

# LABORATORIUM MENTIS

2024 · VOL. 2 NO. 2



Wydawnictwo



Academicum

## **Labolatorium Mentis, 2024**

Some rights reserved by Wydawnictwo Academicon, CC BY-SA 4.0

### **RADA NAUKOWA / SCIENTIFIC BOARD**

**James Robert Brown** (Uniwersytet w Toronto / University of Toronto, Canada)

**Georg Brun** (Uniwersytet w Bernie / University of Bern, Switzerland)

**Adam Grobler** (Uniwersytet Opolski / University of Opole, Poland)

**Robert Piłat** (Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie / Cardinal Stefan Wyszyński University in Warsaw, Poland)

**Tadeusz Szubka** (Uniwersytet Szczeciński / University of Szczecin, Poland)

**Monika Walczak** (Katolicki Uniwersytet Lubelski Jana Pawła II / John Paul II Catholic University of Lublin, Poland)

### **KOMITET REDAKCYJNY / EDITORIAL BOARD**

**Piotr Biłgorajski** (Katolicki Uniwersytet Lubelski Jana Pawła II / John Paul II Catholic University of Lublin, Poland) – zastępca redaktora naczelnego / Deputy Editor-in-Chief

**Marcin Iwanicki** (Katolicki Uniwersytet Lubelski Jana Pawła II / John Paul II Catholic University of Lublin, Poland) – redaktor merytoryczny / Content Editor

**Voin Milevski** (Uniwersytet w Belgradzie / University of Belgrade, Serbia) – redaktor merytoryczny / Content Editor

**Claudia Navarini** (Uniwersytet Europejski w Rzymie / European University of Rome, Italy) – redaktorka merytoryczna / Content Editor

**Robert Kryński** (Wydawnictwo Academicon / Academicon Press) – sekretarz redakcji / Editorial Secretary

**Maciej Sendłak** (Uniwersytet Warszawski / University of Warsaw, Poland) – redaktor merytoryczny / Content Editor

**Marta Soniewicka** (Uniwersytet Jagielloński / Jagiellonian University, Poland) – redaktorka merytoryczna / Content Editor

**Artur Szutta** (Uniwersytet Gdański / University of Gdańsk, Poland) – redaktor naczelny / Editor-in-Chief

**Natasza Szutta** (Uniwersytet Gdański / University of Gdańsk, Poland) – redaktorka merytoryczna / Content Editor

**James Tartagila** (Uniwersytet w Keele / Keele University, United Kingdom) – redaktor merytoryczny / Content Editor

**Maria Sivia Vaccarezza** (Uniwersytet w Genui / University of Genoa, Italy) – redaktor merytoryczny / Content Editor

### **KONTAKT / CONTACT**

Redakcja *Labolatorium Mentis* / *Labolatorium Mentis* editorial office

20-810 Lublin, Poland, ul. Heleny Modrzejewskiej 13

e-mail: [info@labmentis.eu](mailto:info@labmentis.eu)

[www: ojs.academicon.pl/lm](http://www.ojs.academicon.pl/lm)

### **WYDAWCA / PUBLISHER**

Wydawnictwo Academicon / Academicon Press

20-810 Lublin, Poland, ul. Heleny Modrzejewskiej 13

e-mail: [wydawnictwo@academicon.pl](mailto:wydawnictwo@academicon.pl)

[www: omp.academicon.pl](http://www.omp.academicon.pl)

[www: academicon.pl/wydawnictwo](http://www.academicon.pl/wydawnictwo)

### **OPRACOWANIE WYDAWNICZE / PUBLISHING WORK**

Studio DTP Academicon / Academicon DTP Studio

e-mail: [ntp@academicon.pl](mailto:ntp@academicon.pl)

[www: dtp.academicon.pl](http://www.dtp.academicon.pl)

Wersją pierwotną czasopisma jest wersja elektroniczna.

The electronic version of the journal is original version.

# Spis treści

Elżbieta Drozdowska

- 4 Logika kwantowa a empiryczność logiki
- 15 Quantum logic and the question of the empiricity of logic

Krzysztof Cholewa

- 26 Próg wystarczalności a Beverly Hills. Ponowne rozważenie intuicji dotyczących eksperymentu myślowego Rogera Crispa
- 43 Sufficient Level and Beverly Hills: Reexamining Intuitions on Roger Crisp's Thought Experiment

Piotr Biłgorajski

- 58 Czy Bóg klasycznego teizmu posiada poczucie humoru?
- 69 Does the God of classical theism have a sense of humour?

Artur Szutta

- 80 Czy kłamiąc, z konieczności głosimy fałsz?
- 92 Does lying necessarily involve stating falsehoods?





ELŻBIETA DROZDOWSKA

 Katolicki Uniwersytet Lubelski Jana Pawła II (Polska)

 0000-0003-4710-1266

 elzbieta.drozowska@kul.pl

# Logika kwantowa a empiryczność logiki

## Abstrakt:

W artykule podjęty jest problem empiryczności logiki, czyli możliwości zmiany logiki z powodów empirycznych. Omówiony jest argument Hilary'ego Putnama za empirycznością logiki bazujący na odkryciach z zakresu mechaniki kwantowej. Putnam uważał, że należy odrzucić logikę klasyczną i zastąpić ją logiką kwantową. Pogląd ten zwany jest „tezą Putnama”. Spotkała się ona z ostrą krytyką, m.in. Michaela Dummetta i Saula Kripkego. Choć ich krytyka jest w dużej mierze trafna, to nie uderza w ideę empiryczności logiki jako takiej. Celem niniejszego tekstu jest pokazanie, że można przedstawić odkrycie niedystrybutywności sądów kwantowych jako empiryczny fakt mający wpływ na rewizję adekwatności pewnych praw logiki, a więc że empiria może mieć wpływ na logikę.

## Słowa kluczowe:

empiryczność logiki, logika kwantowa, problem przyjmowania logiki

## Jak cytować:

Drozowska, E. (2024). Logika kwantowa a empiryczność logiki [polski oryginał]. *Laboratorium Mentis*, 2(2), 4–14. <https://doi.org/10.52097/lm.9279>

Badania sfinansowane ze środków Narodowego Centrum Nauki, grant numer 2021/41/N/HS1/01319.

Wyobraźmy sobie, że grupa naukowców przeprowadza skomplikowany eksperyment, którego wynik okazuje się dla nich niezrozumiały. Ludzie w białych fartuchach tłoczą się przed ekranem kontrolnym, by osobiście zobaczyć ciąg liczb niedających się wytłumaczyć. Przez tłum przebiega szmer, który powoli narasta, w miarę jak pojawiają się nowe hipotezy wyjaśniające otrzymane wyniki. Czy jest to błąd techniczny? Obliczeniowy? Niespotykana anomalia? Mijają godziny, hipotezy się mnożą, ale żadna się nie sprawdza.

– Nie ma innego wyjaśnienia, drodzy państwo – przemawia w końcu kierownik projektu. – Po prostu obaliliśmy prawa klasycznej logiki.

Czy ta sytuacja nie wygląda absurdalnie? W końcu prawa logiki to prawa myślenia, a nie prawa przyrody. Uważamy je za aprioryczne prawdy, które obowiązują powszechnie i koniecznie, niezależnie od doświadczenia. A jednak taka była teza Hilary’ego Putnama: logika jest nauką, którą można rewidować empirycznie, i to właśnie empiria, a dokładniej mechanika kwantowa, dostarczyła nam powodów do jej rewizji (Putnam, s. 1967).

## Teza Putnama

W słynnym artykule *Is logic empirical?* z 1967 r. Putnam stwierdza, że „logika jest tak samo empiryczna jak geometria” (Putnam, 1967, s. 184). A w jakim stopniu empiryczna jest geometria? Otóż w wystarczającym, by geometria Euklidesowa mogła zostać obalona jako geometria opisująca nasz świat fizyczny i by jej miejsce mogły zająć geometrie nieeuklidesowe związane z formalizmem ogólnej teorii względności (OTW) (Putnam, 1967, s. 176). Wraz z odkryciem krzywizny czasoprzestrzeni opisanej przez OTW okazało się, że wbrew naszym pierwotnym intuicjom liniami prostymi (czyli będącymi najkrótszą ścieżką od jednego punktu do drugiego) mogą być linie, które z punktu widzenia geometrii Euklidesowej prostymi nie są. A ponieważ OTW uzyskała bardzo dobre potwierdzenie eksperymentalne, znaleźliśmy się w sytuacji, w której sąd – zdawałoby się – nieintuicyjny i niemożliwy (tzn. że

taka oto linia krzywa jest tak naprawdę prosta) okazał się prawdziwy (Putnam, 1967, s. 175).

Według Putnama podobny los jak ogólna teoria względności – geometrii, mechanika kwantowa zgotowała logice (Putnam, 1967, s. 179). Sprawa rozbija się o zauważony już przez Garretta Birkhoffa i Johna von Neumanna konflikt między kwantowomechanicznym opisem układów fizycznych a prawem dystrybutywności, czyli rozdzielności koniunkcji i alternatywy, które jest prawem logiki klasycznej (Birkhoff, Neumann, 1936, s. 10).

Charakterystyczną cechą mechaniki kwantowej jest to, że występują w niej wielkości niekompatybilne, tzw. komplementarne, wykluczające się (Putnam, 1967, s. 180). Przykładem wielkości niekompatybilnych są wielkości fizyczne związane zasadą nieoznaczoności Heisenberga, takie jak np. wybrane składowe położenia i pędu. Na takich parach wielkości fizycznych nie da się równocześnie przeprowadzić pomiaru z dowolną dokładnością. Dla logiki kwantowej ma to taki skutek, że przykładowe sądy: (a) „Układ  $s$  znajduje się w położeniu  $a$ ” i (b) „Pęd układu  $s$  wynosi  $b$ ” nie mogą być równocześnie prawdziwe, a co za tym idzie, ich koniunkcja jest *zawsze* fałszywa. Ma to swoje uzasadnienie w matematycznym formalizmie mechaniki kwantowej – podprzestrzenie Hilberta reprezentujące sądy (a) i (b) mają takie własności, że po wykonaniu na nich operacji odpowiadającej koniunkcji sądów uzyskuje się jako wynik przestrzeń pustą, która odpowiada fałszowi (Putnam, 1969, s. 179).

Idąc dalej, załóżmy, iż o naszym układzie  $s$  wiadomo, że jego pęd wynosi  $b$  i że znajdować się może jedynie albo w obszarze zwanym  $a$ , albo w obszarze niebędącym  $a$ , zwanym  $a'$  (które jest ortouzupelnieniem  $a$ ). Zgodnie z prawem dystrybutywności powinno być prawdą, że

$$b \wedge (a \vee a') = (b \wedge a) \vee (b \wedge a'),$$

czyli układ ma pęd  $b$  i znajduje się w położeniu  $a$  lub  $a'$  wtedy i tylko wtedy, gdy albo ma pęd  $b$  i znajduje się w  $a$ , albo ma pęd  $b$  i znajduje się w  $a'$ . Nie możemy jednak stwierdzić zachodzenia tej równości

z powodu zasady nieoznaczoności Heisenberga. Jeśli przyjmiemy, że zdanie  $b$  jest prawdziwe, i zauważymy, iż alternatywa  $a \vee a'$  jest również prawdziwa (ponieważ podprzestrzenie reprezentujące  $a$  i  $a'$  sumują się do całej przestrzeni, która reprezentuje zdanie prawdziwe), to koniunkcja stojąca po lewej stronie będzie prawdziwa. Natomiast prawa strona nie będzie prawdziwa, ponieważ położenie i pęd są wielkościami niekompatybilnymi. Zatem, jak stwierdziliśmy wyżej, obie koniunkcje są fałszywe, a stąd ich alternatywa również jest fałszywa (Birkhoff i von Neumann, 1936, s. 10). Tak więc w opisie układów kwantowych rozdzielność koniunkcji względem alternatywy (czyli dystrybutywność) zawodzi.

Brak dystrybutywności jest właśnie cechą charakterystyczną logiki kwantowej. Według Putnama logikę tę podpowiada nam sama koncepcja przestrzeni Hilberta, stanowiąca podstawę matematycznego formalizmu mechaniki kwantowej (Putnam, 1969, s. 179). Co więcej, Putnam uważa, że to właśnie wcześniejsze nieuwzględnienie logiki kwantowej w opisie zjawisk kwantowych wprowadziło paradoksy, takie jak omówione przez niego komplementarność wielkości fizycznych, superpozycja stanów w eksperymencie dwuszczelinowym czy problem pomiaru (Putnam, 1969, s. 183). Według Putnama wszystkie te trudności można rozwiązać, odrzucając dystrybutywność logiki (Putnam, 1969, s. 184). Co więcej, jest to według niego stosunkowo niewielka zmiana, nieingerująca nawet w znaczenie klasycznych spójników logicznych (Putnam, 1969, s. 190). Innymi słowy, mechanika kwantowa wzięta wraz z logiką klasyczną wymaga niestandardowej metafizyki, niezgodnej z Putnamowskim naukowym realizmem. Jednak mechanika kwantowa w połączeniu z logiką kwantową pozwala rozwiązać kwantowe paradoksy i ocalić realistyczną metafizykę. Cenę stanowi prawo dystrybutywności i Putnam jest gotów ponieść ten koszt (Putnam, 1969, s. 189).

## Wzlot i upadek analogii między logiką a geometrią

Po uświadomieniu sobie różnic między dystrybutywną logiką klasyczną a niedystrybutywną logiką kwantową analogia Putnama między zmianą w geometrii z powodu OTW a zmianą w logice z powodu mechaniki kwantowej rysuje się całkiem wyraźnie. W przypadku ogólnej teorii względności mamy wybór: z jednej strony, możemy pozostać przy geometrii Euklidesowej, co wiązałoby się z tym, że aby opisać ruch ciał po „krzywych” w tej przestrzeni (które są prostymi w odpowiedniej geometrii nieeuklidesowej), należałoby wprowadzić dodatkowe „tajemnicze” siły; z drugiej strony, możemy porzucić geometrię Euklidesową, do której jesteśmy przywiązani, ale za to wtedy opis ruchu się uprości, uzyskamy wyjaśnienie pewnych dotychczasowych anomalii, a wspomniane tajemnicze siły nie będą potrzebne (Putnam, 1969, s. 191). Podobnie zdaje się być z mechaniką kwantową: możemy pozostać przy logice klasycznej, w którą nigdy dotąd nie mieliśmy powodów wątpić, ale mierzyć się dalej z kwantowymi paradoksami, albo możemy ją porzucić, a w zamian uzyskać skuteczną teorię ze zrozumiałą, realistyczną metafizyką, w której wszystkie wielkości fizyczne są w każdej chwili określone (Putnam, 1969, s. 191).

Niestety analogia ta słabnie, gdy weźmie się pod uwagę, że: (1) zaproponowane przez Putnama rozwiązania paradoksów nie są całkiem poprawne (Bell i Hallett, 1982, s. 357; Dummett, 1982, s. 273; Stairs, 2016, s. 25), (2) odrzucenie prawa dystrybutywności jednak ingeruje w znaczenie spójników logicznych (Bell i Hallett, 1982, s. 362; Dummett, 1976, s. 284), a (3) brak dystrybutywności jest nie do pogodzenia z realizmem Putnama (Dummett, 1976, s. 276). Co więcej, Dummett wskazuje, że wprowadzając swoje objaśnienie sensu spójników logicznych i równocześnie przyjmując realistyczną interpretację mechaniki kwantowej, Putnam nie wprowadza spójników logiki kwantowej *zamiast* klasycznych, lecz *obok* nich, „a stąd wynika, że sytuacja ta nie jest analogiczna do przypadku geometrii i nie następuje w niej



porzucenie, w reakcji na doświadczenie czy cokolwiek innego, żadnego uprzednio przyjmowanego prawa logiki” (Dummett, 1976, s. 287). Można by powiedzieć, że następuje co najwyżej *rozszerzenie* istniejącej dotąd logiki o nowe nieklasyczne spójniki. Z kolei gdyby zastąpić logikę klasyczną logiką kwantową, porzucając przy tym realistyczną interpretację mechaniki kwantowej, osłabiłoby to argument za empirycznością logiki, ponieważ „nie można by było dłużej utrzymywać, że odkrycie zawodności prawa dystrybutywności jest odkryciem czegoś o *świecie*” (Dummett, 1976, s. 288).

Putnam podejmował później dalsze prace nad logiką kwantową, opracowując różne jej wersje (Maudlin, 2018, s. 22). Badania te miały jednak głównie na celu obronę realizmu mechaniki kwantowej, charakteryzowanego przez założenie o jednoznacznym określeniu wartości (*value definiteness*), które głosi, że „pomiar jedynie ujawnia pewien fakt na temat układu, nie pełni zaś roli w *tworzeniu* żadnego nowego faktu” (Maudlin, 2018, s. 24), tzn. nie stwarza mierzonej obserwabli ani nie nadaje układowi wyraźnie określonej wielkości, której ten by wcześniej nie posiadał. Jak jednak pokazują różne twierdzenia *no-go*, takie jak np. twierdzenie Kochena–Speckera, założenia tego nie da się w mechanice kwantowej utrzymać (Maudlin, 2018, s. 25; Bell i Hallett, 1982, s. 369; Stairs, 2016, s. 36). Różne dodatkowe problemy, do których doprowadziło stanowisko Putnama, sprawiły, że ostatecznie wycofał się on z promowania logiki kwantowej jako antidotum na problemy mechaniki kwantowej (Maudlin, 2018, s. 38).

## Kwestia empiryczności logiki

Choć program badawczy Putnama okazał się ślepym zaułkiem, fakt ten nie dyskwalifikuje jeszcze potencjalnej empiryczności logiki. Być może analogia do geometrii się nie broni, wciąż jednak pozostaje argument Birkhoffa i von Neumanna za niedystrybutywnością sądów opisujących stany układów w mechanice kwantowej. Wyrosła z pracy wspomnianych autorów tradycja badawcza uznaje logikę kwantową

za logikę nieklasyczną o semantyce algebraicznej (najczęściej) będącej niedystrybutywnymi kratami ortomodularnymi i odrzuca pogląd Putnama o jednoznacznym określeniu wartości wielkości fizycznych. Zgodnie ze wspomnianą wyżej opinią Dummetta w tej sytuacji nie powinno być już mowy o empiryczności logiki, ponieważ jego zdaniem o odrzuceniu dystrybutywności decydują rozważania z zakresu teorii znaczenia, a nie mechaniki kwantowej (Dummett, 1982, s. 288). Teza o zastąpieniu logiki klasycznej logiką kwantową ani nie będzie sądem mechaniki kwantowej, ani nie będzie z takich sądów wyprowadzalna (Dummett, 1982, s. 288). Będzie za to pochodziła, zdaniem Dummetta, z teorii znaczenia, tzn. z poszukiwania „właściwego modelu znaczeń, które nadajemy naszym [...] sądom” (Dummett, 1982, s. 288).

Z kolei Kripke uważa, iż sama idea, że logika mogłaby zostać zrewidowana przez kwestie empiryczne, jest wewnętrznie sprzeczna (Kripke, 2024; Stairs, 2016, s. 29). Autor wychodzi od rozróżnienia pomiędzy *logiką*, a *systemami logiki*, gdzie tę pierwszą rozumie jako zasady rozumowania, którymi wszyscy się spontanicznie posługujemy; systemy logiki to z kolei formalne wytwory, które mają na celu przybliżać, lepiej lub gorzej, rozumowania logiczne (Kripke, 2024, s. 33). O ile systemów logiki jest wiele, o tyle logika jest tylko jedna. Nie ma wielu logik, spośród których można by wybierać. Logika służy m.in. do intuicyjnej oceny, które systemy formalne są „właściwe” (Stairs, 2016, s. 29). Kripke uważa, że czasami pewien system logiki uzyskuje ciekawą interpretację nieformalną, co sprawia, że możemy uznać, iż mamy do czynienia z jakimiś prawami logiki. Czasami może się też zdarzyć, że odkryjemy jakieś nowe spójniki i zidentyfikujemy dotyczące ich prawa logiki (Kripke, 2024, s. 33). Może też być tak, że odkryjemy, iż formuła, którą uważaliśmy za prawo logiki, w rzeczywistości nim nie jest. Według Kripkego nie jest to jednak przyjęcie nowej logiki, lecz raczej odkrycie dotyczące pewnego systemu (Kripke, 2024, s. 33). Co więcej, wszystkie takie odkrycia odbywają się na drodze *rozumowań*, a nie badań empirycznych (Kripke, 2024, s. 34).

Jednak, jak zauważa Stairs, nawet jeśli przyjmiemy rozróżnienie Kripkego na logikę i systemy logiki, to nie wykluczamy tym samym możliwości, żeby czasami kwestia oceny adekwatności pewnych praw czy systemów zależała od tego, jaki faktycznie jest świat, a więc od kwestii empirycznych. Stairs jako jeden z przykładów podaje dyskusję dotyczącą przyszłych zdarzeń przygodnych, determinizmu i prezentyzmu (Stairs, 2016, s. 33). Drugim potencjalnym przykładem może być logika kwantowa (Stairs, 2016, s. 35).

## Jaki wpływ mogłaby mieć empiria na logikę

Na czym by mógł polegać wpływ kwestii empirycznych na ocenę adekwatności praw czy systemów logiki? Chcąc odpowiedzieć na to pytanie, zacznijmy od przypomnienia, że logika bada pewne relacje zachodzące w języku. Na potrzeby tych badań konstruowane są modele języka naturalnego albo jego wycinków, które mają dwa podstawowe komponenty: syntaktyczny i semantyczny (Beall i van Fraassen, 2003, s. 23). Syntaktyka określa składnię języka – wymienia jego alfabet, czyli podstawowe symbole, i podaje reguły tworzenia poprawnie zbudowanych wyrażeń, czyli jego gramatykę. Semantyka z kolei podaje dopuszczalne wartościowania wyrażeń, czyli funkcje przypisujące im pewne wartości, na przykład prawdę czy fałsz. Język składa się z syntaktyki i niepustej klasy jej wartościowań. Rolą logiki jest określenie, które zdania w języku są logicznie prawdziwe i które wnioskowania są poprawne formalnie. System logiki określa syntaktycznie zdefiniowaną relację konsekwencji (wyprowadzalności wniosku ze zbioru przesłanek). Gdy relacja ta pokrywa się z wynikaniem semantycznym (wynikaniem logicznym), system jest silnie pełny względem danego języka. System logiki może być silnie pełny względem wielu języków (van Fraassen, 1979, s. 139).

Semantyka stanowi pomost pomiędzy systemem logiki a światem – nadając zdaniom wartość logiczną i narzucając ograniczenie na to, które spośród możliwych wartościowań są dopuszczalne, wprowadza

możliwość oceny wnioskowań nie tylko pod kątem poprawności formalnej, ale też tego, czy i co można w tym języku powiedzieć o świecie. Zależność może działać też w drugą stronę – możemy modyfikować semantykę tak, by odpowiadała wycinkowi świata, który chcemy modelować i dla którego chcemy znaleźć odpowiedni system logiki.

W klasycznym rachunku zdań dopuszczalne wartościowania przypisują formułom wartości ze zbioru  $\{0, 1\}$ , zachowując określone reguły, takie jak np. to, że formuły  $\varphi$  i  $\neg\varphi$  nie mogą mieć takiej samej wartości logicznej. Interpretowane jako prawda i fałsz są odpowiednio 1 i 0. Okazuje się, że klasyczny rachunek zdań jest również pełny względem języków, w których wartościowania prowadzą do elementów dwuwartościowej macicy, algebry zbiorów albo algebry Boole'a. Nazywamy je odpowiednio semantyką matrycową, teoriomnogościową i algebraiczną. Logika kwantowa nie jest pełna względem tych języków. Jest za to pełna względem klasy (niedystrybutywnych) krat ortomodularnych.

