

Деркач Вадим Леонтійович

*Кандидат філософських наук, доцент
Київський національний економічний університет
імені Вадима Гетьмана (м. Київ, Україна)
<https://orcid.org/0000-0002-0123-3637>
e-mail: vadym_derkach@kneu.edu.ua*

КОНЦЕПТ СПРАВЕДЛИВОСТІ: ТЕОРЕТИКО-ІГРОВА МОДЕЛЬ

Резюме

У статті представлено теоретико-ігрову модель справедливості, що трактує її не як моральну аксіому, а як емерджентний механізм стабілізації кооперації. Описано гру з пороговими внесками, асиметричним ризиком пропонента та ймовірнісною активацією бонусів, у межах якої гравці адаптивно формують коаліції взаємності. Показано, що стійким критерієм справедливості стає помірно асиметричне правило розподілу, засноване на взаємності ролей і гарантіях участі. Модель пояснює виникнення справедливості як інструментової чесноти, що підвищує довгострокову життєздатність спільнот і зберігає свій зміст навіть у ситуаціях розподілу втрат.

Ключові слова: справедливість, теорія ігор, кооперативні моделі, еволюційно стабільна стратегія, соціальна стійкість, критичний дефіцит і розподіл втрат

Вступ

Концепт справедливості (як значення абстрактного терміну) рамково виражає взаємоузгодженість у виправданні рішення в спірному питанні, в якому незадоволений рівень домагань суб'єктів через відповідний конфлікт інтересів. В інший спосіб справедливість можна трактувати як відповідність винагороди чи покарання узгодженим очікуванням. У Арістотеля справедливість — це чеснота співмірності заслуг та винагород, проступків та покарань [1]. Але як встановити цю співмірність? Очевидно, що міру виправданої плати чи відплати неможливо визначити безвідносно до очікувань людей, пов'язаних з їхнім сприйняттям власної гідності й значущості для них спільноти, статус в якій вони прагнуть зберегти. В свою чергу

закріплення певних норм, які регулюють систему винагород і покарань, впливатиме на стійкість спільноти та ефективність соціальних дій.

Внутрішньо, те, що людина оцінюватиме як несправедливе, переживатиметься нею як гіркота образи, приниження та розчарування, породжуватиме гнів з жагою помсти, ревність чи розпач від безсилля та зруйнованих очікувань. Оскільки в основі цих очікувань лежить рівень домагань, то характер реакції на несправедливість (як і сприйняття чогось як справедливого чи ні) сильно залежить від стеничних рис в характері особистості: люди честолюбні, орієнтовані на здобуття високого соціального статусу, з установками домінанта особливо гостро будуть реагувати на приниження і перешкоду своїм домаганням. Натомість люди залежні особливо схильні будуть шукати покровителя, який захистить їхні інтереси, тому пошук «справедливості» переважно експлікується через численні скарги, нарікання тощо.

Однак рівень домагань в значній мірі залежить від культивованої норми взаємин та зразка гідної людини (тобто від моралі). Тому, всупереч поширеним уявленням, не принцип справедливості визначає мораль, а мораль спільноти формує певні очікування щодо допустимого та належного, які закріплюються в структурі образу ідеального себе і таким чином вкладають рівень домагань в деяку норму. Порушення її буде оцінюватися як неприпустима зневага, й відсутність виправлення — як несправедливість, як втрата власного значення в спільноті. Відтак зворотно сама спільнота буде знецінюватися тим, хто зазнає несправедливого до себе ставлення (не відповідає морально виправданому рівню домагань).

Проте в сучасному складноорганізованому суспільстві мораль вже не може працювати як самодостатній механізм соціального контролю в силу розмитості спільнот і високої соціальної мобільності. Але деяка інтуїція очікуваної «справедливості» настільки укорінена в природі людини, що правова система не може бути стійкою, якщо вона тримається лише на залякуванні і виключає повністю солідарність.

Філософи століттями намагалися знайти певну «формулу» справедливості, деякий загальний принцип виправдання рішення, який і забезпечував би таку взаємоузгодженість. Тобто таке правило-критерій, орієнтуючись на яке учасники конфлікту готові були би визнати, що рішення їх задовольняє без приниження котроїсь зі сторін і мусить бути прийнятним. Юрген Габермас, наприклад, пропонує комунікативну модель раціональності, де справедливість постає не як зовнішній закон, а як результат вільного діалогу між учасниками, здатними обґрунтувати свої претензії один перед одним [2]. Подібну ідею розвиває Амартія Сен, наголошуючи, що реальна соціальна справедливість виявляється у практичному наближенні до стану, який суб'єкти визнають кращим за попередній [3]. У свою чергу,

Нансі Фрейзер і Аксель Гоннетт акцентують увагу на факторі визнання, вказуючи, що справедливість неможлива без соціального підтвердження гідності кожного [4].

Пошук універсального принципу справедливості історично розгортався у двох головних напрямках. Перший — деонтологічний, у якому справедливість вбачалася у відповідності дії загальному правилу, незалежно від наслідків (самоцінність справедливості). У модерній теорії Джона Роулза ця ідея набуває процедурної форми: справедливим визнається таке рішення, яке могло б бути ухвалене в умовах «ширми невідання», де кожен учасник позбавлений знання про майбутнє власне становище в «грі за правилами», а тому оцінює не результат гри, а самі правила як засадничо справедливі. «Почуття справедливості» при цьому ґрунтується на деякій моральній інтуїції, а справедливість постає однією з чеснот соціальних установлень, без яких вони не гідні [5].

Другий напрям — функціональний, який розглядає «справедливість» рішень як інструмент, що забезпечує благо для спільноти та процвітання її членів.

Сьогодні пошук «формули» справедливості поступово зміщується від пошуку універсального принципу до аналізу умов можливості взаємного виправдання — тобто до самого того процесу, який на початку був окреслений як рамкова взаємоузгодженість у спірному питанні. Справедливість перестає бути лише нормою розподілу чи моральним ідеалом і постає як жива структура комунікації, у якій суб'єкти не лише відстоюють свої домагання, а й водночас перевіряють, чи можуть їх виправдати перед іншими без втрати взаємної поваги.

Важливо також зауважити, що справедливість не визначається, як часто пишуть, забезпеченням рівності можливостей, оскільки сама ця рівність є відкритим питанням: ми осмислено можемо спитати, чи завжди справедливою є рівність?

