopodobieństwo $p(x \wedge a \wedge \mathbf{b}) < p(x/a \wedge \mathbf{b}) < p(x/a)$. Można łatwo podać taką wielkość $p(x \wedge a \wedge \mathbf{b})$, że dostaniemy $p(x \wedge a \wedge \mathbf{b}) < p(x/a \wedge \mathbf{b}) < t$. Na przykład: 0,05 < 0,45 < 0,5. Załóżmy, że zdanie **b** dostarcza 1 bit informacji. Zatem 1 bit informacji wystarcza nam do tego, aby nie uznać wniosku x. Nawet jeśli informacja dostarczana przez to zdanie jest nieco mniejsza niż 1 bit, to również musimy wycofać się z naszego wcześniejszego wniosku. Jeśli natomiast **b** jest zdaniem fałszywym, to $p(x/a \wedge \neg \mathbf{b}) = p(x \wedge a \wedge \neg \mathbf{b})/p(a \wedge \neg \mathbf{b})$. Prawdopodobieństwo negacji b obliczamy jako $1 - p(\mathbf{b})$. Wtedy również możliwe jest wskazanie takiej wartości $p(\mathbf{b})$, przy której prawdopodobieństwo warunkowe

$$p(x/a \wedge \neg \mathbf{b}) < 0{,}5.$$

Mogło tak się zdarzyć, że pomimo tego, że Jan miał kontakt z osobą chorą na grypę przed dwoma dniami i pomimo tego, że Jan nie zaszczepił się przeciwko grypie przed dwoma tygodniami, nie zachorował dzisiaj na grypę.

Przykład ten pozostaje w zgodzie z naszymi intuicjami i jeszcze raz pokazuje, że we wnioskowaniach supraklasycznych o zachodzeniu relacji konsekwencji nie decyduje forma logiczna przesłanek i wniosku, jak ma to miejsce we wnioskowaniach odwołujących się do klasycznej relacji konsekwencji, a racjonalność wnioskowania może być utrzymana pomimo tego, że wnioskowanie nie spełnia wymogu formalnej poprawności, jakiego spełnienia oczekujemy od wnioskowań czysto dedukcyjnych.

III. WNIOSKOWANIA EMPIRYCZNE UJMOWANE W TERMINACH PRAWDOPODOBIEŃSTWA ALGORYTMICZNEGO

Gdy wnioskujemy o empirycznym świecie, oczekujemy, że nasze wnioskowania będą podstawą przewidywań. Wnioskowania, które polegają na formułowaniu najbardziej trafnej konkluzji na podstawie obserwacji nazywamy wnioskowaniami indukcyjnymi. Jak już o tym wspominaliśmy na początku tego artykułu, obserwacje te zawsze są niezupełne (niekompletne), a tym samym, wyprowadzony na ich podstawie wniosek nigdy nie jest absolutnie prawdziwy. Podobnie jak hipotezy naukowe nigdy nie są absolutnie pewne, tak też nie możemy być nigdy pewni wniosków we wnioskowaniach indukcyjnych, ponieważ wnioski te mówią o czymś czego dotychczas nie poznaliśmy, dotyczą jakiegoś nieznanego dotychczas obszaru. Problem, jaki zawsze pojawia się w przypadku wnioskowań empirycznych dotyczy tego, jak zagwarantować racjonalność naszych przekonań w sytuacji niepewności. Jest to problem, który od dawna towarzyszy myśli filozoficznej: od Epikura do Okhama od Okhama do Hume'a. Najkrótsze sformułowanie nadał mu Hume w pytaniu: „Co jest podstawą wszystkich konkluzji wyprowadzonych z doświadczenia?" Laplace i Bayes byli zainspirowani pracami Hume'a i im zawdzięczamy pierwsze próby formalizacji wnioskowań indukcyjnych. Jednak dopiero od 1964 roku, w którym ukazała się praca „A Formal Theory of Inductive Inference" R. Solomonoffa, możemy mówić o rozwiązaniu problemu indukcji.

III.1. Na czym polega rozwiązanie problemu indukcji?

