Przegląd Filozoficzny – Nowa Seria
R. 22: 2013, Nr 1 (85), ISSN 1230–1493
DOI: 10.2478/pfns-2013-0024

Anna Wójtowicz, Jan Winkowski

Jak nie powinno się obliczać stopnia wiarygodności argumentacji*

Słowa kluczowe: *argumentacja, twierdzenie Coxa, sekwent, twierdzenie Bayesa, wiarygodność wnioskowań, prawdopodobieństwo*

I. Wstęp

Filozofię można uprawiać w różny sposób. Niektórzy mówiąc o problemach filozoficznych posługują się językiem pełnym metafor, bazującym na intuicyjnym wglądzie w dany problem. Dzięki temu u odbiorcy powstają odpowiednie skojarzenia, które czasami mogą dawać poczucie zrozumienia i pełnego przekonania co do głoszonych na dany temat tez. Inni używają języka bardziej sformalizowanego, w którym przynajmniej pewne pojęcia mają jednoznaczny, ściśle zdefiniowany sens, a uzasadnienie tezy sprowadza się do argumentacji o ustalonej strukturze. Każda z tych metod jest właściwa do typu zagadnień, który chcemy analizować.

Zaletą języka formalnego jest niewątpliwie to, że wyrażane za jego pomocą tezy mają jasny sens. W związku z tym są one (na gruncie ustalonej interpretacji) po prostu prawdziwe lub fałszywe. Daje się je również porównać z innymi tezami zapisanymi w podobnym języku. Dzięki temu możemy ocenić ich oryginalność – to, czy wniosły do analizowanego problemu coś nowego. Mówiąc w pewnym uproszczeniu: o takich tezach nie dyskutuje się godzinami w zadymionej kawiarni, ale brutalnie się je sprawdza.

Niewątpliwie w języku częściowo przynajmniej formalnym został napisany tekst Marcina Selingera *Formalna ocena argumentacji*, opublikowany niedawno w „Przeglądzie Filozoficznym”[1]. Autor formułuje kilka tez – poda-

* Praca powstała w ramach grantu NCN 2012/05/B/HS1/01711.

[1] M. Selinger, *Formalna ocena argumentacji*, „Przegląd Filozoficzny – Nowa Seria” 2012, nr 1 (81), s. 89–109.

jąc w szczególności wzór obliczania wiarygodności argumentacji – które aż proszą się o sprawdzenie. Niniejszy artykuł stawia sobie to właśnie zadanie.

Na początku pracy Selinger przedstawia i omawia dwa typy argumentacji: szeregową i równoległą. Samą argumentację rozumie jako ciąg sekwentów. Sekwent <P, *a*> jest to para uporządkowana, składająca się z niepustego i skończonego zbioru P zdań danego języka i pojedynczego zdania *a*. Intencja jest taka, że ów zbiór P jest zbiorem przesłanek danej argumentacji, natomiast zdanie *a* – konkluzją. Mając tak zdefiniowany sekwent, autor definiuje pojęcia spójności, zbieżności czy kolistości argumentacji.

Sedno artykułu dotyczy sposobu oceny tak rozumianych argumentacji. Aby zrobić to ściśle, autor postuluje wprowadzenie miar przekonań (tj. akceptacji lub odrzucenia), które nazywa wartościami epistemicznymi. Te wartości można rozpatrywać bądź z punktu widzenia psychologiczno-opisowego, bądź logiczno-normatywnego. W tym drugim przypadku „wyrażają one obiektywną wiarygodność zdań danego języka dla wyidealizowanego podmiotu poznania, którego racjonalność zakładamy”[2] – i ta perspektywa będzie nas tu przede wszystkim interesować. Autor formułuje dwie interesujące dla nas tezy: na temat własności samego pojęcia wiarygodności i na temat wzoru, jaki pozwala obliczać wiarygodność argumentacji.

Przedstawimy je możliwie wiernie, a następnie skomentujemy.

