

ADAM ELGA

PRZEKONANIA DOTYCZĄCE WŁASNEGO POŁOŻENIA W ŚWIECIE A PROBLEM ŚPIĄCEJ KRÓLEWNY*

Poza niepewnością co do tego, jaki jest świat, można być także niepewnym swojego przestrzennego lub czasowego położenia w świecie. Zamierzam postawić problem wynikający z połączenia tych dwóch rodzajów niepewności, a następnie rozwiązać go i wyciągnąć dwie lekcje z tego rozwiązania.

1. *Problem śpiącej królowny*¹. Grupa badaczy zamierza cię uśpić. Pogrążysz się we śnie na dwa dni, w czasie których obudzą cię na krótko raz albo dwa razy, w zależności od rzutu nieoszukaną monetą (reszka: zostaniesz obudzony raz; orzeł: dwa razy). Po każdej pobudce zostaniesz ponownie uśpiony za pomocą leku, który sprawi, że nie zachowasz żadnych wspomnień dotyczących przebudzenia².

Kiedy budzą cię po raz pierwszy, do jakiego stopnia powinienes uważać, że wypadła reszka?

Prof. ADAM ELGA — profesor filozofii na Uniwersytecie Princeton, USA; obszar jego zainteresowań obejmuje teorię decyzji i gier, epistemologię, filozofię umysłu i filozofię nauki; adres do korespondencji – e-mail: ae@adamelga.net

* Oryginał: „Self-Locating Belief and the Sleeping Beauty Problem”, *Analysis* 60 (2000), issue 2: 143–147. DOI: <https://doi.org/10.1093/analys/60.2.143>. Przekład za zgodą Autora.

¹ Nazwa pochodzi od Roberta Stalnakera (który po raz pierwszy zetknął się z tego rodzaju przykładami w niepublikowanej pracy Arnolda Zuboffa). Omawiany problem pojawia się jako przykład 5 u PICCIONE 1997, który proponuje dwie różne odpowiedzi, nie rozstrzygając między nimi (1997, 12–14). AUMANN et al. 1997 odwołuje się do nieoszukanej loterii, aby przeprowadzić analizę podobnego problemu. Jego analiza, po zastosowaniu do problemu śpiącej królowny, prowadzi do tej samej odpowiedzi, której będę bronił w paragrafie 2. W odróżnieniu od AUMANN et al. 1997 nie odwołam się jednak w swoim argumencie do idei zakładów (*betting*).

² Dokładnie mówiąc, ów lek cofnie twoje przekonania do stanu, w jakim znajdowały się na początku eksperymentu. Jeśli istnienie takiego leku wydaje się zbyt wydumane, to zauważmy, że omawiany problem można sformułować, nie odwołując się do niego — ważne jest tylko to, by uśpiona osoba uważała, że eksperyment przebiegnie tak, jak go przedstawiłem.

Pierwsza odpowiedź: Oczywiście $1/2$! Od początku masz pewność, że moneta nie jest oszukana, a w związku z tym w punkcie wyjścia twój stopień pewności, że wypadnie reszka, wynosił $1/2$. Po przebudzeniu nie uzyskałeś żadnej nowej informacji (cały czas wiedziałeś, że zostaniesz obudzony). Twój stopień pewności, że wypadnie reszka, powinien więc pozostać na poziomie $1/2$.

Druga odpowiedź: Oczywiście $1/3$! Wyobraź sobie, że ten eksperyment jest wielokrotnie powtarzany. Na dłuższą metę około $1/3$ wszystkich przebudzeń byłaby wówczas „przebudzeniami reszkowymi” — to znaczy dochodziłoby do nich po wypadnięciu reszki. Po dowolnym przebudzeniu powinien więc mieć stopień pewności $1/3$, że jest to „przebudzenie reszkowe”, a w związku z tym stopień pewności $1/3$, że wypadła akurat reszka. Względędy te pozostają w mocy również w obecnych okolicznościach, tj. gdy eksperyment przeprowadza się tylko raz.

Będę argumentował, że to druga z tych odpowiedzi jest poprawna.