Wiele wcześniejszych wyników prowadzi do rozwiązania zaproponowanego przez Solomonoffa. Z aksjomatyki prawdopodobieństwa podanej przez Kołmogorowa, którą omawialiśmy wyżej, wyprowadzalna jest tzw. reguła Bayesa. Istotny dla Bayesianizmu jest wynik Cox'a mówiący o tym, że racjonalny system przekonań musi pozostawać w zgodzie ze standardowymi aksjomatami prawdopodobieństwa podanymi przez Kołmogorowa. Jest to istotne dlatego, że interpretacja prawdopodobieństwa przyjmowana przez Bayesa jest interpretacją w duchu subiektywizmu, zgodnie z którą prawdopodobieństwa są rezultatem naszej osobistej przeszłości. To, w co wierzymy dziś zależy od tego, w co wierzyliśmy wczoraj oraz od tego, czego nauczyliśmy się od wczoraj. Z tego wynika więc, że podstawą Bayesianizmu jest racjonalny proces aktualizacji przekonań.

III. 1. 1. Reguła Bayesa

Uaktualnianie przekonań co do przyjętej hipotezy na podstawie danych obserwacji wyraża reguła Bayesa (B):

$$(B): \quad P(A/B) = P(B/A)\, P(A)/\, P(B).$$

Reguła ta interpretowana jest jako mierząca stopień przekonania co do hipotezy H ze względu na skończony ciąg obserwacji **x** równy ilorazowi stopnia przekonania co do ciągu obserwacji **x** w świetle hipotezy H wraz z prawdopodobieństwem (stopniem przekonania) a priori co do hipotezy H przez prawdopodobieństwo ciągu obserwacji **x**, co symbolicznie przyjmuje następującą postać:

$$P(H/\mathbf{x}) = P(\mathbf{x}/H)P(H)/\, P(\mathbf{x}).$$

Formuła ta wyraża przekonanie (prawdopodobieństwo) a posteriori co do hipotezy H ze względu na dane **x**. Nieco inną postać ma formuła, która mówi o tym, jak obliczyć prawdopodobieństwo a posteriori, że następna obserwacja będzie a, jeśli wcześniejsze obserwacje tworzą ciąg **x**. Gdybyśmy odwołali się wyłącznie do prawdopodobieństwa warunkowego, formułę tę zapisalibyśmy jako:

$$P(a/\mathbf{x}) = P(\mathbf{x}a)/P(\mathbf{x}).$$

W teorii Bayesa związek ten przyjmuje następującą postać:

$$\Sigma\, P(H)P(x/H),$$

gdzie P(H) jest prawdopodobieństwem a priori hipotezy H, P(**x**/H) jest rozkładem prawdopodobieństwa dla ciągu doświadczeń **x** wyznaczonym przez hipotezę H, a suma jest po wszystkich hipotezach. Zauważmy, że aby obie formuły Bayesa dawały stopień przekonania wyższy od zera, prawdopodobieństwo a priori danej hipotezy musi być niezerowe.

III. 1. 2. Złożoność Kołmogorowa

Bayesianizm jest tylko jednym z komponentów składających się na teorię Solomonoffa. Innym ważnym komponentem tej teorii jest pojęcie prostoty sformalizowane przez Kołmogorowa. Kołmogorow, opierając się na pojęciu uniwersalnej maszyny Turinga zdefiniował pojęcie złożoności obiektu, która może być utożsamiana z treścią informacji zawartej w dowolnym obiekcie. Istotne w podejściu Kołmogorowa jest to, że opis obiektu zostaje utożsamiony z programem, a zatem z procedurą jednoznacznego kodowania tego obiektu. Tak na przykład obiekt będący tysiącem 8-ek w rzędzie, może być w krótki sposób zakodowany jako program dla uniwersalnej maszyny Turinga:

Dla (i = 0; i < 1000; i++) drukuj („8”).