Таким чином потрібно розрізнити (1) рамкове визначення терміна «справедливість» (його функціональний зміст в мові), (2) конкретне рішення та судження про нього як справедливе чи ні, зі взятим критерієм цієї оцінки й (3) принцип справедливості як загальне правило, на основі якого робилися спроби підкласти під конкретні критерії певну раціональну основу.

Далі я відкину деонтологічний підхід в тому сенсі, в якому справедливість як самоцінність розуміється метафізично; однак деяке почуття справедливості можна гіпотетично розглядати як певну особливість психічної організації людини, вироблену в ході тривалої еволюції (приблизно так само, як кольоровий зір). Проте закріпитися таке почуття могло лише

за умови, що воно функціонально підвищує стійкість та живучість спільнот і має надійну основу для свого відтворення. Тому навіть якщо ми й визначаємо наявність такого вродженого (на рівні молекулярно-генетичного механізму спадковості) або притаманного будь-якій культурі (на рівні культурно-генетичного механізму) деякого універсального почуття справедливості, це не має істотної ролі в розумінні по суті, оскільки апеляція до нього — це апеляція до функції.

В цьому зв'язку найбільш плідним підходом видається теоретико-ігрове моделювання умов соціальних взаємодій, засобами якого можна показати хід відбору певних стратегій поведінки, які максимізують певну цільову функцію.

У цьому напрямі особливе місце посідають моделі з пороговими внесками та кооперативними ефектами, які дають змогу простежити, як індивідуальні рішення впливають на досягнення колективної мети. Зокрема, An, Dong, Wang і Zhang (2023) у дослідженні “Cooperation and Coordination in Threshold Public Goods Games with Asymmetric Players” показали, що у системах, де учасники мають різні початкові ресурси й потенціали, ключову роль відіграє асиметрія внесків. Виявлено, що навіть незначна різниця у вихідних параметрах може зумовити виникнення коаліцій, здатних досягати порогового значення, яке активує бонуси. У таких умовах стратегії кооперації мають шанс на успіх лише за умови, якщо існує механізм координації, що компенсує дисбаланс між гравцями [6].

Подібні висновки розвиває Zhang (2025) у роботі “Co-evolutionary dynamics of threshold PGGs with multiple strategies”, де розглянуто коеволюційний аспект порогових ігор. Моделювання показує, що стійкість кооперації залежить не лише від індивідуальних внесків, а й від динамічної адаптації стратегій, коли учасники переглядають власні рішення, орієнтуючись на успішність попередніх раундів. Відтак, системи з кількома конкуруючими стратегіями виявляють еволюційні цикли, у яких короткостроковий егоїзм може тимчасово витіснити кооперацію, але довгостроково саме спільні дії формують стабільні рівноваги [7].

У межах еволюційно стабільних стратегій (ESS) Bach (2006) у праці “The evolution of n-player cooperation-threshold games” розробив алгоритм визначення стабільних стратегій у багатогравцевих іграх із пороговими умовами. Він показав, що за певних співвідношень вигід і витрат виникають групові рівноваги, де навіть мінімальна частка кооперативних учасників може забезпечити досягнення спільного результату, якщо інші реагують адаптивно [8]. Аналогічно, Chen (2023) у статті “Evolutionary game analysis on decision-making behaviors” виділив п'ять типів еволюційно стабільних стратегій, де рішення кожного гравця змінюється після досягнення критичного порогу колективної дії. Це безпосередньо перегукується з

пропонованою в цій роботі моделлю, у якій гравці конкурують як автори пропозицій, адже їхня стратегія також модифікується на основі досвіду та успіху попередніх пропозицій [9].

Окремий вимір становлять дослідження справедливості у кооперативних іграх, що уточнюють механізми розподілу спільних вигадів. У роботі de Clippel і Rozen (2019) “Fairness through the Lens of Cooperative Game Theory” представлено експериментальні дані, які показують, що реальні учасники схильні до розподілу, який поєднує принцип рівності та внескової справедливості (через значення Шеплі) [10]. Це свідчить про те, що навіть у раціональних умовах рішення гравців включає установковий моральний компонент. У дослідженні Balkanski (2025) “Mechanisms for Fair Attribution” розвинуто цю ідею, запропонувавши формальні критерії існування й унікальності справедливих механізмів винагороди (зокрема, через нуклеолус) [11].

Отож поєднання результатів цих праць дозволяє бачити, що ідея порогових бонусів, конкуренції авторів пропозицій і адаптації уявлень про вигідність вписується у ширший контекст еволюційно-кооперативних моделей. Її евристична цінність полягає не лише у можливості дослідження стабільності певних стратегій, а й у тому, що вона дозволяє моделювати сам процес етичної раціоналізації вигідності. У таких іграх критерій оптимальності не задається ззовні — він формується внаслідок взаємодії гравців, які поступово узгоджують власні очікування, навчаються передбачати реакції інших і таким чином створюють умови для появи спільного правила, яке визнається «справедливим».

З цього погляду, порогові кооперативні моделі дають змогу описати механізм становлення справедливості як колективної рівноваги, а не як абстрактного морального принципу. Коли кожен гравець оцінює свої втрати та вигоди відносно внеску інших, у системі природно виникають пропорційні механізми компенсації, подібні до тих, що формально описуються через значення Шеплі або нуклеолус. Тобто справедливість у моделі не задається — вона емергує як узгоджений баланс між внеском і результатом, між прагматичним інтересом та соціальною стабільністю.

У цьому сенсі ігрове моделювання справедливості має суттєву евристичну перевагу над чисто нормативними підходами: воно дозволяє спостерігати, як із безособових механізмів змагання та взаємної вигоди поступово постає норма, яку суб'єкти визнають морально виправданою. Тим самим теоретико-ігрова перспектива не просто ілюструє етичні уявлення, а пояснює їхню функцію — як механізм стабілізації спільнот у мінливих умовах взаємозалежності.

Методи дослідження.

У роботі застосовано методи теоретико-ігрового моделювання та порівняльного аналізу існуючих моделей кооперативної поведінки, зокрема порогових ігор із публічними благами, еволюційно стабільних стратегій. Методологічна основа спирається на поєднання аналітичного моделювання стратегічних виборів і якісного аналізу їхніх евристичних наслідків для концепції справедливості. Також використано елементи симуляційного підходу для виявлення умов стабілізації кооперації та формування спільних критеріїв вигідності, що дозволяє оцінити, як індивідуальні стратегії поведінки переходять у колективно узгоджені норми.