Można by powiedzieć, że klasyczny rachunek zdań działa w odniesieniu do tych klas przedmiotów, które można potraktować jako elementy zbioru, ponieważ funktory logiczne zachowują się analogicznie do działań na zbiorach. Okazuje się jednak, że istnieją obiekty, które w ten sposób się nie zachowują. Np. stany układów kwantowych w skali mikro zachowują się jak wektory; można wykonywać na nich działania takie jak na wektorach, a nie jak na zbiorach. Widać to szczególnie wyraźnie w przypadku alternatywy, która w logice kwantowej jest interpretowana jako tworzenie supremum podprzestrzeni (dodawanie wektorów), a w logice klasycznej jako sumowanie zbiorów. Odkrycie, że obiekty w skali mikro zachowują się właśnie w taki sposób, ma oczywiście komponent zarówno teoretyczny, jak i empiryczny – sformułowanie teorii jest pracą intelektualną, która nie bazuje wyłącznie na empirii; jest jednak kwestią empirii przetestowanie teorii, by stwierdzić, czy właściwie opisuje ona świat, czy nie. Obecnie w przypadku mechaniki kwantowej nie mamy już wątpliwości, że teoria ta jest skuteczna i prawidłowo opisuje zachowanie mikroobiektów. Stąd wydaje się, że uprawnionym jest powiedzieć, iż testy empiryczne potwierdziły

istnienie klasy obiektów fizycznych (w mikroświecie), do których opisu właściwsza jest teoria przestrzeni Hilberta niż teoria zbiorów, a co za tym idzie, iż znaleźliśmy w świecie obszar, w którym bardziej właściwe jest używanie logiki o semantyce w postaci krat ortomodularnych, a nie o semantyce teoriomnogościowej.

Co to oznacza dla logiki klasycznej? Na pewno nie oznacza, że została „obalona” czy że powinniśmy ją czymkolwiek zastąpić. Jeśli jest tak, jak twierdzi Kripke, że istnieje wiele systemów logiki, które przybliżają różne fragmenty logiki jako takiej, to wystarczy stwierdzić, że odkryliśmy nowy obszar, gdzie pewnego prawa, które uważaliśmy dotąd za prawo logiki, nie da się zastosować, nawet jeśli można je stosować w innych miejscach. Wniosek taki jest zgodny ze stanowiskiem pluralizmu lokalnego, zgodnie z którym różne logiki stosują się do różnych obszarów dyskursu (Czernecka-Rej, 2014, s. 79; Haack, 1978, s. 221).

Powiedzieliśmy, że logika bada pewne relacje zachodzące w języku. Język zaś informuje o świecie. Świat z pewnością ma wpływ na ocenę poprawności materialnej wnioskowań. Nie jest jednak nie do pomyślenia, że to, jakie wnioskowania uznajemy za poprawne, bierze się z tego, jakiego typu własności mają obiekty, z którymi obcujemy na co dzień. W takim razie odkrycie nowego, nieintuicyjnego typu obiektów mogłoby skłonić nas do przemyślenia, jak należy o nich myśleć i w jaki sposób rozumować.

## Bibliografia

Beall, J. C., van Fraassen, B. (2003). *Possibilities and Paradox: an Introduction to Modal and Many-Valued Logic*. Oxford University Press.

Bell, J., Hallett, M. (1982). Logic, Quantum Logic and Empiricism. *Philosophy of Science*, 49(3), 355–379. <http://www.jstor.org/stable/187280>.

Birkhoff, G., von Neumann, J. (1936). The Logic of Quantum Mechanics. *Annals of Mathematics*, 37, 823–843. <https://doi.org/10.2307/1968621>.

Czernecka-Rej, B. (2014). *Pluralizm w logice*. Wydawnictwo KUL.

**Dummett, M.** (1976). Is Logic Empirical? In H. D. Lewis (Ed.), *Contemporary British Philosophy, 4th series* (pp. 45–68). Allen and Unwin.

**van Fraassen, B.** (1979). *Formal Semantics and Logic*. The Macmillan Company. Retrieved data from <https://www.princeton.edu/~fraassen/books/pdfs/Formal%20Semantics%20and%20Logic.pdf>

**Haack, S.** (1978). *Philosophy of Logics*. Cambridge University Press.

**Kripke, S.** (2024). The Question of Logic. *Mind*, 133(529), 1–36. <https://doi.org/10.1093/mind/fzad008>

**Maudlin, T.** (2018). *The Labyrinth of Quantum Logic*. Cornell University. <https://doi.org/10.48550/arXiv.1802.01834>

**Putnam, H.** (1967). Is logic empirical? In R. Cohen, M. Wartofsky (Eds.), *Boston Studies in Philosophy of Science* (vol. 5, pp. 216–241). D. Reidel Publishing Company.

**Stairs, A.** (2016). Could Logic be Empirical? The Putnam-Kripke Debate. In J. Chubb, A. Eskandarian, V. Harizanov (Eds), *Logic and Algebraic Structures in Quantum Computing* (pp. 23–41). Cambridge University Press. <https://doi.org/10.1017/CBO9781139519687.004>



ELŻBIETA DROZDOWSKA



John Paul II Catholic University of Lublin (Poland)



0000-0003-4710-1266



elzbieta.drozdowska@kul.pl

# Quantum logic and the question of the empiricity of logic



## Abstract:

The article undertakes the issue of whether logic is empirical, that is, whether it is possible to modify logic due to empirical reasons. We will call this issue the problem of the empiricity of logic. Hilary Putnam's famous argument based on discoveries in quantum mechanics is discussed. Putnam claimed that classical logic should be discarded and replaced by quantum logic. This view is called the "Putnam's claim." It was met with harsh criticism from Dummett and Kripke, among others. While their criticism is largely valid, it does not dismiss the idea that logic itself could be empirical. The aim of this paper is to show that it is possible to present the discovery of the non-distributivity of quantum propositions as an empirical fact with implications for the revision of the adequacy of certain laws of logic, and thus, that empirical facts can have an impact on logic.



## Keywords:

empiricity of logic, quantum logic, adoption problem



## How to cite:

Drozdowska, E. (2024). Quantum logic and the question of the empiricity of logic [English translation]. *Laboratorium Mentis*, 2(2), 15–25. <https://doi.org/10.52097/lm.9482>

This research was supported by National Science Centre, Poland, grant number 2021/41/N/HS1/01319.

Imagine that a group of scientists conducts a complicated experiment, the result of which proves beyond their comprehension. People in white lab coats gather in front of the control screen to see the string of inexplicable numbers. A murmur sweeps through the crowd, slowly growing as new explanatory hypotheses emerge. Is it a technical error? A computational one? An unprecedented anomaly? Hours pass, hypotheses multiply, but none proves to be correct.

– There is no other explanation, ladies and gentlemen – the project manager finally speaks. – We have simply refuted the laws of classical logic.

Does this situation appear absurd? After all, the laws of logic are the laws of reasoning, not the laws of nature. We consider them a priori truths that are universally and necessarily valid, regardless of experience.

Yet this was Hilary Putnam's claim: that logic is a discipline that can be revised on empirical grounds, and it is the empirical, and more specifically, quantum mechanics, that have provided us with reasons to revise it (Putnam, 1967).

## Putnam's claim

In his famous 1967 article *Is logic empirical?* Putnam states that “logic is as empirical as geometry” (Putnam 1967, 184). And to what extent is geometry empirical? Well, sufficient enough for Euclidean geometry to be disproved as the geometry describing our physical world and for its place to be taken by the non-Euclidean geometry associated with the formalism of General Relativity (GR) (Putnam, 1967, 176). With the discovery of the curvature of space-time described by GR, it turned out that, contrary to our basic intuitions, lines which, from the point of view of Euclidean geometry, are not straight, can be considered straight (i.e., as the lines that are the shortest path from one point to another). And since GR has received excellent experimental confirmation, we have found ourselves in a situation where a proposition that only recently



seemed unintuitive and impossible (i.e., that such a curved line is in fact straight) has turned out to be true (Putnam, 1967, 175).

According to Putnam, quantum mechanics has wrought upon logic a similar fate to the one which General Relativity wrought upon geometry (Putnam, 1967, 179). The issue boils down to the conflict, already noted by Birkhoff and von Neumann, between the quantum mechanical description of physical systems and the law of distributivity of conjunction and disjunction, which is a law of classical logic (Birkhoff, von Neumann, 1936, 10).

A characteristic feature of quantum mechanics is that it involves incompatible quantities, known as complementary or mutually exclusive quantities (Putnam, 1967, 180). An example of incompatible quantities are physical quantities which are tied by Heisenberg's uncertainty principle, such as, for example, selected components of position and momentum vectors. It is impossible to perform simultaneous measurements with arbitrary accuracy on such pairs of physical quantities. The consequence for quantum logic is that propositions such as *a*) "System *s* is at position *a*" and *b*) "The momentum of system *s* is *b*," cannot be simultaneously true, and consequently, their conjunction is *always* false. This is grounded in the mathematical formalism of quantum mechanics: the Hilbert subspaces representing the propositions *a* and *b* have such properties that when the operation corresponding to the conjunction of the propositions is performed on them, the empty space, which corresponds to falsity, is obtained as the result (Putnam, 1969, 179).

Now, suppose that it is known about our system *s* that its momentum is *b*, while about its position, it is known only that it can be located either in a region called *a*, or in a region that is not *a*, called *a'* (which is the orthocomplement of *a*). According to the law of distributivity, it should be true that

$$b \wedge (a \vee a') = (b \wedge a) \vee (b \wedge a'),$$

That is, the system has momentum *b* and is at position *a* or *a'* if and only if either it has momentum *b* and is at *a*, or it has momentum

$b$  and is at  $a'$ . However, we cannot say that this equality is true because of Heisenberg's uncertainty principle. If we assume that sentence  $b$  is true, and observe that the disjunction  $a \vee a'$  is also true (because the subspaces representing  $a$  and  $a'$  sum to the whole space that represents the truth), then the conjunction on the left side is true. The right-hand side, in contrast, is not true, because position and momentum are incompatible quantities, thus, as we stated above, both conjunctions are false, and hence their disjunction is also false (Birkhoff, von Neumann, 1936, 10). Therefore, in the description of quantum systems, the law of distributivity fails.

The failure of distributivity is the characteristic feature of quantum logic. According to Putnam, this logic is derived from the Hilbert space itself, which is the foundation of the mathematical formalism of quantum mechanics (Putnam, 1969, 179). Moreover, Putnam believes that because quantum logic was not previously included in the description of quantum phenomena, many phenomena that are difficult to understand and explain have emerged. He discusses the cases of the complementarity of physical quantities, the superposition of states in a double-slit experiment, and the problem of measurement (Putnam, 1969, 183). According to Putnam, all of these difficulties can be solved in one stroke by simply rejecting the distributivity of logic (Putnam, 1969, 184). Moreover, according to him, this is a relatively minor change, that does not even interfere with the meaning of classical logical connectives (Putnam, 1969, 190). In other words, quantum mechanics combined with classical logic requires a nonstandard metaphysics that is not compatible with Putnam's scientific realism. On the other hand, quantum mechanics combined with quantum logic enables one to resolve quantum paradoxes and salvage a realistic metaphysics. The price is the law of distributivity, and Putnam is willing to pay it (Putnam, 1969, 189).

## The Rise and Fall of the Analogy Between Logic and Geometry

Put in this way, Putnam's analogy between the change in geometry due to GR and the change in logic due to quantum mechanics appears to be quite clear. In the case of General Relativity, we face the following choice: on the one hand, we can retain Euclidean geometry, in which case, in order to describe the motion of bodies along the "curved lines" in this space (which are straight lines in the appropriate non-Euclidean geometry), additional "mysterious" forces have to be introduced; on the other hand, we can abandon Euclidean geometry, which is so intuitive to us, but then the description of motion will be simplified, we will obtain an explanation of a number of anomalies, and no mysterious forces will be needed to describe motion (Putnam, 1969, 191). The same is supposed to be the case with quantum mechanics: we can keep with classical logic, which we have never had any reason to doubt, but continue to face quantum paradoxes, or we can abandon it, and instead obtain an effective theory with a comprehensible, realistic metaphysics, in which all physical quantities are determined at all times (Putnam, 1969, 191).

Unfortunately, the analogy fades away when one considers that 1) Putnam's proposed solutions to paradoxes are not quite correct (Bell & Hallett, 1982, 357; Dummett, 1982, 273; Stairs, 2016, 25), 2) the rejection of the law of distributivity is found to interfere with the meaning of logical connectives (Bell & Hallett, 1982, 362; Dummett, 1976, 284), and 3) the absence of distributivity is incompatible with Putnam's realism (Dummett, 1976, 276). What's more, Dummett points out that if Putnam is to maintain simultaneously his own understanding of logical connectives and his realistic interpretation of quantum mechanics, his quantum connectives will work alongside classical ones, not *replace them*, "and this entails that the situation is not parallel to the geometrical case, and does not involve the abandonment, in response to experience or otherwise, of any logical law formerly held" (Dummett, 1976, 287).

It is at most an *expansion* of logic with some new non-classical connectives. On the other hand, if we were to replace classical logic with quantum logic but abandon the realistic interpretation of quantum mechanics, the argument for the empiricity of logic would be undermined because “it would no longer be possible to claim that the discovery of the invalidity of the distributive law was a discovery *about the world*” (Dummett, 1976, 288).

Putnam has continued his work on quantum logic for several years (Maudlin, 2018, 22). His main goal was to defend the realistic interpretation of quantum mechanics, which he characterized through value definiteness: the idea that ‘a “measurement” merely reveals a pre-existing fact about a system, rather than playing a role in *creating* some new fact’ (Maudlin, 2018, 24). This means that the measurement neither creates the measured observable during the measurement nor vests the system with a sharp value of a quantity which it did not possess beforehand. As is shown by various no-go theorems, such as the Kochen-Specker theorem, value definiteness cannot be upheld in quantum mechanics (Maudlin, 2018, 25; Bell & Hallett, 1982, 369; Stairs, 2016, 36). Various additional difficulties with this approach have led Putnam to abandon his attempts to treat quantum mechanics’s difficulties with quantum logic (Maudlin, 2018, 38).

## The issue of the empiricity of logic

Although Putnam’s research program turned out to be a dead end, this does not yet disqualify the potential empiricity of logic. Perhaps the analogy with geometry doesn’t hold up. But there still remains Birkhoff and von Neumann’s argument for the non-distributivity of quantum propositions. The research tradition that grew out of the work of Birkhoff and von Neumann considers quantum logic to be a non-classical logic with algebraic semantics (usually) of (non-distributive) orthomodular lattices and rejects Putnam’s idea of value definiteness. According to the above-mentioned opinion of Dummett, in this case, there

should no longer be any question of the empiricity of logic, since, in his opinion, the rejection of distributivity is determined by considerations of the theory of meaning, not quantum mechanics (Dummett, 1982, 288). The thesis of replacing classical logic with quantum logic will neither be a proposition of quantum mechanics nor will it be derived from it (Dummett, 1982, 288). Instead, it will be drawn, according to Dummett, from the theory of meaning, i.e., from the search for the “correct model for the meanings which we confer upon our [...] statements” (Dummett, 1982, 288).

Kripke, in turn, believes that the very idea that logic could be revised due to empirical factors is inherently contradictory (Kripke, 2024; Stairs, 2016, 29). Kripke first draws a distinction between *logic* and *systems of logic*. He understands the former as principles of reasoning that we all, spontaneously as it were, follow, while systems of logic are formal creations that aim to approximate, for better or worse, logical reasoning (Kripke, 2024, 33). While there are many systems of logic, there is only one logic. There are not multiple logics to choose from. Among other things, logic is used to make intuitive judgments about which formal systems are “correct” (Stairs, 2016, 29). Kripke believes that sometimes a particular system of logic receives an interesting informal interpretation, which leads us to believe that we are dealing with some laws of logic. Sometimes it may also happen that we discover some new connectives and identify the laws of logic pertaining to them (Kripke, 2024, 33). It may also be the case that we discover that a formula we believed to be a law of logic is in fact not one. However, according to Kripke, this is not an adoption of a new logic but rather a discovery about a particular system of logic (Kripke, 2024, 33). And what’s more, according to Kripke, all such discoveries take place through *reasoning*, not empirical investigation (Kripke, 2024, 34).

However, as Stairs notes, even if we accept Kripke’s distinction between logic and systems of logic, this does not preclude that sometimes the question of assessing the adequacy of particular laws or systems may sometimes depend on what the world is actually like, i.e. on empirical

issues. Stairs gives as one example the discussion of future contingency, determinism and presentism (Stairs, 2016, 33). A second possible example could be quantum logic (Stairs, 2016, 35). How would it work?

## How could the empirical possibly have a bearing on logic?

We will begin by recalling that logic studies certain relations that occur in language. For the purpose of these studies, there are constructed models of natural language or its fragments, which have two basic components: syntax and semantics (Beall & van Fraassen, 2003, 23). Syntax defines the structure of a language: it specifies its alphabet, that is, the basic symbols, and provides the rules for forming well-formed formulas, that is, its grammar. Semantics, on the other hand, provides admissible valuations of the formulas, that is, functions which assign to them certain values, such as truth or falsity. A language consists of its syntax and a non-empty class of its valuations. The role of logic is to determine which propositions of the language are logically true and which inferences are valid. A system of logic defines a syntactically defined relation of consequence (the derivability of a conclusion from a set of premises). When this relation coincides with a semantically defined consequence (logical entailment), the system is strongly complete with respect to a given language. A system of logic can be strongly complete with respect to many languages (van Fraassen, 1979, 139).

Semantics bridges the gap between systems of logic and the world: by providing propositions with a logical value and imposing constraints on the class of admissible valuations, it introduces the possibility of evaluating inferences not only in terms of validity but also in terms of whether and what can be said about the world in this language. On the other hand, we can also modify the semantics to conform to the fragment of the world we want to model and for which we want to find an appropriate system of logic.

In classical propositional calculus, admissible valuations assign the formulas values from the set  $\{0, 1\}$ , observing some specific rules, such as that the formulas  $\varphi$  and  $\neg\varphi$  cannot have the same logical value. 1 and 0 are interpreted as truth and falsity, respectively. It turns out that the classical propositional calculus is also complete with respect to languages in which valuations take as values the elements of a two-valued matrix, an algebra of sets or a Boolean algebra. These semantics are referred to as the matrix semantics, set-theoretic semantics, and algebraic semantics, respectively. Quantum logic is not complete with respect to any of these languages. Instead, it is complete with respect to the class of (non-distributive) orthomodular lattices.

One might say that the classical propositional calculus is appropriate for those objects that can be treated as elements of a set, so that the logical connectives would behave analogously to operations on sets. But as it turns out, there exist some objects that do not behave in this manner. Namely, the states of quantum systems at the microscopic level behave like vectors; one can perform vector operations on them, but not set operations. This is particularly evident in the case of the disjunction, which in quantum logic is interpreted as the supremum of a subspace (vector addition), and in classical logic as the set union. The discovery that micro objects behave in such a way has, of course, both a theoretical and an empirical component: formulating a theory is an intellectual work that is not based solely on empirical evidence; but it is a matter of empirical work, i.e. testing of the theory, to determine whether or not it describes the world correctly. At the moment, in the case of quantum mechanics, we have no doubt that the theory is effective and correctly describes the behavior of micro objects.

Hence, it seems to be justified to say that empirical tests have confirmed that there is a class of physical objects (in the micro-world) for the description of which the Hilbert space theory is more appropriate than set theory. Consequently, we have found that there is a realm of the world in which it is more appropriate to use logic with the semantics of orthomodular lattices rather than set theory.

What does this mean for classical logic? It certainly doesn't mean that it has been "refuted" or that we should replace it. If it is the case, as Kripke argues, that there are various systems of logic which approximate different parts of logic as such, then it is enough to conclude that we have discovered a new domain where a certain law, which we have hitherto considered a law of logic, cannot be applied, even if it can be applied elsewhere. Such a conclusion is consistent with the position of local pluralism, according to which different logics apply to different areas of discourse (Czernecka-Rej, 2014, 79; Haack, 1978, 221).

We said that logic studies certain relations occurring in language. Language, in turn, informs about the world. The world certainly influences the evaluation of the premises in inferences. However, it is not inconceivable that the kinds of inferences we consider valid are derived from the kinds of objects we interact with on a daily basis. If this is the case, the discovery of a new, unintuitive type of objects could lead us to reconsider how we should think and reason about them.

## Bibliography

Beall, J. C., & van Fraassen, B. (2003). *Possibilities and Paradox: an Introduction to Modal and Many-Valued Logic*. Oxford University Press.

Bell, J., & Hallett, M. (1982). Logic, Quantum Logic and Empiricism. *Philosophy of Science*, 49(3), 355–379. <http://www.jstor.org/stable/187280>.

Birkhoff, G., & von Neumann, J. (1936). The Logic of Quantum Mechanics. *Annals of Mathematics*, 37(4), 823–843. <https://doi.org/10.2307/1968621>

Czernecka-Rej, B. (2014). *Pluralizm w logice*. Wydawnictwo KUL.

Dummett, M. (1976). Is Logic Empirical? In H. D. Lewis (Ed.), *Contemporary British Philosophy, 4th series* (pp. 45–68). Allen and Unwin.

Haack, S. (1978). *Philosophy of Logics*. Cambridge University Press.

Kripke, S. (2024). The Question of Logic. *Mind*, 133(529), 1–36. <https://doi.org/10.1093/mind/fzad008>

Maudlin, T. (2018). *The Labyrinth of Quantum Logic*. Cornell University. <https://doi.org/10.48550/arXiv.1802.01834>



**Putnam, H.** (1967). Is logic empirical? In R. Cohen, & M. Wartofsky (Eds), *Boston Studies in Philosophy of Science* (vol. 5, pp. 216–241). D. Reidel Publishing Company.

**Stairs, A.** (2016). Could Logic be Empirical? The Putnam-Kripke Debate. In J. Chubb, A. Eskandarian, & V. Harizanov (Eds), *Logic and Algebraic Structures in Quantum Computing* (pp. 23–41). Cambridge University Press. <https://doi.org/10.1017/CBO9781139519687.004>

**van Fraassen, B.** (1971). *Formal Semantics and Logic*. The Macmillan Company. Retrieved data from <https://www.princeton.edu/~fraassen/books/pdfs/Formal%20Semantics%20and%20Logic.pdf>

**KRZESIMIR CHOLEWA** Uniwersytet Jagielloński (Polska) 0009-0005-2084-810X krzesimir.cholewa@uj.edu.pl

# Próg wystarczalności a Beverly Hills. Ponowne rozważenie intuicji do- tyczających eksperymen- tu myślowego Rogerera Crispa

## Abstrakt:

Prezentowany tekst porusza kwestię sprawiedliwości dystrybucyjnej, bazując na eksperymencie myślowym Rogerera Crispa – przypadek Beverly Hills. Oprócz krótkiego wprowadzenia przedstawiającego podstawowe rozróżnienia teoretyczne dotyczące zagadnienia, artykuł zawiera cztery części. W pierwszej przedstawiono koncepcję Crispa oraz jego oryginalne sformułowanie przypadku Beverly Hills. Następne dwie części zawierają trzy uszczegółowienia tego eksperymentu myślowego oraz analizę wpływu, jaki mają one na nasze pierwotne intuicje. Ostatnia część zawiera propozycje sformułowania pluralistycznego suficyjentyzmu zawierającego dwa progi: statyczny, obiektywny próg zaczerpnięty z koncepcji Crispa oraz dynamiczny – oparty na doktrynie względnej sprawiedliwości.

