

УДК 16:81:81.2

DOI: <https://doi.org/10.17721/2523-4064.2023/9-3/15>

Анатолій КОНВЕРСЬКИЙ, акад. НАН України, д-р філос. наук, проф.

ORCID ID: 0000-0002-3916-4446

e-mail: dean_phd@univ.kiev.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

Наталія КОЛОТИЛОВА, д-р філос. наук, доц.

ORCID ID: 0000-0002-7450-2981

e-mail: nataliakolotilova@knu.ua

Київський національний університет імені Тараса Шевченка, Київ, Україна

ПІДТВЕРДЖУВАЛЬНЕ ВИКРИВЛЕННЯ В ПРОЦЕСАХ АРГУМЕНТАЦІЇ

Вступ. Стаття присвячена вивченню підтверджувального викривлення як когнітивного упередження в рамках сучасної теорії аргументації. У контексті цього дослідження розглядається ефективність методики критичних запитань як стратегії аргументування, спрямованої на зменшення негативного впливу підтверджувального викривлення.

Методи. Для досягнення цілей дослідження використовується метод критичних запитань, який базується на традиційних принципах логіки, діалектики та риторики. Розробляються конкретні переліки критичних запитань для кожної аргументативної схеми.

Результати. Стаття показує, що стандартні критичні запитання до схем практичного міркування, розроблені видатними фахівцями, спроможні мінімізувати негативний вплив підтверджувального викривлення. Це досягається завдяки врахуванню інформації, що може суперечити плану дій агента.

Висновки. Розвідка підкреслює важливість використання критичних запитань як інструмента для мінімізації підтверджувального викривлення. Такий підхід сприяє не лише обґрунтуванню аргументів, але й стимулює самоконтроль і пошук інформації, яка може викликати сумніви щодо власних поглядів. Заключна частина роботи розглядає пропаганду як наємне використання підтверджувального спотворення та вказує на необхідність уваги до мовних моделей, де це відбувається скоріше ненавмисно. У зв'язку зі специфікою досліджуваного феномена пропонується термін "підтверджувальне викривлення" як переклад англійського "confirmation bias".

Ключові слова: підтверджувальне викривлення, когнітивне упередження, теорія аргументації, схема аргументації, критичне запитання, практичне міркування.

Вступ

Починаючи з другої половини XX і особливо на початку XXI ст. спостерігається бурхливий розвиток низки царин гуманітарного знання, що мають насамперед практичний характер і спрямовані на дослідження різних феноменів соціального життя людини. Покликання цієї "практичності" полягає у розробці різноманітних стратегій і тактик досягнення успіху в різних царинах діяльності, передусім у комунікації. Разом із тим, особливо увагу науковців привертають чинники, які заважають людині діяти раціонально. До них належать когнітивні упередження, або ілюзії, під впливом яких людина зазвичай не діє раціонально. Більше того, вони виступають одним із тих джерел, на яких базуються маніпуляції суспільною думкою, що особливо яскраво постають у повномасштабній війні Росії з Україною, зокрема в інформаційній складовій цієї агресії.

Сучасній науці відомо багато різновидів когнітивних ілюзій, одним із яких постає підтверджувальне викривлення (*confirmation bias*). Метою даної статті, що продовжує попереднє дослідження (Kolotilova, 2023), є розгляд цього упередження в контексті стратегії й тактики аргументування задля досягнення переконливості певного повідомлення.

Джерела. Теоретичною базою розвідки є переважно сучасні праці у царині теорії аргументації, філософії, психології, поведінкової економіки таких авторів, як Д. Волтон (D. Walton), Л. Гроарке (L. Groarke), Ф. ван Ємерен (F. van Eemeren), Д. Канеман (D. Kahneman), К. Поппер (K. Popper) та ін. Також були враховані статті енциклопедичного характеру, дотичні до даної тематики, та розробки українських науковців (А. Є. Конверський, В. Надурак та ін.).

Методи

Під час дослідження було використано насамперед загальнонаукові методи та прийоми: аналіз, синтез, узагальнення. Значну увагу було приділено також термі-

нологічному аналізу поняття "підтверджувальне викривлення" для вибору такого терміну як відповідника "confirmation bias".