Результати дослідження

Модель «ідеальної кооперації» та конкуренції «критеріїв справедливості»

Гравці роблять «внесок» — деяку суму грошей, розраховуючи отримати прибуток від участі в цій грі. Цей прибуток залежить від «кооперативного ефекту», який можна змоделювати як «пропозицію» участі в геймі (окремому ігровому епізоді), де, за умови, що вона буде підтримана, то учасники гейму отримують певну суму, більшу, аніж їхні внески, яку потім вони можуть розподілити між собою. Ефект «кооперації» в окремому геймі виникає тоді, коли збереться певна кількість учасників, готових внести деяку порогову суму (порог входження) і за умови, що набрана деяка порогова агрегована сума (порог ефекту). Наприклад, гравець-ініціатор (пропонент) робить пропозицію-заявку, яка спрацьовує за умови, що буде набрано тисячу умовних одиниць і за умови, що знайдеться 10 учасників із внеском не менше 50 одиниць. Якщо ці умови виконаються, то учасники автоматично отримують 200 додаткових одиниць. Оскільки оголошений мінімальний внесок не задовольняє другу умову (це вдвічі менше, ніж потрібно для спрацювання кооперативного ефекту), а в геймі кількість учасників строго визначена — рівно 10, то ефекту буде досягнуто, якщо хтось із гравців внесе більшу суму, але не обов'язково, щоб цей внесок був порівно розподілений між усіма.

Варіантів пропозицій (правил гейму) може бути безліч. Немає також обмежень на сукупну кількість учасників цієї гри — вона невизначена. Мета кожного з гравців — отримати максимальний прибуток як частину тих додаткових одиниць, які генеруються як «кооперативний ефект».

Хід гри полягає в тому, що учасник робить пропозицію (оголошує її), і з деякою ймовірністю (яка визначається додатково як характеристика в моделі, якою можна варіювати) його пропозиція «приймається» ігровим автоматом (незалежно від бажань гравців, шляхом сліпого жеребкування), причому так, що чим вищий прибуток домагається, тим менша ймовірність цього прийняття. Ця ймовірність однак зростає, якщо пропозиція передбачає більшу кількість учасників і більший пороговий внесок. Тобто

ймовірність актуалізації суми бонусу є обернено пропорційно (можливо, не лінійно) від заявленої величини бонусу і прямо пропорційно кількості залучених учасників та величини порогового внеску (сумарного і базового для гравця). Наприклад, пропозиція для двох учасників з пороговим внеском в 1 одиницю і виграшем-бонусом 10000000 одиниць прямує до нуля, але пропозиція на таку суму при тисячі учасників і пороговим внеском 100 одиниць — має істотно вищу ймовірність актуалізації (але не гарантовано прийнята).

Пропонент, однак, не має гарантій, що потрібна кількість учасників підтримає його пропозицію. Після того, як пропозиція актуалізована (випав шанс), потрібно, щоб її підтримали учасники (за правилами пропозиції гейму), тобто зробили свої внески.

Стартова сума в усіх однакова. В одиницю часу гравець може зробити рівно одну пропозицію, але «приєднатися» до інших пропозицій може стільки, скільки в нього вистачатиме коштів. Якщо пропозиція актуалізована, автор-пропонент мусить автоматично її підтримати, оголошуючи при цьому правило розподілу бонусу між учасниками (принцип справедливості). Але якщо ніхто не пристане на його пропозицію, він втрачає суму ставки.

Питання полягає в наступному: які правила розподілу виграшу будуть такими, що така гра буде «стійкою», тобто все більше учасників долучатимуться до неї і матимуть можливість продовжувати її далі і далі? Які пропозиції вигідно робити, щоб ефективно «спокушати» учасників і максимізувати свій дохід?

Розглянемо цю модель спочатку змістовно в контексті концепції справедливості і як методологічний підхід до аналізу, а потім її формалізацію і можливі стратегії комп'ютерної симуляції.

Початкові евристики

Будь-яке стійке соціальне утворення можливе лише за умови, що в ньому досягається кооперативний ефект, який в конкретній екологічній ніші забезпечує вищу живучість (відтворюваність) певного типу організмів, порівняно з тим, якби вони жили осібно. Вигоди соціальної кооперації, однак, можуть бути розподілені не рівномірно таким чином, що частина соціальної групи є донором благ для іншої, тобто їхні зусилля і ризики підвищують в більшій мірі живучість інших членів групи, а не свою власну. В граничному випадку особина жертвує власним життям (байдуже, «добровільно» чи примусово), підвищуючи живучість інших. Однак очевидно, якщо такі жертви знижують відтворюваність групи в цілому, то стійким таке утворення не може бути. Тут ми візьмемо за відправний пункт аналізу ідеалізований граничний випадок, коли такі жертви взагалі відсутні, тобто

характер кооперативного ефекту такий, що всі без винятку члени групи отримують гарантовано вигоду від членства і це точно дає приріст в благах, порівняно з відмовою від участі. Виключимо також примусовий характер членства.

Отже, учасник, який приєднується до гейму, ніколи не може щось втратити: гра для нього не містить ризику. Окрім того, правила розподілу бонусу завжди є гранично прозорими: гравець точно знає, що він отримає, приєднавшись до гейму, і що отримають інші учасники. Ніякого шахрайства чи невизначеності в розподілі блага бути не може; ніхто з учасників не може вплинути на цей розподіл ні до гейму, ні в ході. Пропонент, який заявляє критерій розподілу, робить це до того, як до гейму хтось приєднається.

Зауважмо, що на відміну від принципу «ширми непередбачуваності» Дж. Роулза, тут приймається обернений підхід: гравці точно знають «своє місце» внаслідок участі в грі за оголошеними правилами. Цей принцип тотожний тому, який описують як добровільна інформована участь.

Окрім того гра виключає можливість «нульової суми»: ніхто не може отримати приріст блага для себе за рахунок зменшення його в когось. Здобутки не досягаються за рахунок втрат. Інакше кажучи, для кожного з учасників гра є безпрограшною. Експлуатація людини людиною виключена, як і виключене використання когось без його добровільної згоди в інституційних структурах; правила ніколи не нав'язуються як даність.

Додатково виключена можливість паразитування, отримання вигоди без власного внеску. Виграш для учасників за рахунок кооперативного ефекту можливий лише за умови участі з деяким визначеним наперед в заяві пропонента внеском порогового входження.

Умови також усувають можливість «чистого зла»: ніхто не може нанести шкоду активно один одному як самоціль власних дій.

Нарешті, відсутні хоча б якісь дискримінаційні способи усунути учасника на підставі його особистих рис і всі учасники мають стартові рівні умови (одну й ту саму початкову суму).