## Słowa kluczowe:

sprawiedliwość dystrybucyjna, suficyjentyzm, egalitaryzm, priorytaryzm, przypadek Beverly Hills

**☛ Jak cytować:**

Cholewa, K. (2024). Próg wystarczalności a Beverly Hills. Ponowne rozważenie intuicji dotyczących eksperymentu myślowego Rogera Crispa [polskie tłumaczenie]. *Laboratorium Mentis*, 2(2), 26–42. <https://doi.org/10.52097/lm.9442>

## Wstęp

Jednym z klasycznie wyróżnianych rodzajów sprawiedliwości jest sprawiedliwość rozdzielcza, znana przynajmniej od czasów Arystotelesa (Arystoteles, wyd. 1996, s. 173–174). W najprostszy sposób można powiedzieć, że wszelkie teorie sprawiedliwości dystrybutywnej mają odpowiedzieć na jedno, początkowo zrozumiałe, pytanie: „Jaki jest właściwy sposób rozdzielania korzyści i obciążeń między grupami ludzi lub jednostkami?”. W naszym świecie istnieją ograniczone dobra, które możemy rozdzielać na różne sposoby, jednak niektóre formy takiego podziału są niesłuszne, nieracjonalne lub wręcz niesprawiedliwe.

Dokonując pewnego uproszczenia, można wyróżnić dwie główne grupy poglądów w obszarze zagadnień związanych ze sprawiedliwością dystrybutywną<sup>1</sup>:

(A) utylitaryzm;

(B) poglądy, których nie da się sprowadzić do konsekwencjalistycznego welferyzmu, oraz te, które dotyczą tak czy inaczej rozumianej równości:

(B.1) egalitaryzm;

(B.2) priorytaryzm;

(B.3) suficjentaryzm.

Grupa poglądów (A) jest bardzo dobrze znana pod postacią teorii moralnej. Stosunkowo łatwo można zrozumieć, w jaki sposób założenia utylitarystyczne są aplikowane do problematyki sprawiedliwości

---

<sup>1</sup> Nie twierdzę, że to podział zupełny. Twierdzę jedynie, że wyróżnione grupy poglądów są prawdopodobnie najlepiej rozwinięte w literaturze i najczęściej wykorzystywane w dyskusjach, np. w bioetyce.

dystrybucyjnej. Zgodnie z utylityzmem sprawiedliwe (czy też słuszne) rozdzielanie korzyści polega tym, aby maksymalizować ogólny dobrostan na świecie, przy założeniu, że każdy jednostkowy beneficjent jest równie ważny. Powszechnie wiadomo, że przeciwko utylityzmowi jako ogólnemu stanowisku wysuwanych jest wiele zarzutów (Moore, 1919, s. 95–109; Ross, 2009, s. 43–47). Podobnie jest w przypadku jego aplikacji do zagadnienia sprawiedliwości dystrybucyjnej. Na przykład można wskazać, że prowadzi on do nieintuicyjnych wniosków, takich jak: „Byłoby sprawiedliwe wyleczyć katar u dużej grupy osób zamiast uratować przed przedwczesną śmiercią jakąś wystarczająco mniejszą (choć nadal znaczną) grupę młodych ludzi”.

Chociaż poglądy zawarte w (B) są bardzo różnorodne, możemy tymczasowo, na potrzeby artykułu, traktować je jako zaliczające się do jednej szerokiej grupy. Najprościej możemy przedstawić poszczególne poglądy z tej grupy za pomocą następujących formuł:

(B.1) Równość ma wartość wewnętrzną i dlatego w naszych decyzjach alokacyjnych powinniśmy wybierać te działania, które przybliżą do niej nasz świat (Temkin, 2015).

(B.2) W naszych wyborach alokacyjnych powinniśmy stosować kategorię najgorzej sytuowanych i priorytetowo traktować osoby lub grupy z tej kategorii (Parfit, 2015, s. 123).

(B.3) Nie są ważne priorytetowe traktowanie najgorzej sytuowanych ani nawet równość, lecz osiągnięcie przez jednostki lub ich grupy tzw. poziomu wystarczającego. Musimy rozważyć, czy porównywane grupy lub jednostki mają „wystarczająco dużo”. Jeśli jedna z nich ma, a druga nie, dysponujemy silną racją moralną, aby wesprzeć tę drugą (Frankfurt, 2015; Crisp, 2015).

Wszystkie trzy powyższe poglądy mogą być wspierane albo krytykowane za pomocą różnych argumentów. Jednym z najbardziej znanych jest argument z równania w dół (Parfit, 2015, s. 120–121), stanowiący zarzut przeciwko (B.1) i jednocześnie argument na rzecz (B.2). Można go również odczytać jako wsparcie dla (B.3).

W niniejszym artykule nie będę szczegółowo analizować wszystkich tych poglądów, lecz skupię się jedynie na jednej wersji, (B.3), a dokładnie na konkretnym argumentcie na rzecz (B.3), który pierwotnie miał na celu krytykę (B.2); jak się okaże, może on również stanowić wyzwanie dla (B.1). Tym argumentem jest eksperyment myślowy Rogera Crispa, nazwany przypadkiem Beverly Hills. W pierwszej części artykułu zrekonstruuje argument oraz, na ogólnym poziomie, wersję suficyentaryzmu w ujęciu Crispa. W drugiej części przedstawię trzy sposoby uszczegółowienia oryginalnego eksperymentu myślowego Crispa. Następnie pokażę, że uszczegółowienie i ponowne rozważenie argumentu znacząco zmienia nasze początkowe intuicje. W ostatniej części przedstawię dwa wnioski, które można wyciągnąć na podstawie zmian w naszych intuicjach. Przed przejściem dalej warto poczynić jeszcze jedną uwagę. Dyskusje na temat sprawiedliwości dystrybtywnej mogą mieć wpływ na bardzo różne dziedziny filozofii. Artykuł dotyczy etyki ogólnej, ale przynajmniej jeden z wniosków może odnosić się również do filozofii politycznej.

## Suficyentaryzm Rogera Crispa

Na najbardziej ogólnym poziomie można opisać suficyentaryzm jako pogląd, który obejmuje co najmniej dwa twierdzenia:

- (i) „Posiadanie wystarczająco dużo” ma wartość wewnętrzną, zatem w decyzjach alokacyjnych powinniśmy rozdzielać zasoby tak, aby zapewnić wystarczający ich poziom dla jak największej liczby osób.

(ii) Jeśli ktoś „ma wystarczająco dużo”, wówczas racja moralna, aby pomóc takiej osobie, radykalnie różni się od racji do pomocy w sytuacji, w której owa osoba nie miałaby dostatecznie dużo (Shields, 2020, s. 2).<sup>2</sup>

Na najbardziej ogólnym poziomie wszystkie rodzaje suficyentaryzmu muszą więc stwierdzać, że istnieje pewien próg posiadania dóbr, który jest istotny z moralnego punktu widzenia, i ci, którzy znajdują się powyżej tego progu, powinni być traktowani inaczej w decyzjach dystrybucyjnych od osób znajdujących się poniżej. Każdy, kto znajduje się poniżej, nie posiada bowiem dóbr w stopniu wystarczającym, zaś każdy, kto znajduje się powyżej, ma wystarczająco dużo.

Tak ogólny opis stanowiska jest oczywiście pełen niejasności i musi zostać wypełniony bardziej szczegółową treścią, aby był zrozumiały. Najbardziej oczywiste pytania odnośnie do szczegółów koncepcji dotyczą pojęcia posiadania wystarczająco dużo. Gdzie i jak dokładnie ustalić rzeczony próg oraz jak interpretować różnice w traktowaniu jednostek lub grup znajdujących się poniżej albo powyżej tego progu? Zwolennicy suficyentaryzmu mogliby zaproponować wiele możliwych odpowiedzi na te pytania, tutaj jednak skupimy się na wersji Rogera Crispa. Opiera on swoje stanowisko na kategoriach bezstronnego obserwatora i współczucia.

A zatem pojęcie współczucia w połączeniu z ideą bezstronnego obserwatora może dostarczyć nam podstaw do nakreślenia takich zasad dystrybucji, które uprzywilejowują osoby w gorszym położeniu wtedy i tylko wtedy, gdy sytuacja tych osób jest też zarazem obiektywnie zła. [...] Wskazany w ten sposób próg wyznaczy w sposób

<sup>2</sup> Autor nazywa te twierdzenia (1) tezą pozytywną i (2) tezą przesunięcia: „Pierwsza to teza pozytywna, która głosi, że posiadamy racje, aby zapewnić [komuś] wystarczająco dużo, które to racje są istotne i nieinstrumentalne [...]. Drugą definiującą tezą suficyentaryzmu jest teza przesunięcia, która stwierdza, że następuje zmiana lub przesunięcie nieinstrumentalnych racji przemawiających za przysporzeniem korzyści lub obciążeniem tych osób, które mają zapewnione wystarczająco dużo”.

jednoznaczny pewien poziom dobrostanu, powyżej którego nie będzie podstaw do przyznawania komukolwiek pierwszeństwa, lecz poniżej już tak (Crisp, 2015, s. 178–179).

Crisp twierdzi, że powinniśmy ustalić ów próg, określając poziom dobrostanu, powyżej którego bezstronny obserwator nie odczuwałby współczucia dla danej jednostki. Ten obiektywny poziom dobrostanu, wskazujący, kto zasługuje na współczucie, oddziela tych, którzy mają wystarczająco dużo, od tych, którzy mają zbyt mało.

Zasada współczucia: absolutne pierwszeństwo w podziale korzyści przysługuje jednostkom znajdującym się poniżej granicy wyznaczonej współczuciem dla bliźnich. Poniżej tego progu pomaganie im jest tym ważniejsze, im gorsza jest ich sytuacja, im więcej jest takich jednostek i im znaczniejsze są oferowane korzyści. Powyżej progu, a także w przypadku nieznacznych korzyści poniżej owej granicy, nie ma uzasadnienia dla żadnego pierwszeństwa (Crisp, 2015, s. 179).

W powyższym fragmencie Crisp tłumaczy, jak rozumie różnicę w traktowaniu jednostek w zależności od tego, po której stronie progu się znajdują. Zdaniem autora, jeśli decyzja dystrybucyjna dotyczy wyłącznie osób znajdujących się poniżej progu wystarczalności, powinniśmy zastosować zasadę ważonego pierwszeństwa, a więc brać pod uwagę nie tylko informację, kto jest w gorszej sytuacji, ale także liczbę osób, którym przyniesiemy korzyść, oraz rodzaj tej korzyści. W tego typu decyzjach musimy zatem połączyć założenia priorityzmu i utilityzmu. Kiedy zaś decyzja dotyczy osób znajdujących się po różnych stronach progu, możemy zastosować absolutny priorityzm, ponieważ każda niebłaha korzyść dla osoby poniżej progu jest bezwzględnie ważniejsza niż jakakolwiek korzyść – nawet znacząca i dotycząca dużej populacji – dla tych, którzy są powyżej progu. Wreszcie jeśli decyzja dotyczy wyłącznie osób powyżej progu, powinniśmy powrócić do

klasycznego utilitaryzmu uwzględniającego jedynie liczebność grupy beneficjentów i wielkość korzyści<sup>3</sup>.

Teoria krótko opisana powyżej jest, w pewnym stopniu, skonstruowana w taki sposób, aby uwzględnić argument, który Crisp przedstawił wcześniej w swoim artykule w celu skrytykowania priorytaryzmu. Argument ten opiera się na eksperymencie myślowym o nazwie „Beverly Hills”.

Zastanówmy się jednak nad przypadkiem, który nazwę Beverly Hills i w ramach którego stajemy przed wyborem podarowania znakomitego wina dwóm grupom jednostek znajdujących się w bardzo dobrej sytuacji.

	10 Bogaczy	10 000 Superbogaczy
Status Quo	80	90
Lafite 1982	82	90
Latour 1982	80	92

[...] Wynika stąd nieadekwatność każdego z omówionych wariantów owej zasady [priorytaryzmu]: z chwilą gdy ludzie osiągną dany poziom, nawet jeżeli ich sytuacja okazuje się mniej korzystna w porównaniu z kimś innym, przysparzanie im korzyści nie jest samo w sobie ważniejsze niż pomaganie komu innemu. Wydaje mi się to prawdziwe nawet wówczas, gdy – jak w przypadku Beverly Hills – wartości użyteczności są identyczne. Innymi słowy, nawet jeżeli korzyści dla bogaczy są równie wysokie jak dla superbogaczy i nawet jeżeli ich liczba jest dokładnie taka sama, to i tak nic nie przemawia moim zdaniem za przyznaniem pierwszeństwa grupie „mniej uprzywilejowanej”. Na tym

<sup>3</sup> Podobne interpretacje stanowiska Crispa można znaleźć w: Benbaji (2006, s. 331), Galewicz (2015, s. 26–27).



poziomie liczy się już tylko użyteczność, znika więc problem wyboru między dwoma scenariuszami podziału (Crisp, 2015, s. 175–176).

Dokładniejsza analiza tego, co przekazuje w swoim argumencie Crisp, jest kluczowa gdyż to właśnie na tym eksperymencie myślowym skupię się w następnych częściach artykułu. Mamy dwie bardzo dobrze sytuowane grupy ludzi; pierwsza to dziesięć zamożnych osób z dobrostanem na bardzo wysokim poziomie, reprezentowanym przez 80 jednostek. Druga grupa to 10 tys. osób, które są jeszcze lepiej sytuowane – ich dobrostan reprezentuje 90 jednostek. Musimy podjąć decyzję dystrybucyjną: czy damy wyśmienite wino (Lafite) pierwszej grupie, co przyniesie im symboliczną korzyść o wartości dwóch jednostek, czy damy inne wino (Latour) drugiej grupie, które także przyniesie im korzyść dwóch jednostek. Możemy zapewnić wino tylko jednej grupie, więc jeśli wybierzemy jedną, druga pozostanie na wyjściowym poziomie dobrostanu – bez luksusowego napoju. Zdaniem Crispa jest intuicyjnie oczywiste, że powinniśmy dać wino drugiej, znacznie liczniejszej grupie. Jego teoria uzasadnia tę intuicję. Według Crispa obie grupy znajdują się zdecydowanie powyżej progu wystarczalności, to znaczy, że bezstronny obserwator nie czułby wobec nich współczucia (nie zasługują one na współczucie), a jak już zostało powiedziane, przy porównywaniu grup powyżej rzeczonego progu decydent powinien kierować się zasadą utylitarystyczną, czyli wybrać drugą grupę superbogaczy, ponieważ jest liczniejsza. To zmaksymalizuje ogólny dobrostan (20 tys. jednostek vs 20 jednostek). Zdaniem autora ani egalitaryzm, ani priorytaryzm nie są w stanie uzasadnić intuicji dotyczącej tego przypadku, więc muszą zostać odrzucone. W następnej części tekstu przedstawię trzy uszczegółowienia przypadku Beverly Hills, aby ponownie rozważyć nasze intuicje na bardziej szczegółowych i przejrzystych przykładach.

## Eksperymenty myślowe<sup>4</sup>

Rozważmy trzy różne sytuacje oparte na przedstawionym schemacie przypadku Beverly Hills, w którym mamy 10 010 zamożnych osób, a wszystkie różnice w poziomie ich dobrobytu są niezasłużone:

	10 Bogaczy	10 000 Superbogaczy
Status Quo	80	90
Lafite 1982	82	90
Latour 1982	80	92

### Scenariusz pierwszy

Wszyscy ludzie wspomniani w rozważanym przypadku istnieją w możliwym świecie  $w_a$ <sup>5</sup>, który jest równoważny naszemu rzeczywistemu światu. Oprócz tych 10 010 bardzo zamożnych osób mamy około 8 mld innych ludzi, z których większość znajduje się w gorszej sytuacji – w tym znaczna liczba w okropnych warunkach, cierpiąc głód i tym podobne klęski. Możemy założyć, że decydent również istnieje w świecie  $w_a$  i ma do wyboru jedynie trzy opcje: Lafite, Latour albo nic. Może więc dać wino bogaczom, superbogaczom, albo nie dać nic nikomu.

<sup>4</sup> Niektóre z moich wersji przypadku opisanego przez Crispa są pod pewnymi względami podobne do złożonych przykładów wykorzystujących postaci Andrea i Becky, przedstawionych przez Temkina (2015, s. 201–202). Różnią się jednak znacząco, ponieważ jedynie modyfikują oryginalny przykład Crispa; przykłady z tego artykułu są „prostsze” i, jak się wydaje, ukazują coś innego niż przykłady Temkina.

<sup>5</sup> Na potrzeby eksperymentów przez możliwy świat rozumiemy realne maksymalne obiekty, tak jak w teorii Lewisa (1995). Są to więc realne byty podobne do naszego świata, które nie wchodzą między sobą w relacje czasoprzestrzenne.

### Scenariusz drugi

Wszyscy ludzie wspomniani w rozważanym przypadku istnieją w możliwym świecie  $w_1$ , który jest zamieszkały wyłącznie przez nich. Tak więc w  $w_1$  istnieje dokładnie 10 010 osób. Możemy założyć, że żyją one w jednym miejscu i tworzą jedną (w tej sytuacji dosłownie globalną) społeczność. Decydent istnieje w innym możliwym świecie,  $w_2$ <sup>6</sup>, i może wybrać jedynie Lafite, Latour albo nic.

### Scenariusz trzeci

W możliwym świecie  $w_3$  istnieje 10 bogaczy. Świat ten zamieszkały jest wyłącznie przez nich, a 10 000 superbogaczy istnieje w innym możliwym świecie,  $w_4$ , również zamieszkanym wyłącznie przez nich. Decydent istnieje w kolejnym możliwym świecie,  $w_5$ , i może wybrać jedynie Lafite, Latour albo nic.

## Intuicyjna ocena

Może wydawać się, że Crisp miał na myśli scenariusz pierwszy. Choć w swoim pierwotnym artykule tego nie precyzuje, przypisanie mu przyjęcia tego scenariusza wydaje się najbardziej naturalną interpretacją. Z pewnością Temkin w ten sposób zinterpretował pierwotny przypadek Beverly Hills, twierdząc, że w jego intuicyjnym osądzie lepiej byłoby wybrać nic, niż dać jakiegokolwiek wino bogaczom lub superbogaczom. Gdybyśmy jednak musieli ofiarować wino którejś grupie, według Temkina powinniśmy dostarczyć Lafite bogaczom. Nadal nie byłoby to dobre rozwiązanie, ale lepsze niż przekazanie Latour superbogaczom

<sup>6</sup> Może się wydawać, że w moich wersjach eksperymentu Crispa używam możliwych światów jako bytów podobnych do „rodzynek w cieście”. Mam świadomość, że takie ich rozumienie jest prawdopodobnie całkowicie błędne, ale nie ma to znaczenia dla argumentu. Uważam, że mój sposób użycia tego pojęcia jest zrozumiały i stosuję go jedynie tymczasowo. Nie mam tutaj ambicji, aby przyczynić się do debaty na temat modalności. Więcej na ten temat zob. Stalnaker (1995).

(Temkin, 2015, s. 196–197). Uważam, że intuicje Temkina mogą wydawać się trafne i prawdopodobnie niektórzy z nas je podzielają, szczególnie tę pierwszą. Jednak dla innych czytelników intuicje, które są bliższe teorii Crispa, mogą być bardziej przekonujące, tj. powinniśmy zastosować zasadę utilitarystyczną i wybrać Latour dla superbogaczy. Nie chcę tu jednoznacznie rozstrzygać tej kwestii, szczególnie że to ujęcie przypadku Beverly Hills nie wydaje się wiarygodne.

W odpowiedzi udzielonej Temkinowi Crisp wyraźnie stwierdza, że zamierzał przedstawić ten przykład w sposób który jest zbliżony do scenariusza drugiego (Crisp, 2003, s. 121). W tym scenariuszu świat składa się wyłącznie z naszych 10 010 osób, a według Crispa intuicyjne jest przekonanie, że powinniśmy wybrać Latour dla superbogaczy. Wszyscy ludzie w tym świecie mają „wystarczająco dużo” i, zdaniem Crispa, wszyscy znajdują się powyżej rzezonego progu. Innymi słowy, bezstronny obserwator nie odczuwałby dla nich współczucia. Czy jednak naprawdę osąd, że „posiadanie wystarczająco dużo” to w pełni obiektywna i stała cecha, jest osądem intuicyjnym? Odpowiedź wydaje się negatywna, a scenariusz drugi jasno to ilustruje. Mamy zamożne społeczeństwo<sup>7</sup>, ale w tym społeczeństwie około 0,1% populacji jest wyraźnie i całkowicie niezasłużenie w gorszej sytuacji niż pozostali. Większość z nas prawdopodobnie powiedziałyby, że to niesprawiedliwe. Z tego powodu dostrzegam silną rację moralną na rzecz wyboru opcji Lafite, a nie Latour. Ta racja moralna może zostać przeważona przez inne (na przykład znaczącą różnicę w liczebności grup), jednak owa możliwość nie wpływa na treść analizowanej intuicji. Twierdzę jedynie, że na poziomie intuicyjnym w scenariuszu drugim bogacze mają pewien – niekoniecznie absolutny – priorytet.

---

<sup>7</sup> Oczywiście jego obywatele są zamożni według naszych kryteriów, wywiedzionych z naszego rzeczywistego świata. Wydaje się bowiem oczywiste, że w w1 nasza grupa określana jako „bogacze” nie jest tak naprawdę bogata, ponieważ bogactwo to prawdopodobnie pojęcie względne. Nie ma to jednak większego znaczenia dla argumentu, ponieważ nasi „bogacze” i tak znajdują się na odpowiednim poziomie dobrobytu.

W tym momencie warto poczynić dwie uwagi. Po pierwsze, powyższy argument można odczytać jako przykład zarzutu obojętności, ogólnego argumentu wysuwanego przeciwko suficyntaryzmowi w stylu Crispa (Shields, 2020, s. 3). Zarzut ten opiera się na obserwacji, że każdy suficyntaryzm tego typu implikuje całkowitą nieistotność wszelkich nierówności, nawet niezasłużonych, o ile wszystkie osoby z eksperymentu znajdują się powyżej progu wystraszalności. Kontrintuicyjność tego założenia staje się oczywista w świetle precyzyjnego opisu zawartego w scenariuszu drugim. Po drugie, należy wspomnieć, że sam Crisp prawdopodobnie nie zgodziłby się, że ten wniosek wynikający z jego teorii rzeczywiście jest sprzeczny z intuicją. Jak wspomniałem, wyraźnie stwierdza, że miał na myśli ujęcie, które odpowiada naszemu scenariuszowi drugiemu, i że, jego zdaniem, intuicyjne jest przekonanie, iż bogacze nie mają żadnego priorytetu w tej sytuacji. Ta rozbieżność intuicji prawdopodobnie wynika z dewaluowania przez Crispa względnej sprawiedliwości (Crisp, 2015, s. 167–169). Nie będę rozwijał tego wątku, ponieważ jest on bardzo skomplikowany i wymagałoby to osobnego artykułu. Ograniczę się do zwrócenia uwagi na jedną kwestię. Hipoteza Crispa na temat genealogii intuicji dotyczących względnej sprawiedliwości jest rodzajem argumentu obalającego. Za pomocą tego argumentu chce on unieważnić znaczenie intuicji wskazujących, że względne położenie życiowe jednostek może być etycznie istotne. Co do zasady argumenty obalające nie są niekontrowersyjne – nawet zdecydowani intuicjoniści mogą się przed nimi bronić (Szutta, 2018, s. 343–394) – a co jeszcze ważniejsze, sam argument Crispa jest raczej nieprzekonujący<sup>8</sup>.

---

<sup>8</sup> Przekonujące rozumowanie przeciwne obalającemu argumentowi Crispa można znaleźć u Temkina (2015, s. 193, n. 9). Temkin twierdzi, że idea Crispa, iż względna sprawiedliwość jest historycznie oparta na zazdrości, zuniwersalizowanej przez współodczuwanie, jest nieuzasadniona. Współodczuwanie wpływa na nasze przekonania lub emocje wtedy i tylko wtedy, gdy już uznajemy stanowisko drugiej osoby za rozsądne. Dlatego, jeśli wysuwasz względem kogoś pretensję, ponieważ uważasz, że twoja sytuacja w porównaniu z jego jest niesprawiedliwie gorsza, mogą

Teraz możemy przejść do scenariusza trzeciego, w którym bogacze i superbogacze żyją w różnych możliwych światach. Wydaje się, że ta interpretacja przypadku Beverly Hills pokazuje, iż pomimo tego, co zostało powiedziane o scenariuszu drugim, teoria Crispa posiada pewien intuicyjny urok. Dość naturalne jest przekonanie, że nasz decydent, obywatel  $w_5$ , który może zdecydować o przekazaniu wina bogatym obywatelom  $w_3$  lub superbogatym obywatelom  $w_4$ , powinien wybrać Latour dla superbogaczy. Powinien więc podjąć czysto utylitarystyczną decyzję<sup>9</sup>.