Результати

Сам термін "когнітивне упередження" (*cognitive bias*) належить психології й означає "...систематичні помилки в тому, як люди міркують про світ через суб'єктивне сприйняття реальності" (Eldridge, 2023). У Кембриджському словнику когнітивне упередження визначене таким чином: "те, як конкретна особа розуміє події, факти та інших людей, що ґрунтується на її власних переконаннях і досвіді та може бути нерозумним чи точним" (Cambridge Dictionary, n.d.). Відмітною особливістю таких ілюзій є те, що вони зазвичай не усвідомлюються людиною, проте можуть значно впливати на процеси думання, аргументування та прийняття рішень. У результаті людина викривлено сприймає реальність, її погляди обумовлюють можливо помилкові судження про світ, а відповідно і дії, що є наслідком практичних міркувань, будуть далекими від раціональності.

На сьогодні встановлено чимало видів когнітивних упереджень. Класик у цій царині Д. Канеман зазначав, що навряд чи можливо говорити про подолання когнітивних ілюзій, оскільки людині дуже важко постійно сумніватися у своїх думках, зазвичай вона покладається в прийнятті повсякденних рішень на так зване швидке інтуїтивне мислення, і навіть повільне мислення без додаткових зусиль може просто не знати про помилки. Тому, на його думку, варто говорити про компроміс, тобто навчитися розпізнавати ситуації, коли можливі помилки, і намагатися їх уникнути (Канеман, 2017, с. 35).

У Довіднику, присвяченому когнітивним ілюзіям (Pohl, 2004), вони поділені на три групи: мислення, судження та пам'ять. Підтверджувальне викривлення (*confirmation bias*), яке є предметом цієї розвідки, віднесене до ілюзій мислення. Оскільки мислення в різних ракурсах традиційно вивчалось такими дисциплінами, як логіка,

діалектика, риторика, що стали підґрунтям сучасної теорії аргументації, то вважаємо доцільним залучення здобутків цієї царини для встановлення ефективних стратегій і тактик аргументування, що мають мінімізувати негативний вплив підтверджувального викривлення.

Сам термін "*confirmation bias*" доволі часто перекладають як "*підтверджувальне упередження*" (Боровкова, 2023; Центр стратегічних комунікацій, 2023), або як "*упередженість підтвердження*" (Канеман, 2017, с. 93), або як "*упередження підтвердження*" (Надурак, 2022, с. 138). Проте, на нашу думку, щодо цієї ілюзії більш доречним був би термін "*підтверджувальне викривлення*", питання ж стосовно можливого застосування терміну "*упередження*" для когнітивних ілюзій загалом потребуватиме, напевно, окремого дослідження. Згідно з академічним тлумаченням Словником української мови за ред. І. К. Білодіда (хоча він і був виданий у 1970–1980 рр., однак залишається фундаментальним виданням у цій царині), *упередження* – це: 1) хибна думка, яка складається щодо кого-, чого-небудь наперед, без ознайомлення, та пов'язане з нею відповідне ставлення; 2) негативна, несправедлива думка, яка складається стосовно кого-, чого-небудь наперед, і пов'язане з нею відповідне ставлення, а *упередженість* як властивість від "упереджений" у значенні: який склався щодо кого-, чого-небудь наперед, до ознайомлення з ним; необ'єктивний (Словник української мови, 1970–1980). Тобто характерною ознакою *упереджень* постає те, що вони формуються "наперед". У випадку ж підтверджувального викривлення все відбувається ніби навпаки: уже після формування, прийняття певної (або бажаної) думки людина схильна шукати ту інформацію, яка її підтверджує, та ігнорувати те, що їй суперечить. Тому для позначення цієї когнітивної ілюзії вважаємо доречнішим використання терміну "*викривлення*" від дієслова "*викривляти*", "*викривити*", переносним значенням якого є переколювати, спотворювати, показувати у неправильному вигляді (Словник української мови, 1970–1980). У такий спосіб найяскравіше "схоплюється" специфіка вищезазначеної ілюзії, оскільки відбувається переколювання, викривлення інформації в бік тієї, що підкріплює думку, на протипагу тій, що її заперечує.

Підтверджувальне викривлення, як зазвичай неусвідомлена обробка інформації, обумовлюється кількома чинниками психологічного характеру. У межах даної статті зосередимося на тому, яку ж небезпеку містить ця ілюзія саме як ілюзія мислення. Це стосується не лише царини науки, адже і в повсякденному житті людина доволі часто конструє та перевіряє гіпотези щодо реальності. Підтверджувальне викривлення, як зазначають дослідники, ніби створює імунітет для гіпотези, адже систематично інформація, що її підтверджує, отримує більшу вагу, оцінюється менш критично або запам'ятовується краще, ніж дані, що не підтверджують цю гіпотезу (Oswald, & Grosjean, 2004, р. 94). Підступність цієї ілюзії, як і багатьох інших когнітивних упереджень, криється в тому, що навіть її усвідомлення не гарантує уникнення помилкових суджень і рішень. Як зауважує Д. Канеман, часто самі науковці попри рекомендації іншим перевіряти гіпотези шляхом їхнього спростування, шукають інформацію, що відповідатиме їхнім поточним віруванням (Канеман, 2017, с. 93).