2. Załóżmy, że pierwsze przebudzenie ma miejsce w poniedziałek, a drugie (o ile do niego dochodzi) we wtorek. Po przebudzeniu zatem masz pewność, że znajdujesz się w jednej z trzech sytuacji:

R_1 — reszka i jest poniedziałek;

O_1 — orzeł i jest poniedziałek;

O_2 — orzeł i jest wtorek.

Zauważmy, że różnica między byciem w O_1 a byciem w O_2 nie dotyczy tego, który świat możliwy jest aktualny, lecz twojego czasowego położenia w świecie. (Przy bardziej technicznym podejściu moglibyśmy przyjąć schemat podobny do zasugerowanego przez Lewisa 1983, według którego podstawowe opcje, między którymi rozdzielony jest twój stopień pewności, nie są po prostu możliwymi światami, ale ześrodkowanymi (*centered*) możliwymi światami — to znaczy takimi możliwymi światami, z których każdy zawiera wyróżnione indywiduum i czas. W takim schemacie R_1 , O_1 , O_2 byłyby reprezentowane przez odpowiednie zbiory ześrodkowanych światów).

Niech P będzie funkcją stopnia pewności, jaki powinienes mieć po pierwszym przebudzeniu. Po pierwszym przebudzeniu masz pewność co do tego, że znajdujesz się w sytuacji R_1 wtedy i tylko wtedy, gdy wypadła reszka. Z tego zaś wynika, że wystarczy obliczyć $P(R_1)$, aby rozwiązać problem śpiącej królowy. Będę najpierw argumentował, że $P(O_1) = P(O_2)$, a następnie, że $P(R_1) = P(O_1)$.

Jeśli (po pierwszym przebudzeniu) dowiedziałbyś się, że wypadł orzeł, to wiedziałbyś dzięki temu, że znajdujesz się albo w O_1 , albo w O_2 . Skoro te dwie sytuacje są subiektywnie nieodróżnialne i skoro w obu prawdziwe są dokładnie te same sądy, to nawet bardzo restrykcyjna zasada obojętności (*principle of indifference*) wskazuje, że w obu powinieneś mieć jednakowy stopień pewności. Twój zaś stopień pewności, że jesteś w sytuacji O_1 , po tym jak dowiedziałeś się, że wypadł orzeł, powinien być równy warunkowemu stopniowi pewności $P(O_1|O_1 \text{ lub } O_2)$ — i podobnie dla O_2 . A zatem $P(O_1|O_1 \text{ lub } O_2) = P(O_2|O_1 \text{ lub } O_2)$, z czego wynika, że $P(O_1) = P(O_2)$.

Grupa badaczy ma rozstrzygnąć, czy obudzić cię raz, czy dwa razy, na podstawie rzutu nieoszukaną monetą. Mogą wywiązać się z tego zadania na dwa sposoby:

- (1) najpierw rzucić monetą, a następnie obudzić cię raz lub dwa razy w zależności od wyniku rzutu;
- (2) obudzić cię raz, a następnie rzucić monetą, by ustalić, czy budzić cię powtórnie.

(Po przebudzeniu) twój stopień pewności w sprawie tego, że wypadła reszka, powinien pozostać taki sam niezależnie od tego, której z tych dwóch metod użyją badacze. Nie tracąc nic z ogólności naszych rozważań, załóżmy więc, że używają oni — za twoją wiedzą — drugiej metody.

Otóż jeśli (po przebudzeniu) dowiedziałbyś się, że jest poniedziałek, to na tej podstawie mógłbyś wiedzieć, że jesteś albo w R_1 , albo w O_1 . Twój stopień pewności, że znajdujesz się w R_1 , byłby wówczas stopniem pewności, że w zbliżającym się rzucie nieoszukaną monetą wypadnie reszka. Nie ważne, że zostaniesz obudzony następnego dnia wtedy i tylko wtedy, gdy wypadnie orzeł — w tych okolicznościach twój stopień pewności, że wypadnie reszka, powinien wynosić $1/2$. Twój stopień pewności, że wypadnie reszka (po tym jak dowiedziałeś się, że jest poniedziałek), powinien być jednak taki sam, jak warunkowy stopień pewności $P(R_1|R_1 \text{ lub } O_1)$. A zatem $P(R_1|R_1 \text{ lub } O_1) = 1/2$, z czego wynika, że $P(R_1) = P(O_1)$.