## W stronę konkluzji

Czy dwa wyżej przedstawione intuicyjne osądy – inny w drugim scenariuszu (bogacze mają pewien priorytet w otrzymaniu wina), a inny w trzecim (bogacze nie mają żadnego priorytetu w otrzymaniu wina) są ze sobą niespójne? Wydaje się, że nie. W teorii Crispa

---

ci współczuć tylko wtedy, gdy najpierw uznam twoją opinię za rozsądną, a nie odwrotnie. Mówiąc inaczej, nie mogę współodczuwać zazdrości, którą uznają za nieuzasadnioną.

<sup>9</sup> Jest wysoce prawdopodobne, że Crisp zgodziłby się z tym wnioskiem. Pisze bowiem: „Narzuca się tu rzecz jasna oczywiste pytanie: ile to jest wystarczająco wiele? Czy moja sugestia, jakoby położenie osób z przykładu Beverly Hills było wystarczająco dobre, nie jest sama w sobie wynikiem zbyt skromnego rozumienia istoty dobrostanu? Wyobraźmy sobie, że nasz bezstronny obserwator zdaje sobie sprawę z istnienia bilionów istot o standardzie życia nieporównanie wyższym od najlepszej choćby sytuowanych ludzi na naszej planecie. Czy zajmie on takie samo jak ja stanowisko wobec Beverly Hills, czy może próg jego współczucia okaże się znacznie wyższy? Trudno przewidzieć odpowiedzi na takie pytania, lecz intuicja i refleksja podpowiadają mi, że – powiedzmy – osiemdziesiąt lat dobrej jakości życia na tej planecie to wystarczająco wiele, niewykluczone nawet, że więcej niż dosyć dla jakiegokolwiek istoty” Crisp (2015, s. 184). Co Crisp ma na myśli w przywołanym fragmencie, jest dalece niejednoznaczne. Prawdopodobnie uważa, że można ustalić obiektywny i statyczny próg dla wszelkiego rodzaju porównań, nawet tych dokonywanych między światami możliwymi. Rozwinę ten wątek w dalszej części tekstu głównego.

mamy jeden obiektywny i statyczny próg, który całkowicie zmienia wagę jednostek i racje brane pod uwagę przy decyzjach alokacyjnych (absolutny priorytet, jeśli porównanie zachodzi między poziomami) i może być używany we wszystkich typach porównań. Dlatego argumentacja z jego wykorzystaniem jest podatna na zarzut obojętności i inne zarzuty z kontrintuicyjnych konsekwencji, które są trudne do zaakceptowania nawet dla zwolenników suficyjentyzmu (Benjabi, 2006, s. 331–333; Shields, 2020, s. 2; Axelsen i Nielsen, 2015, s. 416–417). Możliwa jest jednak modyfikacja tego poglądu, chroniąca go przed zaprezentowanymi kontrargumentami i wspierająca obie powyższe intuicje. W tym celu musimy jednak do pewnego stopnia zgodzić się z ideą względnej sprawiedliwości, a co za tym idzie, przyjąć pluralizm, czyli, mówiąc inaczej, wrócić do doceniania równości.

Wydaje się, że pogląd Crispa, iż obiektywny i statyczny próg wyznaczający absolutny priorytet oraz definiowany przez połączenie pojęć bezstronnego obserwatora i współczucia jest pożądany i wart zachowania. Niemniej jednak powinniśmy dodatkowo wprowadzić drugi próg – dynamiczny, który nie jest linią absolutnego priorytetu<sup>10</sup>.

---

<sup>10</sup> Moja propozycja różni się od tej przedstawionej przez Benbajiego (2006, s. 338–343). Benbaji proponuje wprowadzenie trzech progów, z których najwyższym jest „próg luksusu”. Nie jest to oczywiste, ale wydaje się, że Benbaji uważa, iż wszystkie osoby z przypadku Beverly Hills znajdują się powyżej tego progu, zatem prawdopodobnie powinny być traktowane zgodnie z zasadami utylitarystycznymi. Jak stwierdziłem podczas omawiania drugiego scenariusza, nie zgadzam się z tym. Moja propozycja różni się także od skomplikowanego rozwiązania przedstawionego przez Axelsena i Nielsena (2015). Proponują oni rewizję suficyjentyzmu w sposób uwzględniający wiele progów, jednak nie pionowych, lecz poziomych. Rezygnują z jednej skali, takiej jak „dobrobyt,” i koncentrują się na wielu różnych obszarach, w których należy dbać o znajdowanie się powyżej progu. Ich interpretacja przypadku Beverly Hills jest niestety daleka od jasności. Niemniej wydaje się stosunkowo oczywiste, że w drugim scenariuszu zarówno bogacze, jak i super-bogacze znajdują się powyżej jakiegokolwiek progu (aby to założyć, wystarczy stwierdzić, że są wolni od jakiegokolwiek istotnego ucisku, nawet relatywnego czy politycznego). Jeśli tak jest, to zgodnie z tą reinterpretacją nierówności między tymi dwiema grupami są całkowicie nieistotne, co, moim zdaniem, jest zupełnie

Na przykład gdy pierwszy obiektywny próg zostanie ustawiony na poziomie dobrobytu 25, ale mamy społeczeństwo niesprawiedliwie nierówne, w którym najsłabiej sytuowane grupy znajdują się na poziomie 30, a najlepiej sytuowane na poziomie 50, możemy powiedzieć, że wszyscy w tym społeczeństwie mają wystarczająco dużo w pierwszym sensie, ale nie w drugim. Drugi próg można ustawić zaś np. na poziomie 40 i wówczas każdy człowiek znajdujący się poniżej tego poziomu powinien mieć nieabsolutny priorytet w jakichkolwiek decyzjach dystrybucyjnych. Ów próg jest dynamiczny, więc konsekwentnie prowadzi do równości, ponieważ np. jeśli w tym społeczeństwie wszyscy z czasem znajdą się powyżej poziomu 40, próg przesunie się na poziom 45 i tak dalej.

Możemy uznać, że pierwsza intuicja (odnosząca się do drugiego scenariusza) została wstępnie wyjaśniona, przejdźmy zatem do drugiej oraz do różnicy między nimi. Jeśli w mojej propozycji istnieją dwa progi – pierwszy statyczny a drugi dynamiczny (zależny od początkowego stanu porównywanych grup) – jak mogę uzasadnić wykorzystywanie drugiego progu w scenariuszu drugim, a pomijanie go w trzecim? Możemy odpowiedzieć, że powinniśmy używać drugiego progu tylko wtedy, gdy porównujemy jednostki lub grupy, które dają się sensownie porównywać. Dwa całkowicie różne możliwe światy, pozbawione między sobą relacji czasoprzestrzennych, z dwiema populacjami, które nic o sobie nie wiedzą, itd., są prawdopodobnie najbardziej oczywistym przykładem grup, których nie można sensownie porównać, a pojęcie względnej sprawiedliwości jest w ich przypadku bezsensowne. Bogacz z  $w_3$  nie znajduje się w stosunkowo niesprawiedliwej sytuacji względem superbogacza z  $w_4$ . Natomiast bogacz z  $w_1$  jest w stosunkowo niesprawiedliwej sytuacji w porównaniu z superbogaczem z  $w_1$ . Ustanowienie drugiego progu jest uzasadnione tylko wtedy, gdy grupy są porównywalne; w  $w_1$  mamy do czynienia z jednym społeczeństwem,

---

nieintuicyjne. Zob. Axelsen, Nielsen (2015, s. 413–418, 422–424). Moja propozycja jest prawdopodobnie najbliższa eklektycznej propozycji Shieldsa (2020, s. 6–8), choć wydaje się od niej nieco prostsza.



w  $w_3$  i  $w_4$  istnieją zupełnie odrębne grupy bez żadnego związku ze sobą. W pierwszym przypadku względna sprawiedliwość ma znaczenie, więc możemy używać drugiego progu; w drugim przypadku sensowne orzekanie o względnej sprawiedliwości jest niemożliwe, więc nie możemy go stosować<sup>11</sup>.

Na koniec możemy powrócić z abstrakcyjnego poziomu analiz do naszego rzeczywistego świata. W ramach etyki normatywnej często mamy skłonność do myślenia jedynie w uniwersalnych kategoriach, odnoszących się do wszystkich ludzi lub, jeszcze szerzej, do wszystkich istot czujących. Idea wstępnie zarysowana powyżej może i prawdopodobnie powinna być stosowana w ten sposób. Jednak konieczne są dalsze badania nad odpowiednim ograniczeniem zbioru sensownie porównywalnych grup. Niewykluczone, że na gruncie filozofii politycznej poprawniejsze byłoby ustanowienie i stosowanie drugiego progu raczej w obrębie państw narodowych lub niektórych jednostek ponadnarodowych (takich jak Unia Europejska, Unia Afrykańska czy BRICS) niż globalnie. Na przykład porównywanie Szwajcarii i Sierra Leone, oraz próba wyciągania wniosków na temat płacy minimalnej na podstawie tego porównania byłoby całkowicie pozbawione sensu.

## Bibliografia

**Arystoteles.** (1996). *Dzieła wszystkie* (t. 5), (D. Gromska, tłum.). PWN. (Oryginał został opublikowany ok. 300 p.n.e.)

**Axelsen D. V., Nielsen L.** (2015). Sufficiency as Freedom from Duress. *The Journal of Political Philosophy*, 23(4), 406–426. <https://doi.org/10.1111/jopp.12048>

**Benbaji, Y.** (2006). Sufficiency or Priority? *European Journal of Philosophy*, 14(3), 327–348. <https://doi.org/10.1111/j.1468-0378.2006.00228.x>

---

<sup>11</sup> Moja propozycja wydaje się połączeniem suficyentaryzmu oraz tezy egalitarystycznej. Najbardziej pasującą tu tezę egalitarystyczną jest ta, którą Parfit nazwał „egalitaryzmem deontycznym” – koncentruje się on na sprawiedliwości porównawczej i jest zdolny ograniczać swój zakres do jednej społeczności. Zob. Parfit (2015, s. 116–119, 131–132).

- Crisp, R. (2003). Egalitarianism and Compassion. *Ethics*, 114(1), 119–126.  
<https://doi.org/10.1086/377088>
- Crisp, R. (2015). *Równość, pierwszeństwo i współczucie* (I. Zwiech, tłum.). W: W. Galewicz (red.), *Antologia Bioetyki. Sprawiedliwość w medycynie. Cz. 2: Dystrybucja zasobów w opiece zdrowotnej* (t. 4, s. 163–185). Universitas.
- Frankfurt, H. (2015). *Równość jako ideał moralny* (I. Zwiech, tłum.). W: W. Galewicz (red.), *Antologia Bioetyki. Sprawiedliwość w medycynie. Cz. 2: Dystrybucja zasobów w opiece zdrowotnej* (t. 4, s. 59–87). Universitas.
- Galewicz W. (2015). *Spór o zasady alokacji zasobów zdrowotnych*. W: W. Galewicz (red.), *Antologia Bioetyki. Sprawiedliwość w medycynie. Cz. 2: Dystrybucja zasobów w opiece zdrowotnej* (t. 4, s. 9–34). Universitas.
- Lewis, D. (1995). *Światy możliwe* (U. Żegleń, tłum.). W: T. Szubka (red.), *Metafizyka w filozofii analitycznej* (s. 127–135). Towarzystwo Naukowe Katolickiego Uniwersytetu Lubelskiego.
- Moore, G. E. (1919). *Zasady etyki* (C. Znamierowski, tłum.). M. Arcta.
- Parfit, D. (2015). *Równość i pierwszeństwo* (I. Zwiech, tłum.). W: W. Galewicz (red.), *Antologia Bioetyki. Sprawiedliwość w medycynie. Cz. 2: Dystrybucja zasobów w opiece zdrowotnej* (t. 4, s. 111–132). Universitas.
- Ross, W. D. (2009). *The Right and the Good*. Oxford University Press.
- Shields, L. (2020). Sufficientarianism. *Philosophy Compass*, 15, article 12704.  
<https://doi.org/10.1111/phc3.12704>
- Stalnaker, R. C. (1995). *Światy możliwe* (U. Żegleń, tłum.). W: T. Szubka (red.), *Metafizyka w filozofii analitycznej* (s. 137–154). Towarzystwo Naukowe Katolickiego Uniwersytetu Lubelskiego.
- Szutta, A. (2018). *Intuicje moralne. O poznaniu dobra i zła*. Wydawnictwo Academicon.
- Temkin, L. S. (2015). *W obronie egalitaryzmu* (I. Zwiech, tłum.). W: W. Galewicz (red.), *Antologia Bioetyki. Sprawiedliwość w medycynie. Cz. 2: Dystrybucja zasobów w opiece zdrowotnej* (t. 4, s. 187–210). Universitas.



KRZESIMIR CHOLEWA

 Jagiellonian University (Poland)

 0009-0005-2084-810X

 krzesimir.cholewa@uj.edu.pl

# Sufficient Level and Beverly Hills: Reexamining Intuitions on Roger Crisp's Thought Experiment

## Abstract:

The presented article addresses the issue of distributive justice based on Roger Crisp's thought experiment, the Beverly Hills case. Besides a brief introduction outlining key theoretical distinctions, the article is divided into four parts. The first part introduces Crisp's perspective and the original formulation of the Beverly Hills case. The following two parts explore three specifications of this thought experiment, analyzing their impact on initial intuitions. The final part includes a proposal of pluralist sufficientarianism, incorporating two thresholds: a static, objective threshold grounded in Crisp's view and a dynamic threshold derived from the concept of relative fairness.

## Key words:

distributive justice, sufficientarianism, egalitarianism, prioritarianism, Beverly Hills case

## How to cite:

Cholewa, K. (2024). Sufficient Level and Beverly Hills: Reexamining Intuitions on Roger Crisp's Thought Experiment [English original]. *Laboratorium Mentis*, 2(2), 43–57. <https://doi.org/10.52097/lm.9481>

## 1. Introduction

One of the traditionally recognized types of justice, namely its distributive form, has been known at least since Aristotle (Aristotle, 2009, pp. 84–86). In a probably most simplified manner, we can say that all kinds of theories of distributive justice are made to answer one, at first glance, understandable question: What is the proper way to distribute benefits and burdens among groups of people or individuals? In our world, there are limited resources that can be distributed in various ways; however, some of these ways are unfair, irrational, or outright unjust.

Risking some simplification, we can distinguish two main groups of views in the area of distributive justice:<sup>1</sup>

- (A) Utilitarianism;
- (B) Views that cannot be reduced to consequential welfarism and, in some sense, are associated with somehow understood equality:
  - (B.1) Egalitarianism;
  - (B.2) Prioritarianism;
  - (B.3) Sufficientarianism.

View (A) is very well known as a general moral theory, and it is relatively simple to understand how utilitarian assumptions are applied to the issue of distributive justice. According to this view, distributing benefits in a just way (or maybe the “right way”) consists in distributing them so as to maximize general welfare in the world, assuming that each individual welfare recipient weighs equally. It is commonly known that utilitarianism, as a general view, can be attacked with

---

<sup>1</sup> I do not claim that this classification is in any way complete. However, the highlighted views are probably best developed in literature and primarily used in discussions, e.g., in bioethics.

many arguments (Moore, 2002, pp. 147–160; Ross, 2009, pp. 43–47). The same can be said of its application to the issue of distributive justice. For example, it can be indicated that utilitarianism has non-intuitive conclusions such as “it would be just to cure runny nose in some huge group instead of saving some sufficiently smaller (but still significant) group of young people from premature death.”

Although the views included in (B) are very diverse, we can temporarily, for the purposes of this article, treat them as belonging to one broad group. Most simply, we can present particular views from this group as follows:

(B.1) Equality has intrinsic value, and because of that, in our allocative decisions, we should choose the action that will bring our world closer to equality (Temkin, 2003);

(B.2) In our allocative choices, we should use the “worse off” category and prioritize the worse-off individuals or groups (Parfit, 1997, pp. 213);

(B.3) What is important is not whether the worse off are prioritized, not even equality, but reaching a sufficient level. We must consider whether the compared groups or individuals “have sufficient”; if one of them does and the other does not, we have a strong moral reason to benefit the latter (Frankfurt, 1987; Crisp, 2003a).

All these views can be supported or attacked with many different arguments. One of the best-known is the leveling down argument (Parfit, 1997, pp. 210–211), which serves as a critique of (B.1) and at the same time as an argument in favor of (B.2). It can also be interpreted as support for (B.3).

In this article, I will not examine in detail all these views; I will focus only on one version of (B.3), and to be more precise, I am most

interested in one particular argument supporting (B.3) that originally was aimed to criticize (B.2), but also turns out to challenge (B.1). I have in mind Roger Crisp's thought experiment named Beverly Hills Case. In the first part, I will reconstruct the argument and, on a general level, Crisp's version of sufficientarianism. In the second part, I will present three ways Crisp's original thought experiment can be specified. Next, I will show that those specifications and reexaminations significantly change our initial intuitions. In the final part, I will show two conclusions that can be inferred from our changes in intuitions. Before I proceed, it is worth clarifying one more thing. Distributive justice discussions can have some impact on very different branches of philosophy. My primary concern is general ethics, but at least one of my conclusions could also apply to political philosophy.

## 2. Roger Crisp's sufficientarianism

At the most general level, we can describe sufficientarianism as a view that includes at least two claims:

(i) "Having enough" is intrinsically valuable; so in the allocative decisions, we should distribute resources to secure a sufficient level of it for as many people as possible.

(ii) If someone "has enough," the moral reason to benefit such a person radically differs from the situation where a person would not have enough (sufficiently) (Shields, 2020, pp. 2)<sup>2</sup>

---

<sup>2</sup> The author calls these characteristics (1) positive and (2) shift thesis: "The first is the positive thesis, which states that we have reasons to secure enough that are weighty and non-instrumental [...] The second defining thesis of sufficientarianism is the shift thesis, which states that there is a change or shift in our non-instrumental reasons to benefit or burden persons once they have secured enough."

Therefore, at the most general level, all kinds of sufficientarianism must state that there exists a morally important threshold, and individuals above this threshold should be treated differently in distributive decisions than those below it. Anyone below doesn't have enough, while anyone above does.

Such a general description is ambiguous and must be filled with more particular content. The most apparent questions regarding this description refer to "having enough." Where and how precisely should we set the threshold? And how should we interpret the differences in treating individuals or groups below and above this threshold? Advocates of sufficientarianism may offer various answers to these questions, but our focus will be on Roger Crisp's perspective. He bases his view on the categories of impartial spectator and compassion.

The notion of compassion, then, used in conjunction with the notion of an impartial spectator, may provide us with the materials for an account of distribution which allows us to give priority to those who are worse off when, and only when, these worse off are themselves badly off. [...] Such a view will incorporate an absolute threshold above which priority does not count but below which it does (Crisp, 2003a, p. 757).

Crisp claims that we should set the threshold by determining the level of well-being above which an impartial spectator would not feel compassion for a given individual. This objective well-being level, indicating who is worthy of compassion, separates those who have enough and those who haven't.

The Compassion Principle: absolute priority is to be given to benefit to those below the threshold at which compassion enters. Below the threshold, benefiting people matters more the worse off those people are, the more of those people there are, and the greater the size of the benefit in question. Above the threshold, or in cases concerning only trivial benefits below the threshold, no priority is to be given (Crisp, 2003a, p. 758).

The above quote shows how Crisp understands the difference between treating individuals depending on which side of the threshold they are on. He bases his view on various methods and theories that, in a nutshell, come down to the following formula: if a distributive decision concerns only persons that are below the sufficiency threshold, we should use the principle of weighted priority, taking into account not only who is worse off, but also the number of individuals that will benefit, and what kind of benefit it will be. Thus, for these types of decisions, we must combine prioritarian and utilitarian assumptions. When the decision involves individuals on either side of the threshold, we can, in a sense, adopt absolute prioritarianism, as any non-trivial benefit to someone below the threshold is deemed unequivocally more important than even significant benefits, regardless of their scale or impact on a large population, for those above the threshold. Finally, if the decision involves only individuals above the threshold, we should revert to classical utilitarianism, which considers only the size of the beneficiary group and the magnitude of the benefit.<sup>3</sup>

To some extent, the theory briefly described above is designed to address the argument that Crisp made earlier in his article to criticize prioritarianism. The argument is based on the thought experiment called the Beverly Hills case.

But now consider what I shall call the Beverly Hills case, in which you can offer fine wine to different groups of well-off individuals

	10 Rich	10,000 Super-rich
Status Quo	80	90
Lafite 1982	82	90
Latour 1982	80	92

---

<sup>3</sup> Similar interpretations of Crisp's view are found [in:] Benbaji (2006, p. 331); Galewicz (2015, pp. 26–27).



[...] This implies that any version of the priority view must fail: when people reach a certain level, even if they are worse off than others, benefiting them does not, in itself, matter more. And this seems to me to be true even if, in a Beverly Hills case, the utilities are equal. That is, even if the benefits to each of the Rich and the Super-rich are identical and their numbers are the same, there still seems to me nothing to be said for giving priority to the “worse off.” At this level, only utilities matter, so there would be nothing to choose between the two distributions (Crisp, 2003a, p. 755).

A more detailed presentation of what Crisp claims in his argument is crucial here, as this thought experiment will be the focus of the subsequent sections of this article. We have two very well-positioned groups; the first contains ten wealthy persons with a very high level of general welfare, symbolically represented by 80 units. The second group is 10 thousand people who are even in a better position—90 units symbolically represent their welfare. We have to make a distributive decision: whether we give an excellent wine (Lafite) to the first group that will benefit them in some symbolic two units, or whether we give another wine (Latour) to the second group that will also benefit them by two units. We can provide wine only to one group, so if we choose one, the other will stay on the starting level of welfare—without a luxurious drink. The Crisp view is as follows: intuitively, it is evident that we should give wine to the second, more numerous group. His theory justifies this intuition. In his terms, these two groups are surely above the sufficiency threshold, so in other words, an impartial spectator would not feel compassion for them—they are not worthy of it—and as it was said, when comparing groups above the threshold, decision-maker should base on a utilitarian rule, i.e., choose the group of super-rich because it is larger. It would maximize general welfare (20,000 units vs 20 units). Crisp thinks that because neither egalitarianism nor prioritarianism can justify the intuition about this case, they must be rejected. In the next part, I will present three Beverly Hills case specifications to reexamine our intuitions using more specified and clear examples.

### 3. Thought experiments<sup>4</sup>

Let us consider three different scenarios based on the presented framework of the Beverly Hills case, where we have 10,010 wealthy individuals, and all the differences in their levels of well-being are undeserved:

	10 Rich	10,000 Super-rich
Status Quo	80	90
Lafite 1982	82	90
Latour 1982	80	92

#### First scenario

All the people mentioned in the considered case exist in a possible world  $w_a$ <sup>5</sup> that is equivalent to our actual world. In addition to these 10,010 very wealthy individuals, there are approximately 8 billion other people, most of whom are in worse situations—some in moderately worse conditions, but a significant number live in dire circumstances, suffering from hunger and similar hardships. We can assume that the decision-maker also exists in  $w_a$ , and he can only choose Lafite, Latour, or nothing. Thus, he can give wine to the rich or the super-rich, or he can give nothing to anyone.