Мотивацію участі в цій грі ми братимемо як константу і винесемо за дужки.

Сукупність всіх цих умов можна інтуїтивно визнати справедливими за замовчуванням, вважаючи, що лише відхилення від них має якимось виправдовуватися. Тобто всі ці відхилення можуть визнаватися справедливими лише після обґрунтування; в межах цих рамок таке обґрунтування зайве.

Це дає змогу виокремити «чисту» якість справедливості, яка буде визначатися лише рівнем домагань гравця співвідносно з інформацією як змінюється його ставище, порівняно з такими змінами для інших гравців.

Причому його реакція виражатиметься прямо у відмові від участі в спільних зусиллях чи у бажанні їх підтримувати, тобто прямо впливатиме на стійкість соціальних утворень.

Ця гра має очевидне обмеження, оскільки ми не розглядаємо тут елімінацію «гравця» зі спільноти як активну соціальну дію й примус гравця «до правил» як умову збереження його статусу (зрештою, і життя). В цій грі немає зовнішніх покарань; єдине, в чому може гравець зазнати невдачі, це те, що він втратить свій «капітал», якщо буде робити пропозиції або малореалістичні (з низькою ймовірністю актуалізації), або вони будуть проігноровані учасниками як неприйнятні. Причому ця «неприйнятність» виникатиме або через конфлікт домагань, або через те, що гравець має кращі пропозиції (з якихось його міркувань) з боку інших пропонентів. Відтак особливістю гри є динамічна зміна установок гравців залежно від їхнього досвіду і корекції очікувань й, окрім того, конкуренція пропонентів.

Фактично, ризикує в цій грі лише пропонент, оскільки він може втратити заявлений внесок, якщо гейм буде актуалізований, але не підтриманий.

Аналіз критерію справедливості в конкретних життєвих умовах (а в них ніколи цей ідеалізований випадок не виконується) можна, таким чином, розбити на два етапи: побудову об'єктивної оптимізованої стратегії максимізації соціальної стійкості з врахуванням фактора «чистої» справедливості для ідеальних умов і розбору допустимого відхилення у визнанні справедливими паразитування, експлуатацію, дискримінацію, обман, примус, агресію, стартову нерівність тощо за умови, що в конкретних умовах якийсь із цих чинників підвищуватиме сукупну пристосованість спільноти й одночасно в ній надійно відтворюватиметься ядро бенефіціарів, для яких критерій справедливості вигідний тому, що вони мають вагомні гарантії утриматися в цьому ядрі, не зруйнувавши відтворення спільноти (критерій справедливості, прямо протилежний Роулзовій ширмі непередбачуваності та принципу мінімізації диференціації).

При цьому, для ідеальних умов, виправдання критерію справедливості зводиться лише до ілюстрації того, що «справедливість» в розподілі бонусів кооперативного ефекту закріплює деяку соціальну коаліцію, яка орієнтується на цей критерій як на взаємно очікуване узгодження, що максимізує їх виграш на тому рівні домагань, який вони приймають для себе. Такий розподіл вони вважатимуть гідним для себе, а частота конфліктів через оцінку плати як неспівмірної з огляду на те, що отримують інші — мінімальною.

Методологічно, ця модель, хоча вона й орієнтована на конкретні емпірично виявлені умови й історично сформовані погляди на справедливість, але ця емпірична складова вагома лише в другій частині аналізу.

Але на рівні ідеалізації конкретні стратегії можна проробити засобами комп'ютерної симуляції, оскільки в самій моделі абстрагована вичерпно суть «чистої» справедливості як відношення.

Формалізація моделі

1. Основні об'єкти

Набір «потенційних» гравців: I (може бути великою або фактично нескінченною множиною; у реалізації — скінченна, N).

Дискретний час ($t=0,1,2,\dots$). На кожному кроці кожен гравець має однаковий початковий капітал (W) (можна інтерпретувати як «банк»).

В кожному кроці гравець може виконати одну з двох ролей: (i) зробити пропозицію (стати пропонентом) або (ii) не пропонувати і/або приєднуватися до існуючих відкритих пропозицій (join).

2. Пропозиція (епізод гейму)

Пропозиція — це кортеж параметрів

$$\pi=(n,c_{\min},C_{\text{tot}},B,f),$$

де n — потрібна кількість учасників (жорстко визначено для цього гейму);

c_{\min} — мінімальний індивідуальний поріг внеску (мінімальна сума, яку має внести кожен учасник);

C_{tot} — поріг агрегованої суми (може бути альтернативою/додатком до n); у базовому випадку $C_{\text{tot}}=n \cdot c_{\min}$ але допускаються відмінності;

B — розмір бонусу (додаткових одиниць), що генерується, якщо пропозиція актуалізована та підтримана;

f — правило розподілу бонусу між учасниками (функція виплат), f : (вектори внесків, B) \rightarrow вектор алокацій $(\alpha_1, \dots, \alpha_n)$ з умовою $\sum_i \alpha_i = B$.

Автор-пропонент одночасно є потенційним учасником і, якщо пропозиція актуалізована, автоматично має підтримати її (тобто внести свій оголошений внесок).

3. Стохастична актуалізація (ігровий автомат)

Перш ніж збір учасників почнеться, пропозиція проходить стадію «жеребкування/активації» — тобто належить подія A (актуалізація).

Ймовірність активації позначимо

$$p=P(A|\pi)=\Phi(B,n,c_{\min},C_{\text{tot}}),$$

де Φ — задана функція: монотонно спадна за B (чим більший заявлений бонус, тим менший шанс активації), і монотонно зростає за n і c_{\min} (чим більше учасників / вищий поріг — тим вище p). Функція може бути нелінійною; приклад:

$$\Phi(B,n,c)=\sigma(n,c) \cdot (1+\lambda B)^\gamma$$

де $\sigma(n,c)$ — зростаюча у (n,c) нормалізована величина у $[0,1]$, $\lambda, \gamma > 0$ — параметри «несприйнятливості» великих бонусів.

4. Фаза збору підтримки

Якщо А (актуалізація) відбулася, пропозиція відкривається для збору учасників. Кожен гравець (i) може (за наявності достатніх коштів) обрати: приєднатися до конкретної відкритої пропозиції л і зробити внесок $c_i \geq c_{\min}$ (може перевищувати c_{\min} ; або не приєднуватися).