Протягом тривалого історичного розвитку проблематика мислення цікавила насамперед філософів, особливо тих, хто розробляв такі філософські дисципліни, як логіка, діалектика, певною мірою риторика, що стали підґрунтям сучасної теорії аргументації. Привертанням

уваги до помилок нашого мислення, способів побудови переконливої аргументації ми завдячуємо насамперед софістам. У контексті стратегій аргументування, які можуть мінімізувати негативну дію того самого підтверджувального викривлення, варто згадати одного з фундаторів софістики, а саме Протагора. Як свідчив Діоген Лаєрцій (IX, 51), Протагор першим зазначив, що в кожному питанні є дві сторони, протилежні одна одній, він проводив суперечки в цьому напрямі. Важливим також було мистецтво запитування як складова маєстички іншого відомого античного філософа Сократа, завдяки якому проходили випробування, "тестувалися" думки співрозмовників щодо того, чи виражають вони справжнє знання, чи є обманом (Plato. Theaetetus, p. 151).

Отже, від доби Античності до нас дійшла стратегія аргументації, що полягає у розгляді аргументів "за" і "проти" певного твердження, і тактика постановки відповідних запитань, завдяки якій можливим було встановлення слабкості певних аргументів. У сучасній теорії аргументації, формування якої в середині минулого століття свідчить про повернення до тріади з логіки, діалектики, риторики в їхньому первісному тлумаченні (Kolotilova, 2022), відбувається подальша розробка ефективних стратегій і тактик аргументування. На межі століть відомий фахівець Ф. ван Ємерен для характеристики стану справ у вивченні аргументації користувався терміном "*мистецтво*" (art) (Emeren, 2001, р. 11). Мистецтво переконання (риторика) тісно пов'язане з мистецтвом обґрунтування (логіка) і мистецтвом ведення суперечок (діалектика). Щоправда, логіка згодом зайняла виняткову позицію, оскільки на початку XX ст. завдяки збагаченню математичним апаратом перетворилася в сувору формальну науку, яка, по суті, втратила зв'язок з реальними аргументативними практиками. Таке витлумачення логіки виключало вивчення того, як люди мислять, а отже, і фіксацію тих когнітивних упереджень, під впливом яких робляться необґрунтовані висновки. Саме тому Ф. ван Ємерен указував на те, що хоча поки не існує єдиної теорії аргументації, проте більшість підходів у цій царині є розвитком ідей риторики та діалектики (Emeren, 2001, р. 11).

Сучасна теорія аргументації містить низку різноманітних підходів, серед яких доволі цікавим є прагматична, яку якраз і розробляли представники нідерландської школи на чолі з уже згадуваним Ф. ван Ємереном. Одним із вагомих джерел цього підходу, як зазначав сам Ф. ван Ємерен (Emeren, 2001, р. 15), були ідеї філософії критичного раціоналізму, що в минулому столітті розвинув відомий філософ К. Поппер. У передмові до англійського видання (1959 р.) праці "*Логіка наукового дослідження*" (*The Logic of Scientific Discovery*) К. Поппер указував на, так би мовити, "єдиний метод", яким є метод раціональної або критичної дискусії (Popper, 2002, р. XIX). При цьому він ототожнював "раціональну" і "критичну" позицію (attitude). Сутність методу полягає в тому, що "коли ми пропонуємо вирішення проблеми, ми повинні якомога більше намагатися спростувати наше рішення, а не захищати його. Мало хто з нас, на жаль, виконує цю заповідь..." (Popper, 2002, р. XIX). Разом із тим, що є найголовнішою стратегією аргументування, при цьому на психологічному рівні має працювати повільне мислення на протипагу інтуїтивному, яке схиляється до менш затратних рішень. Д. Канеман підкреслював, що для застосування повільного мислення (Системи 2) потрібний самоконтроль, який виснажує, що спричиняє принаймні часткову втрату мотивації (Канеман, 2017, с. 50–51). Якраз відсутність мотивування щодо пошуку

тих положень, які спростовують наші думки, є характерною ознакою підтверджувального викривлення як когнітивного упередження: "Справа не в тому, що люди нездатні генерувати аргументи, які суперечать їхнім переконанням, скоріше люди не мотивовані це робити" (Casad, & Luebering, 2023).