W przeciwieństwie do powyższego programu, program, który koduje 1000 losowo wybranych 0 i 1 nie będzie taki krótki, gdyż trzeba będzie wymienić kolejność tych 0 i 1. Podobna idea zawarta jest w zdefiniowanej przez Kołmogorowa złożoności serii (ciagu) K(**x**):

$$K(\mathbf{x}) = \min\{\text{długość}(p) : U(p) = \mathbf{x}\}.$$

Definicja ta mówi, że złożoność serii[1] (ciągu) jest minimalnym (najkrótszym) programem, który drukuje **x** na uniwersalnej maszynie Turinga. W prosty sposób można też zdefiniować warunkową złożoność za względu na dodatkowy ciąg (serię) **y**:

$$K(\mathbf{x}/\mathbf{y}) = \min\{\text{długość}(p) : U(\mathbf{y}, p) = \mathbf{x}\}.$$

Korzystając z powyższych pojęć, Solomonoff definiuje prawdopodobieństwo algorytmiczne a priori jako:

$$P(H) = 2^{-K(H)}$$

Ponieważ program p jest binarną serią, więc możemy zdefiniować uniwersalne prawdopodobieństwo algorytmiczne a priori M ciągu **x** według następującego wzoru, który mówi, że to, co drukuje uniwersalna maszyna Turinga jest ciągiem **x**, jeśli na jej wejściu jest wynik rzutu rzetelną monetą (czyli program p):

[1] W przypadku serii operacją złożenia jest konkatenacja. Ontologicznie seria i ciąg są różnymi obiektami.

$$M(\mathbf{x}) = \Sigma\, 2^{-\text{długość}(p)}$$

Suma jest w tym wzorze nad wszystkimi zatrzymującymi się programami p, dla których uniwersalna maszyna Turinga ma na wyjściu ciąg **x**. Jest to suma nieskończona, a więc nie jest skończenie obliczalna. Inaczej mówiąc, nie wiemy, czy dany program p zatrzyma się. Tym samym algorytmiczne prawdopodobieństwo a priori M(**x**) nie jest obliczalne, lecz może być podane z dolnym przybliżeniem (obliczone asymptotycznie). Stosuje się obecnie wiele różnych technik uzyskiwania tych przybliżeń, z których chyba najbardziej znaną jest technika nazywana „minimum długości opisu” (MDL). U jej podstaw leży zasada, że większe prawdopodobieństwo mają te dane, które minimalizują kompresję z danymi wcześniejszymi. Jeśli mierzymy złożoność **x** w terminach złożoności Kołmogorowa, to widzimy, że serie najprostsze mają największe prawdopodobieństwo M(**x**). Teoria Solomonoffa jest więc współczesną formalizacją zasady Okhama.

III. 1. 3. Przykład hipotezy

Pokażmy jak teoria Solomonoffa może być stosowana do konkretnego przykładu. Załóżmy, że naszą wyjściową hipotezą jest

H = Wszystkie kruki są czarne.

Zacznijmy od sformalizowania alfabetu naszych obserwacji. Będzie to zbiór, w którym znajdą się następujące predykaty:

1. czarny kruk: CK;
2. nie-czarny kruk: C’K;
3. czarny nie-kruk: CK’;
4. nie-czarny nie-kruk: C’K’.

Oznaczmy ten alfabet przez A = {CK, C’K, CK’, C’K’}. Każdemu predykatowi przypiszemy odpowiednią proporcję, jaką jego zakres zajmuje w całej populacji, co oznaczymy przez **P** = {p(CK), p(C’K), p(CK’), p(C’K’)}. Klasę modeli będziemy tym samym oznaczali jako klasę wektorów prawdopodobieństwa M:

$$M = \{ \mathbf{P} \in [0, 1]: p(CK) + p(C'K) + p(CK') + P(C'K') = 1\}.$$

Przy tych oznaczeniach nasza hipoteza H = { **P** ∈ M: p(C’K) = 0). Prawdopodobieństwo a priori dla wektora stanowiącego rozkład prawdopodobieństwa **P** obliczymy jako:

$$P(\mathbf{P}) = 2^{-K(\mathbf{P})}$$

Prawdopodobieństwo a priori hipotezy H będzie więc większe od zera:

$$P(H) = \Sigma\, 2^{-K(\mathbf{P})} > 0.$$

Suma jest tu po wszystkich ciągach (wektorach) **P.**

Tym samym również M(H/**x**) będzie zdążało do 1, co pokazuje, że M potwierdza hipotezę H o ile nie zaobserwowano nie-czarnego kruka.

