

Anna Wójtowicz

Czy pojęcie racjonalności wnioskowania ma charakter normatywny, czy deskryptywny?*

WSTĘP

Wnioskowania dzielą się na *niezawodne* (dedukcyjne) i *zawodne* (niededukcyjne). We wnioskowaniach niezawodnych prawdziwość przesłanek gwarantuje prawdziwość wniosku. Aby sprawdzić, czy mamy do czynienia z wnioskowaniem niezawodnym, wystarczy ustalić, czy jego schemat jest prawem odpowiedniej logiki. Choć zadanie to bywa trudne, a wybór konkretnej logiki wymaga uzasadnienia, to procedura taka pozwala jednoznacznie rozstrzygnąć, czy mamy do czynienia z wnioskiem pewnym. We wnioskowaniach zawodnych prawdziwość przesłanek z definicji nie gwarantuje prawdziwości wniosku. Wnioski takie mogą być tylko lepsze (bardziej racjonalne) lub gorsze (mniej racjonalne), nigdy natomiast nie są pewne. Ponieważ na co dzień wnioskujemy w większości wypadków zawodnie, pojawia się bardzo praktyczny problem, jak w takich wnioskowaniach odróżnić konkluzje lepsze od gorszych?

Na ten temat powstało wiele wzajemnie wykluczających się teorii. Różnice między nimi polegają przede wszystkim na wyborze języka, w którym wnioskowanie jest formalizowane. To samo rozumowanie sformułowane w języku potocznym można bowiem na różne sposoby przełożyć na język o charakterze formalnym. Najogólniej biorąc, problem polega na tym, że musimy zdecydować, które z wyrażenń uznamy za istotne przy ustalaniu związku między przesłankami a wnioskiem (w szczególności — które uznamy za stałe logiczne) i jakie dokładnie przyporządkujemy im znaczenie. Gdy już się z tym uporamy, rozstrzygnięcie, czy wnioskowanie jest racjonalne, ma w dużej mierze charakter techniczny i nie budzi kontrowersji.

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

Ocenę racjonalności wnioskowania można więc sprowadzić do oceny systemu, w którym dane wnioskowanie jest formalizowane.

Ponieważ istnieje wiele takich systemów i brak jest jasnego kryterium wyboru między nimi, pojawia się pokusa stosowania „argumentu z człowieka”, który można sformułować w następujący sposób:

Wnioskowanie należy formalizować w taki sposób, w jaki zrobiliby to ludzie, i uznać za właściwy ten wniosek, który ludzie na podstawie danych przesłanek wyciągają.

1. JAK UZASADNIĆ ARGUMENT Z CZŁOWIEKA?

Skąd pomysł, aby traktować zachowania ludzi jako wzór racjonalności? Można wskazać trzy podstawowe strategie, które pośrednio nas do tego przekonują.

Pierwsza z nich związana jest z tym, że w miarę rozwoju techniki pojawiła się potrzeba podania kryterium, które musi spełnić maszyna (program komputerowy), żeby można ją było określić mianem „sztucznej inteligencji”. Chcąc stwierdzić, czy zachowanie maszyny jest inteligentne, możemy zbadać, czy ma ono określone cechy. Zwykle wymienia się tu umiejętność uczenia się, twórczego rozwiązywania problemów i stawiania sobie nowych, nieprzewidzianych przez programistę zadań. Trudno sprecyzować jednak taką listę cech i uniezależnić od typu środowiska, w którym maszyna działa. Z tego powodu Alan Turing (1995) zaproponował, aby za kryterium inteligencji maszyny uznać po prostu to, czy przejdzie ona tzw. test Turinga, a więc będzie umiała zachować się w danej sytuacji tak jak człowiek:

Uznamy, że maszyna jest inteligentna, jeśli jej zachowanie byłoby dla testera nieodróżnialne od zachowania człowieka.

Jednym z problemów, które stawia się przed testowaną maszyną, jest zadanie polegające na wywnioskowaniu czegoś z danych przesłanek. Tester ocenia takie wnioskowanie i klasyfikuje jednostkę jako człowieka, jeśli wnioskowanie to będzie w jego mniemaniu racjonalne¹. Innymi słowy, z powodu braku uniwersalnego i łatwo stosownego kryterium utożsamiamy wnioskowanie racjonalne z wnioskowaniem, które przeprowadza człowiek. Człowiek staje się więc wzorem racjonalności, ponieważ nie umiemy wskazać efektywnie innego wzoru.

Do uzasadnienia argumentu z człowieka możemy również wykorzystać idee zawarte w pracach Davidsona (por. np. 1982). Celem Davidsona była konstrukcja teorii znaczenia. Problem przypisywania znaczenia terminom pojawiającym się w wypowiedziach innych użytkowników języka starał się rozwiązać za pomocą Zasady Życzliwości (*Principle of Charity*), zgodnie z którą wypowiedzi ludzi należy inter-

¹ Trudno wyobrazić sobie sytuację, w której tester uznaje, że jednostka jest człowiekiem, ponieważ zachowała się w sposób nieracjonalny.

pretować w taki sposób, aby jak najwięcej z nich można było uznać za prawdziwe, a dokonywane wnioskowania — za racjonalne. Dzięki założeniu, że ludzie mówią prawdę i wnioskujeją racjonalnie, możemy ustalić, jakie znaczenie przypisują użytym terminom. W tym wypadku przyjmujemy, że człowiek jest wzorem racjonalności, ponieważ pozwala nam to zbudować teorię o pożądanym własnościach.

Inne, ewolucyjne uzasadnienie argumentu z człowieka można znaleźć w szeroko omawianej pracy Robina Foxa (1992), która zapoczątkowała cały nurt analiz dotyczących użyteczności uproszczeń obecnych we wnioskowaniach ludzi. Były one z punktu widzenia normatywnych modeli wnioskowania uznawane za błędy i wyraz nieracjonalności. Fox broni przekonania, że są one niezbędne do tego, aby człowiek mógł przeżyć w skomplikowanej rzeczywistości, wykraczającej poza jego możliwości poznawcze. Różnego typu kategoryzacje (np. dzielenie obiektów według wielkości lub ludzi według ras) i utożsamianie obiektów należących do jednej kategorii jest sposobem upraszczania rzeczywistości, a co się z tym wiąże — usprawniania wnioskowań. Gdyby człowiek chciał przy wnioskowaniu wykorzystywać wszystkie dostępne mu informacje, to działałby niewydajnie i nieskutecznie (zostałby na przykład pożarty przez drapieżnika). Osobniki posługujące się heurystykami² przetrwały, a więc heurystyki są dobrym, uzasadnionym przez ewolucję sposobem wnioskowania. Mając więc do wyboru uznanie wnioskowania za racjonalne na tej podstawie, że jest ono zgodne z jakimś modelem normatywnym, albo na tej podstawie, że tak postępują ludzie, autor opowiada się za tym drugim rozwiązaniem. Człowiek jest wzorem racjonalności, ponieważ stanowi to wytłumaczenie jego sukcesu ewolucyjnego.