Отже, аби уникнути небажаних наслідків підтверджувального викривлення, необхідно ретельно продумувати свою аргументацію. Першим важливим стратегічним кроком є зосередження не лише на аргументах, що підкріплюють власну позицію, але й на можливих запереченнях. Прихильники прагматодіалектики вказували, що під час підготовки до аргументації навіть якщо достеменно невідомо, хто буде опонентами, варто поставити такі запитання: Якою може бути їхня позиція? Які аргументи вони можуть використовувати? Які заперечення вони можуть висунути проти наведених аргументів? (Emereh et al., 2002, p. 172).

Сама стратегія критичних запитань є доволі плідною й широко застосовується в царині теорії аргументації. Конкретні переліки здебільшого визначаються тим, за якими схемами відбувається аргументація. Поняття "схеми аргументації" (*argument scheme*) активно розробляється сучасними дослідниками, хоча поки що важко говорити про його остаточне визначення на відміну від засадничого для логіки поняття форми міркування. Ф. ван Емерен (Emereh, 2001, p. 19-20) зауважував, що схеми аргументації є тим, що забезпечує перенесення прийнятності з експліцитного засновку на точку зору, в якій людина намагається переконати інших, оскільки кожний тип аргументації залежить від того, яка аргументативна схема використовувалась, а також передбачає відповіді на певні критичні запитання. У межах неформальної логіки, як ще одного впливового підходу в теорії аргументації, такі схеми постають як повторювані шаблони (*patterns*) міркувань. На думку Л. Грорке (2021), найчастіше розгляд схеми аргументації поєднує відповідний шаблон з набором критичних запитань або більше: відповіді на критичні запитання вбудовані в "повне" визначення схеми, що є необхідною передумовою переконливих прикладів такої схеми. Тобто кожна аргументативна схема характеризується власним набором критичних запитань.

На сьогодні досліджено чимало різних типів аргументів і, відповідно, аргументативних схем. Зупинимось докладніше на прикладі так званих практичних міркувань, оскільки вони відіграють важливу роль у ситуаціях прийняття рішень, вибору певної дії. У вітчизняній літературі проблематика практичного силлогізму стосовно логіки розглядалася А. Т. Ішмуратовим, який подавав зокрема такий неформальний запис його: "Х бажає, щоб було А. Х вірить, що якщо не буде С, то не буде й А. Х має зробити С" (Ішмуратов, 1997, с. 156).

З погляду поняття схем аргументації, що докладно висвітлено в публікації провідних фахівців Ф. Макагно, Д. Волтона та К. Ріда (Macagno et al., 2017), найпростіша форма практичного міркування постає як економна евристика для швидкого висновку, який пізніше може бути переглянутим, оскільки агент як раціональний суб'єкт здатний змінювати свої дії та цілі залежно від відповідей на критичні запитання. Аргумент від практичного міркування має такий вигляд: Більший засновок: Я хочу досягти цілі **G**. Менший засновок: Виконання дії **A** є засобом реалізації цілі **G**. Висновок: отже, я маю виконати дію **A**. Дослідники вказують на п'ять стандартних критичних запитань до цієї схеми: 1) які у мене є інші цілі, що можуть суперечити **G**? 2) які альтернативні дії стосовно **A**,

спрямовані на досягнення **G**, потрібно розглянути? 3) серед **A** та цих альтернативних дій, які є найефективнішими? 4) які їхні підстави суперечать практичній можливості для мене виконати **A**? 5) які наслідки мого виконання **A** також слід узяти до уваги? (Macagno et al., 2017, p. 2523–2524).

Відповідно до вищенаведеного переліку критичних запитань, усі вони мінімізують негативний вплив підтверджувального викривлення, адже передбачають урахування насамперед тієї інформації, яка може йти врозрід із планом дій агента. При цьому відповідь на останнє критичне запитання або запитання побічних ефектів, як зазначають фахівці, у випадку виявлення негативних наслідків здатна відкликати висновок, сформульований за схемою аргументації, та стимулювати розгляд альтернативних дій, які не мають цих наслідків (Macagno et al., 2017, p. 2524). Тобто такі міркування мають модифікувальний (*defeasible*) характер, оскільки передбачають ніби попереднє прийняття плану дій і можливої зміни за умов виникнення нових обставин. Хоча у подальшому відбувається розгляд як аргументу позитивних наслідків, що обґрунтовують продовження запланованих дій, так і аргументу негативних наслідків, що подають причину проти вчинення дії (Macagno et al., 2017, p. 2525). Загалом може виникнути спокуса під впливом підтверджувального викривлення зосередитися лише на першому, нехтуючи другим варіантом аргументу.