---

<sup>4</sup> Some of my specifications of Crisp's case are in some respects similar to the complex examples involving the characters Andrea and Becky, presented by Temkin (2003, pp. 775–776). They are, however, significantly different because they only modify the original Crisp's example; they are more straightforward, and in my opinion, they show something else than Temkin's.

<sup>5</sup> We can assume, for the purposes of the experiments, that by the possible world, we understand real maximal objects like in Lewis's theory—Lewis (1973). Thus, they are real beings similar to our world, and they don't enter into spatiotemporal relations with one another.

### Second scenario

All the people mentioned in the case exist in a possible world  $w_1$  populated only by them. So, in  $w_1$ , exactly 10,010 persons exist. We can assume that they live in one place and are one (in this situation, literally global) society. The decision maker exists in another possible world,  $w_2$ ,<sup>6</sup> and can only choose Lafite, Latour, or nothing.

### Third scenario

10 Rich exist in a possible world  $w_3$  that is populated only by them, and 10,000 Super-rich exist in another possible world  $w_4$  that is populated only by them. The decision maker exists in another possible world  $w_5$ , and he can only choose Lafite, Latour, or nothing.

## 4. Intuitive assessment

At first glance, the first scenario seems to be the one the one Crisp had in mind. Although in his original article, he doesn't clarify that, the first scenario seems to be the most natural specification. Clearly, it was for Temkin, who read the original Beverly Hills case this way and claims that in his intuitive judgment, he would rather choose nothing than give any wine to the rich or super-rich. If we had to give the wine to some group, Temkin thought we should provide Lafite to the rich. It would not be a good solution, but it would be less evil than giving Latour to the super-rich less evil than giving Latour to the super-rich (Temkin, 2003, p. 771). I think Temkin's intuitions can be compelling, and probably some readers share them with him, especially the first

---

<sup>6</sup> It may seem that, in my specifications of the original thought experiment, I use possible worlds as entities similar to "raisins in a pudding." I am aware that this kind of understanding is probably incorrect, but this is irrelevant to the argument. I think my usage is understandable and I use it only provisionally. I don't have any ambitions here to contribute to the debate about modalities. See more: Stalnaker (1984).

one. However, intuitions that align more closely with Crisp's theory might be more compelling for others. I.e., we should apply a utilitarian principle and choose Latour for the super-rich. I do not want to settle this here explicitly, especially since this specification of the Beverly Hills case is highly questionable.

In his reply to Temkin, Crisp says that he intended to present this example similarly to what is specified in the second scenario (Crisp, 2003b, p. 121). In this scenario, the world consist only of 10,010 people, and in Crisp's opinion, it is intuitive to think that we should give Latour to the super-rich. All the people in this world have enough, and in Crisp's opinion, they are all above the threshold. In other words, they are not compassion-worthy; impartial spectator would not feel compassion for them. But is it really so intuitive to think that "having enough" is an entirely objective and constant attribute? The answer seems negative; precisely, our second scenario plainly illustrates this. We have a well-off society,<sup>7</sup> but in this society, approximately 0,1% of the population is noticeably and completely undeservedly worse off than the others. I think that most of us would say that it is unjust. Because of that, I see a strong moral reason for choosing Lafite and not Latour. This moral reason may be outweighed by others (like a significant difference in the size of the group), but this possibility does not affect this intuition. I only claim that on an intuitive level, in the second scenario, the rich have some—not necessarily absolute—priority.

At this moment, it is worth to notice two more things. First, my above argument can be read as an exemplification of the indifference objection, a general argument against Crisp-type sufficientarianism (Shields, 2020, p. 3). This objection is based on the observation that this kind of sufficientarianism implies that all inequalities, even undeserved

---

<sup>7</sup> Of course, they are well-off in our terms, derived from our actual world. It seems evident that in  $w_1$  our group named "rich" is not really rich, because richness is probably a strictly relative concept. It is, however, irrelevant to the argument because our "rich" are already at an adequate level of well-being

ones, are entirely irrelevant if only everyone is above the threshold. How counterintuitive this is becomes evident in the precise description provided in the second scenario. Second, it is necessary to mention that Crisp himself probably would disagree that this is a counterintuitive consequence of his theory. As I mentioned, he explicitly says that he had in mind this specification and that it is intuitive to him that the rich don't have any priority in this situation. This difference in intuitions probably arises from Crisp's devaluation of relative fairness (Crisp, 2003a, pp. 748–750). I would not elaborate on this thread because it is very complicated and would require another article. I will limit myself to pointing out one thing. Crisp's hypothesis about the genealogy of intuitions regarding relative fairness is a kind of debunking argument. Through this argument, he wants to cancel out the significance of the intuition that relative positions of individuals could be ethically meaningful. Debunking arguments are generally not uncontroversial—even robust intuitionists can defend themselves against them (Szutta A., (2018, pp. 343–394)—and, more importantly, Crisp's argument itself is rather unconvincing.<sup>8</sup>

Now, we can proceed to our third scenario, where the rich and super-rich live in different possible worlds. In my opinion, this specification of the Beverly Hills case shows that despite what was said about the second scenario, Crisp's theory possesses some intuitive appeal. I think that, in this situation, it is quite natural to believe that our decision-maker—a citizen of  $w_5$ , who can choose to give wine to the rich citizens of  $w_3$  or the super-rich citizens of  $w_4$ —should opt for

---

<sup>8</sup> Compelling reasoning against Crisp's debunking argument can be found in Temkin (2003, p. 769); footnote 8. Temkin argues that Crisp's idea – that comparative fairness is historically based on envy, generalized by sympathy—is unjustified. Sympathy only impacts our beliefs or emotions when we already consider the proposition of the other person reasonable. Therefore, if you complain to someone because you think that your position compared to him is unfair, I can feel sympathy for you only if I first regard your opinion as reasonable, not conversely. I.e., I cannot empathize with envy that I consider unjustified

Latour for the super-rich. Therefore, the decision should have been based purely on utilitarianism.<sup>9</sup>

## 5. Towards a conclusion

But were these two intuitive judgments—different in the second (rich have some priority in getting wine) and the third (rich haven't any priority in getting wine) scenarios inconsistent? I think not, and now I will try to justify it. In Crisp's theory, we have one objective and static threshold that changes everything (absolute priority if comparison is between levels) and can be used in all types of comparisons. Because of that, Crisp becomes open to indifference and other counterintuitive objections that are hard to accept even by proponents of sufficientarianism (Benjabi 2006, pp. 331–333; Shields, 2020, p. 2; Axelsen, Nielsen, 2015, pp. 416–417). I believe it is possible to modify sufficientarianism to protect it from the presented counter-arguments and justify both of the above intuitions. Nevertheless, to do so, we must agree to some extent with the idea of relative fairness and, consequently, adopt a pluralist view—in other words, return to appreciating equality.

In my opinion, Crisp's view that an objective and static threshold, determining absolute priority and defined by the conjunction of the

---

<sup>9</sup> It is highly likely that Crisp himself would also agree with this conclusion “Then, of course, the obvious question is: How much is enough? Might my suggestion that those in the Beverly Hills case have enough itself be based on too narrow a conception of welfare? Imagine that the impartial spectator knows that the universe contains trillions of beings whose lives are at a much higher level of welfare than even the best off on this planet. Will he or she take the same view of the Beverly Hills case, or might his or her threshold for compassion be set at a much higher level? It is hard to know how to answer such questions, but, on reflection, my own intuition is that, say, eighty years of high-quality life on this planet is enough, and plausibly more than enough, for any being” (Crisp 2003a, p. 762). What Crisp meant in this quote is very ambiguous. He probably thinks that objective and static threshold is established for all kinds of comparisons, even those made between possible worlds. I would elaborate it further in the main text.

notions of an impartial spectator and compassion, is desirable and worth keeping. Nevertheless, we should also introduce a dynamic second threshold that isn't an absolute priority line.<sup>10</sup> Thus, for example, when the first objective threshold is set at the level of welfare 25, but we have an undeservedly unequal society with the worst-off group on the level of 30 and the best-off on the level 50, we can say that all people in this society have enough in first sense, but not in second. For example, the second threshold could be set at level 40, and anyone below this level should have some priority in distributive policies. This second threshold is dynamic, ensuring continuous movement toward equality. For example, if at some point in this society all groups surpass 40, the threshold will shift to 45, and so on.

We can consider the first intuition (the second scenario) to have been preliminarily explained, let us move to the second one and the difference between the two intuitions. Thus, the question arises: if my proposal includes two thresholds—one static and the other dynamic (dependent on the initial state of the compared groups)—how can I justify using the second threshold in the second scenario but not

---

<sup>10</sup> My proposal differs from that of Benbaji (2006, pp. 338–343). Benbaji proposes to introduce three thresholds, and the highest of them is the “luxury threshold”. It may not be obvious, but it seems that Benbaji thinks that all persons from the Beverly Hills case were above the luxury threshold, and, therefore they should be treated according to utilitarian principles. As I stated while discussing the second scenario, I disagree with this. My proposal also differs from the complicated solution by Axelsen, Nielsen (2015). They propose a revision of sufficientarianism by introducing many thresholds, not vertically but horizontally; they abandon a single scale like “welfare” and focus on many distinct areas where being above the threshold should be maintained. Their interpretation of the Beverly Hills case is far from clarity. Still, it is relatively clear that in the second scenario, both the rich and the super-rich are probably above any threshold (to assume this, we must only state that they are free from any essential duress, even relative or political). If so, on the grounds of this reinterpretation, inequalities between these two groups are also completely irrelevant, which is, in my opinion, very unintuitive. See: Axelsen, Nielsen (2015, pp. 413–418, 422–424). My proposition is probably most similar (but it seems a bit simpler) to the eclectic proposition of Shields (2020, pp. 6–8).

in the third? I think that we should use the second threshold only if we compare individuals or groups that can be reasonably compared. Two completely different possible worlds without spatiotemporal relations between them, with two populations that don't know about each other, are possibly the most evident example of groups that cannot be reasonably compared and to which the notion of relative fairness does not apply. The rich man from  $w_3$  is not in a relatively unfair situation compared to the super-rich man from  $w_4$ . However, the rich man from  $w_1$  is in a relatively unfair situation compared to the super-rich man from  $w_1$ . Setting a second threshold is justified only when groups are comparable; in  $w_1$ , there is one society; in  $w_3$  and  $w_4$ , we deal with entirely distinct groups with no association. In the first case, relative fairness is relevant; therefore, we can use the second threshold. In the second case, it is impossible to make meaningful claims about relative fairness; consequently, we shouldn't use it.<sup>11</sup>

Finally, we can return from the abstract level to our actual world. In normative ethics, we often tend to think solely in universal terms, applying to all humans or, more broadly, to all sentient beings. The idea initially outlined above can and probably should be used that way. However, further research is necessary to determine the appropriate limitations for the set of meaningfully comparable groups. It is possible that, within the framework of political philosophy, it would be more appropriate to establish and apply the second threshold at the level of nation-states or certain supranational entities (such as the European Union, the African Union, or BRICS) rather than globally. For example, comparing Switzerland and Sierra Leone and attempting to draw conclusions about the minimum wage based on such a comparison would be entirely meaningless.

---

<sup>11</sup> My proposition seems to be some conjunction of sufficientarianism and some egalitarianism thesis. Probably the closest egalitarian thesis is that which was named by Parfit "Deontic Egalitarianism," which focuses on comparative justice and can narrow scope to one society. See Parfit (1997, pp. 207–210, 220–221).



## Bibliography:

- Aristotle. (2009). *The Nicomachean Ethics* (W. D. Ross, tłum.). Oxford University Press.
- Axelsen, D. V., Nielsen, L. (2015). Sufficiency as Freedom from Duress. *The Journal of Political Philosophy*, 23(4), 406–426. <https://doi.org/10.1111/jopp.12048>
- Benbaji, Y. (2006). Sufficiency or Priority? *European Journal of Philosophy*, 14(3), 327–348. <https://doi.org/10.1111/j.1468-0378.2006.00228.x>
- Crisp, R. (2003a). Equality, Priority, and Compassion. *Ethics*, 113(4), 745–763. <https://doi.org/10.1086/377088>
- Crisp, R. (2003b). Egalitarianism and Compassion. *Ethics*, 114(1), 119–126.
- Frankfurt, H. (1987). Equality as a Moral Ideal, *Ethics*, 98(1), 21–43.
- Galewicz, W. (2015). Spór o zasady alokacji zasobów zdrowotnych. W: W. Galewicz (red.), *Antologia Bioetyki. Sprawiedliwość w medycynie. Cz. 2: Dystrybucja zasobów w opiece zdrowotnej* (t. 4, s. 9–34). Universitas.
- Lewis, D. (1973). Possible Worlds. In D. Lewis, *Counterfactuals* (pp. 84–90). Malden.
- Moore, G. E. (2002). *Principia Ethica*. Cambridge University Press.
- Parfit, D. (1997). Equality and priority. *Ratio*, 10(3), 202–221.
- Ross, W. D. (2009). *The Right and the Good*. Oxford University Press.
- Shields, L. (2020). *Sufficientarianism*. *Philosophy Compass*, 15, article 12704. <https://doi.org/10.1111/phc3.12704>
- Stalnaker, R. (1984). Possible Worlds. In R. Stalnaker, *Inquiry* (pp. 43–58). Bradford Books.
- Szutta, A. (2018). *Intuicje moralne. O poznaniu dobra i zła*. Wydawnictwo Academicon.
- Temkin, L. S. (2003). Egalitarianism Defended. *Ethics*, 113(4), 764–782.



## PIOTR BIŁGORAJSKI

 Katolicki Uniwersytet Lubelski Jana Pawła II (Polska)

 0000-0001-5139-3455

 piotr.bilgorajski@kul.pl

# Czy Bóg klasycznego teizmu posiada poczucie humoru?

### Abstrakt:

W artykule rozważam, czy atrybut poczucia humoru jest zgodny z teistyczną koncepcją Boga, rozumianego jako istota wszechmocna, doskonała i wszechwiedząca. Rozpaczynam od nakreślenia obecnych w języku potocznym i fikcji literackiej intuicji dotyczących przypisywania Bogu poczucia humoru. Następnie omawiam popularne teorie humoru – wyższości, uwalniania napięcia i niezgodności – wskazując na ich niezgodność z teistyczną koncepcją Boga. Na koniec zastanawiam się, dlaczego ludzie mają poczucie humoru, a odpowiedź na to pytanie wyznacza główną tezę artykułu: chociaż Bóg wie, co jest zabawne, nie ma poczucia humoru.

### Słowa kluczowe:

teizm, filozofia dowcipu, śmiech, filozofia Boga

### Jak cytować:

Biłgorajski, P. (2024). Czy Bóg klasycznego teizmu posiada poczucie humoru? [polski oryginał]. *Laboratorium Mentis*, 2(2), 58–68. <https://doi.org/10.52097/lm.9280>

## Wprowadzenie

Rozważania na temat Bożego poczucia humoru najlepiej rozpocząć od żartu. Aby jednak nie odbierać materii rozważań odpowiedniego ciężaru, żart, który przytoczę, będzie się odnosił do spraw bardzo poważnych. Mam nadzieję, że nikogo nie urażę.

Ocalony z Holocaustu umiera śmiercią naturalną w podeszłym wieku i idzie do nieba. Spotyka tam Boga i opowiada Mu żart o Holocaustu. Bóg, usłyszawszy żart, odpowiada, że nie widzi w nim nic śmiesznego. Ocalony wzrusza ramionami i mówi: „Chyba trzeba było tam być, aby to zrozumieć”.

Chociaż ta anegdota pojawia się w wielu miejscach po prostu jako żart, można ją również potraktować jako przypowieść filozoficzną, pewien eksperyment myślowy. Filozoficznym kontekstem są tutaj problem zła i próba zbudowania teodycei po Holocaustu. W opowieści osoba, która przeżyła zagładę, opowiada Bogu żart dotyczący swoich doświadczeń. Chociaż jego treść nie jest nam znana, to żarty z Holocaustu są powszechnie uważane za tabu i Bóg potwierdza to tabu, informując, że nie widzi w tym nic śmiesznego. Na co Ocalały udziela protekcyjnej odpowiedzi: „Bóg nie zrozumiał żartu, ponieważ nie doświadczył wydarzeń, o których ten opowiada”.

Odczytany w ten sposób żart przedstawia w syntetycznej formie jedną z wersji teodycealnego usprawiedliwienia. Bóg nie rozumie żartu, ponieważ nie doświadczył okrucieństwa Holocaustu, a nie posiadając na jego temat wiedzy, nie jest odpowiedzialny za dopuszczenie zła. Taka odpowiedź, gdyby ją potraktować poważnie, byłaby jednak niezgodna z klasyczną koncepcją teizmu, wedle której Bóg jest istotą wszechwiedzącą. Konsekwentny teista, chcąc ocalić klasyczną koncepcję Boga, będzie musiał przyznać, że jeśli Bóg jest wszechwiedzący, to żart Go

nie śmiesz nie z tego powodu, że tego żartu nie rozumie, ale dlatego, że ten żart po prostu nie jest śmieszny<sup>1</sup>.

Hipotetyczna odpowiedź teisty wydaje się rozstrzygać tytułowe pytanie. Skoro bowiem Bóg potrafi rozpoznać i ocenić humor (w tym przypadku – negatywnie), to Bóg wie, co jest śmieszne, a co nie. Czy jednak posiadanie takiej wiedzy lub, innymi słowy, posiadanie prawdziwej teorii humoru jest tożsame z posiadaniem poczucia humoru? Czy samo przypisywanie Bogu jakiegokolwiek rodzaju poczucia nie stoi w sprzeczności z teistycznym obrazem Boga? Być może samo pytanie o humor Boga jest nie tylko niepoważne, ale też źle postawione?

Artykuł zaczynam od uwag metodologicznych dotyczących rozumienia rozbawienia i humoru. Następnie przedstawiam popularne intuicje na temat przypisywania Bogu humoru obecne w języku potocznym oraz w fikcji literackiej. W kolejnej części dokonuję przeglądu podstawowych koncepcji humoru: wyższości, uwalniania napięcia oraz niezgodności, wskazując na ich niekompatybilność z teistyczną koncepcją Boga. W zakończeniu zastanawiam się, po co ludziom poczucie humoru, odpowiedź zaś na to pytanie pozwala na postawienie tezy, że Bóg nie ma poczucia humoru.

## Humor a śmiech

Na początek warto odróżnić od siebie dwa zjawiska: rozbawienie i poczucie humoru. Rozbawienie, czyli śmiech, jest reakcją emocjonalną, która nie musi być z konieczności związana z poczuciem humoru, czyli z dyspozycją do rozpoznawania komizmu. Śmiech bywa często reakcją na bodźce fizjologiczne (np. podczas łaskotania), może być wyrazem zakłopotania (tzw. nerwowy śmiech), a u niektórych osób towarzyszy

---

<sup>1</sup> Teista, gdyby udzielił takiej odpowiedzi, ocaliłby atrybut wszechwiedzy z klasycznej koncepcji Boga, ale kosztem uwikłania się w dużo trudniejszy problem zła.

poczuciu zwycięstwa (śmiech triumfalny). W powyższych przykładach nie potraktujemy rozbawienia jako reakcji na sytuację komiczną.

Śmiech może być także wskaźnikiem, co dana osoba uważa za zabawne. Innymi słowy, humor może pełnić funkcję diagnostyczną, co oznacza, że na podstawie tego, co nas bawi, można wnioskować o rodzaju poczucia humoru osoby rozbawionej. Gdy kogoś na przykład śmieszy sytuacja, gdy ktoś się przewraca na skórce od banana, lub bawią go żarty skatologiczne, powiemy, że ma niewyrobione, infantylne poczucie humoru. Jeśli ktoś opowiada w towarzystwie żarty rasistowskie lub seksistowskie, to w sposób niezamierzony informuje innych o swoich uprzedzeniach. Humor paradoksalny lub autoironiczny są zwykle odbierane jako przejaw wyrobionego poczucia humoru, mówi się wtedy o humorze jako o oznace inteligencji. W konsekwencji niezrozumienie żartu bywa odbierane jako przejaw naiwności lub braku inteligencji. Niezrozumiały żart można jednak komuś wytłumaczyć, co nie musi wcale odebrać żartowi potencjału komicznego. Stąd, jak się wydaje, poczucie humoru to dyspozycja, którą można w sobie doskonalić, podobnie do smaku lub gustu.

Wprowadzę jeszcze jedno rozróżnienie. Czym innym jest wiedzieć, co jest zabawne (czyli mieć teorię humoru), czym innym zaś jest posiadać poczucie humoru (dyspozycję do rozpoznawania zabawnych treści). Fakt, że są to dwie różne sprawy uwidacznia się w sytuacji, gdy pytamy kogoś, dlaczego się śmieje. Większość osób ma problem, aby od razu dostarczyć spójnego wyjaśnienia, co tak właściwie ich śmieszy w konkretnej komicznej scenie czy danym żarcie. Najczęściej wskazuje się, że śmieszność jest po prostu odczuwana. Gdyby jednak uparcie indagować rozbawioną osobę o to, co w usłyszanym żarcie jest takiego śmiesznego, może ona po chwili refleksji podjąć (niechętnie) próbę wyjaśnienia. Kogoś na przykład może śmieszyć, że osoba o wysokim statusie społecznym znalazła się w nietypowej dla siebie, przyziemnej sytuacji, czyli że zabawne jest, gdy ktoś, kto wcześniej stał wyżej od nas w hierarchii społecznej, nagle znalazł się niżej. Komuś innemu zabawna może wydawać się niezgodność między przewidywanym

finałem sytuacji a jej faktycznym, zaskakującym rezultatem. Te i tym podobne wyjaśnienia można uznać za załączek teorii humoru. Dlatego też przyjmuję, że poczucie humoru jest odbiciem teorii humoru, jaką dana osoba (nie zawsze świadomie) wyznaje.

## Powszechne intuicje dotyczące Boskiego poczucia humoru

Wydaje się, że w powszechnym mniemaniu Bóg posiada poczucie humoru, które z punktu widzenia człowieka nie zawsze jest zrozumiałe. W sytuacjach codziennych przejawem Boskiego poczucia humoru nazywane są niezwykle zbiegi okoliczności czy tajemnicze, na pozór zagadkowe wydarzenia o doniosłych skutkach. O boskim poczuciu humoru wnioskuje się także na podstawie wyglądu egzotycznych zwierząt lub roślin – usłyszymy wtedy, że Bóg musi mieć poczucie humoru, skoro stworzył takie gatunki. Czasem nawet samo istnienie człowieka bywa traktowane jako przykład Bożego humoru. Kurt Vonnegut w jednym ze swoich opowiadań spekulował, że Bóg stworzył człowieka i obdarzył go wolną wolą, aby dostarczyć samemu sobie rozrywki. Świat rządzony deterministycznymi prawami jest dla wszechwiedzącej istoty przewidywalny i nudny, natomiast istnienie istot posiadających wolną wolę wymyka się tej przewidywalności i wprowadza do świata element zaskakujący, a przez to zabawny.

Powszechnie akcentuje się także kruchość człowieka wobec Boskiej wszechwiedzy, co wyraża powiedzenie, że człowiek planuje, a Bóg się śmieje. Woody Allen trafnie wyraził tę myśl w jednym z aforyzmów: „Jeśli chcesz rozśmieszyć Boga, opowiedz mu o swoich planach”. W tym kontekście postrzega się Boga jako istotę, która czerpie przyjemność z poczucia wyższości i dla której ludzkie zamiary oraz plany są zabawne w swojej daremności.