Podsumowując, traktowanie zachowania ludzi jako wzorca racjonalności ma trojaki podstawy:

- nie widać innego *efektywnego* podejścia;
- jest to założenie stanowiące warunek konieczny budowy innych teorii dotyczących ludzkich aktywności;
- jest ono uzasadnione przez osiągnięty przez człowieka sukces ewolucyjny.

Oddzielną kwestią, do której wrócę w dalszej części artykułu, jest to, czy mówiąc o racjonalności ludzi, mamy na myśli *dowolnego* człowieka, czyli człowieka,

² Warto zwrócić uwagę, że termin „heurystyka” (w kontekście wydawania sądów) stopniowo zmienia swoje znaczenie. W pracach Tversky’ego i Kahnemana występuje najczęściej z przydawkami: heurystyka dostępności, heurystyka zakotwiczenia, heurystyka reprezentatywności. Takie heurystyki są utożsamiane z błędami poznawczymi, czyli uproszczonymi metodami wnioskowania często prowadzącymi do nieracjonalnych wniosków. Obecnie — np. w pracach Gigerenzera i jego zespołu — przez heurystykę rozumie się metodę, w której *świadomie* rezygnuje się z wykorzystania przy wydawaniu sądów wszystkich dostępnych informacji. Może to być uzasadnione ograniczeniami czasowymi i obliczeniowymi, a w rezultacie przynosić wnioskującemu więcej zysków niż strat. W tym znaczeniu heurystyki nie są *a priori* uznawane za błędne uproszczenia. Tak też rozumiem termin „heurystyka” w tym artykule.

którego zachowanie jest reprezentatywne dla *większości* ludzi, czy człowieka, którego zachowanie jest reprezentatywne dla jakiejś *wybranej grupy* ludzi.

Uznanie argumentu z człowieka za dopuszczalny zmienia istotnie perspektywę badawczą przy opisie i ocenie wnioskowań zawodnych. Zmusza nas bowiem nie tylko do analizy własności formalnych systemu, którego używamy do wyciągania wniosków, lecz także do sprawdzenia (na drodze doświadczenia), czy w danej sytuacji również ludzie się nim posługują. Można powiedzieć, że degraduje pojęcie racjonalności z pozycji apriorycznej i normatywnej do deskryptywnej.

W artykule spróbuję przeanalizować, gdzie takie deskryptywne podejście do problemu oceny wnioskowań zawodnych może się pojawić, jakie rodzi problemy i czy możemy uznać je za obiecujące.

2. JAK WNIOSKUJĄ LUDZIE? CZTERY KLASYCZNE PRZYKŁADY

Jako materiał badawczy przeanalizujemy cztery przykłady wnioskowań często opisywane w literaturze przy różnych okazjach.

PRZYKŁAD 1. FEMINISTKA LINDA

W 1971 roku Amos Tversky i Daniel Kahneman opisali eksperyment, w którym przedstawili badanym następujący problem:

Linda ma 31 lat, jest otwartą, inteligentną i niezamężną kobietą. Ukończyła filozofię. Jako studentka poświęcała dużo czasu problemom sprawiedliwości społecznej i dyskryminacji, uczestniczyła też w demonstracjach antynuklearnych. Co jest bardziej prawdopodobne?

H1. Linda pracuje w banku.

H2. Linda pracuje w banku i jest aktywną działaczką ruchu feministycznego.

Ku ich nieopisanemu zdumieniu 85% badanych wybrało odpowiedź H2 (Tversky, Kahneman 1971). Jest to zachowanie w tym sensie nieracjonalne, że zdanie H2 jest koniunkcją zdania H1 i dodatkowego stwierdzenia, że Linda jest feministką. Ponieważ zawsze jest tak, że prawdopodobieństwo zdania koniunkcyjnego jest nie większe niż prawdopodobieństwo każdego z członów koniunkcji (symbolicznie: $P(p \wedge q) \leq P(p)$), to nie można twierdzić, że H2 jest rzeczywiście bardziej prawdopodobne. Autorzy eksperymentu uznali jego wynik za argument za tym, że ludzie w większości nie wnioskują zbyt rozsądnie (dodajmy, że osoby, na których eksperyment przeprowadzono, były studentami wyższych lat lub doktorantami mającymi za sobą kursy logiki i statystyki).

Na temat tego eksperymentu (i innych zbudowanych według podobnego schematu) napisano bardzo wiele prac, a błąd, który według autorów popełnili badani,

nazwano błędem koniunkcji. Od razu pojawiły się jednak głosy, że może ludzie nie tyle popełniają błąd, ile inaczej, niż to sobie wyobrażają twórcy eksperymentu, interpretują przedstawione dane. Moro (2009) przeanalizował tezę, zgodnie z którą ludzie wnioskujeją racjonalnie, ale ich rozumienie przynajmniej niektórych sformułowań pojawiających się w historii o Lindzie odbiega od zamierzonych przez eksperymentatorów. Nazwał ją „tezą o niewłaściwym rozumieniu”. Niewłaściwe rozumienie może dotyczyć:

- koniunkcji (która mogła być interpretowana jako alternatywa);
- zwrotu: „jest bardziej prawdopodobne” (któremu przypisywano znaczenie inne niż matematyczne);
- samej hipotezy H1 (która miałaby zawierać ukryte założenie, że Linda nie jest feministką);
- celu, w jakim zostało zadane pytanie (badani mogli formułować wniosek, maksymalizując jego wartość informacyjną, a nie po prostu prawdopodobieństwo).

Współcześnie (por. numer „Synthese” 184(1) z 2012 r. z pracami Schupbacha, Shogenjiego czy Atkinsona) dużo uwagi poświęca się problemowi formalnych założeń (dotyczących różnych związków zachodzących między informacjami pojawiającymi się w eksperymencie z Lindą), które należałoby przyjąć, aby wnioskowanie tych 85% badanych można było uznać za racjonalne. Jest to niewątpliwie zastosowanie argumentu z człowieka: prowadzone badania opierają się na założeniu, że błąd popełniany przez tak wielu ludzi nie może być błędem. W wyniku tych analiz model milcząco zakładany przez eksperymentatorów, w którym do opisu wnioskowania o Lindzie stosowano zwykły rachunek prawdopodobieństwa i logikę klasyczną, został zastąpiony odmiennym modelem, bardziej skomplikowanym i subtelnym. Można powiedzieć, że znaczenie przypisywane użytym we wnioskowaniu terminom — w duchu Zasady Życzliwości — zostało dobrane tak, aby maksymalizować prawdziwość i racjonalność tez formułowanych przez badanych.

PRZYKŁAD 2. JAK LEKARZE STAWIAJĄ DIAGNOZY?

Rozważmy teraz wnioskowanie, które było przedmiotem testu przeprowadzonego na dużej grupie amerykańskich i niemieckich lekarzy w 1982 roku (Westover, Westover, Bianchi 2011). Lekarzom zostały przedstawione następujące dane kliniczne: szansa wystąpienia raka piersi w danej grupie kobiet wynosi 1%, badanie mammograficzne wykazuje zmianę, mimo że kobieta *de facto* nie ma raka, w 9,6% przypadków (są to tzw. wyniki fałszywie dodatnie), natomiast badanie mammograficzne nie wykazuje zmiany, mimo że kobieta faktycznie ma raka, w 20% przypadków (są to tzw. wyniki fałszywie ujemne). Następnie przedstawiono lekarzom wnioskowanie:

- (*) **Przesłanki:** Badanie mammograficzne jest dobrym testem na to, czy kobieta jest chora na raka; badanie mammograficzne u kobiety z danej grupy wykryło zmianę;

Wniosek: Jest bardzo prawdopodobne, że dana kobieta jest chora na raka,

a potem poproszono ich o jego ocenę, czy rzeczywiście jest bardzo (a konkretnie: jak?) prawdopodobne, że kobieta jest chora na raka.