На нашу думку, сам процес відповідей на критичні запитання здатний зменшити спотворювальний ефект досліджуваної когнітивної ілюзії, оскільки запитання має особливу форму серед інших типів висловлювання, що неодноразово підкреслювалося у межах логічного знання. Наприклад, А. Є. Конверський зазначає, що запитання постає як така "...думка, в якій зафіксована вимога або прохання поновити наявну інформацію з метою усунення або зменшення пізнавальної невизначеності" (Конверський, 2020, с. 120). У свою чергу, відповідь на запитання постає як певна дія, і у випадку використання переліку критичних запитань, що стимулюють дії-відповіді, це може додатково мотивувати шукати різні аргументи, а не лише ті, які підтверджують наші погляди.

Дискусія і висновки

Для подальших досліджень когнітивних упереджень загалом і підтверджувального викривлення зокрема важливим є те, що вони активно використовуються у пропаганді. Тому потрібно й надалі розробляти методи, які сприятимуть виявленню та зменшенню шкоди подібних когнітивних ілюзій, адже їхня небезпека пов'язана насамперед з тим, що вони зазвичай не усвідомлюються людиною, проте на них ґрунтуються маніпуляції суспільною думкою, що особливо яскраво постає в інформаційній війні як складовій агресії Росії проти України.

Л. Грорке та К. Тіндейл указують на небезпеку когнітивних упереджень в аргументації, які вони називають "неправомірними упередженнями" (*illegitimate biases*). Такі необ'єктивності впливають негативно як на наш спосіб аргументування, так і заважають нашій здатності сприймати аргументи тих, хто висуває іншу точку зору. Однією з технік спотворення аргументації є техніка "викривлення пропуском" (*slanting by omission*), яка, по суті, базується на підтверджувальному викривленні: "У процесі прийняття рішення про те, що буде, а чого не буде повідомлено та підкреслено, сперечальнику відносно легко повідомити ті факти й деталі, які сприяють бажаному враженню. У цьому процесі сперечальник може применшити або зовсім пропустити ті факти, які наводять на альтернативний висновок. Сперечальник дійсно подає "нічого, крім правди", але не в змозі надати "всю

правду", уникаючи аспектів ситуації, які можуть викликати сумніви щодо його точки зору" (Groarke, & Tindale, 2004, p. 117). Техніки "викривлення" майже завжди наявні в тих процесах, коли людина чи група людей просуває певну ідею, продукт тощо. Звісно, що найочевиднішим спотворення є у пропаганді. Щоправда, дослідники підкреслюють "іронію долі", яка, на їхню думку, полягає в тому, що екстремальне викривлення, характерне для пропаганди, часто робить відсутність об'єктивності прозорою, а аргументи пропагандистів радше непереконалими, ніж переконливими (*convincing*) (Groarke, & Tindale, 2004, p. 122). На жаль, другий рік повномасштабної російсько-української війни і дев'ятий рік незаконної анексії українських територій засвідчує те, що, з одного боку, російська пропаганда працює, так би мовити, у "крайніх традиціях", а з іншого – те, що кінцеві споживачі сприймають її за "чисту правду", тобто відсутність об'єктивності буде очевидною для людини як стороннього спостерігача, а не для тих, хто охоплений пропагандою.

Одним із останніх видань, що аналізує специфіку російської пропаганди, є аналітичний посібник "Гібридна війна проти України. Як перемогти на інформаційному фронті?" (2023), розроблений командою Центру стратегічних комунікацій та інформаційної безпеки, створеного при Міністерстві культури та інформаційної політики України, спільно з Центром демократії та верховенства права, в якому розглядаються в тому числі й механізми російської пропаганди, серед яких підкреслено дію підтверджувального викривлення. Дія цієї когнітивної ілюзії як тенденція вірити інформації, яка підтверджує вже наявні переконання, продемонстрована на прикладі того, що за опитуванням великий відсоток росіян вірить у те, що метою "спеціальної військової операції" є захист мирного населення Донбасу, запобігання нападу на Росію та "денацифікація" України (Центр стратегічних комунікацій, 2023, с. 11). Використання підтверджувального викривлення належить до вияву такої ознаки російської пропаганди, як швидкість, безперервність і повторюваність, коли споживач більш схильний вірити тому, що почув раніше. І навіть якщо пізніше може бути подана правдива новина, то він все одно триматиметься думки, яку почув раніше кілька разів (Центр стратегічних комунікацій, 2023, с. 10).