Teza 1

Niech Z będzie zbiorem zdań danego języka, X – zbiorem wartości epistemicznych, w – funkcją częściową ewaluacji, prowadząca ze zbioru Z w zbiór X. X i w spełniają następujące warunki:

(a) X jest zbiorem uporządkowanym.
(b) Jeśli a jest zdaniem w pełni akceptowalnym (np. tautologią), to $w(a)$ jest elementem największym zbioru X.
(c) Jeśli a jest zdaniem całkowicie odrzuconym (np. kontrtautologią), to $w(a)$ jest elementem najmniejszym zbioru X.
(d) Jeśli a jest zdaniem, co do którego nie jesteśmy w stanie rozstrzygnąć, czy jest prawdziwe, czy fałszywe, to $w(a)$ jest wartością odpowiadającą niezdecydowaniu.
(e) Zbiór X jest gęsty (dla dowolnych dwóch wartości epistemicznych istnieje między nimi wartość pośrednia).

Według autora, można bez straty ogólności założyć, że X jest domkniętym przedziałem <0, 1> liczb wymiernych. Wartości niezdecydowania odpowiada 1/2. Można ją traktować jako próg akceptacji, które rozsądne konkluzje powin-

[2] Tamże, s. 98.

ny przekraczać. Zamiast o funkcji ewaluacji *w* dalej będziemy mówić o funkcji wiarygodności: $w(p)$ to wiarygodność zdania *p*.

Aby można było obliczyć wiarygodność konkluzji w danej argumentacji, musimy znać wartości wiarygodności przesłanek i wiedzieć, jak przekładają się one na wartość konkluzji przy połączeniach szeregowych, a jak przy połączeniach równoległych. Dzięki temu możemy krok po kroku obliczyć wartość konkluzji pośrednich aż do wartości konkluzji głównej. Aby uprościć cały wywód, będziemy na początku zajmować się przypadkami, gdzie zbiór przesłanek jest jednoelementowy (i zamiast P będziemy pisać *p*), a w związku z tym nie ma znaczenia, czy mamy do czynienia z argumentacjami równoległymi, czy szeregowymi.

Teza 2

Wiarygodność konkluzji *w* argumentacji $<p, a>$, co symbolicznie będziemy oznaczać przez $w(p, a)$, wyraża następujący wzór:

$$w(p, a) = w(a/p) \times w(p),$$

gdzie $w(a/p)$ jest miarą tego, „jak silny jest związek inferencyjny między przesłanką a konkluzją. Współczynnik ten postaramy się wyrazić bazując na intuicjach związanych z pojęciem warunkowego prawdopodobieństwa logicznego”[3]. Jest to według autora wartość epistemiczna, jaką należy przyporządkować konkluzji *a*, jeśli wiadomo, że $w(p) = 1$.

II. Ocena

Pozornie wszystko wygląda bardzo dobrze: teza 1 aksjomatycznie charakteryzuje pojęcie wiarygodności, a teza 2 formułuje wzór będący nieomalże Świętym Graalem filozofów zajmujących się pojęciem racjonalności wnioskowań zawodnych: pozwala wyliczać wiarygodność wniosków. Mając taki wzór, zawsze potrafilibyśmy rozstrzygać, jakie wnioski na podstawie danych przesłanek należy formułować: te mianowicie, które maksymalizują wyliczoną według wzoru wartość wiarygodności.

Po dokładniejszym przyjrzeniu okazuje się jednak, że podana charakterystyka pojęcia wiarygodności jest tylko mniej subtelną wersją założeń twierdzenia R.T. Coxa, sformułowanego w 1946 roku[4]. Szeroko cytowana i komen-

[3] Tamże, s. 101.

[4] R.T. Cox, *Probability, frequency and reasonable expectation*, „American Journal of Physics” 1946, nr 17, s. 1–13.

towana praca K.S. Van Horna z 2003 roku[5] omawia dokładnie te założenia. Przytoczymy je krótko w celu porównania z tym, co mamy zawarte w tezie 1, i stwierdzenia, czego w niej brakuje.