Wnioskowanie (*) jest sformułowane w języku, w którym występują potoczne, nieostre pojęcia: „jest dobrym testem na”, „jest bardzo prawdopodobne, że”. Ocena wniosku zależy od tego, jak te pojęcia zostaną sprecyzowane i które z dostępnych danych będą wykorzystane. Można na przykład przełożyć je na język ilościowy w następujący sposób:

() Przesłanki:** Jeśli kobieta nie jest chora na raka, to badanie mammograficzne z dużym prawdopodobieństwem (równym $0,904$)³ nie wykáže zmiany; badanie mammograficzne u danej kobiety wykryło zmianę;

Wniosek: Jest bardzo prawdopodobne, że dana kobieta jest chora na raka.

Takiego przekładu dokonują w szczególności zwolennicy systemu opisu wnioskowań zawodnych nazywanego testem hipotezy zerowej (NHST, *Null Hypothesis Significance Test*). Według nich jest to naturalne uogólnienie (na pojęcia związane z niepewnością) schematu wniosku niezawodnego *modus tollens*⁴:

(*) Przesłanki:** $\sim\alpha$ implikuje, że bardzo prawdopodobne jest $\sim\beta$;
Zaszło β ;

Wniosek: Jest bardzo prawdopodobne, że α .

Wydaje się, że według tego schematu wnioskowali prawie wszyscy lekarze, ponieważ ocenili szansę, że kobieta jest chora na ponad 80%.

Z takim rozwiązaniem nie zgadzają się zwolennicy systemu opisu wnioskowań opartego na twierdzeniu Bayesa, którzy przedstawiliby (*) w następujący sposób:

(**) Przesłanki:** Prawdopodobieństwo, że badanie wykáže zmiany pod warunkiem, że kobieta nie jest chora na raka, wynosi $0,096$; prawdopodobieństwo, że badanie nie wykáže zmiany pod warunkiem, że kobieta jest rzeczywiście chora, wynosi $0,2$; prawdopodobieństwo *a priori*, że dana kobieta jest chora na raka, wynosi $0,01$, badanie mammograficzne u danej kobiety wykryło zmianę;

³ Ponieważ prawdopodobieństwo wyniku fałszywie dodatniego wynosi $0,096$ (występuje w $9,6\%$ przypadków), to prawdopodobieństwo, że u osoby zdrowej badanie nie wykáže zmiany, wynosi $(1 - 0,096) = 0,904$ (czyli występuje w $90,4\%$ przypadków).

⁴ Schemat *modus tollens* ma postać: $[(\sim\alpha \Rightarrow \sim\beta) \wedge \beta] \Rightarrow \alpha$.

Wniosek: Prawdopodobieństwo, że dana kobieta jest chora na raka, wyliczamy ze wzoru Bayesa i wynosi ono 0,078⁵.

Ponieważ prawdopodobieństwo to nie jest wysokie, wnioskowanie (*) należy uznać za nieracjonalne (co oczywiście nie wyklucza podjęcia dalszych badań diagnostycznych).

Zasadnicza różnica między tymi dwoma podejściami polega na innej interpretacji terminu „prawdopodobieństwo”. Zwolennicy NHST są tzw. częstościowcami, a zwolennicy twierdzenia Bayesa — subiektywistami. Pierwsi z nich we wnioskowaniach nie są skłonni opierać się na danych dotyczących tego, jak wysokie jest prawdopodobieństwo *a priori* hipotezy, traktując je bądź jako niewiarygodne, bądź pozbawione sensu „częstościowego” i będące wyrazem tylko subiektywnego przekonania wnioskującego na temat szans na prawdziwość hipotezy. A to właśnie te dane odgrywają istotną rolę przy oszacowaniu prawdopodobieństwa, że kobieta, u której wykryto zmianę, faktycznie jest chora⁶. Zamiast więc opierać się na podejrzanych (choć dostarczonych przez twórców eksperymentu) danych, lekarze mogą brać pod uwagę ryzyko złej diagnozy (tzw. błędu pierwszego rodzaju, czyli uznania osoby chorej za zdrową) i wykorzystać heurystykę diagnostyczną (wnioskowanie szybkie i świadomie pomijające niektóre informacje). Skoro ze względu na niewiarygodne dane nie da się zastosować twierdzenia Bayesa, trzeba szacować prawdopodobieństwo inaczej.

Krytyczna ocena (formułowana przez metodologów badających wnioskowanie lekarzy) na temat racjonalności wnioskowania typu (*), a więc również (**), jest zasadna wyłącznie na gruncie podejścia Bayesowskiego. Zwolennicy testu hipotezy zerowej czy heurystyk mogą twierdzić, że to, jak wnioskowali lekarze, jest jak najbardziej rozsądne (Koehler 1996, Gigerenzer 2008) i, co więcej, w ten właśnie sposób należy stawiać diagnozy⁷. Zastosowanie argumentu z człowieka rzuca światło na to, w jaki sposób formułuje się diagnozy medyczne, ale też bardziej ogólnie — jak podejmuje się decyzje w sytuacji niepewności. Zestawienie modelu Bayesowskiego i częstościowo-heurystycznego skłania nas do przyjęcia tego drugiego.

⁵ Można to policzyć ze wzoru Bayesa:

$$P(h/e) = P(e/h) \times P(h) / [P(e/h) \times P(h) + P(e/nie-h) \times P(nie-h)],$$

gdzie $P(h/e)$ to $P(\text{jest rak/wykryto zmianę})$ i wiemy, że:

$$P(e/h) \text{ to } P(\text{wykryto zmianę/jest rak}) = 1 - 0,2 = 0,8;$$

$$P(e/nie-h) \text{ to } P(\text{wykryto zmianę/nie ma raka}) = 0,096;$$

$$P(h) \text{ to } P(\text{jest rak}) = 0,01;$$

$$P(nie-h) \text{ to } P(\text{nie ma raka}) = 0,99.$$

⁶ Gdyby np. szansa *a priori*, że kobieta jest chora, wynosiła 0,5, to prawdopodobieństwo, że jest chora, o ile badania wykazały zmianę, wynosi 0,89.

⁷ Zdaniem Gigerenzera, jeśli tylko zorientujemy się, że w izbie przyjęć jakiegoś szpitala lekarze zaczynają szacować prawdopodobieństwo *a posteriori* tego, że jesteśmy chorzy, opierając się na wzorze Bayesa, to powinniśmy natychmiast stamtąd uciekać, por. np. <<http://www.gocognitive.net/interviews/medical-decision-making>>.