Якщо в пропаганді використання негативної дії підтверджувального викривлення відбувається навмисно, то в мовних моделях штучного інтелекту на зразок ChatGPT можемо спостерігати скоріше ненавмисну шкоду цієї ілюзії. ChatGPT було запущено 30 листопада 2022 р., в Україні він доступний з 18 лютого 2023 р., після тривалих перемовин із представниками OpenAI, компанією-розробником чат-боту. Однак ще за рік до приголомшливої презентації цього чат-боту група дослідників великих мовних моделей звертала увагу в тому числі й на спотворювальну дію підтверджувального викривлення, що виявляється в навчанні цих моделей. А саме: велика кількість вебтексту сприймається як те, що начебто репрезентує "все людство" і веде до "увічнення" домінуючої точки зору без належного розгляду інших позицій через те, що в базах навчальних даних великих мовних моделей зменшено вплив даних від недостатньо представлених груп людей (Bender et al., 2021, p. 613-614). Дослідники зазначали, що мовна модель – "...це система для випадкового зшивання послідовностей лінгвістичних форм, які вона спостерігала у своїх величезних навчальних даних, відповідно до ймовірнісної інформації про те, як вони поєднуються, але

без будь-якого посилання на значення: стохастичний папуга" (Bender et al., 2021, p. 617).

На підставі вищенаведеного розгляду підтверджувального викривлення як когнітивного упередження у контексті сучасної теорії аргументації та його виявів у пропаганді й мовних моделях можемо зробити такі висновки: по-перше, підтверджувальне викривлення (як й інші когнітивні ілюзії) зазвичай не усвідомлюється людиною, проте негативні дії його можуть використовуватися як навмисно (у пропаганді), так і ненавмисно (в мовних моделях). По-друге, зменшити спотворювальний ефект цього упередження можливо завдяки аргументативній стратегії (постановка критичних запитань) і тактиці (опанування схем аргументації з переліком відповідних критичних запитань). Саме такі техніки, які мотивують шукати й розглядати першочергово ту інформацію, яка суперечить нашим поглядам, дозволяють не лише оцінювати ті аргументи, які нам пропонують у різних комунікативних практиках, але й самим конструювати переконливі аргументи й, у підсумку, сприяють вдосконаленню нашого мислення. І по-третє, задля формування української наукової термінології з урахуванням специфіки досліджуваного феномена вважаємо доцільним термін "підтверджувальне викривлення" як переклад терміну "confirmation bias".

Внесок авторів: Анатолій Конверський – наукове керівництво дослідженням, валідація даних, наукове редагування тексту рукопису; Наталія Колотілова – проведення розвідки, збір, обробка та узагальнення даних, опрацювання джерел, методологія, написання основного тексту рукопису, формулювання висновків.