IV. KONKLUZJE

Analizowaliśmy relacje konsekwencji występujące we wnioskowaniach, w których wniosek nie jest konsekwencją klasyczną, lecz jest wynikiem takiej a nie innej informacji zawartej w przesłankach. Okazuje się, że zdecydowana większość naszych wnioskowań, jakie wykonujemy w życiu codziennym, nie jest wnioskowaniami czysto dedukcyjnymi, lecz odwołującymi się do szeroko rozumianego pojęcia informacji, wykraczającego poza informację zawartą w przesłankach. Wnioskowania takie umożliwiają wyprowadzenie niebanalnych wniosków w takich przypadkach, w których nie pozwala nam na to klasyczna relacja konsekwencji. Pomimo tego, że nie spełniają one wymogu formalnej poprawności, to jednak wyróżniają się własnościami, które czynią je wnioskowaniami racjonalnymi.

Analizowaliśmy również pojęcie informacji, do jakiego mogłyby odwoływać się tego rodzaju wnioskowania. Zwróciliśmy uwagę na trzy aspekty informacji takie jak redukcja niepewności, treść sądu i aspekt zaskoczenia. Te trzy aspekty informacji odgrywają istotną rolę w potocznych wnioskowaniach supraklasycznych. Przykładem takiego wnioskowania może być rozważane w tym artykule wnioskowanie przyczynowe. We wnioskowaniu tym odwoływaliśmy się do supraklasycznej i niemonotonicznej relacji konsekwencji zdefiniowanej w terminach prawdopodobieństwa. Wybór tej a nie innej relacji konsekwencji w rozważanym tu przykładzie spowodowany był tym, że również pojęcie informacji, jakie odgrywa istotną rolę w tym wnioskowaniu, w jednym z klasycznych swych ujęć opisywane jest również w terminach prawdopodobieństwa. Nie tylko wnioskowania przyczynowe są domeną supraklasycznych i niemonotonicznych relacji konsekwencji, lecz również wnioskowania, jakie prowadzą do zmiany przekonań i aktualizacji naszej wiedzy.

Trzecia część artykułu poświęcona była problemowi indukcji i jego rozwiązaniu, jakie odwołuje się z jednej strony do teorii Bayesa, a z drugiej strony do pojęcia algorytmicznego prawdopodobieństwa, które zostało skonstruowane przez Solomonoffa i wykorzystane przez niego do rozwiązania problemu indukcji, a tym samym problemu racjonalności wnioskowań opartych na przesłankach będących wynikami naszych obserwacji.

LITERATURA

1. Adriaans, P., J.van Benthem (eds), 2008, *Philosophy of Information*, Amsterdam: Elsevier.
2. Applebaum, D., 1996, *Probability and Information. An Integrated Approach* , Cambridge: Cambridge University Press.
3. van Benthem, J., M. Martinez, 2008, „The Stories of Logic and Information”, [w:] *Philosophy of Information*, ibid., s. 217 – 280.
4. Dretske, F., 2008, “Epistemology and Information”, [w:] *Philosophy of Information*, ibid., s. 29 – 47.
5. Hintikka, J., 1967, “The Varieties of Information and Scientific Explanation” [w:] van Rotseelar, B., Staal, J. (eds), *Logic, Methodology, and Philosophy of Science*, Amsterdam: North-Holland.
6. Jago, M., 2006, “Imagine the Possibilities. Information without Overload”, *Logique et Analyse*, 196, 345 – 370.
7. Lombardi, O. 2004, “What is Information?”, *Foundations of Science*, 9, 105 – 134.
8. Makinson, D., 2005, *Bridges from Classical to Nonmonotonic Logic*, London: King’s College Publications.
9. Rott, H., 2008, “Information Structures in Belief Revision”, [w:] *Philosophy of Information*, ibid., s. 457 – 482.