PRZYKŁAD 3. TWIERDZENIE COXA

Twierdzenie sformułowane przez Richarda Coxa w 1946 roku miało stanowić teoretyczne uzasadnienie dla opisu wnioskowań zawodnych w języku rachunku prawdopodobieństwa. W literaturze można znaleźć jego różne sformułowania, ale dla naszych celów wygodnie skoncentrować się na następującym:

Dowolna formalizacja wnioskowań zawodnych (spełniająca pewne naturalne założenia) jest równoważna z rachunkiem prawdopodobieństwa.

Aby formalnie udowodnić twierdzenie Coxa, musimy przede wszystkim ustalić, na czym polegają owe naturalne założenia. Ich najlepszą formalną eksplikację można znaleźć w artykule Van Horna (2003). Autor zadaje następujące pytanie:

Jakie minimalne założenia na temat pojęcia wiarygodności, używanego potocznie przez ludzi w odniesieniu do przekonań w obrębie danej wiedzy, należy przyjąć, aby — już nieuchronnie i z matematyczną ścisłością — dało się dowieść, że wiarygodność jest pojęciem równoważnym pojęciu prawdopodobieństwa charakteryzowanemu przez aksjomaty Kołmogorowa?

Istnieje pięć takich założeń. Pierwsze cztery, choć oczywiście można nad nimi dyskutować (por. Shafer 1976), mają postać pewnych rozstrzygnięć normatywnych. Dotyczą ogólnych własności systemu, w którym opisujemy pojęcie wiarygodności:

1. Wiarygodność A^8 na podstawie X (symbolicznie: $W(A/X)$) daje się wyrazić za pomocą jednej liczby rzeczywistej (jest funkcją prowadzącą ze zbioru przekonań w zbiór liczb rzeczywistych).
2. Znaczenie spójników logicznych jest klasyczne.
3. Zbiór wartości, które przyjmuje funkcja wiarygodności, jest gęsty.
4. Wiarygodności zdań niezależnych nie wpływają na siebie.

Ze względu na interesujący nas problem ważne jest założenie ostatnie:

5. Wiarygodność koniunkcji zdań A i B (na podstawie niesprzecznego X), jest funkcją wiarygodności A na podstawie B i X oraz wiarygodności B na podstawie X , czyli:

$$W(A \wedge B/X) = F(W(A/B, X), W(B/X)).$$

Zauważmy, że z czysto teoretycznego punktu widzenia, żeby ustalić, jaka jest wiarygodność koniunkcji A i B na podstawie X , możemy ogólnie skorzystać z wartości różnych funkcji wiarygodności: $W(A/X)$, $W(B/X)$, $W(A/B, X)$, $W(B/A, X)$ i ich kombinacji (jest ich 15). Tylko one mają wpływ na to, jaka jest wiarygodność koniunkcji zdań A i B , a więc wiarygodność koniunkcji jest jakąś ich funkcją. Zasto-

⁸ Wszędzie dalej za pomocą A , B będziemy oznaczać zdania, a za pomocą X — zbiór zdań utożsamiany z wiedzą podmiotu wnioskującego.

sujemy najpierw wszystkie rozsądne redukcje tego piętnastoelementowego zbioru. Przede wszystkim wyeliminujemy ze względu na symetryczność koniunkcji 6 kombinacji, następnie ze względu na różne niepożądane konsekwencje formalne dalszych 5⁹. Zostają nam 4 możliwości:

- I. $W(A \wedge B/X) = F(W(A/B, X), W(B/X))$
- II. $W(A \wedge B/X) = F(W(A/B, X), W(B/X), W(A/X))$
- III. $W(A \wedge B/X) = F(W(A/B, X), W(B/X), W(B/A, X))$
- IV. $W(A \wedge B/X) = F(W(A/B, X), W(B/X), W(A/X), W(B/A, X))$

Nie istnieje żaden formalny powód, żeby którąś z nich uznać za właściwą.

Przytaczany przez Van Horna argument za wyborem (I) opiera się na założeniu, że ta kombinacja jest najprostszą i tak właśnie szacowałby wiarygodność $A \wedge B$ na podstawie X *człowiek*¹⁰. Najpierw sprawdziliby, czy w sytuacji, gdy zajdzie B, zaszło również A. Jeśli okaże się, że A nie zaszło, to nie ma co dalej sprawdzać: wiemy, że koniunkcja jest fałszywa. Jeśli natomiast przy zajściu B zaszło A, to trzeba jeszcze sprawdzić, czy faktycznie zaszło B. Oto przykład, który może przybliżyć te abstrakcyjne rozważania:

Mamy ocenić, czy wiarygodne jest to, że pewien biegacz przebiegnie z punktu x do punktu y (oznaczymy to stwierdzenie przez B) i z powrotem z punktu y do punktu x (oznaczymy to stwierdzenie przez A). Aby zastanawiać się nad wiarygodnością stwierdzenia o postaci $A \wedge B$, a więc że biegacz dobiegł do y i wrócił, musimy najpierw zbadać wiarygodność stwierdzenia, że dobiegł do y (czyli B). Jeśli tak się stało, to badamy następnie, jaka jest wiarygodność, że wrócił (a więc po dobiegnięciu do y ma siły wrócić do x); por. analogiczne wnioskowanie w (Jaynes 1988: 4).

Według autora tak właśnie rozumie wiarygodność koniunkcji *człowiek* (choć żadnych badań na ten temat nie przeprowadzono).

⁹ Na przykład, posługując się argumentem z symetrii koniunkcji, należy odrzucić ewentualność, że $W(A \wedge B/X) = F(W(A))$ (czyli wiarygodność koniunkcji jest funkcją wiarygodności tylko jej pierwszego członu), ponieważ ze względu na to, że $W(A \wedge B/X) = W(B \wedge A/X)$, mielibyśmy $W(A) = W(B)$, co nie musi być prawdą. Z kolei z uwagi na niepożądane konsekwencje należy odrzucić np. to, że $W(A \wedge B/X) = F(W(A/X))$. Załóżmy bowiem, że A jest zdaniem tautologicznym, a B dowolnym zdaniem atomowym. Ponieważ w takim wypadku $W(A \wedge B/X) = W(B/X)$, a wiarygodność tautologii A ma wartość stałą, oznaczałoby to, że wiarygodność B jest również stała, co jest sprzeczne z tym, że B jest przypadkowym zdaniem atomowym. Szczegółowe analizy pozostałych przypadków można znaleźć w (Van Horn 2003: 13).

¹⁰ „At this point we have to admit that there is no compelling reason for choosing any particular one of the four remaining candidates. However, (I) seems intuitively appealing to many people (it has not engendered any controversy of which we are aware)” (Van Horn 2003: 14).

Jest to zastosowanie kryterium deskryptywnego — argumentu z człowieka. Bez wyboru uzasadnionego w ten sposób (I) nie da się udowodnić twierdzenia Coxa.

PRZYKŁAD 4. PROBLEM MONTY'EGO HALLA

Gracz w teleturnieju ma wybrać jedną z trzech kopert A, B lub C. W jednej z nich jest cenna nagroda. Tylko prowadzący teleturniej wie, w której kopercie jest nagroda. Gracz wskazuje jedną kopertę (załóżmy, że to A). Ponieważ nie ma żadnych informacji, wybiera ją losowo. Prowadzący otwiera teraz jedną z pozostałych dwóch kopert — tę, o której wie, że jest pusta (załóżmy, że to koperta B), a następnie pyta gracza: „Pozostajesz przy swoim wyborze koperty A czy zmieniasz na kopertę C?”.