Підтримка вважається успішною, якщо одночасно виконуються умови пропозиції, наприклад:

$$\#\{i: c_i \geq c_{\min}\} \geq n_i \text{ і } \sum c_i \geq C_{\text{tot}}.$$

Пропонент сам вносить свій вклад c_p (записаний у пропозиції або оголошений під час підтримки).

5. Виплати (payoffs)

Якщо пропозиція не була активована (подія А не відбулася): пропонент або може втратити частину ставки (залежить від протоколу); в простому варіанті — пропонент нічого не втрачає (або втрачає фіксований cost); якщо А істинне, але підтримки не набрано — пропонент втрачає всю чи частину обіцяної суми (sP) (stake).

Якщо А істинне і підтримка набрана (успіх): кожен учасник (i) має payoff

$$u_i = W_i - c_i + \alpha_i,$$

де W_i — початковий капітал (можна опустити як константу), c_i — внесок, $c_i = f_i(c, V)$ — частка бонусу відповідно до правила (f). Пропонент також отримує свою (α_p).

Якщо А істинне, але підтримки не набрано: кожен, хто приєднався, повертає свій внесок.

6. Стратегічний простір

Для кожного гравця: стратегія s — це правило вибору ролі та поведінки на кожному кроці, тобто

$s: (\text{історія}) \rightarrow \{\text{пропонувати л, або не пропонувати; для join: вибрати л}^*, c_i\}$.

Простір стратегій включає: як визначається параметри пропозиції л (для пропонента), як обирати, чи приєднатися, як обирати вклад (c_i), та як оновлювати поведінку на основі попереднього досвіду (learning rule).

Критерії стійкості (формально)

Нехай P_t — частка населення, що беруть участь у грі (серед активних пропозицій) на кроці t. Ми називаємо гру стійкою, якщо існує атрактор у динаміці стратегій s, для якого P_t монотонно зростає (або залишається високим) при перших малих збуреннях популяційної структури; формально — існує стійка рівновага стратегій s^* така, що для сусідніх станів $s \approx s^*$, $\lim_{t \rightarrow \infty} P_t(s) \geq P_0$.

Розглянемо деякі евристичні міркування щодо стратегій гри, які мають конкурентну перевагу перед іншими і потенційно є еволюційно стійкими (жодна інша не може їх витіснити).

Очевидно, що пропонент, який обирає максимально обережний підхід — заявляти невеликий бонус при мінімальній кількості учасників, але з великим пороговим внеском (що максимізує ймовірність актуалізації), може не отримати підтримки жодного, якщо пропонує різко асиметричний розподіл невеликого бонусу. Але якщо він пропонує ділити бонус порівну, то потрапляє в невідгідне становище, бо свою частку він отримує з ризиком, а інші учасники, що приєдналися — ні. Виникає колізія, коли великий пороговий внесок пропонента робить його ризиковано залежним від згоди партнерів, що змушує обіцяти їм більшу частку, але саме це робить тепер вже для нього менш привабливим рішенням робити пропозицію.

Виходить так, що приєднуватися до пропозицій вигідніше, ніж робити їх самому (бо таке приєднання не містить ризику взагалі). Але це унеможливує отримати взагалі винагороду.

Натомість удача — заявка на дуже великий бонус при невеликих внесках, успіх актуалізації якої є малоімовірний, — якщо вже так трапилося, відкриває можливість спокушування згідливих за умови, що сума винагороди доволі висока, але й сам пропонент може її розподіляти різко асиметрично на свою користь.

В цьому сенсі для учасника стратегічно вигідно стимулювати інших робити заявки й ризикувати, погоджуючись на асиметричність розподілу виграшу на їх користь, тобто на накопичувану нерівність, розраховуючи, що при цьому виникатимуть умови для власного стабільного зростання доходу без ризику. Але для пропонента все одно залишається ризик того, що пропозиція не буде актуалізована, а якщо буде, то він може потрапити в «ігнор». Це в свою чергу стимулює робити заявки на значні бонуси, але тепер вже через об'єктивно низьку ймовірність такі заявки будуть холостими. Й ті, хто орієнтується на більш реалістичні варіанти, отримують стратегічну перевагу. Однак їм потрібно максимізувати гарантії участі.

Очевидно, що утворення «груп взаємності» й створює такі гарантії. Якщо деяка кількість гравців стабільно підтримують пропозиції один одного, тобто всі вони в кожен такт гри зроблять заявку як пропонент, який заявляє цю кількість учасників і доволі низький пороговий внесок, але й сам робить внесок при всіх пропозиціях інших, то загальний успіх буде визначатися лише ймовірністю активації. А вона зростає при невеликих домаганнях і збільшенні кількості учасників такої коаліції. При цьому уніфікація правила розподілу (я пропоную такий розподіл бонусу, на який погоджуюся, якщо його пропонує інший, і всі роблять так само) усуває спокусу диктату власника «джекпоту» й водночас ризику ігнорування з боку

тих, хто не прийме його ультиматум. Однак не вигідно, щоб таку ризиковану заявку робили всі в коаліції. Це може стабільно робити один, але завжди пропонуючи розподіл великого бонусу на всіх, або по черзі всі так само. В чому ж тоді вигода ділитися великим вигрaшем пропонентові? Очевидно, що такий поділ веде до приросту стартової суми в усіх, а отже дозволяє підвищувати регулярні винагороди (невеликі й реалістичні) за рахунок підвищення ймовірності актуалізації (вона росте, якщо порогова сума зростає). При цьому успішним спільнотам з досвідом такого ведення гри стає вигідно об'єднуватися з іншими такими спільнотами. За рахунок цього ще більше зростає реалістичність трохи вищого кооперативного ефекту й разом з тим ймовірність зірвати джекпот.

Коаліції мають еволюційну перевагу, бо вони вирівнюють ризики, вони уніфікують правило розподілу, вони створюють предиктивність результату, вони генерують більший сумарний дохід, і вони видають стабільні сигнали надійності, що робить їх привабливими для інших гравців.

У термінах еволюційної теорії ігор: *коаліція взаємності є еволюційно стабільною стратегією (ESS) проти індивідів-опортуністів.*

В цій моделі конкуренція, ризик і випадковість призводять до того, що: індивідуалістичні стратегії або високоризикові, або нежиттєздатні; стратегія «я приєднуюсь, але не пропоную» може працювати лише паразитично і не може домінувати; коаліції взаємності з уніфікованим розподілом бонусу випереджають будь-які інші стратегії; коаліції, що діляться «джекпотом» симетрично, мають перевагу в експоненційному зростанні капіталу, а з ним — і в успіху пропозицій. Таким чином, у моделі найуспішнішою є не асиметрія, а симетрична, довгостроково вирівняна кооперація.