Założenia tw. Coxa (na podstawie pracy Van Horna)

Niech dany będzie zbiór zdań Z danego języka i funkcja wiarygodności w, prowadząca ze zbioru Z w zbiór X. X i w spełniają następujące warunki:

(A) X jest zbiorem liczb rzeczywistych (jest uporządkowany i gęsty).

(B) Funkcja wiarygodności w zachowuje znaczenie spójników klasycznych, tzn.:

1. jeśli a i b są równoważne logicznie, to $w(a) = w(b)$;
2. jeśli a jest tautologią, to $w(a)$ jest elementem największym zbioru X;
3. jeśli a jest kontrtautologią, to $w(a)$ jest elementem najmniejszym zbioru X;
4. $w(a/b \wedge c) = w(a/b,c)$ (koniunkcja działa klasycznie);
5. jeśli $w(\sim a)$ jest różne od elementu maksymalnego w X, to $w(a)$ jest różne od elementu minimalnego w X.

(C) Jeśli a i b są zdaniami niezależnymi, to $w(a)$ nie wpływa na $w(b)$ (mogą niezależnie przyjmować różne wartości).

(D) $w(a \wedge b)$ jest funkcją $w(a/b)$ i $w(b)$ (czyli $w(a \wedge b) = f(w(a/b), w(b))$.

Bez straty ogólności możemy założyć, że w(tautologia) = 1, w(kontrtautologia) = 0, a więc X jest domkniętym przedziałem <0, 1> liczb rzeczywistych.

Są dwie zasadnicze różnice między założeniami tw. Coxa a tezą 1.

Po pierwsze, Selinger nie rozstrzyga jawnie, czy funkcja w zachowuje znaczenie spójników klasycznych. Można próbować to wyczytać w stwierdzeniach, że wiarygodność przypisuje zdaniom w pełni racjonalny podmiot i że przy wnioskowaniach dedukcyjnych wiarygodność schematu wnioskowania jest równa 1. Wnioskowania dedukcyjne to takie, w których wniosek wynika logicznie z przesłanek – co oznacza, że wynika wyłącznie na mocy znaczenia stałych logicznych użytych we wnioskowaniu. Zwykle, jeśli jawnie nie zaznaczy się inaczej, stałe logiczne to po prostu klasyczne spójniki i kwantyfikatory.

Po drugie i bardziej istotne, nie ma odpowiednika założenia D, tzn. rozstrzygnięcia, jak należy liczyć wiarygodność koniunkcji zdań, które nie są od siebie niezależne. W efekcie nie wiadomo, jak liczyć wiarygodność schematu wnioskowania $w(a/p)$. Selinger tylko sugeruje, że należy go traktować

---

[5] K.S. Van Horn, *Constructing a logic of plausible inference: a guide to Cox's Theorem*, „International Journal of Approximate Reasoning" 2003, nr 1 (34), s. 3–24.

podobnie jak prawdopodobieństwo warunkowe: wiarygodność zdania *a* przy założeniu, że zdanie *p* jest w pełni wiarygodne. Jak zobaczymy, będzie to miało swoje konsekwencje przy ocenie tezy 2. Przypomnijmy też, że podstawowa idea przedstawionych wyżej założeń jest taka, że po ich przyjęciu można formalnie udowodnić, że pojęcie wiarygodności jest równoważne pojęciu prawdopodobieństwa, a więc zamiast o funkcji *w* można po prostu mówić o prawdopodobieństwie (w rozumieniu Kołmogorowa). Innymi słowy, jeśli Selinger zgodziłby się, że jego warunki należy uzupełnić o te, które podał Cox, musiałby również uznać, że zdefiniowana przez niego funkcja wiarygodności *w* to po prostu funkcja przypisująca zdaniom ich prawdopodobieństwo.