Przyjmując chrześcijańską koncepcję Boga, a tym samym założenie, że Bóg stworzył człowieka na swój obraz i podobieństwo, łatwo postawić tezę, iż jeśli człowiek posiada poczucie humoru, to Bóg

także je posiada. Tak twierdzi na przykład Rik Peels (2015). Człowiek posiada jednak bardzo wiele cech, których klasyczny teista nie byłby skłonny przypisywać Bogu. Z tezy, że człowiek został stworzony na podobieństwo Boga, zwykle wyciąga się ostrożny wniosek, iż chodzi o ludzką racjonalność, natomiast kwestię, czy istnieją inne podobieństwa, pozostawia się w sferze domysłów. Peels argumentuje, że jeśli jakieś działanie nie jest w sposób konieczny związane z moralnym złem, to może znajdować się w zasięgu Bożych aktów. Wynika jednak z tego zaledwie, że jest możliwe, iż Bóg posiada poczucie humoru, o ile Jego humor byłby wyłącznie moralnie dobry. Nie wynika natomiast, na czym miałyby polegać.

Wspólnym mianownikiem przywołanych intuicji oraz ludowych porzekadeł jest ludzkie poczucie braku kontroli nad siłami przyrody, stąd zaskakujące i nieprzewidywalne zjawiska przyjęło się nazywać „igraszkami losu”. To odczucie w połączeniu z religijną tezą o podobieństwie ludzi i Boga sprawia, że powstaje pokusa, aby wykorzystać humor jako narzędzie diagnostyczne do zbadania jednego z aspektów natury Boga. Jeśli bowiem przyjmie się istnienie celowości w świecie, to subiektywne poczucie, że zachodzące zjawiska mają charakter zabawny (nawet jeśli jest to czarny humor), skłania do spekulacji, jakie poczucie humoru ma istota za to odpowiedzialna. Obraz, jaki obecny jest w antropomorficznych wyobrażeniach Istoty Najwyższej trudno jednak pogodzić z obrazem Boga prezentowanym w klasycznym teizmie, gdzie zakłada się, że Bóg jest bytem w najwyższym stopniu dobrym.

Wydaje się, że powszechnie intuicje na temat Boga więcej nam mówią o ludzkich przeświadczeniach niż o samym Bogu, dlatego spróbuję do sprawy podejść z innej strony. W kolejnej części przedstawię popularne koncepcje humoru, aby sprawdzić, która z nich jest najbardziej zgodna z teistycznym pojęciem Boga.

## Teorie humoru

Pierwsze próby wyjaśnienia humoru można znaleźć już u Arystotelesa. O humorze pisali też Kant, Kartezjusz, Bergson oraz Schopenhauer. We współczesnej literaturze dotyczącej filozofii humoru wymienia się przede wszystkim trzy konkurencyjne teorie: wyższości, rozładowania napięcia oraz niezgodności.

Zgodnie z teorią wyższości przyjemność związana z humorem dotyczy poczucia wyższości względem innych lub względem przeszłej wersji samego siebie (Scruton, 1987). W tym ujęciu śmiech bierze się z poczucia triumfu, czyli z satysfakcji z bycia lepszym. Jeśli natomiast przyjmiemy założenie, że wyjaśniając zjawisko humoru, traktujemy rozbawienie jako reakcję na rozpoznany komizm, to w świetle teorii wyższości musielibyśmy uznać, iż poczucie bycia lepszym jest nie tylko przyjemne, lecz również zabawne. Teoria wyższości nie wyjaśnia jednak, co jest zabawnego w triumfie. To główny powód, dla którego teoria wyższości nie jest dzisiaj zbyt popularna (Carroll, 2014). Także z punktu widzenia etyki triumfalne rozbawienie wydaje się moralnie podejrzane. Natomiast humor tego typu założony jest w powiedzeniach o śmiechu Boga w reakcji na ludzkie plany. Chociaż interesujące byłoby zbadanie genezy takich przedstawień Boga w ludowej psychologii, to z punktu widzenia filozoficznego teizmu nie można przecież powiedzieć o Bogu, że ma wyższościowe poczucie humoru, czyli że raduje Go cudzy upadek lub że porównuje się On z gorszą wersją samego siebie. Stałoby to w sprzeczności z tradycyjnym poglądem, że Bóg jest bytem niezmiennym – niezmiennie doskonałym oraz absolutnie dobrym.

Zgodnie z teorią uwalniania napięcia ludzie przez śmiech wyzwala ją się z nagromadzonych napięć psychicznych, podobnie jak garnek pod ciśnieniem uwalnia skumulowaną parę. Niektórzy badacze spekulują, że w zaginionej księdze *Poetyki* Arystoteles mógł analizować komedię analogicznie do tragedii, wskazując, że również ta pierwsza służy osiągnięciu *katharsis* (Watson, 2012). W tym ujęciu gdy słyszymy żart, w naszym umyśle tworzy się napięcie, które dopiero uwalnia



usłyszana puenta. Zejściu napięcia towarzyszy przyjemność wyrażana jako śmiech. Krytycy tej teorii wskazują na przestarzałe, „hydrauliczne” rozumienie umysłu, w którym gromadzić by się miały tajemnicze „napięcia psychiczne”. Takie problematyczne założenia dotyczące umysłu są głównym powodem podważania tej teorii (Carroll, 2014). Można jednak bronić jej w osłabionej wersji, ponieważ dobrze ujmuje zjawisko używania humoru do „rozładowania” nerwowych sytuacji społecznych. Chociaż więc można na jej przykładzie wskazać terapeutyczne lub społeczne funkcje humoru, to trudno pogodzić z klasycznym obrazem Boga koncepcję, wedle której używa On humoru, aby rozładować tkwiące w Nim napięcia. W klasycznym teizmie Bóg jest bytem prostym, a więc nie występują w Nim skonfliktowane naprężenia różnych części, oraz wszechmogącym, a więc nie grozi Mu trwanie w stanie napięcia, które domaga się rozprężenia.

Zgodnie z teorią niezgodności istotą humoru, czyli przyczyną rozbawienia, jest dostrzeżenie niezgodności między pewnym stanem rzeczy a naszymi przeświadczeniami lub wyobrażeniami względem tego, jaki ten stan rzeczy zwykle jest lub jaki powinien być (Clark, 1987). W tej koncepcji śmiech to reakcja na zaskoczenie w obliczu nietypowej sytuacji. Oczywiście tak szerokie ujęcie humoru jest narażone na kontrprzykłady – można wskazać mnóstwo sytuacji, w których zaskoczenie wcale nie wywołuje śmiechu, a raczej irytację (np. niespodziewane odwołanie lotu). Natomiast istotne jest to, że każdy przejaw komizmu można zinterpretować w kategoriach niezgodności. Więc chociaż jest to koncepcja „nieszczęlna”, to jednocześnie pozostaje najbardziej ogólna, ponieważ za pomocą jednej kategorii dostarcza narzędzia wyjaśniającego wszystkie zjawiska, które tylko częściowo są tłumaczone przez alternatywne koncepcje (Carroll, 2014).

Współcześnie teoria niezgodności znalazła silne wsparcie w perspektywie ewolucjonistycznej (Hurley i in., 2011). Można tutaj jeszcze raz przywołać podobieństwo między humorem i smakiem. Śmiech jest przyjemnym doznaniem, podobnie jak przyjemnym smakiem jest słodycz. Jednak z ewolucyjnego punktu widzenia cukier ma przyjemny

smak nie dlatego, że jest słodki, ale dlatego, że jest (dla nas) dobry. Nasze kubki smakowe odczuwają smak cukru jako przyjemny, ponieważ cukier jest niezbędny dla naszego organizmu – dostarcza nam energii. Analogicznie przyjemna reakcja na komizm nie wynika z jakiejś fundamentalnej śmieszności ukrytej w żarcie, lecz raczej śmiech jest przyjemnym uczuciem towarzyszącym zidentyfikowaniu niezgodności między tym, co widzimy, a tym, o czym byliśmy wcześniej przeświadczeni. Innymi słowy, śmiech jest nagrodą za dostrzeżenie błędu. Jeśli zaś umiejętność identyfikacji błędów i niezgodności jest korzystna dla naszego przetrwania, to poczucie humoru, jako detektor niespójności, pełni istotną funkcję poznawczą.

Chociaż teoria niezgodności zdaje się najlepiej tłumaczyć humor, to w sposób oczywisty nie pasuje do analizy ewentualnego poczucia humoru Boga. Bóg, którego raduje dostrzeganie błędów, nie będzie posiadał atrybutu wszechwiedzy. Istota wszechwiedząca nie tylko nie popełnia błędów, ale też nie daje się niczym zaskoczyć. Kończąc przegląd teorii humoru, zostajemy z konstatacją, że skoro żadna teoria humoru nie pasuje do teistycznego pojęcia Boga, to dopóki nie pojawi się lepsza teoria, jedynym możliwym wnioskiem jest, iż Bóg nie posiada poczucia humoru.

## Konkluzja

Z perspektywy ewolucjonistycznej sztuka opowiadania dowcipów pełni funkcję nie tylko społeczną, ale też poznawczą – wyrabia nawyk wyszukiwania błędów w naszych przeświadczeniach. Być może jest to pozostałość z dawnych czasów, gdy ludzie, zgromadzeni wokół ogniska, opowiadali sobie rozmaite historie. Opowiadany dowcip, podobnie jak pierwotne mity, został gdzieś wcześniej zasłyszany i bywa powtarzany z nadzieją, że inni usłyszą go po raz pierwszy.

Konwencja dowcipu zakłada, że nikt nie przedstawia siebie jako autora żartu – autor jest zawsze nieznanym. Isaac Asimov w opowiadaniu *Kawalarz* bawi się ideą, że jeśli dowcip musi mieć autora, a łańcuch

kolejnych powtórzeń żartu nie może być w nieskończoność, to autorem pierwszego dowcipu jest z konieczności Bóg. Bohaterowie opowiadania próbują odtworzyć pierwszy opowiedziany przez Boga dowcip, który – jak podejrzewają – posłużył Bogu do wywyższenia małpy na poziom człowieka. W opowiadaniu Asimova uczeni rozwiążą zagadkę, ale wraz z poznaniem pierwszego dowcipu znika też tajemnica humoru – od tej pory nic już nie jest śmieszne<sup>2</sup>. W kontekście teorii niespójności puenta tej historii brzmi jak ostrzeżenie – wraz z utratą umiejętności dziwienia się światem być może stracimy też powody do śmiechu.

Artykuł dotyczył pytania, czy Bóg klasycznego teizmu posiada poczucie humoru. Przegląd popularnych teorii humoru doprowadził zaś do odpowiedzi negatywnej. Teoria wyższości jest trudna do pogodzenia z Bożą doskonałością, teoria uwalniania napięcia nie zgadza się z założeniem o Bożej prostocie i wszechmocy, natomiast teoria niezgodności nie pasuje do pojęcia Boga jako istoty wszechwiedzącej. Jeśli zgodzimy się, że teoria niezgodności stanowi koncepcję wyjaśniającą poczucie humoru w sposób najbardziej przekonujący, to stanie się oczywiste, iż Bóg nie potrzebuje „detektora niespójności” – dla istoty wszechwiedzącej nic nie jest przecież zaskakujące.

Jednak poczucie humoru i teoria humoru to dwie różne rzeczy. Teoria humoru nie tylko wyjaśnia poczucie humoru, lecz również pozwala oceniać poszczególne żarty jako śmieszne lub nie, podobnie jak teoria

---

<sup>2</sup> Czytelnik opowiadania nie dowiaduje się, jak brzmiał pierwszy dowcip, ale Slavoj Žižek ironicznie wskazuje, że chrześcijanie skądinąd wiedzą, jaki żart opowiedział Bóg pierwszemu człowiekowi, chociaż nie jest oczywiste, co jest w nim takiego zabawnego. Ów dowcip, czyli pierwsza w Biblii wypowiedź Boga skierowana bezpośrednio do człowieka, brzmiał tak: „Z wszelkiego drzewa tego ogrodu możesz spożywać według upodobania; ale z drzewa poznania dobra i zła nie wolno ci jeść, bo gdy z niego spożyjesz, niechybnie umrzesz” (Žižek, 1992); tłumaczenie cytatu za: Zespół biblistów polskich z inicjatywy Benedyktynów Tynieckich (red., 1991). *Pismo Święte Starego i Nowego Testamentu w przekładzie z języków oryginalnych*. Wydawnictwo Pallottinum.

moralna pozwala oceniać, co jest dobre. Jeśli Bóg klasycznego teizmu istnieje, to (raczej) nie ma poczucia humoru, natomiast wie, co jest, a co nie jest zabawne. Zaskakującą konsekwencją teizmu jest więc teza, że humor to zjawisko o charakterze obiektywnym. W tym kontekście opowiadanie dowcipów stanowi ćwiczenie się w sztuce wykrywania fałszu, co jest przecież warunkiem poznania. Tak więc humor, jako siła napędowa umysłu, to poważna sprawa, dotyczy wszak kwestii prawdy. Tutaj nie ma się z czego śmiać.

## Bibliografia

Carroll, N. (2014). *Humour: A Very Short Introduction Get*. Oxford University Press.

Clark, M. (1987). Humor and Incongruity. In J. Morreall (Ed.), *The Philosophy of Laughter and Humor* (pp. 139–155). State University of New York Press.

Hurley, M. M., Dennet, D., Adams, R. (2011). *Inside Jokes: Using Humor to Reverse-engineer the Mind*. The MIT Press.

Peels, R. (2015). Does God Have a Sense of Humor? *Faith and Philosophy: Journal of the Society of Christian Philosophers*, 32(3), Article 2. <https://doi.10.5840/faithphil20157742>

Scruton, R. (1986). Laughter. In J. Morreall (Ed.), *The Philosophy of Laughter and Humor* (pp. 156–171). State University of New York Press.

Watson, W. (2012). *The Lost Second Book of Aristotle's Poetics*. University of Chicago Press.

Žižek, S. (1992). *Looking Awry. An Introduction to Jacques Lacan through Popular Culture*. The MIT Press.



## PIOTR BIŁGORAJSKI



The John Paul II Catholic University of Lublin (Poland)



0000-0001-5139-3455



piotr.bilgorajski@kul.pl

# Does the God of classical theism have a sense of humour?



### Abstract:

In this paper, I discuss whether the attribute of a sense of humour is compatible with the theistic conception of God, understood as an omnipotent, perfect and omniscient being. I begin by outlining the intuitions present in common language and in fiction regarding the attribution of humour to God. I then discuss popular theories of humour—superiority, release and incongruity—pointing out their incompatibility with a theistic conception of God. I conclude by asking why humans have a sense of humour, the answer to which marks the main thesis of the paper: although God knows what is funny, he does not have a sense of humour.



### Keywords:

theism, philosophy of humour, laughter, philosophy of God



### How to cite:

Biłgorajski, P. (2024). Does the God of classical theism have a sense of humour? [English translation]. *Laboratorium Mentis*, 2(2), 69–79. <https://doi.org/10.52097/lm.9483>

## Introduction

A consideration of God's sense of humour is best begun with a joke. However, so as not to deprive the subject of its proper weight, the joke I am going to quote will relate to serious matters. I hope I will not offend anyone.

A Holocaust survivor dies a natural death at an advanced age and goes to Heaven. There, he meets God and tells Him a joke about the Holocaust. Having heard the joke, God replies that He sees nothing funny in it. The survivor shrugs his shoulders and says: 'I guess you had to be there to get it.'

Although this story appears in many places simply as a joke (I couldn't find the author, but then again, do jokes need a bibliographical footnote?), it could be considered a philosophical parable, a thought experiment. The philosophical context of the story is the problem of evil and the attempt to construct a theodicy after the Holocaust. In the story, a person who experienced the Holocaust tells God a joke about it. Although the content of this particular joke is not known to us, jokes about the Holocaust are generally considered taboo, and God confirms this taboo by saying that He sees nothing funny in it. In response to this, the Survivor gives a condescending reply: God did not understand the joke because He did not experience the events that the joke is about.

Read in this way, the joke synthetically presents one version of theodical justification. God does not understand the joke because He did not experience the atrocities of the Holocaust, and having no knowledge of them, He is not responsible for allowing this evil. Such an answer, however, if taken seriously, is incompatible with the classical concept of theism, according to which God is an omniscient being. A consistent theist, wishing to save the classical conception of God, will have to admit that if God is all-knowing, then a joke is not funny to

Him, not because God does not understand the joke, but because the joke is simply not funny.<sup>1</sup>

The theist's hypothetical answer seems to settle the title question. Since God can recognise and judge humour (in this case, negatively), God knows what is and what is not funny. But is having such knowledge or, in other words, having a true theory of humour the same as having a sense of humour? Does not the very attribution of some kind of sense to God contradict the theistic image of God? Perhaps, then, the very question of God's humour is not only non-serious but also misplaced?

I begin the paper with methodological remarks on understanding amusement and humour. I will then present popular intuitions about the attribution of humour to God present in common language and in fiction. In the next section, I will review the basic theories of humour—superiority, release and incongruity—pointing out their incompatibility with a theistic conception of God. Finally, I will consider why people need a sense of humour; the answer to this question will establish the main thesis of the article: God has no sense of humour.

## Humour and laughter

It is worth distinguishing between two phenomena: amusement and humour. Laughter (amusement) is an emotional response that is not necessarily linked to a sense of humour, that is, to a disposition to recognise humour. Laughter is often a reaction to physiological stimuli (e.g. when tickled), it can be an expression of embarrassment (so-called nervous laughter), and in some people, it accompanies a sense of victory (triumphant laughter). We will not treat laughter (or amusement) of these kinds as a reaction to a comic situation.

---

<sup>1</sup> The theist, if he were to give such an answer, would save the attribute of omniscience in the classical conception of God from the charge that God did not understand the joke, but at the cost of becoming entangled in the much more difficult problem of evil.

Laughter can also be an indicator of what a person finds funny. In other words, what amuses us can reveal the nature of our sense of humour. For instance, if someone finds it funny when a person slips on a banana peel or finds scatological jokes amusing, we will say that they have an unrefined, infantile sense of humour. If someone tells racist or sexist jokes, they are unintentionally informing others of their prejudices. Paradoxical or self-deprecating humour is usually perceived as a sign of a cultivated sense of humour, humour is then said to be a sign of intelligence. Consequently, not getting a joke is sometimes perceived as a sign of naivety or lack of intelligence. On the other hand, a misunderstood joke can be explained to someone, which does not necessarily take away the comic potential of the joke. Hence, a sense of humour seems to be something that can be developed in oneself, similar to taste or gusto.

Let me make another distinction. It is one thing to know what is funny (that is, to have a theory of humour) and another to have a sense of humour (a disposition to experience a given piece of humour as funny). We can see that these are two different things when we ask someone why they are laughing. Most people have trouble immediately providing a coherent explanation of what actually makes them laugh. The most common answer indicates that the funny is something one simply feels. However, if one were to persistently inquire as to what is so funny here, after a moment's reflection there might be a (reluctant) attempt at an explanation. Someone, for example, might find it funny when a person of high social status has found himself in an unusual, mundane situation, e.g., when someone who previously stood higher than us in the social hierarchy suddenly finds himself lower. Someone else might find humour in the disparity between how we expected a situation to unfold and its actual, surprising outcome. These and similar explanations can be seen as the seeds of a theory of humour. Therefore, I assume that a sense of humour reflects the theory of humour that a person (not always consciously) embraces.



## Common intuitions about God's sense of humour

It seems to be a common perception that God has a sense of humour, which, from a human point of view, is not always understandable. In everyday situations, unusual coincidences or mysterious, seemingly puzzling events with significant consequences are called manifestations of the divine sense of humour. The divine sense of humour is also inferred from the appearance of exotic animals or plants, in which case we hear that God must have a sense of humour if he created such species. Sometimes even the very existence of humans is taken as an example of God's humour. Kurt Vonnegut speculated in one of his short stories that God created man and endowed him with free will in order to provide himself with entertainment. A world governed by deterministic laws is predictable and boring to an omniscient being, whereas the existence of beings with free will would escape this predictability and introduce an element of surprise, and thus amusement, into the world.

Common intuitions also emphasise the fragility of man in the context of divine omniscience, as expressed by the saying that man plans and God laughs or, as Woody Allen stated in one of his aphorisms—if you want to make God laugh, tell him your plans. In this context, God is seen as a being who derives pleasure from a sense of superiority and to whom human intentions and plans are amusing in their futility.

In the context of the Christian conception of God, it is easy to put forward the thesis that if God created man in his own image and likeness, then if man has a sense of humour, God also has one (See, for example, Peels, 2015). On the other hand, man possesses a great many properties that a classical theist would not be willing to attribute to God. From the thesis that humans were created in the likeness of God, the cautious conclusion is usually drawn that we are similarly rational, while whether other likenesses exist is left to speculation. Peels argues that if an action is not necessarily associated with moral evil, it cannot

be excluded that it is within the scope of God's acts. It merely follows that it is possible for God to have a sense of humour provided that His humor is exclusively morally good or at least neutral. However, it does not follow what such humor would consist of.

Common to these intuitions and folk proverbs is a sense of lack of control over the forces of nature, hence surprising and unpredictable phenomena have come to be known as tricks of fate. This, combined with the religious thesis of the similarity between humans and God, makes it tempting to apply the diagnostic function of humour to God. For if one accepts the existence of intentionality in the world, the subjective sense that the phenomena taking place is funny (even if it is black humour), prompts speculation about what kind of sense of humour the being responsible for it might have. The image present in anthropomorphic depictions of the supreme being is, however, difficult to reconcile with the image of God presented in classical theism, which assumes that God is a being of the highest moral goodness.

It seems that popular intuitions about God tell us more about human beliefs than about God himself, so I will try to approach the issue from a different angle. In the next section, I will present well-established theories of humour to see which is most compatible with a theistic conception of God.

## Theories of humour

The first attempts to explain humour can already be found in Aristotle. Kant, Descartes, Bergson and Schopenhauer wrote about humour. The contemporary literature on the philosophy of humour mainly mentions three competing approaches: the superiority theory, the release theory, and the incongruity theory.

According to superiority theory, the pleasure associated with humour is based on a sense of superiority to others or to a past version of oneself (Scruton 1987). In this view, laughter derives from a sense of triumph or the satisfaction of being superior. If, on the

other hand, we assume that, in explaining the phenomenon of humour, we treat amusement as a reaction to recognised comedy, then, in the light of superiority theory, we would have to acknowledge that the feeling of superiority is not only pleasurable but also funny. However, superiority theory does not explain what is funny about triumph. This is the main reason why superiority theory is not very popular (Carroll 2014). Also from an ethical point of view, such triumphal amusement seems morally suspect. This kind of intuition is also present in sayings about God's laughter in response to human plans. Although it would be interesting to explore the origins of such representations of God in folk psychology, from the point of view of philosophical theism, it cannot be said of God that he has a superior sense of humour, that he rejoices in others' failings or that he compares himself with an inferior version of himself. This would contradict the traditional view that God is an unchanging being - unchangeably perfect and absolutely good.

According to the release theory, people release accumulated psychological tensions through laughter, just as a pressurised pot releases accumulated steam. Some scholars speculate that in the lost book of *Poetics*, Aristotle may have analysed comedy analogously to tragedy, indicating that it serves to achieve catharsis (Watson 2012). In this view, when we hear a joke, tension is created in our mind, which is only released by the punchline. The release of tension is accompanied by pleasure expressed as laughter. Critics of this theory point to an outdated 'hydraulic' understanding of the mind, in which mysterious 'mental tensions' would accumulate. Such problematic assumptions about the mind are the main reason for challenging the theory (Carroll 2014). However, one could defend a weakened version of this theory, on the grounds that it captures well the phenomenon of using humour to 'defuse' nervous social situations. Nevertheless, while the therapeutic or social functions of humour can be identified, the vision of a God who uses humour to relieve the tensions inherent in Him seems difficult to reconcile with the classical image of God. In classical theism, God is a simple and omnipotent being, which means there are no conflicting

tensions between different parts in Him, and He is never in a state of tension demanding resolution.

According to incongruity theory, the essence of humour, or the cause of amusement, is to perceive an incongruity between a certain state of affairs and our beliefs about what that state of affairs usually is or should be (Clark 1987). In this conception, laughter is a reaction of surprise in the face of an unusual situation. Of course, such a broad account of humour is vulnerable to counter-examples—one can point to plenty of situations in which surprise does not provoke laughter at all, but rather irritation (e.g. an unexpected flight cancellation). On the other hand, the important point is that any display of comedy can be interpreted as incongruity. Thus, although somewhat ‘leaky’ this account is at the same time the most general, because, with a single category, it provides an explanatory tool for all phenomena that are only partially explained by alternative theories (Carroll 2014).