Список використаних джерел

- Боровкова, Х. (2023). Підтверджувальне упередження: як ми бачимо те, що хочемо бачити. <https://nauka.ua/article/nauka-proty-propagandy>
- Викривляти. *Словник української мови*: в 11-ти т. (1970–1980). АН Української РСР, Ін-т мовознав. імені О. О. Потебні; редкол.: І. К. Білодід (голова) та ін. Наук. думка. <http://sum.in.ua/s/vykrivlyaty>
- Гібридна війна Росії проти України. Як перемогти на інформаційному фронті? (2023). Центр стратегічних комунікацій
- Ішмуратов, А. Т. (1997). *Вступ до філософської логіки*. Абрис.
- Канеман, Д. (2017). *Мислення швидко й повільно*. Пер. з англ. М. Яковлева. Наш формат.
- Конверський, А. Є. (2020). *Критичне мислення*. (2-ге вид.). Центр учбової літератури.
- Надурак, В. (2022). Критичне мислення: поняття та практика. *Філософія освіти*, 28(2), 129–147. <https://doi.org/10.31874/2309-1606-2022-28-2-7>
- Упередженість. *Словник української мови*: в 11-ти т. (1970–1980). АН Української РСР, Ін-т мовознав. імені О. О. Потебні; редкол.: І. К. Білодід (голова) та ін. Наук. думка. <http://sum.in.ua/s/uperedzhenistj>
- Упередження. *Словник української мови*: в 11-ти т. (1970–1980). АН Української РСР, Ін-т мовознав. імені О. О. Потебні; редкол.: І. К. Білодід (голова) та ін. Наук. думка. <http://sum.in.ua/s/uperedzhennja>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, Sh. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *FAccT '21, March 3–10, 2021: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Privacy* (pp. 610–623). <https://doi.org/10.1145/3442188.3445922>
- Casad, B. J., & Luebering, J. E. (2023). Confirmation bias. In *Encyclopedia Britannica*. <https://www.britannica.com/science/confirmation-bias>
- Cognitive bias. In *Cambridge Dictionary*. 2023. <https://dictionary.cambridge.org/dictionary/english/cognitive-bias>
- Diogenes Laertius (1925). *Lives of Eminent Philosophers, Volume II: Books 6–10*. R. D. Hicks (Transl.). Loeb Classical Library 185. Harvard University Press.
- Eldridge, S. (2023). Cognitive bias. In *Encyclopedia Britannica*. 2023. <https://www.britannica.com/science/cognitive-bias>
- van Eemeren, F. H. (2001). The State of the Art in Argumentation Theory. In: F. H. Van Eemeren (Ed.). *Crucial Concepts in Argumentation Theory* (pp. 11–26). Amsterdam University Press.
- van Eemeren, F. H., Grootendorst, R., & Snoeck Henkemans, A. F. (2002). *Argumentation. Analysis, evaluation, presentation*. Lawrence Erlbaum Associates, Inc.
- Groarke, L. A., & Tindale, Ch. W. (2004). *Good Reasoning Matters! A Constructive Approach to Critical Thinking*. Oxford University Press.
- Groarke, L. (2021). Informal Logic. In *Stanford Encyclopedia of Philosophy*. Retrieved July 31, 2023. <https://plato.stanford.edu/entries/logic-informal/>

Kolotilova, N. (2022). Argumentation theory in the 20th century: a return to the origins of logic, dialectics, rhetoric. In *International Scientific Conference "The Days of Science of the Faculty of Philosophy – 2022"* (pp. 176–179). <https://doi.org/10.17721/dsphd.2022/45>

Kolotilova, N. (2023). Confirmation bias from the point of view of argumentation theory. In *International Scientific Conference "The Days of Science of the Faculty of Philosophy"* (pp. 247–249).

Macagno, F., Walton, D., & Reed, C. (2017). Argumentation Schemes. History, Classifications, and Computational Applications. *IFCoLog Journal of Logics and their Applications*, 4(8), 2493–2556.

Oswald, M. E., & Grosjean, S. (2004). Confirmation bias. In R. F. Pohl (Ed.). *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory* (pp. 79–96). Psychology Press.

Plato (1921). *Theaetetus. Sophist*. Harold North Fowler (Transl.). Loeb Classical Library 123. Harvard University Press.

Pohl, R. F. (Ed.). (2004). *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*. Psychology Press.

Popper, K. (2002). *The Logic of Scientific Discovery*. Routledge Classics. (Original work published in English 1959).

References

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, Sh. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *FACCT '21, March 3–10, 2021: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). <https://doi.org/10.1145/3442188.3445922>

Bias. In *Dictionary of the Ukrainian language*. Retrieved July 27, 2023 [in Ukrainian]. <http://sum.in.ua/s/uperedzhenistj>

Biases. In *Dictionary of the Ukrainian language*. Retrieved July 27, 2023. [in Ukrainian]. <http://sum.in.ua/s/uperedzhenija>

Borovkova, Ch. (2023). *Confirmation bias: how we see what we want to see*. Retrieved July 26, 2023 [in Ukrainian]. <https://nauka.ua/article/nauka-proty-popandy>

Casad, B. J., & Luebering, J. E. (2023). Confirmation bias. In *Encyclopedia Britannica*. Retrieved July 25, 2023. <https://www.britannica.com/science/confirmation-bias>

Cognitive bias. In *Cambridge Dictionary*. Retrieved July 25, 2023. <https://dictionary.cambridge.org/dictionary/english/cognitive-bias>

Diogenes Laertius (1925). *Lives of Eminent Philosophers, Vol. II: Books 6–10*. R. D. Hicks (Transl.). Loeb Classical Library 185. Harvard University Press.

Eldridge, S. (2023). Cognitive bias. In *Encyclopedia Britannica*. Retrieved July 25, 2023. <https://www.britannica.com/science/cognitive-bias>

Groarke, L. (2021). Informal Logic. In *Stanford Encyclopedia of Philosophy*. Retrieved July 31, 2023. <https://plato.stanford.edu/entries/logic-informal/>

Groarke, L. A., & Tindale, Ch. W. (2004). *Good Reasoning Matters! A Constructive Approach to Critical Thinking*. Oxford University Press.