Aby teza 2 rzeczywiście miała takie znaczenie, jakie chcemy jej przypisać, powinna w dokładniejszy sposób charakteryzować pojecie wiarygodności schematu wnioskowania. Jeśli nie jest dany ogólny przepis, jak liczymy $w(a/p)$, cały wzór traci swoją moc[6]. Potraktujmy poważnie sugestię autora (i twierdzenie Coxa) i spróbujmy zastosować wzór Bayesa:

(*) $w(a/p) = w(a \wedge p) \,/\, w(p)$

Oznacza to, że wiarygodność schematu jest stosunkiem między wiarygodnością prawdziwości zarazem konkluzji i przesłanki, i wiarygodnością przesłanki. Liczymy po prostu, ile wśród zdarzeń potwierdzających *p* potwierdza również *a*.

Jeśli teraz podstawimy to do wzoru z tezy 2, otrzymamy:

(**) $w(p, a) = w(a/p) \times w(p) = (w(a \wedge p) \,/\, w(p)) \times w(p) = w(a \wedge p)$.

W dowolnym wnioskowaniu redukcyjnym z wniosku *a* wynikają logicznie przesłanki, więc:

$w(a \wedge p) = w(a)$.

W efekcie więc mamy:

$w(p, a) = w(a)$,

---

[6] Istnieją teorie (por. np. N. Pfeifer, G.D. Kleiter, *Framing human inference by coherence based probability logic*, „Journal of Applied Logic" 2010, nr 2 (7), s. 206–217), w których przyjmuje się, że pojęcie prawdopodobieństwa warunkowego $P(B/A)$ jest pojęciem pierwotnym. Nie definiuje się go jako $P(A \wedge B) \,/\, P(A)$, ale nakłada się aksjomatyczne warunki na funkcję przypisującą wartości od razu zdaniom warunkowym (a nie zdaniom „absolutnym"). Jak należy się domyślać z kontekstu, autor nie taką jednak teorię ma na myśli – nie powołuje się zresztą na żadną związaną z tym literaturę.

Ustaliliśmy więc, że zgodnie z tezą 2, jeśli przyjmiemy wzór (*), wiarygodność konkluzji a otrzymanej w wyniku przeprowadzenia dowolnej argumentacji redukcyjnej jest zawsze taka sama:

$$w(p, a) = w(p', a)$$

dla dowolnych przesłanek p i p'. Jest to oczywiście wniosek absolutnie dyskwalifikujący dla całej koncepcji.

Czy można się jakoś przed takimi konsekwencjami tezy 2 bronić? Należałoby może uznać, że co innego *wiarygodność aprioryczna przesłanki p* (do której odwołanie pojawia się we wzorze (*)), a co innego *stopień akceptacji, który danej przesłance przypisujemy w konkretnym sekwencie <p, a>*. Oznaczmy w związku z tym przez $W(a \wedge p)$ i $W(p)$ aprioryczne wartości wiarygodności konkluzji a i przesłanki p (utożsamiane np. z ich prawdopodobieństwem), a przez $w(p)$ – stopień wiarygodności przesłanki, przypisywany jej w danej argumentacji (np. związany z zaufaniem do źródła, które przekazuje nam tę przesłankę). Zamiast wzoru (*) (i niezgodnie z tw. Coxa) powinniśmy stosować wzór:

$$(***)\ w(a/p) = W(a \wedge p) / W(p),$$

Dzięki takiemu zabiegowi nie będziemy mogli po prostu skrócić $W(p)$ z $w(p)$, jak to zrobiliśmy w (**), i cała koncepcja wydaje się bronić.

Rozważmy jednak następujące dwa wnioskowania.

Przykład 1

Losujemy z pełnej talii kart jedną kartę. Na podstawie tego, że zobaczyliśmy, że wylosowana karta jest czarna, argumentujemy, że wylosowaliśmy pika.

Intuicyjnie widać, że stopień niezawodności takiego schematu wynosi 1/2 (bo czarna karta to z równym prawdopodobieństwem pik albo trefl). Dodatkowo przyjmijmy, że nie widzimy dokładnie karty i nie jesteśmy na 100% pewni, że karta jest czarna. W złym oświetleniu nasze spostrzeżenie uznajemy tylko w 9/10 za wiarygodne. Tak rozumiana wiarygodność nie jest więc równa prawdopodobieństwu apriorycznemu, że wylosujemy czarną kartę (jest ono równe 1/2). Mając więc dane: $w(a/p) = 1/2$, $w(p) = 9/10$, wyliczamy zgodnie z tezą 2, że wiarygodność konkluzji wynosi: $1/2 \times 9/10 = 9/20$.