Однак яким має бути оце уніфіковане правило розподілу (критерій справедливості, якщо таке правило приймається як очікувана взаємність)?

З аналізу представленої моделі, це правило можна представити в наступних пунктах.

1. Симетрія при рівних внесках

Якщо два учасники роблять однаковий мінімальний внесок і однаково «підтримують» пропозиції інших, розподіл не повинен робити їх нерівними. Суть: ніякого «статусного» привілею.

2. М'яка асиметрія за ініціативу

Пропонент має отримувати трохи більше, але не настільки, щоб інші вважали це диктатом або ризиком ігнорування. Суть: винагорода за ризик — але в межах прийнятності.

3. Позитивність для всіх у довгій грі

Розподіл повинен збільшувати стартові фонди всіх учасників (у середньому), бо саме за рахунок цього коаліція може: підвищувати пороги, зменшувати ризики, стабілізувати доходи.

4. Стабільність: жоден учасник не має стимулу вийти з коаліції

5. Прозорість і передбачуваність

Правило повинно бути простим, щоб кожен легко перевіряв: що він отримає, що отримає будь-хто, як зміна порогу або числа учасників вплине на всіх.

Звідси, уніфіковане правило розподілу, яке задовольняє всі ці вимоги: нормована пропорційність із премією за ініціативу»

Формула-принцип:

1. Всі члени коаліції отримують:

$$\text{Базова частка} = B \setminus N + \alpha$$

де B — бонус (у т.ч. джекпот), N — кількість учасників, α — ваговий коефіцієнт ініціативи.

2. Пропонент отримує премію за ризик і ініціативу:

$$\text{Частка пропонента} = \alpha B \setminus N + \alpha$$

Таким чином: коли $\alpha \geq 1$, пропонент отримує трохи більше, але зростання α має бути лінійним, а не експоненційним, тобто пропонент не може захопити 50 — 80% бонусу.

Бо воно: не дає мотиву виходити з коаліції (кожен отримує більше в середньому), не створює мотиву домінувати, адже пропонент не може сильно збільшити свою частку, зменшує стимул робити нереалістичні заявки, бо прибуток зростає через частоту активацій, а не величину бонусу, уніфіковане: кожен погоджується на таке, бо знає, що за тим же правилом отримає премію, коли сам стане пропонентом

Це операційне правило справедливості, яке: не редукується до рівності, але й не допускає експлуатації, будується на взаємності ролей, створює інституційну довіру.

Це не «справедливість результату», а «справедливість ролей»: *те, що ти приймаєш як вигідне для інших, ти готовий прийняти для себе, коли сам будеш у їх ролі.*

Модель вводить критерій справедливості як операційний принцип взаємності, що виникає як оптимальна стратегія в повторюваній кооперативній грі з ризиком для пропонента і можливою асиметрією розподілу благ. Це дозволяє зіставити її з основними традиціями.

У Дж. Роулза справедливість — це установлення правил гри шляхом гіпотетичного договору за «ширмою непередбачуваності», де раціональні агенти максимізують мінімальний вигравш (максимін).

В представленому підході критерій справедливості не задається умовою рівності, а виникає як стабілізуюча стратегія в реальній динаміці взаємодій. Ключовим стає не «максимін», а стабільність коаліції та відсутність стимулів до виходу з неї. Нерівність допустима, якщо вона зміцнює кооперацію і підвищує імовірність спільного успіху. Справедливість не є базовим моральним принципом, а еволюційним продуктом стратегічної взаємодії, що не потребує ширми непередбачуваності.

Утилітаризм вводить принцип максимізація сумарної корисності.

В представленій моделі максимізація сумарного вигравшу трапляється природно, але як побічний ефект стабільної коаліції, а не початкова моральна вимога. Одиниця аналізу — не поведінка індивіда, а інтерактивний процес, де вигравші залежать від поведінки інших. Справедливість — це правило, що мінімізує ризики руйнування кооперації. Ефективність такої кооперації залежить більше від технологічного рівня та людського капіталу, однак максимум добробуту для кожного окремо є наслідком стабілізації кооперативного ефекту. Однак в не ідеальних умовах, критерій справедливості допускає, що коаліція може усувати частину членів задля збереження максимального кооперативного ефекту.

Роулз виходить з того, що максимізація сумарної корисності не може справедливо допускати те, що це досягається ціною жертв будь-якого індивіда. В нашій моделі це виконується для умов ідеальної кооперації (жоден учасник нічим не жертвує примусово). Однак модель не допускає як справедливе те, що вирівнювання добробуту також відбувається примусово (бо це руйнує стійкість коаліції і опосередковано знижує загальний добробут).

Егалітаризм трактує рівність як первинну норму (ресурсів, можливостей). Однак представлена тут концепція не розглядає рівність як цінність саму по собі. Помірна нерівність — інструментальна, вона винагороджує ризик і стимулює ініціативу. Те, що егалітарист назвав би «несправедливістю», тут є функціонально необхідною умовою стабільності. Проте характер критерію справедливості як побічний ефект сприяє зниженню різкого розшарування, проте не потребує якихось компенсаційних заходів для усунення нерівності між соціальними групами. В практичному сенсі, якщо індивід не проявляє просоціальної ініціативи (не здатен підтримувати взаємодії згідно з оптимальним критерієм розподілу благ і не здатен робити відповідний пороговий внесок), створення для нього рівних умов просто не можливе: він функціонально або паразит, або експлуататор. Це не виключає того, що спільнота може культивувати деякий виправданий рівень милосердя і жалості; але до справедливості це не має стосунку, оскільки тут

йдеться про розподіл деякого не критичного надлишку, який можливий лише у високоефективних суспільствах. В цьому сенсі справедливість лише опосередковано може сприяти милосердю, але як така справедливість не є проявом милосердя.

В марксистській концепції справедливості базовим є принцип недопустимості панування одних людей над іншими через забезпечення рівності доступу до засобів виробництва і усунення форм експлуатації (але не власне «грабунок награвованого», урівнювання в доступі до благ, як часто думають). В представленій моделі цей принцип закладений початково, оскільки в ній не можлива гра з нульовою сумою. Однак, навіть якщо ми послабимо цю вимогу, то це не змінить принципово природу кооперативної динаміки: диктатори демотивують коаліцію і стратегічно програють на довгій дистанції; експлуатація не вигідна серед своїх в довгій грі. Проте там, де виникають умови для відчуження й вигоди від експлуатації «людської худоби», такі суспільні форми тривкі лише допоки немає альтернативних версій організації кооперативного ефекту. Але це не робить експлуатацію людини людиною справедливою. Це означає лише те, що в певних умовах забезпечити справедливість неможливо.