Co ma zrobić gracz? Wydaje się, że nie ma znaczenia, czy gracz pozostanie przy swoim pierwotnym wyborze, czy nie — sytuacja wygląda na całkowicie symetryczną. Gdyby strategia zmiany była dobra i na początku gracz wybrał kopertę C, to teraz musiałby zmienić ją na A. Zamiana nie ma więc sensu.

Tak wnioskuje zdecydowana większość ludzi postawionych przed tym problemem (według badań Granberga i Browna (1995) jest ich 87%). Co ciekawe, znajdują się wśród nich osoby z wykształceniem formalnym. Okazuje się jednak, że zmiana decyzji jest sensowna. Jeśli gracz otworzy kopertę A, którą wybrał na początku, prawdopodobieństwo wygranej wynosi $1/3$. Natomiast jeśli zmieni kopertę na C, prawdopodobieństwo wygranej wzrośnie do $2/3$. Zmiana dwukrotnie podnosi jego szansę na wygraną. To, że tak jest, najłatwiej ustalić, przeprowadzając zwykły eksperyment: sprawdzając, jak często wygrywamy, stosując strategię pozostawiania przy pierwotnym wyborze, a jak często, stosując strategię zmiany koperty. Można powiedzieć, że w tym wypadku sama rzeczywistość (niezależnie od jakichkolwiek interpretacji występujących we wnioskowaniu pojęć) daje nam jednoznaczne rozstrzygnięcie problemu: wniosek *nie warto zmieniać koperty* jest nieracjonalny.

Problem Monty'ego Halla traktowany jest jako jeden z wielu argumentów na rzecz tezy, że ludzie nie radzą sobie dobrze z problemami, które wymagają oszacowania prawdopodobieństwa wystąpienia danego zdarzenia.

3. ANALIZA WNIOSKÓW Z PRZYKŁADÓW

W przykładach 1-3 argument z człowieka pojawiał się na różnych poziomach. W przykładzie z Lindą wskazywał na to, że wnioskowanie o pozornie jednoznacznej postaci można rozumieć na wiele różnych sposobów, a przy jego ocenie odgrywają rolę niuanse niedoceniane przez formułujących problem. Z jednej strony świadczy to o tym, że ludzie analizują sytuacje, które są zdecydowanie bardziej złożone, niż to zaplanowali eksperymetatorzy. W rezultacie przypisują tym samym wyrażeniom różne znaczenie w zależności od — zdawałoby się niewielkiej — zmiany kontekstu. Z drugiej strony niepokojący jest fakt, że nie robią tego konsekwentnie. Moro (2009) przedstawił cały szereg pomysłowych eksperymentów, które przekonująco pokazują, że nie

da się utrzymać tezy o niewłaściwym rozumieniu: ludzie w analogicznych kontekstach dobrze rozumieją, kiedy prawdziwa jest koniunkcja, używają właściwie pojęcia prawdopodobieństwa i widzą różnicę między hipotezą H1 i hipotezą: $H1 \wedge$ Linda nie jest aktywną feministką¹¹. W tym wypadku argument z człowieka ma więc zastosowanie w bardzo konkretnych okolicznościach i skłania nas do przyjęcia określonych interpretacji pewnych szczególnych terminów. Dlatego można powiedzieć, że ma zasięg lokalny.

W przykładzie dotyczącym diagnozy, aby zachować zgodność z wnioskiem sformułowanym przez większość lekarzy, należy nie tyle posłużyć się wzorem Bayesa, ile rozszerzyć na warunki niepewności wnioskowanie o strukturze *modus tollens*. Jest to heurystyka w tym sensie, że nie bierze się tu pod uwagę całej dostępnej wiedzy (w szczególności tego, jakie jest podane przez eksperymentatorów prawdopodobieństwo *a priori*, że dana kobieta jest chora). Dzięki temu wniosek można wyciągnąć szybko: nie trzeba dokonywać żadnych obliczeń, lecz wystarczy posłużyć się diagnostycznością badań. Z punktu widzenia praktyki lekarskiej jest to własność pożądana i racjonalna. W tym wypadku zastosowanie argumentu z człowieka uzasadnia mniej lub bardziej świadomą rezygnację z modelu wnioskowania teoretycznie optymalnego (czyli uwzględniającego całość dostępnych informacji) na rzecz modelu prostszego i wygodniejszego w praktyce. To, że w wyniku wnioskowań (***) i (****) otrzymujemy tak rozbieżne wyniki, świadczy o tym, jak istotny jest wybór systemu, w którym formalizujemy dane wnioskowanie, i w związku z tym, jak znaczące konsekwencje może dać zastosowanie argumentu z człowieka¹². Ma on więc w tym wypadku zasięg dużo szerszy, wpisując się w spór prowadzony przez zwolenników uniwersalnej stosowalności twierdzenia Bayesa ze zwolennikami częstościowej interpretacji pojęcia prawdopodobieństwa (prowadzącej do ograniczenia stosowalności twierdzenia Bayesa).

W przykładzie 3 argument z (hipotetycznego) człowieka służy do uzasadnienia przyjęcia określonego sposobu opisu wnioskowania: jeśli chcemy otrzymywać jak

¹¹ Badania Mora wskazują przy okazji na jeszcze jeden ciekawy fakt: subtelne różnice przy formułowaniu problemu Lindy w sposób istotny wpływają na odsetek ludzi popełniających błąd koniunkcji. Na przykład sformułowanie pytania tak, aby dotyczyło cech Lindy w przyszłości (co jest bardziej prawdopodobne: czy to, że Linda za 10 lat będzie pracownikiem banku, czy to, że za 10 lat będzie pracownikiem banku i aktywną działaczką ruchu feministycznego) spowodowało, że większość badanych podała właściwą (tzn. zgodną z rachunkiem prawdopodobieństwa) odpowiedź.

¹² Oceniając to wnioskowanie i zastanawiając się, czy *rzeczywiście* jest tak, że kobieta jest prawie na pewno zdrowa, czy prawie na pewno chora, musimy pamiętać, że odgrywają tu rolę dwa czynniki:

- i. ustalenie, jakie jest *rzeczywiste* prawdopodobieństwo *a priori*, że kobieta jest chora;
- ii. rozstrzygnięcie, czy wniosek dotyczy wyłącznie matematycznego prawdopodobieństwa, czy wartości użyteczności decyzji.

Popelnienie błędu i uznanie osoby chorej za zdrową jest bardzo kosztowne, tak więc maksymalizując wartość użyteczności naszej decyzji, nawet liczbowo niskie prawdopodobieństwo uznajemy za „użytecznościowo” wysokie i traktujemy kobietę tak, *jakby* prawie na pewno była chora.

najbardziej wiarygodnie (w zdefiniowanym sensie tego słowa) wnioski, to powinniśmy posługiwać się rachunkiem prawdopodobieństwa. Bez przyjęcia wzoru (I), uzasadnionego przez odwołanie się do tego, jak wnioskuje ludzie, nie możemy w ogóle takiej implikacji udowodnić. Zauważmy, że na takie zastosowanie tego argumentu powinni zgodzić się zarówno Bayesowcy, jak i zwolennicy NHST (częstościowcy). Ich spór nie dotyczy bowiem tego, czy stosować rachunek prawdopodobieństwa, lecz tego, jaki jest dokładnie zakres jego stosowalności.