Ishmuratov, A. (1997). *Introduction to philosophical logic*. Abris [in Ukrainian].

Kahneman, D. (2017). *Thinking, Fast and Slow*. M. Yakovlev (Transl.) Nash format [in Ukrainian].

Kolotilova, N. (2022). Argumentation theory in the 20th century: a return to the origins of logic, dialectics, rhetoric. In *International Scientific Conference "The Days of Science of the Faculty of Philosophy"* (pp. 176–179) [in Ukraine]. <https://doi.org/10.17721/dsphd.2022/45>

Kolotilova, N. (2023). Confirmation bias from the point of view of argumentation theory. In *International Scientific Conference "The Days of Science of the Faculty of Philosophy"* (pp. 247–249) [in Ukraine].

Konverskyi, A. (2020). *Critical thinking*. (2nd ed.). Center for Educational Literature [in Ukrainian].

Macagno, F., Walton, D., & Reed, C. (2017). Argumentation Schemes. History, Classifications, and Computational Applications. *IFCoLog Journal of Logics and their Applications*, 4(8), pp. 2493–2556.

Nadurak, V. (2022). Critical thinking: concept and practice. *Philosophy of Education*, 28(2), pp. 129–147 [in Ukrainian]. <https://doi.org/10.31874/2309-1606-2022-28-2-7>

Oswald, M. E., & Grosjean, S. (2004). Confirmation bias. In R. F. Pohl (Ed.). *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory* (pp. 79–96). Psychology Press.

Plato (1921). *Theaetetus. Sophist*. Harold North Fowler (Transl.). Loeb Classical Library 123. Harvard University Press.

Pohl, R. F. (Ed.). (2004). *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*. Psychology Press.

Popper, K. (2002). *The Logic of Scientific Discovery*. Routledge Classics. (Original work published in English 1959).

Russia's hybrid war against Ukraine. *How to win on the information front?* (2023). Center for Strategic Communications [in Ukrainian].

Slant. In *Dictionary of the Ukrainian language*. Retrieved July 27, 2023. [in Ukrainian]. <http://sum.in.ua/s/vykrivljaty>

van Eemeren, F. H. (2001). The State of the Art in Argumentation Theory. In: F. H. van Eemeren (Ed.). *Crucial Concepts in Argumentation Theory* (pp. 11–26). Amsterdam University Press.

van Eemeren, F. H., Grootendorst, R., & Snoeck Henkemans, A. F. (2002). *Argumentation. Analysis, evaluation, presentation*. Lawrence Erlbaum Associates, Inc.

Отримано редакцією журналу / Received: 01.12.23

Прорецензовано / Revised: 11.12.23

Схвалено до друку / Accepted: 18.12.23

Anatolii KONVERSKYI, Active Member of the National Academy of Sciences of Ukraine, DSc (Philos.), Prof.

ORCID ID: 0000-0002-3916-4446

e-mail: dean_phd@univ.kiev.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Nataliia KOLOTILOVA, DSc (Philos.), Assoc. Prof.

ORCID ID: 0000-0002-7450-2981

e-mail: nataliakolotilova@knu.ua

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

CONFIRMATION BIAS IN ARGUMENTATION PROCESSES

Background. The article is devoted to the study of confirmatory distortion as a cognitive bias within the framework of the modern theory of argumentation. In the context of this study, the effectiveness of the critical questioning technique as an argumentation strategy aimed at reducing the negative impact of confirmatory bias is considered.

Methods. To achieve the goals of the research, the method of critical questions is used, which is based on the traditional principles of logic, dialectics and rhetoric. Specific lists of critical questions are developed for each argumentative scheme.

Results. The article demonstrates that standard critical questions for schemes of practical reasoning, developed by eminent specialists, are able to minimize the negative impact of confirmation bias. This is achieved by taking into account information that may conflict with the agent's action plan.

Conclusions. The study highlights the importance of using critical questions as a tool to minimize confirmation bias. Such an approach not only contributes to the substantiation of arguments, but also stimulates self-control and the search for information that may cause doubts about one's own views. The final part of the article examines propaganda as the deliberate use of confirmatory distortion and points to the need for attention to language patterns where it occurs rather unintentionally. In connection with the specificity of the studied phenomenon, the term "confirmation bias" is proposed as a translation of the English "confirmation bias".

Keywords: confirmation bias, cognitive bias, argumentation theory, argument scheme, critical question, practical reasoning.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.