Przykład 2

Losujemy z pełnej talii kart jedną kartę. Tym razem na podstawie tego, że zobaczyliśmy, że wylosowana karta to czarny walet, wnioskujemy, że wylosowaliśmy pika.

I znowu możemy intuicyjnie ocenić stopień niezawodności takiego schematu: wynosi on 1/2 (bo wśród dwóch czarnych waletów jeden to pik). Ponieważ oświetlenie jest równie złe jak poprzednio, musimy założyć, że wzrok może nas mylić. Ale ponieważ tym razem twierdzimy coś więcej niż poprzednio, wiarygodność przesłanki spada do 8/10. Mając więc dane: $w(a/p) = 1/2$, $w(p) = 8/10$, wyliczamy zgodnie z tezą 2, że wiarygodność konkluzji wynosi: $1/2 \times 8/10 = 8/20$.

Przedstawione w powyższych przykładach konkluzje mają zgodnie z zaproponowanym wzorem inną wiarygodność. Trudno się jednak z tym zgodzić – to, że wylosowana karta jest czarnym waletem, czy jest jakąkolwiek inną czarną kartą o dowolnej wysokości, nie powinno wpływać na ocenę konkluzji. Wysokość karty nie jest istotna z punktu widzenia rozstrzygnięcia, czy jest to pik. A więc wzór z tezy 2 daje wyniki ewidentnie niezgodne z naszymi przekonaniami. Zaprezentowany tutaj argument nie jest zresztą niczym nowym – został on przez Davida Atkinsona[7] nazwany efektem nieistotnej koniunkcji i użyty, aby pokazać, że czymkolwiek miałaby być wiarygodność konkluzji, nie może ona zależeć wprost od wiarygodności przesłanki. Przedstawmy jeszcze raz – tym razem ogólnie – jego ideę: dodając do przesłanki jakiś nieistotny z punktu widzenia wnioskowania szczegół (niewpływający na wiarygodność schematu i niezależny od istotnej informacji, która jest zawarta w przesłance), zmieniamy jej wiarygodność. Gdyby wiarygodność konkluzji zależała wprost od wiarygodności przesłanki, to zmieniałoby to jej wartość. Co sprzeczne z tym, że założyliśmy, że wzmocnienie przesłanki (dodanie tego nieistotnego szczegółu) jest z punktu widzenia wnioskowania i konkluzji nieistotne.

Na zakończenie jeszcze jedna uwaga szczegółowa, która pokazuje, że również inne dystynkcje i proponowane wzory obecne w artykule można podważyć. Jak już powiedzieliśmy, autor dzieli argumentacje na szeregowe i równoległe. W argumentacjach szeregowych usunięcie którejkolwiek przesłanki sprawiłoby, że związek inferencyjny między przesłankami a konkluzją zostałby zerwany, w argumentacjach równoległych usunięcie którejś z przesłanek spowoduje tylko osłabienie takiego związku[8].

---

[7] D. Atkinson, *Confirmation and justification. A commentary on Shogenji's measure*, „Synthese" 2012, nr 184, s. 49–61.

[8] M. Selinger, *Formalna ocena...*, dz. cyt., s. 91.

Rozważmy wobec tego następujący klasyczny przykład argumentacji indukcyjnej.

Przykład 3

Przesłanki:

p_1: Student *A* zda egzamin z logiki.
p_2: Student *B* zda egzamin z logiki.
...
p_8: Student *H* zda egzamin z logiki.

Konkluzja: Wszyscy studenci ze zbioru $\{A,..., H\}$ zdadzą egzamin z logiki.