Nowadays, the theory of incongruity has found strong support from the evolutionist perspective (Hurley, Dennett, Adams 2011) The similarity between humour and taste can be recalled here once again. Laughter is a pleasant sensation, just as sweetness is. However, from an evolutionist point of view, sugar has a pleasant taste not because it is sweet, but because it is beneficial to us. Our taste buds perceive the taste of sugar as pleasant because sugar is essential for our bodies—it provides us with energy. Similarly, the pleasurable reaction to comedy is not due to some fundamental funniness hidden in the joke, but rather laughter is a pleasurable sensation that accompanies the identification of an incongruity between what we see and what we were previously convinced of. In other words, laughter is the reward for spotting an error. And if the ability to identify errors and inconsistencies is beneficial to our survival, then a sense of humour, as a detector of inconsistencies, has an important cognitive function, too.

Nevertheless, while the theory of incongruity seems to best explain humour, it clearly does not fit into an analysis of God’s possible sense of humour. A God who finds delight in spotting mistakes will

not possess the attribute of omniscience. An omniscient being not only does not make mistakes, but nothing would surprise him. Thus, concluding the overview of theories of humour, we are left with the constation that since no theory of humour fits the theistic notion of God, until a better theory emerges the only possible conclusion is that God does not possess a sense of humour.

## Conclusion

From an evolutionist perspective, the art of telling jokes has not only a social function but also a cognitive one—it develops the habit of finding errors in our beliefs. Perhaps this is a remnant from the old days when people, gathered around a campfire, and told each other various stories. The joke being told, much like the original myths, was heard somewhere before and is repeated in the hope that others will hear it for the first time.

The convention of the joke is that no one presents themselves as the author of the joke—the author is always unknown. Isaac Asimov, in his short story, ‘The Jokester’ plays with the idea that if a joke must have an author, and the chain of successive repetitions of a joke cannot run indefinitely, then the author of the first joke is necessarily God. The protagonists of the story attempt to recreate the first joke told by God, which they suspect was used by God to elevate the ape to the level of man. In Asimov’s story, the scientists solve the riddle, but with the knowledge of the first joke, the mystery of humour disappears—from then on, nothing is funny anymore.<sup>2</sup> In the context of the

---

<sup>2</sup> The reader of the story does not find out what the first joke sounded like, but Slavoj Žižek ironically points out that Christians would otherwise know what joke God told the first man, although it is not obvious what is so funny about it. This joke—that is, God’s first statement in the Bible addressed directly to man—would be: „You are free to eat from any tree in the garden; but you must not eat from the tree of the knowledge of good and evil, for when you eat from it you will certainly die” (Žižek, 1992).

incongruity theory, the punchline of this story sounds like a warning—as we lose the ability to wonder at the world, we may also lose the reasons to laugh.

The paper addressed the question of whether the God of classical theism has a sense of humour and a discussion of popular theories of humour led to a negative answer. The superiority theory is difficult to reconcile with God's perfection. The release theory does not fit with the assumption of God's simplicity and omnipotence. Likewise, the incongruity theory does not fit with the notion of God as an omniscient being. If we agree that the incongruity theory most convincingly explains the sense of humour, then it is clear that God does not need a 'detector of inconsistencies'—after all, nothing is surprising to an omniscient being.

On the other hand, a sense of humour and a theory of humour are two different things. A theory of humour not only explains a sense of humour but it also allows one to judge particular jokes as funny or not, just as a moral theory allows one to judge what is good. If the God of classical theism exists, then He (probably) has no sense of humour. But He knows what is and what is not funny. A surprising consequence of theism is that humour is objective. Telling jokes may be an exercise in the art of detecting falsehood, which is, after all, a prerequisite for knowledge. So humour, as a driving force of the mind, is a serious matter; after all, it concerns the question of truth. There is nothing to laugh at.

## Bibliography

Carroll, N. (2014). *Humour: A Very Short Introduction*. Oxford University Press.

Clark, M. (1987). Humor and Incongruity. In J. Morreall (Ed.), *The Philosophy of Laughter and Humor* (pp. 139–155). State University of New York Press.

Hurley, M. M., Dennet, D., & Adams, R. (2011). *Inside Jokes: Using Humor to Reverse-engineer the Mind*. The MIT Press.

Peels, R. (2015). Does God Have a Sense of Humor? *Faith and Philosophy: Journal of the Society of Christian Philosophers*, 32(3), Article 2. <https://doi.10.5840/faithphil20157742>

Scruton, R. (1986). Laughter. In J. Morreall (Ed.), *The Philosophy of Laughter and Humor* (pp. 156–171). State University of New York Press.

Watson, W. (2012). *The Lost Second Book of Aristotle's Poetics*. University of Chicago Press.

Žižek, S. (1992). *Looking Awry. An Introduction to Jacques Lacan through Popular Culture*. The MIT Press.



ARTUR SZUTTA

 Uniwersytet Gdański (Polska) 0000-0002-1062-4808 artur.szutta@ug.edu.pl

# Czy kłamiąc z konieczności, głosimy fałsz?

## Abstrakt:

Artykuł podejmuje pytanie, czy kłamstwo wymaga ze swojej istoty wyrażania obiektywnego fałszu. Podczas gdy tradycyjne ujęcia subiektywistyczne skupiają się na rozbieżności między przekonaniami nadawcy a jego wypowiedziami, niektórzy współcześni teoretycy kłamstwa twierdzą, że kłamstwo wymaga wypowiedzi, które są obiektywnie fałszywe. Proponuję krytyczną analizę sześciu eksperymentów myślowych, aby naświetlić intuicyjność stanowiska subiektywistycznego. Wyjaśniając intuicyjność tradycyjnego ujęcia kłamstwa, wskazuję na istotowo moralny charakter kłamstwa, tj., że jest ono zakorzenione w samych aktach woli, a nie w zewnętrznych skutkach.

## Słowa kluczowe:

kłamstwo, subiektywizm, obiektywizm, eksperyment myślowy

## Jak cytować:

Szutta, A. (2024). Czy kłamiąc z konieczności, głosimy fałsz? [polski oryginał]. *Laboratorium Mentis*, 2(2), 80–91. <https://doi.org/10.52097/lm.9484>



## Wstęp

Jednym z diskutowanych problemów związanych z pojęciem kłamstwa jest to, czy kłamstwo koniecznie obejmuje głoszenie sądu obiektywnie fałszywego. Zgodnie z tradycyjnym stanowiskiem z kłamstwem mamy do czynienia, gdy (przy założeniu, że spełnione są inne jeszcze warunki kłamstwa) kłamiący wyraża sąd, który uważa za fałszywy. W tym przypadku decydujący jest fakt niezgodności między twierdzeniem a przekonaniem posiadanym przez twierdzącego. Stąd tradycyjne stanowisko w kwestii omawianego tutaj warunku jest określane mianem subiektywizmu.

W ostatnich latach pojawiają się publikacje, których autorzy głoszą stanowisko obiektywistyczne, zgodnie z którym jednym z warunków koniecznych kłamstwa jest głoszenie obiektywnego fałszu. Oznacza to, że gdy np. jakaś osoba błędnie przekonana, że w jej domu w danej chwili nikogo nie ma, powie: „W moim domu ktoś jest”, nie będzie kłamała, jeśli w jej domu rzeczywiście ktoś jest (w momencie wypowiedzi). Na rzecz tego stanowiska formułowanych jest kilka argumentów. Niektórzy (Turi i Turi, 2015; Turi, 2021; Carson, 2006) odwołują się do intuicji, inni – do samej natury kłamstwa, twierdząc, że kłamstwo z definicji domaga się posiadania pewnego potencjału do skutecznego oszustwa, a tylko wygłaszając (lub implikując) obiektywny fałsz, mamy szansę zwieść adresata kłamstwa (zob. Grimaltos i Rosell, 2013; Kallestrup, 2023). Jeszcze inni argumentują, że kłamstwo musi polegać na głoszeniu obiektywnych fałszów – na mocy reguł konstytutywnych dla praktyki twierdzenia (zob. np. Teichmann, 2024).

W niniejszym artykule skupię się na jednym wątku wspomnianej powyżej debaty, mianowicie na pytaniu, które z wymienionych stanowisk, subiektywizm czy obiektywizm, jest stanowiskiem bardziej intuicyjnym. W celu sformułowania przekonującej odpowiedzi zaproponuję rozwiązanie sześciu eksperymentów myślowych, z których dwa, stanowiące punkt wyjścia, będą eksperymentami zaproponowanymi przez Neriego Marsilio (2021); pierwszy odwołuje się do tzw. scenariusza złego demona, drugi ukazuje możliwość kłamstwa w sytuacji niepewności. Kolejne

eksperymenty będą miały na celu uwyrażnienie intuicyjności subiektywistycznego stanowiska w kwestii kłamstwa. Ostatecznym wyjaśnieniem tych intuicji wydaje się fakt, że terminy takie jak „kłamstwo” są istototwo moralne i odnoszą się przede wszystkim do aktów woli, w drugiej zaś kolejności (i nie zawsze koniecznie) do zewnętrznych skutków.

## Badania nad intuicjami ludzi

Zanim przejdę do prezentacji zapowiedzianych eksperymentów, poświęcę kilka słów argumentacji, wedle której obiektywistyczne stanowisko jest bardziej zgodne z intuicjami większości zwykłych ludzi. John i Angela Turi przeprowadzili badania empiryczne polegające na przedstawieniu ich uczestnikom scenariusza, w którym pewna osoba, Karen, na pytanie pracowników służb rządowych, gdzie się znajduje jej znajoma Mary, odpowiada, że jest w sklepie. Karen mówi tak, mimo że myśli, iż jej znajomej tam nie ma. Następnie okazuje się, że Mary jednak jest we wskazanym miejscu (Turi i Turi, 2015; Turi i Turi, 2021; Turi, 2021).

W jednej z wersji eksperymentu badacze zadawali uczestnikom proste pytania z możliwością odpowiedzi „tak” albo „nie”, np.: „Czy bohaterka historii (Karen) skłamała?”. Większość odpowiadała zgodnie z intuicjami subiektywistycznymi. Badacze postawili hipotezę, że być może subiektywistyczny wynik był zdeterminowany dwoma czynnikami: (i) nieświadomym utożsamianiem się badanych z perspektywą bohaterki scenariusza oraz (ii) nieodróżnianiem kłamstwa od jego nieudanej próby. W zmodyfikowanych wersjach badania odpowiedzi, które mogli wybrać respondenci były bardziej złożone: (a) „Skłamał i powiedział obiektywny fałsz”; (b) „Próbował skłamać, ale powiedział prawdę”. W tym przypadku większość wybierała odpowiedź obiektywistyczną.

Badania Turich poddali krytycznej ocenie Wiegmann i jego współpracownicy (Wiegmann, Samland i Waldmann, 2016; Wiegmann i Meibauer, 2019; Wiegmann 2023), którzy przeprowadzili własne badania empiryczne dotyczące intuicji zwykłych ludzi na temat kłamstwa. W jednym z eksperymentów grupa badawcza Wiegmannna zadała

uczestnikom pytania, do których przygotowali dwie opcje odpowiedzi o podobnym schemacie, jak w badaniach Turich: (a) „Próbował i rzeczywiście tego dokonał”; (b) „Próbował, ale obiektywnie tego nie dokonał”. W przypadku interesującego nas eksperymentu pytania dotyczyły złożenia obietnicy. Bohater historii prezentowanej badanym składał obietnicę, której – w zależności od scenariusza – udawało mu się dotrzymać albo nie.

Co ciekawe, w przypadku scenariusza, w którym bohater nie dotrzymywał słowa, na pytanie, czy bohater rzeczywiście złożył obietnicę, czy jedynie próbował ją złożyć (zauważmy, że w pytaniu nie padało sformułowanie „dotrzymać obietnicy”), większość badanych odpowiadała, że jedynie próbował złożyć obietnicę. Wiegmann interpretuje ten wynik jako efekt torowania wywołany przez konstrukcję pytań, która mogła wprowadzać uczestników w błąd, skłaniając ich do wyboru opcji: „Próbował, ale obiektywnie tego nie dokonał”, nawet w sytuacjach, gdy bohater rzeczywiście złożył obietnicę. Możliwe zatem, że badani w eksperymentach Turich również ulegali tego rodzaju mechanizmowi, a więc ich odpowiedzi nie mogą stanowić wiarygodnej przesłanki dla tezy, że obiektywistyczne stanowisko jest bardziej intuicyjne dla zwykłych ludzi.

Inne eksperymenty Wiegmann wprost wspierają tezę subiektywistyczną. Wiegmann przeprowadził badania, w których uczestnicy byli proszeni o proste odpowiedzi na pytania typu: „Czy to jest kłamstwo?”. Scenariusze przedstawiały sytuacje, w których mówca miał intencję oszukania, ale nie wiadomo było, czy wypowiedziane zdanie jest obiektywnie prawdziwe, czy fałszywe. W każdym przypadku większość uczestników uznawała wypowiedź za kłamstwo, jeśli tylko mówca miał intencję wprowadzenia w błąd. Ocena pozostawała przy tym niezależna od obiektywnej prawdziwości owej wypowiedzi. Badania te sugerują, że intuicje uczestników dotyczące kłamstwa uwzględniają głównie subiektywne intencje mówcy, nie zaś obiektywny stan rzeczy. Wiegmann podkreśla, że prostota pytań oraz brak ich dodatkowego kontekstu minimalizowały ryzyko wpływu torowania, co wzmacnia

wiarygodność wyników jako wsparcia dla stanowiska subiektywistycznego. Eksperymenty te zdają się podważać wynik badań Turich.

Z faktu, że większość ludzi posiada subiektywistyczne intuicje w kwestii kłamstwa nie wynika jeszcze, że teza subiektywistyczna jest intuicyjna, zwłaszcza, że niektórzy filozofowie sami uważają obiektywizm za stanowisko językowo intuicyjne (np. Carson, 2006). Dlatego ważne jest zaproponowanie przekonujących przykładów eksperymentów myślowych, które pomogą uchwycić prawdziwość tezy, że zaistnienie kłamstwa nie wymaga głoszenia obiektywnego fałszu, lecz wystarczy głoszenie twierdzenia, które nie jest uważane za prawdziwe przez osobę formułującą to twierdzenie. Dzięki tym eksperymentom czytelnik będzie w stanie sam ocenić, które stanowisko w prezentowanym sporze jest bardziej przekonujące.

## Eksperymenty myślowe

Wydaje się, że w obecnie dostępnej literaturze najbardziej przekonujące przykłady eksperymentów myślowych dotyczących interesującego nas problemu zaproponował Neri Marsili (2022). Oto dwa z nich.

### Eksperyment 1. Złośliwy demon

Pinokio, żyjący w naszym świecie, widział, jak Lucignolo ukradł jabłko. Świadomie twierdzi: „Eugenio ukradł jabłko”, zatem świadomie kłamie. Jego bliźniak Bucocchio żyje w świecie „złego demona”, który mami zmysły osób zamieszkujących jego świat, i doświadcza tych samych zjawisk, co Pinokio. Również twierdzi: „Eugenio ukradł jabłko”. Niestety nie może wiedzieć, że jego twierdzenie jest fałszywe, ale intuicyjnie wydaje się, że obaj kłamią.

## Eksperyment 2. Nowa dziura w spodniach

Pinokio ma w kieszeni 20 lirów i mówi: „Nie mam 15 lirów”, wiedząc, że to fałsz. Bucocchio, jego bliźniak, ma nową dziurę w kieszeni i część z jego 20 lirów wypadła, ale nie wie on o tym. Także wypowiada słowa: „Nie mam 15 lirów”. Intuicja nam podpowiada, że obaj kłamią.

Oba przykłady sugerują nie tylko to, co twierdzi Marsili, czyli że kłamiący nie musi wiedzieć, iż wypowiada obiektywny fałsz, ale także to, że obiektywny fałsz nie musi być *de facto* wypowiedany, aby dany akt komunikacji był kłamstwem. Bohaterami obu eksperymentów są postacie posiadające wadę kłamania. Eksperyment pierwszy możemy zinterpretować jako przedstawiający umieszczenie konsekwentnego kłamcy w dwóch różnych światach:  $W_1$  i  $W_2$ . W  $W_1$  kłamca ma możliwość uzyskania wiedzy, dzięki której byłby zdolny do formułowania obiektywnie fałszywych twierdzeń. W  $W_2$  jest on pozbawiony możliwości zdobycia wiedzy, ale nie jest pozbawiony swojej wady kłamcy.

Jeśli będziemy rozumieć wadę kłamcy jako sprawność do niezawodnego formułowania kłamstw i zgodzimy się, że zarówno Pinokio, jak i jego (moralny) bliźniak Bucocchio korzystają ze swojej sprawności do kłamania, to musimy przyjąć, iż obaj kłamią. Uznanie obiektywizmu prowadziłoby do nieintuicyjnego twierdzenia, że Bucocchio żyje w świecie, którego okoliczności (zewnętrzne względem intencji i decyzji Bucocchio) nie pozwalają jego wadzie kłamcy się przejawiać w jego wypowiedziach.

W odpowiedzi na powyższy komentarz możliwa jest kontrargumentacja, że wada kłamania przejawia się nie tylko w przypadkach kłamstwa, ale także w podejmowanych próbach kłamania, zatem eksperymenty myślowe w rodzaju złośliwego demona mogą okazać się niewystarczające dla ukazania intuicyjności stanowiska subiektywistycznego. Mogą być konieczne dalsze eksperymenty, niemniej Eksperyment 2. pokazuje inną jeszcze wadę tezy obiektywistycznej. Pokazuje on – wbrew stanowisku obiektywizmu – że pojęcie kłamstwa stosujemy także wówczas, gdy nie jest pewne, jak rzeczy, których

dotyczy kłamliwa wypowiedź, obiektywnie się mają. Nie ma pewności, czy Bucochio w Eksperymentie 2. wypowiada obiektywny fałsz. Mógł zgubić więcej niż 5 lirów, wówczas nie głosiłby obiektywnego fałszu, a zatem według obiektywistów nie kłamałby. Mógł też zgubić mniej niż 5 lirów, zatem głosiłby fałsz a zarazem kłamał według obiektywistycznego stanowiska. Mamy jednak (jak zakładam za Marsilim) intuicję, że nie potrzebujemy rozstrzygnięcia co do faktów, aby uznać, że Bucochio kłamie. Tymczasem stanowisko obiektywistyczne wyklucza możliwość orzekania o tym, czy faktycznie kłamstwo ma miejsce – aż do momentu ostatecznego stwierdzenia, że fakty nie są zgodne z daną wypowiedzią.

### **Eksperyment 3. Wszzechmocny złośliwy demon**

Wróćmy do scenariusza ze złośliwym demonem. Eksperyment 1. możemy wzmocnić, zmieniając nieco rolę, jaką odgrywa demon. Założmy, że nie tyle bliźniak Pinokio ulega iluzjom generowanym przez demona, ile demon (korzystając z prawie boskich atrybutów), wiedząc, co zamierza powiedzieć Bucocchio, dostosowuje za pomocą swoich mocy rzeczywistość do wypowiedzi kłamcy.

Gdy na przykład Bucochio mówi: „Mam w kieszeni 100 lirów”, a pamięta wyraźnie, że miał ich 20 ostatnim razem, gdy zaglądał do kieszeni, demon sprawia (na 500 milisekund przed otwarciem ust przez Bucocchio, czyli zanim kłamstwo wypowie, ale zarazem tuż po ujawnieniu się decyzji, aby skłamać, lub równocześnie z nią), że w jego kieszeni znajduje się 100 lirów. Czy w takich okolicznościach powiemy, że Bucochio próbował jedynie kłamać, ale mu się nie udało? Wydaje się, że raczej powiemy, iż skłamał i jest kłamcą w wyniku wypowiedzenia swoich kłamstw – niezależnie od tego, czy demon zainterweniował, czy nie. Jeżeli powyższa modyfikacja eksperymentu ze złośliwym demonem nie wystarcza do zneutralizowania argumentu przez brak wyraźnego odróżnienia między kłamstwem a nieudaną próbą kłamstwa, proponuję rozważyć kolejne eksperymenty.

## Eksperyment 4. Zdrada małżeńska

Pani X podejrzewa swojego męża o to, że zamierza ją zdradzić. Przybierając inną tożsamość, nawiązuje z nim kontakt przez internet. Umawiają się na randkę w kawiarni. Tutaj musimy założyć, że jakimś cudownym sposobem pani X udaje się zmienić swoją powierzchowność (do wyboru czytelnika: korzysta ona z magii albo wyrafinowanej technologii). Po jakimś czasie oboje udają się do hotelu, w którym dochodzi do wiadomej czynności, po której pani X, „zdejmuje maskę” i krzyczy: „Mam cię zdrajco!”.

Analogicznie do przypadku kłamstwa Bucocchio, działanie męża pani X zdaje się nie spełniać warunku obiektywności, ponieważ nie polega ono na obcowaniu z osobą inną niż jego żona, pani X. Można powiedzieć, że mężowi pani X jedynie się wydaje, że obcuje z inną kobietą niż jego żona. Jednak czy wówczas intuicje nie podpowiadają nam, żeby przyznać rację pani X, że jej mąż ją zdradził?

## Terminy istotowo moralne

Co mogłoby wyjaśniać intuicje, że kłamanie, czy też zdrada, nie wymagają zaistnienia pewnych zewnętrznych stanów rzeczy (obiektywnego fałszu, faktycznego obcowania fizycznego z osobą niebędącą naszym partnerem)? Dobrym kandydatem na takie wyjaśnienie wydaje się fakt, że oba te terminy „kłamanie” i „zdrada” mają ściśle moralny charakter, są istotowo moralne. Z etycznego punktu widzenia nie zawsze interesuje nas zewnętrzny skutek działania. Przynajmniej w niektórych przypadkach – szczególnie tych, w których zewnętrzne skutki czynu mogą być poza kontrolą podmiotu – czynnikiem decydującym dla moralnej oceny jest sama intencja zaktualizowana w decyzji i podjętych działaniach. Działanie staje się zdradą już w momencie wdrożenia zamiaru w czyn. Moment ten albo jest tożsamy z momentem decyzji, albo następuje bezpośrednio po podjęciu decyzji i dotyczy działania wolnej (jak zakładamy) woli.

## **Eksperyment 5. A jeśli żyjemy w symulacji?**

Dla wsparcia powyższej tezy proponuję kolejny eksperyment. Załóżmy, że żyjemy zanurzeni w symulacji komputerowej. Liczne nasze akty mające charakter kłamstwa, zdrady, kradzieży, morderstw, etc., ostatecznie, gdyby przyjąć sposób rozumowania obiektywistów, musiałyby okazać się nieudanymi próbami tychże działań. Należałoby uznać, że de facto nie było żadnych aktów zdrady, kłamstwa, etc.

Obiektywista mógłby się bronić przed powyższą konkluzją, twierdząc, że przecież nawet nieudana próba ma negatywną wartość moralną, na przykład przez powodowanie jeszcze innych konsekwencji niż te zamierzone. Np. nieudana zdrada także boli niedoszłego zdradzonego; nieudane kłamstwo, gdy wyjdzie na jaw nasza intencja, także rani; nieudane morderstwo, choć nie zabija, podkopuje poczucie bezpieczeństwa, budzi chęć odwetu, etc. Stąd powyższe scenariusze niekoniecznie wspierają tezę subiektywistyczną.