Argumentu z człowieka używa się więc do realizacji sprzecznych celów: z jednej strony stosujemy go, żeby uzasadnić stosowalność rachunku prawdopodobieństwa (przykład 3), a z drugiej, aby pokazać, że czasami lepiej nie używać rachunku prawdopodobieństwa (przykłady 1 i 2). W przykładzie z Lindą argument ten ma wskazywać na subtelności istotne przy wnioskowaniu, a w przykładzie diagnozy — na konieczność jego upraszczania. Bayesowcy powinni go zaakceptować w odniesieniu do przykładu 3 i odrzucić w odniesieniu do przykładu 2. Nie da się więc uznać argumentu z człowieka za trafny *en bloc*.

Nie możemy również zapominać o wymowie przykładu 4. Pokazuje on, że gdybyśmy chcieli stosować argument z człowieka bezkrytycznie i uniwersalnie, to po prostu popełnimy błąd: uznamy za racjonalne wnioskowanie, które niezależnie od interpretacji takie nie jest. Zauważmy jednak, że ten ostatni przykład jest pod pewnymi względami wyjątkowy, ponieważ istnieje w nim proste kryterium rozstrzygnięcia, kiedy decyzja jest racjonalna. Tak jest wtedy, gdy decyzja częściej prowadzi do wybrania koperty, w której jest nagroda, niż do wybrania koperty, w której nagrody nie ma. Mamy tu również do czynienia z dobrze scharakteryzowanym pojęciem sukcesu (zdobyć nagrodę) i jednoznacznie określonymi cechami środowiska (trzy nieodróżnialne koperty, w dokładnie jednej z nich jest koperta itd.).

Wcześniejsze przykłady niestety nie mają tych własności. W przykładzie z Lindą wprawdzie intencją eksperymentatorów było, by wnioskiem racjonalnym był wniosek bardziej prawdopodobny, niekoniecznie jednak tak rozumieli to badani. Co uznamy za sukces (za racjonalny wniosek) w przykładzie dotyczącym diagnozy medycznej? Spadek liczby zgonów na raka, ograniczenie kosztów niepotrzebnie wykonywanych badań czy też dokładne zmierzenie, jakie jest prawdopodobieństwo, że pewna konkretna kobieta jest chora? Nie wiemy również, jakie dokładnie są cechy środowiska (na ile rzetelnie obliczona jest diagnostyczność badań) i czy w przypadku kobiety zgłaszającej się na badania mammograficzne można faktycznie ustalić prawdopodobieństwo *a priori* tego, że jest chora: wiąże się to z dokładnym ustaleniem jej grupy odniesienia (czy są to kobiety w danym wieku, w danym wieku i danej rasy, w danym wieku, danej rasy i określonej historii innych chorób itd.¹³). To sprawia, że wprawdzie trudno wykazać, iż wnioskujący zachowali się nieracjonalnie, ale też trudno odróżnić ich decyzję od zwykłego błędu (który są skłonni popełniać w przykładzie 4).

¹³ Niejednoznaczność wyboru grupy odniesienia jest jednym z podstawowych problemów wskazywanych przez zwolenników częstościowej interpretacji pojęcia prawdopodobieństwa.

4. KRYTYKA ARGUMENTU Z CZŁOWIEKA

Argument z człowieka ma oczywiste słabości metodologiczne. Są one związane z:

I. Problemem motywacji badanych. Badani wypełniają ankiety z zadaniami, mając poczucie, że są to tylko testy. Ich motywacja nie jest zbyt wysoka, co przekłada się na ograniczenie wysiłku wkładanego w znalezienie optymalnego rozwiązania.

II. Problemem sztuczności sytuacji. Inaczej zachowuje się człowiek podczas badania, a inaczej w realnej sytuacji, gdy od tego, jaki wniosek wyciągnie, faktycznie coś realnego zależy. Testowanie ludzi w rzeczywistych sytuacjach z kolei powodowałoby, że trudno byłoby ustalić takie same warunki eksperymentu dla wszystkich i mieć dostatecznie reprezentatywną i liczną próbkę.

III. Problemem właściwej identyfikacji celu. Wnioskującemu zależy czasami nie tylko na prawdziwości wniosku, lecz także na informacji, której on dostarcza, na uniknięciu błędu, który chociaż nawet jest mało prawdopodobny, to ma potencjalnie istotne konsekwencje, czy wreszcie na spełnieniu oczekiwań testera.

IV. Problemem oddzielenia kryteriów przedmiotowych od metakryteriów, ze względu na które dokonywany jest wybór. Wnioskujący mogą np. bać się kompromitacji związanej z podaniem fałszywego wniosku, mogą nie chcieć wyróżnić się wśród innych badanych itp.

V. Problemem braku wyraźnego odgraniczenia środowiska, w którym dokonywane jest wnioskowanie. Zarówno czynniki wpływające na wniosek, jak i analiza konsekwencji poszczególnych wniosków nie są dobrze zdefiniowane: nie jest podany ani horyzont skutków, ani powiązań przesłanek z wiedzą podmiotu wnioskującego.

Są to oczywiście problemy, przed którymi stają wszyscy prowadzący badania empiryczne, ale zwróćmy uwagę, że nabierają one w tym kontekście szczególnego znaczenia, ponieważ na ich podstawie miałyby być odrzucane (tak jak w przykładach 1 i 2) lub przyjmowane (jak w przykładzie 3) określone modele rozumowań. Jeśli chcemy na podstawie tego, *jak ludzie wnioskują*, stwierdzić, *jaki jest racjonalny sposób wnioskowania*, to ustalenie, *jak faktycznie w rzeczywistych warunkach* ludzie wnioskują, jest bardzo istotne.

Na oddzielne omówienie zasługuje problem reprezentatywności próbki. W stwierdzeniu (niezależnie od tego, czy formułujemy je jako tezę, czy jako założenie), że człowiek wnioskuje racjonalnie, wyjaśnienia wymaga to, czy mamy na myśli wszystkich ludzi, czy tylko jakiś ich wybrany podzbiór. Ten problem dotyczy więc zarówno wyboru próbki w przykładach 1-4, jak i trzech sposobów uzasadnienia argumentu z człowieka.