Zgodnie z przytoczoną dystynkcją argumentację taką należałoby zaliczyć do równoległej – ponieważ przesłanki wspierają wniosek niezależnie od siebie. Według autora, taka argumentacja rozpada się na szereg sekwentów postaci: $<p_1, a>$, $<p_2, a>$,...,$<p_8, a>$. Policzmy teraz, jaka jest wiarygodność konkluzji *a* w takich sekwentach (zakładając, że wiarygodność wszystkich przesłanek jest równa 1 i że w równym stopniu wspierają one konkluzję):

$$w(p_i, a) = w(a/p_i) \times w(p_i) = 1/8 \times 1 = 1/8.$$

Oznacza to, że żaden z sekwentów nie wspiera konkluzji na tyle mocno, żeby przekroczyć próg 1/2 (czyli próg niezdecydowania, poniżej którego konkluzji nie można uznać za akceptowalną). Nie można w związku z tym użyć zaproponowanego przez autora na s. 103 wzoru, pozwalającego liczyć, jak niezależne od siebie przesłanki w sumie wspierają konkluzję. Nie da się więc ocenić argumentacji z przykładu 3 – co jest ewidentnie niezgodne z intuicją, że jest to po prostu argumentacja dedukcyjna i konkluzja jest pewna.

Autor mógłby próbować się bronić, że powyższa argumentacja jest szeregowa – tylko pozornie mamy do czynienia z 8 różnymi przesłankami i faktycznie należy je traktować jako koniunkcję: $p_1 \wedge ... \wedge p_8$. Wtedy – jeśli wszystkie przesłanki mają stopień wiarygodności równy 1, i dzięki temu, że schemat wnioskowania jest dedukcyjny (jest to indukcja enumeracyjna zupełna), po podstawieniu do wzoru:

$$w(p_1 \wedge ... \wedge p_8, a) = w(a/p_1 \wedge ... \wedge p_8) \times w(p_1 \wedge ... \wedge p_8)$$
$$= w(a/p_1 \wedge ... \wedge p_8) \times w(p_1) \times ... \times w(p_8) = 1$$

otrzymujemy oczekiwany wniosek, że konkluzja tej argumentacji jest pewna.

Zauważmy jednak, że takie rozwiązanie prowadzi do niepożądanych konsekwencji, jeśli przyjmiemy, że wiarygodność każdego p_i jest duża, ale różna od 1 i wynosi np. 0,9. Wtedy mamy:

$$w(p_1 \wedge \ldots \wedge p_8, a) = w(a/p_1 \wedge \ldots \wedge p_8) \times w(p_1) \times \ldots \times w(p_8)$$
$$= 1 \times 0{,}9^8 = 0{,}43.$$

A więc konkluzja nie przekracza progu akceptacji i należy ją odrzucić. Nie jest to wniosek, który chcielibyśmy przyjąć.

III. Podsumowanie

Marcin Selinger poruszył w swoim artykule problemy ważne i szeroko na świecie dyskutowane. Nie można w związku z tym przejść obojętnie wobec zaproponowanych przez niego rozwiązań. Brak reakcji sugerowałby, że je akceptujemy. Powyższa analiza wskazuje jednak, że koncepcja ta nie wytrzymuje krytyki.

Streszczenie

Artykuł jest krytyczną analizą pracy Marcina Selingera *Formalna ocena argumentacji* i dotyczy sposobu oceny wiarygodności argumentacji. Ponieważ omawiany problem jest bardzo ważny i szeroko współcześnie dyskutowany, nie można podanej przez autora propozycji pozostawić bez komentarza. Przedstawiamy zwięzłe streszczenie tekstu i rekonstrukcję dwóch najważniejszych jego tez. Analizujemy różnice między pierwszą z nich a twierdzeniem Coxa, co prowadzi do krytyki drugiej z tez. Zastosowanie twierdzenia Bayesa ujawnia jej niepożądane konsekwencje. Podejmujemy próbę obrony zaproponowanej przez Selingera tezy, jednak tzw. efekt nieistotnej koniunkcji pokazuje, czemu nie może się ona udać. Artykuł kończy dodatkowy kontrprzykład dotyczący obecnego w tekście rozróżnienia sposobów obliczania wiarygodności argumentacji szeregowych i równoległych.