В етичних концепціях Д. Юма і А. Сміта мислителі розглядали критерій справедливості як правило, що виникає з потреби обміну, формування довіри та з розрахунку на взаємну вигоду. В представленій моделі ця інтуїція доповнюється формалізацією «гри», яка ілюструє, чому саме певна форма взаємності стабільна і показує, що навіть за високої асиметрії ризиків кооперація може утворити самопідтримувані групи взаємності, які вирівнюють ризики і доходи.

Застереження та проблемні моменти моделі

Представлена модель розгортається в умовах певної ідеалізованої кооперації, і важливо підкреслити, що в реальному житті ці умови майже ніколи не виконуються. Ключові обмеження, які переводять модель у зону абстракції, стосуються двох припущень.

1. Неможливість втрати базового капіталу через пасивність. У моделі передбачається, що гравець, діючи без ризику або утримуючись від пропозицій, не може втратити вже наявні ресурси. В реальності ж базовий капітал: може зменшуватися незалежно від дій агента, піддається дії зовнішніх факторів (хвороба, старіння, економічні кризи), не має «гарантованої стартової величини» у сильному сенсі. Тобто в реальному світі пасивність не є нейтральною стратегією, а часто сама по собі становить ризик.

2. Припущення про односторонній рух: лише приріст благ, а не втрата. У моделі передбачається, що результат кожного такту гри — це або приріст блага, або збереження статус-кво. Проте реальні ситуації включають:

втрати (матеріальні, соціальні, фізичні), незворотні події (смерть, інвалідність, руйнування життєвих планів), дефіцит неподільних ресурсів, які не можна розширити кооперативним способом. Саме в таких «негативних» умовах питання справедливості стають найбільш гострими.

3. Модель у ситуаціях від'ємного результату (розподіл тягарів).

У термінах моделі від'ємні результати — це негативні бонуси, тобто ситуації, коли гравці не отримують щось додаткове, а навпаки втрачають певне благо. Інший спосіб формалізації — розглядати благо не як приріст, а як запобігання втраті: неотримання блага → зниження базового капіталу → можлива елімінація агента. У реальному житті саме такі ситуації і визначають моральний сенс справедливості.

4. Чи змінюється критерій справедливості у випадку втрат?

Інтуїтивно здається, що розподіл втрат має принципово іншу логіку, ніж розподіл вигравів. Однак це враження пов'язане з нашою психологічною асиметрією щодо болю та вигоди, а не з конструкцією самої справедливості. Якщо припустити, що всі індивіди мають певний мінімальний запас життєстійкості, нижче якого вони не можуть існувати, то: абсолютні розбіжності благ не мають вирішального значення (діапазон життя й так обмежений біологічно), реальною стає лише загроза виходу з гри, тобто фатальний або критичний колапс капіталу. Це означає, що в обох випадках (виграші і втрати) ключовим є не рівність чи милосердя, а стійкість спільноти як коаліції здатних діяти індивідів.

5. Модельний приклад: кораблетроща.

Ситуація з кораблетрощею — це різко загострена версія дефіциту. Є обмежена кількість рятувальних засобів, і ті, хто виживе, не отримають допомоги ззовні. Така постановка дає змогу прозоро висвітлити структуру критеріїв справедливості.

Несумісність з рівністю (жеребкуванням): якщо застосовувати рівність як жереб, то в шлюпку потраплять немовлята, хворі, кволі, люди з вузькими спеціальностями, які не підвищують шанси групи на виживання. Результат: спільнота виживає тимчасово, але не має шансів на довгострокову репродукцію.

Несумісність з милосердям: милосердя як критерій віддає перевагу найслабшим. У критичних умовах це еквівалентно самознищенню групи.

Оптимальність критеріїв, близьких до коаліційної логіки моделі

Той самий принцип, що робить коаліцію стабільною в грі з бонусами, працює й тут. У виживання мають пріоритет: ті, хто має максимальний прогнозований запас життєстійкості, ті, хто володіє пластичними здібностями, ті, хто може забезпечити відтворення і функціонування спільноти. Це не гарантує успіху, але максимізує шанс виживання спільноти як системи.

6. Висновок: критерій справедливості не змінюється, але змінюється роль стартового капіталу

Критерій справедливості, введений у моделі (стабільність коаліції, симетрія взаємних очікувань, вироблення стійких груп взаємності), не знає принципової зміни у випадку втрат. Однак в умовах критичного дефіциту: стартова величина капіталу стає визначальною, а капітал включає не лише фізичні ресурси, а й соціальний капітал спільноти, її згуртованість, навички та здатність до колективної дії.

Тобто в кризових ситуаціях справедливість не стає «іншою» — просто питання входу в гру та здатності зберігати себе як агента стають первинними.

Але оскільки справедливість як така опосередковано веде (як показно) до формування коаліцій, які здатні накопичувати базовий капітал і нарощувати кооперативний ефект (вища частота актуалізацій з вищим бонусом), то це утворює контур позитивного зворотного зв'язку, коли розширення таких спільнот закріплює сам критерій справедливості.

Еволюція критерію справедливості в цій моделі формує уніфіковане правило як самопідтримувану норму. Після кількох циклів стає стратегічно вигідним прийняти правило розподілу R , яке: забезпечує високий рівень довіри; мінімізує ризики пропонента; не провокує ультиматумів і шантажу; гарантує всім членам групи участь у виграшах.

Таке правило стає фактичним критерієм справедливості:

$$R^* = \arg \max_{i \in G} \min E_i(R).$$

Тобто справедливість тут — це правило, яке максимізує мінімальний очікуваний виграш у групі, забезпечуючи її довгострокову стійкість.

Розширення коаліцій. Коли група демонструє стабільність і передбачуваний приріст капіталу, вона стає привабливою для інших гравців.

Нехай G_1, G_2 — дві стабільні групи з однаковим правилом R^* ; тоді їх об'єднання збільшує загальний пороговий потенціал, імовірність великих актуалізацій, швидкість накопичення капіталу. Таким чином, справедливість, що виникла як локальна норма групи, еволюційно масштабується.