Zbierając uzyskane rezultaty, okazuje się, że $P(R_1) = P(O_1) = P(O_2)$. Skoro te stopnie pewności sumują się do 1, to $P(R_1) = 1/3$.

3. Niech R będzie sądem, że wynikiem rzutu monetą jest reszka. Przed uśpieniem twój stopień pewności w sprawie R wynosił $1/2$. Argumentowałem, że po tym, jak zostaniesz obudzony w poniedziałek, ten stopień pewności powinien się zmienić na $1/3$. Ta zmiana przekonań jest nietypowa. Nie jest ona rezultatem uzyskania przez ciebie nowych informacji — już wcześ-

niej miałeś pewność, że zostaniesz obudzony w poniedziałek³. (Możemy nawet założyć, że już w punkcie wyjścia eksperymentu wiedziałeś, jakie dokładnie doświadczenia zmysłowe będą ci towarzyszyć po przebudzeniu w poniedziałek). Ta zmiana przekonań nie jest również wynikiem jakichś poznawczych zawirowań, do których dochodzi w międzyczasie — pamiętajmy, że lek powodujący zapomnienie zostaje podany długo po pierwszym przebudzeniu. Co zatem uzasadnia tę zmianę przekonań?

Odpowiedź jest taka, że twoja sytuacja uległa zmianie: z sytuacji, w której uznawałeś swoje położenie czasowe za nieistotne dla prawdziwości R , przeszedłeś do sytuacji, w której uznajesz je za istotne dla prawdziwości R . Załóżmy, na przykład, że na początku eksperymentu nie miałeś pewności, czy jest godzina 1:01 czy 1:02. W tym czasie uznawałeś swoje miejsce w czasie za nieistotne dla prawdziwości R : twój stopień pewności w sprawie R , zakładając, że jest 1:01, wynosił $1/2$, podobnie jak twój stopień pewności w sprawie R , zakładając, że jest 1:02.

Z drugiej strony (jeśli racjonalnie uaktualniasz swoje przekonania), kiedy zostajesz obudzony w poniedziałek, uznajesz swoje obecne miejsce w czasie za istotne dla prawdziwości R : twój stopień pewności co do R , zakładając, że jest poniedziałek, wynosi $1/2$, ale twój stopień pewności co do R , zakładając, że jest wtorek, wynosi 0. W poniedziałek twój bezwarunkowy stopień pewności w sprawie R różni się od $1/2$, ponieważ jest to średnia ważona tych dwóch warunkowych stopni pewności — tj. średnia ważona $1/2$ i 0.

Nic dziwnego, że sposób, w jaki podmiot uznaje swoje miejsce w czasie za istotne dla prawdziwości pewnych sądów, może z czasem ulegać zmianie. Zadziwiający jest natomiast fakt, że ten rodzaj zmiany może się przydarzyć doskonale racjonalnemu podmiotowi w czasie, w którym nie uzyskuje on żadnych nowych informacji ani nie podlega żadnym zawirowaniom poznawczym — i to jest nasza pierwsza lekcja.

Na początku eksperymentu twój stopień pewności w sprawie R wynosił $1/2$. Ale byłeś także pewien, że po przebudzeniu w poniedziałek twój stopień pew-

³ (W moim rozumieniu) powiedzieć, że podmiot uzyskał nową informację, to powiedzieć, że uzyskał nowe racje uzasadniające (*evidence*), pozwalające wykluczyć możliwe światy niewykluczone przez dotychczas posiadane racje uzasadniające. Inaczej mówiąc, podmiot uzyskuje nową informację, kiedy dowiaduje się o prawdziwości sądu wyrażalnego przez zdanie wieczne jakiegoś odpowiednio bogatego języka (QUINE 1960, 191).