W teście Turinga tester porównuje zachowanie jednostki ze swoją wizją zachowania człowieka. Skoro jednostka ma być uznana za inteligentną, to ma nie tyle naśladować *wszystkie* zachowania człowieka, ile zachowania *pewnego typu*. Przejawia

się to z jednej strony tym, że pytania zadawane przez testera mają swoją specyfikę: dotyczą kwestii, których rozstrzygnięcie wymaga zaangażowania w sposób istotny inteligencji (a niekoniecznie innych charakterystycznych cech człowieka, takich jak np. empatia czy zdolności werbalne¹⁴). Z drugiej strony, tester po uzyskaniu odpowiedzi porównuje ją ze swoim przekonaniem, jaka odpowiedź na dane pytanie jest racjonalna. Innymi słowy, aby przejść test, testowana jednostka musi być przynajmniej (czy może — dokładnie) tak inteligentna, jak tester. Tester nie jest jednak dowolnym człowiekiem, lecz — ze względu na samo sformułowanie problemu — osobą jakoś wyróżnioną. Tak więc, jeśli w teście Turinga przyjmujemy założenie, że człowiek jest wzorcem racjonalności, to mowa jest o wyróżnionych ludziach — takich, którzy nadają się na testerów.

Sytuacja przedstawia się bardzo podobnie, gdy uzasadniamy racjonalność człowieka tym, że jest to warunek konieczny możliwości skonstruowania teorii znaczenia. Zauważmy, że posługując się Zasadą Życzliwości, można w sposób jednoznaczny (lub przynajmniej niesprzeczny) ustalić znaczenia terminów należących do języka osoby, której poglądy są spójne. Jeśli np. Davidson miałby ustalić znaczenie terminu „szansa na wylosowanie koperty zawierającej nagrodę” na podstawie zachowania ludzi biorących udział w badaniu opisanym w (Granberg, Brown 1995), to wydaje się, że taka teoria zawierałaby tezy wzajemnie się wykluczające. Zakresem stosowności Zasady Życzliwości nie jest więc dowolny człowiek, lecz taki, który skądinąd zachowuje się niesprzecznie i np. w przykładzie 4 nie popełni błędu (przypomnijmy, że takich ludzi jest mniej niż 20%). Tak więc z punktu widzenia Davidsona wzorem racjonalności są ludzie, do których da się zastosować Zasadę Życzliwości z korzyścią dla teorii znaczenia.

Uzasadnienie argumentu z człowieka w duchu testu Turinga lub Davidsonowskiej Zasady Życzliwości nie opiera się na realnych badaniach, lecz na postulatach metodologicznych. Przyjmujemy, że człowiek zachowuje się racjonalnie, ponieważ jest to założenie wygodne dla nas ze względu na inne realizowane cele. Można powiedzieć, że nie mamy tu do czynienia z realnym człowiekiem, lecz z pewną jego idealizacją: idealnym testerem czy konsekwentnym użytkownikiem języka. Argumentacja Foga odwołująca się do sukcesu ewolucyjnego ma pozornie bardziej empiryczny charakter. Nie rozstrzyga jednak, czy sukces ewolucyjny jest zasługą wszystkich ludzi, czy też zawdzięczmy go tylko niektórym (mówiąc w pewnym uproszczeniu — tym, któ-

¹⁴ Pierwowzorem testu Turinga była zabawa, w której tester miał rozpoznać, nie widząc, z kim rozmawia, czy ma do czynienia z kobietą, czy z mężczyzną. Można się domyślać, że zadawał pytania inne niż w teście, który nas interesuje. Specyficznym testem Turinga jest tzw. konwersacyjny test Turinga. Uczestnicząca w nim jednostka ma prowadzić rozmowę z testerem na dowolne tematy. Maszyna przejdzie ten test, jeśli tester uzna ją za człowieka (zdobędzie wtedy tzw. nagrodę Loebnera, do tej pory jeszcze nieprzyznaną). Pytania zadawane w teście konwersacyjnym badają nie tyle (a przynajmniej nie przede wszystkim) inteligencję, ile umiejętności werbalne i zapewne jakiś rodzaj kompetencji społecznych. Stąd też muszą się różnić od pytań z testu, o którym mowa w tym artykule i który pierwotnie miał na myśli Turing.

rzy w przykładzie 4 dokonają zamiany koperty). Niewątpliwie w każdym społeczeństwie można wskazać ludzi bardziej i mniej twórczych (inteligentnych, sprawnych decyzyjnie) i tym samym w większym stopniu odpowiedzialnych za sukces ewolucyjny. Wydaje się więc, że w stwierdzeniu Foxa należy również użyć kwantyfikatora szczegółowego: wytłumaczeniem tego, że ludzie jako gatunek osiągnęli sukces, jest to, że *niektórzy* z nich działali racjonalnie¹⁵.

Skoro jednak zgodzimy się, że postulując racjonalność człowieka, mamy na myśli nie wszystkich ludzi, lecz jakiś ich wybrany podzbiór, kolejnym problemem, przed którym stają zwolennicy argumentu z człowieka, jest wybór odpowiedniej próbki, której zachowanie będzie dla nas wiążące. Czy za racjonalne należy uznać wnioski wyciągane przez *badanych* w przykładach 1 i 2, czy raczej te wnioski, które spodziewali się otrzymać *eksperymentatorzy*? Jeśli teraz nieopatrnie zażądamy, by po prostu dobierać próbkę składającą się z osób dostatecznie racjonalnych, to znowu wylamy się w nieskończony regres, ponieważ musimy podać kryteria bycia osobą racjonalną. Problem ten nie ma dobrego rozwiązania.

5. WNIOSKI

Na podstawie przedstawionych rozważań można sformułować następujące wnioski:

1) Argument z człowieka stosowany jest na różnych poziomach i służy do uzasadnienia niezgodnych ze sobą tez. Nie można go więc stosować uniwersalnie.

2) Uzasadnienia argumentu z człowieka nie odwołują się do zachowania dowolnej grupy ludzi, lecz grupy w pewien sposób wyróżnionej.

3) Zastosowanie argumentu z człowieka rodzi poważne problemy metodologiczne. W szczególności trudno podać kryterium doboru reprezentatywnej próbki, której zachowanie należy uznać za wiążące przy rozstrzygnięciu problemu racjonalności wnioskowania.

Jeśli zgodzimy się z takimi wnioskami, będzie to oznaczać, że argument z człowieka może być wykorzystywany *co najwyżej* w kontekście odkrycia, a nie uzasadnienia. Gdy zachodzi duża rozbieżność między praktyką a modelem zakładanym do opisu i oceny wnioskowania, pozwala on postawić dodatkowe pytanie, czy model jest rzeczywiście dobrze dobrany do środowiska, w którym wnioskowanie jest przeprowadzane. Tak traktuje argument z człowieka Gigerenzer (Goldstein, Gigerenzer 2002, Gigerenzer, Brighton 2009, Gigerenzer, Hoffrage 1995), formułując program tzw. ekologicznej racjonalności. Według Gigerenzera dotychczasowe próby opisu

¹⁵ Zauważmy, że przy takiej interpretacji teza, zgodnie z którą heurystyki są dobrymi sposobami wnioskowania, ponieważ tak wnioskują większość ludzi, przestaje być uzasadniona ewolucyjnym sukcesem człowieka

wnioskowań zawodnych i kryteriów ich racjonalności nie uwzględniały specyfiki środowiska, w którym wnioskowanie jest przeprowadzane. Co istotne, przez środowisko rozumie on nie tylko zewnętrzny wobec wnioskującego świat, rządzący się określonymi prawami, lecz także sprzężone z nimi cechy podmiotu. Podmiot nie jest idealnym demonem Laplace'a, ale podlega różnym ograniczeniom, np. obliczeniowym lub czasowym. Zamiast posługiwać się teoretycznie optymalnym, ale skomplikowanym modelem jest skazany na heurystyki. Heurystyka nie jest tu wynikiem zaniedbania, lenistwa czy braku wiedzy, lecz jedynym *dostępnym praktycznie* rozwiązaniem. Takie podejście pozwala sformułować pytanie, na które odpowiedź będzie miała już charakter normatywny, a nie deskryptywny:

Niech dane będzie pewne wnioskowanie zawodne. Jakie cechy musiałoby mieć środowisko, żeby to wnioskowanie można było uznać w nim za racjonalne?