Описаний механізм дозволяє дати неklasичну, але концептуально сильну інтерпретацію: справедливість не є абстрактною моральною цінністю. Вона — форма організації кооперації, яка мінімізує ризики та стабілізує очікування. Справедливість виникає не у свідомості, а в практиці повторних дій. Це не умоглядна угода, а наслідок еволюційного відбору стратегій.

Коаліції зі справедливим розподілом мають еволюційну перевагу. Їхня пропускну здатність щодо актуалізацій вища, тому вони накопичують ресурси швидше.

Справедливість самопідсилюється. Коли норма стає джерелом стабільного зростання, її прийняття стає стратегічно раціональним.

Насамкінець справедливість стає інституційною основою масштабних спільнот. Правило, що виникло як локальна взаємність, еволюційно постає нормою великої кооперації.

Висновки

Запропонована модель дозволяє показати, що інтуїція справедливості не є ані первинною моральною аксіомою, ані наслідком абстрактної рівності, а постає як функціональний механізм стабілізації соціальної взаємодії. На рівні індивідуальних рішень гравці діють у межах раціонального егоїзму, керуючись прагненням максимізувати власний прибуток і мінімізувати ризик. Проте саме ці локальні інтереси генерують умови, за яких кооперація стає вигіднішою за ізольовані чи конфліктні стратегії. Утворення стабільних коаліцій і прийняття уніфікованого правила розподілу бонусу є не моральним імперативом, а емерджентним ефектом адаптивної поведінки.

Евристика моделі полягає в тому, що вона демонструє: критерій справедливості формується як правило підтримки кооперативної рівноваги, а не як попереднє нормативне припущення. Саме такий критерій — помірно асиметричний, заснований на взаємності ролей і гарантіях участі — сприяє довгостроковій стійкості системи, оскільки забезпечує зростання базового капіталу всіх її членів та збільшення частоти успішних кооперативних ефектів. Цей позитивний зворотний зв'язок пояснює, чому справедливість є самопідсилювальним явищем: спільноти, що приймають її як принцип розподілу, стають продуктивнішими й привабливішими для нових учасників, що, у свою чергу, закріплює сам критерій.

Модель також виявляє важливе обмеження: справедливість як критерій розподілу зберігається навіть тоді, коли йдеться не про приріст благ, а про розподіл втрат або тягарів. Умови дефіциту, загрози чи критичних рішень (на зразок «рятівної шлюпки») не змінюють функціональної структури справедливості, хоча загострюють залежність результату від стартового капіталу — матеріального й соціального. Це підкреслює, що справедливість не тотожна рівності й не може бути зведена до милосердя: її роль полягає у підтриманні життєздатності спільноти в умовах ризику, а не у випадковому або сентиментальному розподілі ресурсів.

Таким чином, модель показує, що справедливість набуває сенсу як інструментова чеснота, що забезпечує стабільність соціальної взаємодії, максимізацію довгострокового блага та підтримку кооперативних структур. Її раціональність не потребує метафізичних припущень: достатньо розуміння стратегічних умов, у яких актори, переслідуючи власні інтереси, зрештою

обирають правило, яке стає спільним та здатним відтворюватися. У цьому полягає як новизна запропонованої моделі, так і її пояснювальна сила.

Список посилань / References

1. Aristotle. (2009). *Nicomachean ethics* (W. D. Ross, Trans.; Revised ed.). Oxford University Press.
2. Habermas, J. (1990). *Moral consciousness and communicative action* (C. Lenhardt & S. W. Nicholsen, Trans.). MIT Press.
3. Sen, A. (2009). *The idea of justice*. Harvard University Press.
4. Fraser, N., & Honneth, A. (2003). *Redistribution or recognition? A political – philosophical exchange* (J. Golb, J. Ingram, & C. Wilke, Trans.). Verso.
5. Rawls, J. (1971). *A theory of justice*. Harvard University Press.
6. An, X., Dong, Y., Wang, X., & Zhang, B. (2023). Cooperation and coordination in threshold public goods games with asymmetric players. *Games*, 14(6), Article 76. <https://doi.org/10.3390/g14060076>
7. Zhang, Z., Yang, C., Han, W., & Xia, C. (2025). Co-evolutionary dynamics of threshold public goods games with collective-risk environment feedback. *Physica A: Statistical Mechanics and its Applications*, 658, 130275. <https://doi.org/10.1016/j.physa.2024.130275>
8. Bach, L. A., Helvik, T., & Christiansen, F. B. (2006). The evolution of n-player cooperation – threshold games and ESS bifurcations. *Journal of Theoretical Biology*, 238(2), 426 – 434. <https://doi.org/10.1016/j.jtbi.2005.06.007>
9. Chen, D., & Chen, B. (2023). Evolutionary game analysis on decision-making behaviors of participants in mega projects. *Humanities and Social Sciences Communications*, 10, Article 921. <https://doi.org/10.1057/s41599-023-02436-w>
10. de Clippel, G., & Rozen, K. (2022). Fairness through the lens of cooperative game theory: An experimental approach. *American Economic Journal: Microeconomics*, 14(3), 810 – 836. <https://doi.org/10.1257/mic.20200015>
11. Balkanski, E., & Singer, Y. (2015). Mechanisms for fair attribution. In T. Roughgarden, M. Feldman, & M. Schwarz (Eds.), *Proceedings of the Sixteenth ACM Conference on Economics and Computation* (pp. 529 – 546). ACM. <https://doi.org/10.1145/2764468.2764505>

Vadym Derkach

PhD in Philosophy, Associate Professor

Kyiv National Economic University named after Vadym Hetman (Kyiv, Ukraine)

<https://orcid.org/0000-0002-0123-3637>

e-mail: vadym_derkach@kneu.edu.ua

THE CONCEPT OF JUSTICE: A GAME-THEORETIC MODEL

Abstract

The article presents a game-theoretic model of justice that interprets it not as a moral axiom but as an emergent mechanism for stabilizing cooperation. It introduces a threshold-contribution game with asymmetric proposer risk and probabilistic bonus activation, within which players adaptively form coalitions of reciprocity. The analysis shows that a moderately asymmetric allocation rule — grounded in role reciprocity and participation guarantees — emerges as a stable criterion of justice. The model explains the genesis of justice as an instrumental virtue that enhances the long-term viability of communities and preserves its significance even in situations involving the distribution of losses.

Keywords: justice; game theory; cooperative models; evolutionarily stable strategy (ess); social stability; critical scarcity and loss allocation**

Стаття надійшла до редакції 17.06.25

© Деркач В. Л., 2025