## **Eksperyment 6. Solipsystyczna symulacja**

W odpowiedzi na taki argument warto rozważyć jeszcze jeden scenariusz. Załóżmy, że każda realna osoba ma swoją własną symulację, w której jest jedynym autentycznym podmiotem z umysłem zdolnym do doznawania doświadczeń. W obiektywistycznym sensie wszystkie moralnie istotne akty analogiczne do kłamstwa w takiej symulacji nie byłyby skuteczne. Mimo to wydaje się, że nawet w takiej sytuacji – skoro podejmowane (i realizowane) były decyzje, aby stwierdzić coś, co uznaliśmy za fałsz, zdradzić kogoś, kogo uznaliśmy za realną osobę, z kimś, kogo również uznajemy za realną osobę – mamy podstawę, by powiedzieć, że zaszły autentyczne kłamstwa i autentyczne zdrady. Terminy „kłamstwo” czy „zdrada” dotyczą bowiem tego, co dzieje się w samym wnętrzu człowieka i na tym polega ich ściśle moralny charakter, przytoczone zaś w tym artykule przykłady odwołujące się do naszych intuicji zdają się tę tezę potwierdzać.



## Wnioski

Przedmiotem niniejszych refleksji był problem, czy kłamstwo wymaga obiektywnej fałszywości wypowiedzi. Analiza przeprowadzona w artykule, bazująca na eksperymentach myślowych, sugeruje, że obiektywna fałszywość nie jest koniecznym warunkiem kłamstwa. Przykłady omówione w tekście pokazują, że subiektywne nastawienie nadawcy wobec treści wypowiedzi może wystarczyć do uznania jej za kłamstwo, nawet jeśli wypowiedź ta jest obiektywnie prawdziwa lub jej obiektywna prawdziwość jest nieznaną.

Fakt, że do zaistnienia kłamstwa nie jest konieczna obiektywna fałszywość wypowiedzi, da się, być może, wyjaśnić tym, że pojęcie kłamstwa (jak wiele innych pojęć, np. pojęcie zdrady) jest z istoty swojej pojęciem moralnym, odnosi się przede wszystkim do aktów woli i postawy moralnej podmiotu, dla których oceny ich obiektywne skutki nie zawsze odgrywają centralną rolę.

## Bibliografia:

- Św. Augustyn. (2022). *O kłamstwie* (Ł. Libowski, tłum.). Wydawnictwo Scriptum.
- Benton, M. (2018). Lying, Belief, and Knowledge. In J. Meibauer (Ed.), *The Oxford Handbook of Lying* (pp. 120–133). Oxford University Press.
- Bok, S. (1978). *Lying: Moral Choice in Public and Private Life*. Pantheon Books.
- Bolzano, B. (2007). Über die Wahrhaftigkeit. In E. Morscher, K. Strasser, *Erbauungsreden der Studienjahre 1804/05 bis 1819/20* (pp. 290–303). Frommann-Holzboog.
- Carson, T. (2006). The Definition of Lying. *Noûs*, 40, 284–306. <https://doi.org/10.1111/j.0029-4624.2006.00610.x>
- Carson, T. (2010). *Lying and Deception: Theory and Practice*. Oxford University Press.
- Chisholm, R. M., Feehan, T. D. (1977). The Intent to Deceive. *The Journal of Philosophy*, 74, 143–159. <https://doi.org/10.2307/2025605>
- Chudy, W. (2003), *Filozofia kłamstwa Kłamstwo jako fenomen zła w świecie osób i społeczeństw*. Oficyna Wydawnicza.

- Fallis, D. (2009). What is Lying? *The Journal of Philosophy*, 106(1), 29–56. <https://doi.org/10.5840/jphil200910612>
- Grimaltos, T., Rosell, S. (2013). *On Lying: A Conceptual Argument for the Falsity Condition*. Universitat de València.
- Holguín, B. (2021). Lying and Knowing. *Synthese*, 198(4), 5351–5371. <https://doi.org/10.1007/s11229-019-02407-2>
- Kallestrup, J. (2023). The Myth of True Lies. *Theoria*, 89(4), 451–466. <https://doi.org/10.1111/theo.12466>
- Kucharski, J. (2014). *Usprawiedliwione kłamstwo we współczesnej etyce stosowanej*. Akademia Ignatianum Wydawnictwo WAM.
- Maitra, I. (2018). Lying, Acting, and Asserting. In E. Michaelson, A. Stokke (Eds), *Lying. Language, Knowledge, Ethics, and Politics* (pp. 65–82). Oxford University Press.
- Mahon, J. E. (2015). The definition of lying and deception. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Retrieved data from <https://plato.stanford.edu/entries/lying-definition>
- Marsili, N. (2014). Lying as a Scalar Phenomenon. In S. Cantarini, W. Abraham, E. Leiss (Eds), *Certainty-uncertainty-and the attitudinal space between* (pp. 153–173). John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.165.09mar>
- Marsili, N. (2022). Lying: Knowledge or Belief? *Philosophical Studies*, 179(5), 1445–1460. <https://doi.org/10.1007/s11098-021-01713-1>
- Meibauer, J. (2018). The linguistics of lying. *Annual Review of Linguistics*, 4(1), 357–375. <https://doi.org/10.1146/annurev-linguistics-011817-045634>
- Stokke, A. (2014). Insincerity. *Noûs*, 48(3), 496–520. <https://doi.org/10.1111/nous.12001>
- Stokke, A. (2018). *Lying and Insincerity*. Oxford University Press.
- Teichmann, R. (2024). Assertion, Lying and the Norm of Truth. *Topoi*, 43(2), 459–467. <https://doi.org/10.1007/s11245-023-09956-9>
- Turri, J. (2021). Objective falsity is essential to lying: An argument from convergent evidence. *Philosophical Studies*, 178(6), 2101–2109. <https://doi.org/10.1007/s11098-020-01525-9>
- Turri, A., Turri, J. (2015). The Truth About Lying. *Cognition*, 138, 161–168. <https://doi.org/10.1016/j.cognition.2015.01.007>
- Turri, A., Turri, J. (2021). Lying, fast and slow. *Synthese*, 198(3), 757–775. <https://doi.org/10.1007/s11229-018-02062-z>

Viebahn, E. (2024). True lies and attempted lies. *Inquiry*, 1–12. <https://doi.org/10.1080/0020174X.2024.2365878>

Wiegmann, A., Meibauer, J. (2019). The folk concept of lying. *Philosophy Compass*, 14(6), article 12620. <https://doi.org/10.1111/phc3.12620>

Wiegmann, A., Samland, J., Waldmann, M. R. (2016). Lying despite telling the truth. *Cognition*, 150, 37–42. <https://doi.org/10.1016/j.cognition.2016.01.017>

Wiegmann, A. (2023). Does lying require objective falsity? *Synthese*, 202(2), 52. <https://doi.org/10.1007/s11229-023-04291-3>

Williams, B. (2002). *Truth and Truthfulness*. Princeton University Press.



ARTUR SZUTTA

- University of Gdańsk
- 0000-0002-1062-4808
- artur.szutta@ug.edu.pl

# Does lying necessarily involve stating falsehoods?

## Abstract:

This article addresses the question of whether lying inherently requires the expression of objective falsehood. While traditional subjectivist approaches focus on the divergence between a speaker's beliefs and their statements, some contemporary theorists of lying argue that lying necessitates objectively false utterances. I offer a critical analysis of six thought experiments to illuminate the intuitiveness of the subjectivist claim. In explaining the intuitiveness of the traditional view of lying, I point to its fundamentally moral nature—namely, that lying is rooted in acts of will rather than external outcomes.

## Keywords:

lying, subjectivism, objectivism, thought experiment

## How to cite:

Szutta, A. (2024). Does lying necessarily involve stating falsehoods? [English translation]. *Laboratorium Mentis*, 2(2), 92–102. <https://doi.org/10.52097/lm.9485>

## Introduction

One of the debated issues concerning the concept of lying is whether lying necessarily involves making an objectively false statement. According to the traditional view, lying occurs when (assuming other conditions for lying are met) the liar expresses a statement they believe to be false. Decisive, in this case, is the mismatch between the statement and the speaker's belief. Hence, the traditional position on this condition is referred to as subjectivism.

In recent years, publications have emerged advocating for an objectivist position, according to which one of the necessary conditions for lying is the assertion of an objectively false statement. This means that, for example, if a person mistakenly believing that no one is currently in their house says, "There is someone in my house," they would not be lying if, in fact, someone is in their house at the time of the statement. Several arguments have been proposed in favour of this position. Some (Turi & Turi, 2015; Turi, 2021; Carson, 2006) appeal to intuition, while others refer to the very nature of lying, claiming that lying, by definition, requires a certain potential for effective deception, which can only be achieved by asserting (or implying) an objectively false statement (see Grimaltos & Rosell, 2013; Kallestrup, 2023). Still others argue that lying must involve the assertion of objectively false statements by virtue of the constitutive rules of the practice of assertion (see, e.g., Teichmann, 2024).

In this article, I focus on one aspect of the aforementioned debate: the question of which position, subjectivism or objectivism, is more intuitive. To formulate a compelling answer, I propose considering six thought experiments. Two of these, serving as a starting point, are experiments introduced by Neri Marsili (2021): one referencing the so-called "evil demon scenario" and the other illustrating the possibility of lying in a situation of uncertainty. The subsequent experiments aim to emphasize further the intuitiveness of the subjectivist position on lying. The ultimate explanation for these intuitions seems to lie in the

fact that terms such as “lying” are inherently moral and primarily pertain to acts of will while only secondarily (and not always necessarily) to external outcomes.

## Empirical research

Before presenting the announced thought experiments, it is worth dedicating a few words to the argument that the objectivist position is more aligned with the intuitions of most ordinary people. John and Angela Turi conducted empirical studies in which participants were presented with a scenario involving a person named Karen. When asked by government officials where her acquaintance Mary was, Karen replied (believing Mary to be elsewhere) that Mary was at the store, only for it to turn out that Mary was indeed at the store (Turi & Turi, 2015, 2021; Turi, 2021).

In one version of the experiment, participants were asked simple yes-or-no questions, such as whether the protagonist (Karen) lied. The majority of respondents answered in line with subjectivist intuitions. The researchers hypothesized that two factors might have influenced the subjectivist results: i) participants unconsciously identifying with the protagonist’s perspective and ii) their failing to distinguish lying from a failed attempt to lie. In modified versions of the experiment, the questions allowed for more nuanced answers. Instead of simply choosing “yes” or “no,” participants could select options such as (a) “She lied and stated an objective falsehood” or (b) “She attempted to lie but stated the truth.” In these cases, the majority chose objectivist answers.

Wiegmann and his colleagues (Wiegmann, Samland & Waldmann, 2016; Wiegmann & Meibauer, 2019; Wiegmann, 2023) critically evaluated Turi & Turi’s studies by conducting their own empirical research on ordinary people’s intuitions about lying. In one experiment, Wiegmann’s team presented participants with response pairs structured similarly to those in Turi & Turi’s research: “He tried and succeeded” versus “He tried but objectively failed.” In the experiment

relevant to this discussion, the questions pertained to making a promise. The protagonist in the story presented to participants made a promise, which they either managed to keep or failed to keep, depending on the scenario.

Interestingly, in scenarios where the protagonist failed to keep his promise, participants were asked whether the protagonist had actually made a promise or merely attempted to make one (notably, the question did not include the phrase “keep the promise”). Most participants responded that the protagonist merely attempted to make a promise. Wiegmann interprets this result as a priming effect caused by the construction of the question, which may have misled participants into choosing the “tried but failed” option, even in situations where the protagonist had indeed made a promise.

This raises the possibility that participants in Turi & Turi’s experiments were similarly influenced by priming effects. As a result, their responses may not provide reliable evidence for the claim that the objectivist position is more intuitive for ordinary people.

The fact that most people hold subjectivist intuitions about lying does not necessarily mean that the subjectivist thesis itself is intuitive, especially given that some philosophers consider objectivism linguistically intuitive (e.g., Carson, 2006). Therefore, it is important to propose compelling examples of thought experiments that help illustrate the truth of the thesis that, for lying to occur (in addition to other conditions not discussed here), it is not necessary to assert an objectively false statement but rather a statement that the person making it does not believe to be true. These thought experiments aim to enable the reader to assess for themselves which position in the presented debate is more convincing.

## Thought Experiments

It seems that the most compelling examples of thought experiments in the currently available literature have been proposed by Neri Marsili (2022). Here are two of them:

### Experiment 1. The Vicious Demon

Pinocchio, living in our world, witnessed Lucignolo stealing an apple. He knowingly asserts, “Eugenio stole the apple,” and thus knowingly lies. His twin, Bucocchio, lives in a “vicious demon” world, where the demon deceives the senses of its inhabitants, and he experiences the same phenomena as Pinocchio. Bucocchio also asserts, “Eugenio stole the apple.” Unfortunately, he does not know that his statement is false, but intuitively, both seem to be lying.

### Experiment 2. A New Hole

Pinocchio has 20 liras in his pocket and says, “I don’t have 15 liras,” knowing this is false. Bucocchio, his twin, has a new hole in his pocket, and some of his 20 liras have fallen out without him realizing it. He also says, “I don’t have 15 liras.” Intuitively, it seems that both are lying.

Both examples suggest (not only, as Marsili claims, that a liar does not need to know they are stating an objective falsehood, but also) that an objective falsehood does not necessarily have to be uttered for a communicative act to be considered a lie. The protagonists of both experiments are characters possessing the vice of lying. The first experiment can be interpreted as depicting the placement of a consistent liar in two different worlds,  $W_1$  and  $W_2$ . In  $W_1$ , the liar has access to knowledge that enables him to make objectively false statements. In  $W_2$ , he is deprived of the possibility of knowledge but not of his vice as a liar.



If we understand the vice of lying as a disposition to make lies reliably and agree that both Pinocchio and his moral twin, Bucocchio, are employing their disposition to lie, we should conclude that both are lying. Adopting objectivism would lead to the counterintuitive conclusion that Bucocchio lives in a world where external circumstances (beyond Bucocchio's intentions and decisions) prevent his vice as a liar from manifesting in his statements.

In response to the above comment, one could counterargue that the vice of lying manifests not only in actual cases of lying but also in attempts to lie. Thus, thought experiments like the vicious demon scenario may be insufficient to demonstrate the intuitiveness of the subjectivist position. Further experiments might be necessary. Nevertheless, Experiment 2 reveals another flaw in the objectivist thesis. It shows that we apply the concept of lying—contrary to objectivist claims—even in cases where the facts concerning the content of the lie are uncertain.

It is unclear whether Bucocchio in Experiment 2 is making an objectively false statement. He may have lost more than 5 liras, in which case his statement would not qualify as an objective falsehood, and, according to objectivists, he would not be lying. Conversely, if he lost less than 5 liras, objectivists would consider his statement a lie. However, as I assume with Marsili, we have the intuition that determining whether Bucocchio is lying does not require resolving the discussion about the facts. The objectivist position, by contrast, insists that lying cannot be judged until it is definitively established that the statement conflicts with reality.

### Experiment 3. The Omnipotent Vicious Demon

Let us return to the malicious demon scenario. Experiment 1 can be reinforced by slightly altering the demon's role. Let us assume that Pinocchio's twin does not merely fall prey to the illusions generated by the demon. Instead, the demon, endowed with almost god-like attributes,

adjusts reality to match the liar's statement, knowing in advance what Bucocchio intends to say.

For example, if Bucocchio says, "I have 100 liras in my pocket," while clearly remembering that he last saw only 20 lire there, the demon intervenes (say, 500 milliseconds before Bucocchio opens his mouth to speak—or precisely at the moment his decision to lie manifests) and ensures that 100 liras appear in his pocket. Under such circumstances, would we say that Bucocchio merely attempted to lie but failed? Or rather, should we conclude that he lied and is a liar as a result of his lies, regardless of the demon's intervention?

If the above modification of the vicious demon experiment is insufficient to address the argument concerning the lack of a clear distinction between lying and a failed attempt to lie, I propose considering additional experiments.

## Experiment 4. Marital Infidelity

Mrs. X suspects that her husband is planning to cheat on her. Taking on another identity, she contacts him online. They arrange to meet at a café. Here, we must assume that Mrs. X, by some miraculous means, manages to change her appearance (the choice is left to the reader: she either uses magic or advanced technology). After some time, they both go to a hotel, where the expected act occurs. Then, Mrs. X "removes her mask" and exclaims, "Got you, cheater!"

Analogously to Bucocchio's case of lying, Mr. X's actions seem not to fulfil the condition of objectivity; they do not involve being with someone other than Mr. X's wife. One could say that Mr. X merely believes he is with a woman other than his wife. However, would we not agree with Mrs. X's claim that her husband has indeed cheated on her?

## Essentially Moral Terms

What could explain the intuition that lying or infidelity does not require certain external states of affairs (objective falsehood, actual physical interaction with someone other than one's partner)? A strong candidate for such an explanation seems to be that both terms have a distinctly moral character—they are essentially moral. From a moral perspective, the external outcome of an action does not always matter. At least in some cases, particularly those where the external consequences of an act may be beyond the agent's control, the decisive factor for moral evaluation is the intention realized in the decision and the actions taken. An act becomes infidelity at the very moment the intention is carried out in an act, which either coincides with the moment of decision or follows directly afterwards and involves the act of a free (as we assume) will.

### Experiment 5. What if We Live in a Simulation?

To support the above thesis, I propose another thought experiment. Let us assume that we are living immersed in a computer simulation. All our actions—such as lying, infidelity, theft, or murder—if we were to follow the objectivist reasoning would ultimately have to be considered failed attempts at these actions. One would have to conclude that, in fact, no acts of infidelity, lying, or similar behaviours ever occur within the simulation.

An objectivist might argue that even a failed attempt can carry negative moral value, for example, through unintended consequences. For instance, a failed act of infidelity can still hurt the would-be betrayed partner; a failed lie, when the intent is exposed, can also cause harm; and a failed murder, although it does not result in death, undermines a sense of security and provokes a desire for revenge, etc. Thus, it would be claimed, the above scenarios do not necessarily support the subjectivist thesis.

## Experiment 6. Solipsistic Simulation

In response to such an argument, it is worth considering yet another scenario. Let us assume that each real person has their own simulation, in which they are the only genuine subject with a mind capable of experiences. In the objectivist sense, all essentially moral acts analogous to lying in such a simulation would not be actualized. However, it seems that even in this situation, since decisions were made (and carried out) to state something believed to be false, to betray someone perceived as a real person, or to engage with someone also believed to be real, there is a basis to claim that authentic lies and authentic acts of infidelity occurred. It is so because the terms “lying” and “infidelity” pertain to what occurs within a person’s inner self, which constitutes their distinctly moral character. The examples presented in this article, appealing to our intuitions, seem to support this thesis.

## Conclusions

The focus of this reflection was the question of whether lying requires the objective falsehood of a statement. The analysis conducted in this article and the thought experiments suggest that objective falsehood is not necessary for lying. The examples discussed demonstrated that the speaker’s subjective attitude toward the content of their statement can suffice to classify it as a lie, even if the statement is objectively true or its objective truth is unknown.

The fact that lying does not require the objective falsehood of a statement could be explained by recognizing that the concept of lying—like many other concepts, such as that of infidelity—is inherently moral. It pertains primarily to acts of will and the moral stance of the subject, for which their objective consequences do not always play a central role.

## Bibliography:

- Augustine.** (2013). De mendacio. In A. Städele (Ed.), *Opera/Werke* (vol. 50, pp. 60–149). Schönigh. (Original work published 1810)
- Benton, M.** (2018). Lying, Belief, and Knowledge. In J. Meibauer (Ed.), *The Oxford Handbook of Lying* (pp. 120–133). Oxford University Press.
- Bok, S.** (1978). *Lying: Moral Choice in Public and Private Life*. Pantheon Books.
- Bolzano, B.** (2007). Über die Wahrhaftigkeit. In E. Morscher, K. Strasser, *Erbauungsreden der Studienjahre 1804/05 bis 1819/20* (pp. 290–303). Frommann-Holzboog. (Original work published 1810)
- Carson, T.** (2006). The Definition of Lying. *Noûs*, 40(2), 284–306. <https://doi.org/10.1111/j.0029-4624.2006.00610.x>
- Carson, T.** (2010). *Lying and Deception: Theory and Practice*. Oxford University Press.
- Chisholm, R. M., & Feehan, T. D.** (1977). The Intent to Deceive. *The Journal of Philosophy*, 74(3), 143–159. <https://doi.org/10.2307/2025605>
- Chudy, W.** (2003). *Filozofia kłamstwa. Kłamstwo jako fenomen zła w świecie osób i społeczeństw*. Oficyna Wydawnicza.
- Fallis, D.** (2009). What is Lying? *The Journal of Philosophy*, 106(1), 29–56. <https://doi.org/10.5840/jphil200910612https://doi.org/10.5840/jphil200910612>
- Grimaltos, T., & Rosell, S.** (2013). *On Lying: A Conceptual Argument for the Falsity Condition*. Universitat de València.
- Holguín, B.** (2021). Lying and Knowing. *Synthese*, 198(4), 5351–5371. <https://doi.org/10.1007/s11229-019-02407-210.1007/s11229-019-02407-2>
- Kallestrup, J.** (2023). The Myth of True Lies. *Theoria*, 89(4), 451–466. <https://doi.org/10.1111/theo.12466https://doi.org/10.1111/theo.12466>
- Kucharski, J.** (2014). *Usprawiedliwione kłamstwo we współczesnej etyce stosowanej*. Akademia Ignatianum Wydawnictwo WAM.
- Maitra, I.** (2018). Lying, Acting, and Asserting. In E. Michaelson, A. Stokke (Eds), *Lying. Language, Knowledge, Ethics, and Politics* (pp. 65–82). Oxford University Press.
- Mahon, J. E.** (2015). The definition of lying and deception. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Retrieved data from <https://plato.stanford.edu/entries/lying-definition>

- Marsili, N. (2014). Lying as a Scalar Phenomenon. In S. Cantarini, W. Abraham, & E. Leiss (Eds), *Certainty-uncertainty-and the attitudinal space between* (pp. 153–173). John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.165.09mar>
- Marsili, N. (2022). Lying: Knowledge or Belief? *Philosophical Studies*, 179(5), 1445–1460. <https://doi.org/10.1007/s11098-021-01713-1>
- Meibauer, J. (2018). The linguistics of lying. *Annual Review of Linguistics*, 4(1), 357–375. <https://doi.org/10.1146/annurev-linguistics-011817-045634>
- Stokke, A. (2014). Insincerity. *Noûs*, 48(3), 496–520. <https://doi.org/10.1111/nous.12001>
- Stokke, A. (2018). *Lying and Insincerity*. Oxford University Press.
- Teichmann, R. (2024). Assertion, Lying and the Norm of Truth. *Topoi*, 43(2), 459–467. <https://doi.org/10.1007/s11245-023-09956-9>
- Turri, J. (2021). Objective falsity is essential to lying: An argument from convergent evidence. *Philosophical Studies*, 178(6), 2101–2109. <https://doi.org/10.1007/s11098-020-01525-9>
- Turri, A., & Turri, J. (2015). The Truth About Lying. *Cognition*, 138, 161–168. <https://doi.org/10.1016/j.cognition.2015.01.007>
- Turri, A., & Turri, J. (2021). Lying, fast and slow. *Synthese*, 198(13), 757–775. <https://doi.org/10.1007/s11229-018-02062-z>
- Viebahn, E. (2024). True lies and attempted lies. *Inquiry*, 1–12. <https://doi.org/10.1080/0020174X.2024.2365878>
- Wiegmann, A., & Meibauer, J. (2019). The folk concept of lying. *Philosophy Compass*, 14(6), article 12620. <https://doi.org/10.1111/phc3.12620>
- Wiegmann, A., Samland, J., & Waldmann, M. R. (2016). Lying despite telling the truth. *Cognition*, 150, 37–42. <https://doi.org/10.1016/j.cognition.2016.01.017>
- Wiegmann, A. (2023). Does lying require objective falsity? *Synthese*, 202, 52. <https://doi.org/10.1007/s11229-023-04291-3>
- Williams, B. (2002). *Truth and Truthfulness*. Princeton University Press.