⁴ Powiedzieć, że podmiot uznaje swoją miejsce w czasie za istotne dla prawdziwości danego sądu, to powiedzieć, że jest taka chwila t , w której przekonania podmiotu są zgodne z jego umiejscowieniem w t , a jego stopień pewności co do danego sądu, zakładając, że znajduje się w t , różni się od jego nieuwarunkowanego stopnia pewności co do danego sądu.

ności co do R będzie wynosił $1/3$ — mimo że byłeś pewien, iż w międzyczasie nie uzyskasz żadnych nowych informacji ani nie będziesz podlegać żadnym zawirowaniom poznawczym. W związku z tym przykład śpiącej królowny stanowi nowy rodzaj kontrprzykładu dla „zasady równowagi” (*Reflection Principle*) Basa van Fraassena (1984, 244; 1995, 19). Albowiem nawet opatrzona wieloma zastrzeżeniami wersja tej zasady pociąga następujące twierdzenie:

Dowolny podmiot, który jest pewny, że jutro będzie miał stopień pewności x w sprawie sądu R (i nie uzyska nowych informacji ani nie będzie w międzyczasie podlegał zawirowaniom poznawczym), powinien *teraz* mieć stopień pewności x w sprawie R .⁵

David Lewis, zapytany pewnego razu, „co stanie się z teorią decyzji, jeśli [zamienimy przestrzeń możliwych światów na przestrzeń ześrodkowanych możliwych światów]?”, odpowiedział: „Niewiele” (LEWIS 1983, 149). Drugą lekcją, jaka płynie z rozważań nad problemem śpiącej królowny, jest to, że jednak coś się stanie — pojawi się nowe pytanie, jak racjonalny podmiot powinien aktualizować swoje przekonania z biegiem czasu⁶.

Z języka angielskiego przełożył Krzysztof Nowicki

REFERENCJE

- AUMANN, Robert J., Sergiu HART i Motty PERRY. 1997. „The forgetful passenger”. *Games and Economic Behavior* 20: 117–20.
- LEWIS, David. 1983. „Attitudes *de dicto* and *de se*”. W: IDEM. *Philosophical Papers*. Tom I, 133–159. New York: Oxford University Press.
- PICCIONE, Michele, i Ariel RUBINSTEIN. 1997. „On the interpretation of decision problems with imperfect recall”. *Games and Economic Behavior* 20: 3–24.
- QUINE, Willard Van. 1960. *Word and Object*. Cambridge, Mass.: The MIT Press.
- VAN FRAASSEN, Bas C. 1984. „Belief and the will”. *Journal of Philosophy* 81: 235–56.
- VAN FRAASSEN, Bas C. 1995. „Belief and the problem of Ulysses and the sirens”. *Philosophical Studies* 77: 7–37.

⁵ Jestem wdzięczny Nedowi Hallowi za zwrócenie uwagi, że odpowiedź $1/3$ kłóci się z zasadą równowagi.

⁶ Moje podziękowania niech przyjmą następujące osoby: Jamie Dreier, Gary Gates, Ned Hall, Vann McGee, Robert Stalnaker, Roger White, Sarah Wright, uczestnicy konferencji w Brown University w 1999 r. (na której przedstawiono wcześniejszą wersję tego artykułu), oraz anonimowy recenzent.

PRZEKONANIA DOTYCZĄCE WŁASNEGO POŁOŻENIA W ŚWIECIE
A PROBLEM ŚPIĄCEJ KRÓLEWNY

Streszczenie

Poza niepewnością co do tego, jaki jest świat, można być także niepewnym swojego przestrzennego lub czasowego położenia w świecie. Celem artykułu jest postawienie problemu wynikającego z połączenia tych dwóch rodzajów niepewności, a następnie rozwiązanie go i wyciągnięcie dwóch lekcji z tego rozwiązania.

SELF-LOCATING BELIEF
AND THE SLEEPING BEAUTY PROBLEM

Summary

In addition to being uncertain about what the world is like, one can also be uncertain about one's own spatial or temporal location in the world. The aim of the paper is to pose a problem arising from the interaction between these two sorts of uncertainty, solve the problem, and draw two lessons from the solution.

Słowa kluczowe: paradoks; problem śpiącej królowny.

Key words: paradox; Sleeping Beauty problem.

Information about Author: Prof. ADAM ELGA — Department of Philosophy at the Princeton University, USA; areas of his interest include decision and game theory, epistemology, philosophy of mind, and philosophy of science; address for correspondence — e-mail: ae@adamelga.net

Information about Translator: KRZYSZTOF NOWICKI, MA — PhD student at the Faculty of Philosophy of the John Paul III Catholic University of Lublin; address for correspondence — e-mail: nowicki.krz@gmail.com