Jednoznaczna ocena wnioskowania w przykładzie 4 wynika stąd, że cechy środowiska są w nim ustalone jednoznacznie (jest to środowisko bardzo proste). Jeśli natomiast zadamy takie pytanie odnośnie do wnioskowania (**) z przykładu 2, to musimy przyznać, że środowisko, w którym należy stwierdzić wysokie prawdopodobieństwo tego, że kobieta jest chora na raka, ma następujące cechy:

- lekarzowi nie są dostępne rzetelne dane na temat prawdopodobieństwa *a priori* tego, że kobieta jest chora;
- lekarz nie ma możliwości zdobycia takich danych w rozsądnym czasie;
- lekarz musi podjąć natychmiast decyzję;
- celem lekarza jest uniknięcie błędu polegającego na zakwalifikowaniu osoby chorej jako zdrowej;
- koszty podjęcia dodatkowych badań diagnostycznych nie są brane pod uwagę.

Jeżeli rzeczywiście środowisko ma takie cechy, to decyzja lekarza jest racjonalna. Jeśli natomiast środowisko opisujemy tak:

- lekarzowi znane jest prawdopodobieństwo *a priori* tego, że kobieta jest chora, i uważa, że te dane są rzetelne;
- lekarz umie zastosować twierdzenie Bayesa;
- celem lekarza jest podanie wyłącznie prawdopodobieństwa *a posteriori*, że kobieta jest chora,

to wnioskowanie (**) nie jest racjonalne, niezależnie od tego, ilu lekarzy będzie innego zdania. Argument z człowieka nie działa więc jako uzasadnienie.

Oczywiście, powstaje problem, jak ocenimy wnioskowanie w środowisku, w którym lekarzowi jest znane prawdopodobieństwo *a priori*, jego celem jest podanie prawdopodobieństwa *a posteriori*, ale nie umie zastosować twierdzenia Bayesa, za to mylnie utożsamia prawdopodobieństwo tego, że *jeśli kobieta jest zdrowa, to badanie nie wykáže zmiany*, z prawdopodobieństwem, że *jeśli badanie wykazało zmianę, to*

jest chora (czyli popełnia tzw. błąd odwrócenia). Jeśli jego słabe umiejętności są traktowane jako cechy środowiska, to wnioskowanie może być uznane za racjonalne¹⁶. Aby uniknąć takich niezamierzonych konsekwencji, należałoby jakoś odróżnić „racjonalne” cechy podmiotu od cech „nieracjonalnych”, co wydaje się nas wikłać w nieskończony regres definicyjny. Jest to niewątpliwa słabość stanowiska Gigerenzera i całego nurtu wykorzystującego argument z człowieka w kontekście uzasadnienia, jego analiza wykracza jednak poza ramy tego artykułu.

Na pewno możemy stwierdzić, że argument ten należy stosować z dużą ostrożnością, mając świadomość jego słabości i ograniczeń.

BIBLIOGRAFIA

- Atkinson D. (2012), *Confirmation and Justification. A Commentary on Shogenji's Measure*, „Synthese” 184(1), 49-61.
- Cox R. T. (1946), *Probability, Frequency and Reasonable Expectation*, „American Journal of Physics” 14(1), 1-13.
- Davidson D. (1982), *Two Paradoxes of Irrationality* [w:] *Philosophical Essays on Freud*, R. Wollheim, J. Hopkins (red.), Cambridge (MA): Cambridge University Press, 289-305.
- Fox R. (1992), *Prejudice and the Unfinished Mind. A New Look at an Old Failing*, „Psychological Inquiry” 3(2), 137-152.
- Gigerenzer G., Hoffrage U. (1995), *How to Improve Bayesian Reasoning without Instruction. Frequency Formats*, „Psychological Review” 102(4), 684-704.
- Goldstein D. G., Gigerenzer G. (2002), *Models of Ecological Rationality. The Recognition Heuristic*, „Psychological Review” 109(1), 75-90.
- Gigerenzer G., Brighton H. (2009), *Homo Heuristicus. Why Biased Minds Make Better Inferences*, „Topics in Cognitive Science” 1(1), 107-143.
- Gigerenzer G. (2008), *Rationality for Mortals. How People Cope with Uncertainty*, New York (NY): Oxford University Press.
- Granberg D., Brown T. A. (1995), *The Monty Hall Dilemma*, „Personality and Social Psychology Bulletin” 21(7), 711-723.
- Jaynes E. T. (1988), *How Does the Brain Do Plausible Reasoning?* [w:] *Maximum-Entropy and Bayesian Methods in Science and Engineering 1*, G. J. Erickson, C. R. Smith (red.), Dordrecht: Kluwer, 1-24.
- Koehler J. (1996), *The Base Rate Fallacy Reconsidered. Descriptive, Normative, and Methodological Challenges*, „Behavioral and Brain Sciences” 19(1), 1-53.
- Moro R. (2009), *On the Nature of the Conjunction Fallacy*, „Synthese” 171(1), 1-24.
- Schupbach J. N. (2009), *Is the Conjunction Fallacy Tied to Probabilistic Confirmation?*, „Synthese” 184(1), 13-27.
- Shafer G. (1976), *A Mathematical Theory of Evidence*, Princeton (NJ): Princeton University Press.

¹⁶ Idąc tym tropem, można by też uznać za racjonalne wnioskowanie z przykładu 4, zakładając, że cechą podmiotu jest to, że nie umie obliczyć prawdopodobieństwa warunkowego. Rozumując w ten sposób, należałoby wszystkie wnioskowania uznawać za racjonalne, co spowodowałoby, że pojęcie to — wbrew intencjom — nie miałyby sensu wartościującego.

- Shogenji J. (2012), *The Degree of Epistemic Justification and the Conjunction Fallacy*, „Synthese” 184(1), 29-48.
- Turing A. (1995), *Maszyna licząca a inteligencja* [w:] *Filozofia umysłu*, B. Chwedeńczuk (red.), Warszawa: Alethea, 271-300 [*Computing Machinery and Intelligence*, „Mind” 59(236), 433-460].
- Tversky A., Kahneman D. (1971), *Belief in the Law of Small Numbers*, „Psychological Bulletin” 76(2), 105-110.
- Van Horn K. S. (2003), *Constructing a Logic of Plausible Inference. A Guide to Cox's Theorem*, „International Journal of Approximate Reasoning” 34(1), 3-24.
- Westover M. B., Westover K. D., Bianchi M. T. (2011), *Significance Testing as Perverse Probabilistic Reasoning*, „BMC Medicine” 9:20, DOI: 10.1186/1741-7015-9-20.