Computers in Human Behavior: Artificial Humans 6 (2025) 100236

Contents lists available at ScienceDirect

Computers in Human Behavior: Artificial Humans

journal homepage: www.journals.elsevier.com/computers-in-human-behavior-artificial-humans

Research paper

Quantitative fairness—A framework for the design of equitable cybernetic societies

Kevin Riehl *, Anastasios Kouvelas, Michail A. Makridis

Traffic Engineering Group, Institute for Transport Planning and Systems, ETH Zürich, Stefano-Franscini-Platz 5, 8093, Zürich, Switzerland

ARTICLE INFO

Keywords:
Resource allocation
Equitable societies
Distributive fairness
Procedural fairness
Algorithmic fairness

ABSTRACT

Advancements in computer science, artificial intelligence, and control systems have catalyzed the emergence of cybernetic societies, where algorithms play a pivotal role in decision-making processes shaping nearly every aspect of human life. Automated decision-making for resource allocation has expanded into industry, government processes, critical infrastructures, and even determines the very fabric of social interactions and communication. While these systems promise greater efficiency and reduced corruption, misspecified cybernetic mechanisms harbor the threat for reinforcing inequities, discrimination, and even dystopian or totalitarian structures. Fairness thus becomes a crucial component in the design of cybernetic systems, to promote cooperation between selfish individuals, to achieve better outcomes at the system level, to confront public resistance, to gain trust and acceptance for rules and institutions, to perforate self-reinforcing cycles of poverty through social mobility, to incentivize motivation, contribution and satisfaction of people through inclusion, to increase social-cohesion in groups, and ultimately to improve life quality. Quantitative descriptions of fairness are crucial to reflect equity into algorithms, but only few works in the fairness literature offer such measures; the existing quantitative measures in the literature are either too application-specific, suffer from undesirable characteristics, or are not ideology-agnostic. This study proposes a quantitative, transactional, and distributive fairness framework based on an interdisciplinary foundation that supports the systematic design of socially-feasible decision-making systems. Moreover, it emphasizes the importance of fairness and transparency when designing algorithms for equitable, cybernetic societies, and establishes a connection between fairness literature and resource allocating systems.

1. Introduction

Technology is increasingly employed for the automation of processes that were previously performed by humans. This shift has expanded from early applications in agriculture, manufacturing, and mechanical processes to now encompass planning, decision-making, and control processes (Xu et al., 2018).

Especially the advancements in computer science, artificial intelligence, and control systems of the recent have catalyzed the emergence of cybernetic societies, where algorithms play a significant role in decision-making processes that affect the daily life of humans in almost every aspect. This algorithmic decision-making is becoming more prevalent across industries, from finance and healthcare to media, retail and customer service, in the life-reality of citizens of smart and mega cities, and it also involves the design and operations of energy and transportation networks. Algorithms even influence the very fabric of our social interactions, personal relationships, and communication using digital media and social networks. What is more, automated processes are increasingly employed even in law-enforcement, budget allocation, and planning at the governmental level (Ashby, 1956; Friedman, 2019; Larsson, 2022).

The automation offers numerous potential benefits, such as increased efficiency and objectivity of decisions, improved enforcement of legislation, reduced corruption, acceleration of bureaucratic processes, standardization of processes, and the dehumanization of repetitive tasks that negatively affect health of (human) workers (Abbott et al., 2024). At the same time, misspecified or purposefully misused technologies harbor the threat to create societal inequities through systematic discrimination, and enable even more corrupt systems through mass surveillance technology and extreme restriction of individual freedom (Hossain et al., 2020).

Together with efficiency, fairness plays a crucial role when designing and implementing philanthropic, cybernetic systems, that serve and benefit humans (Friedman, 2019).

* Corresponding author.
E-mail address: kriehl@ethz.ch (K. Riehl).

https://doi.org/10.1016/j.chbah.2025.100236
Received 21 September 2025; Received in revised form 11 November 2025; Accepted 15 November 2025
Available online 19 November 2025

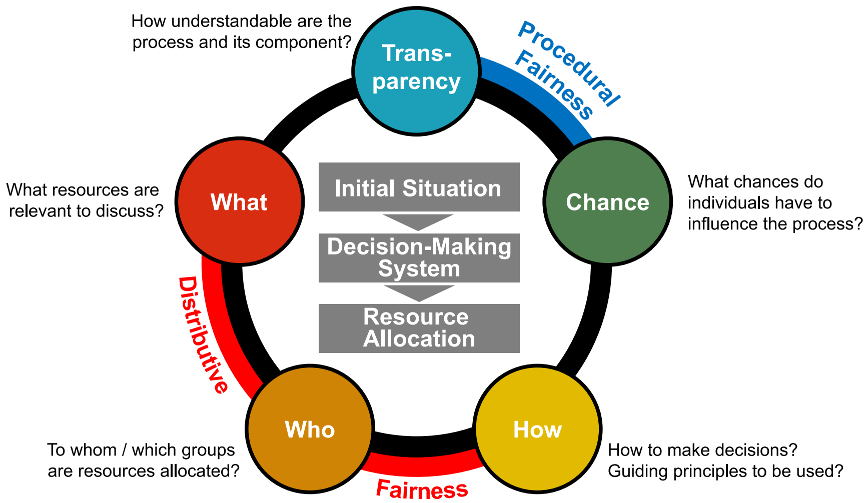


Fig. 1. The Fairness of Resource-Allocating Decision-Making Systems.
Decision-Making is depicted as a resource allocating process, that takes an initial situation as an input, makes a decision, and provides a resource allocation as an output. The fairness discussion of decision-making therefore covers procedural and distributive fairness. For procedural fairness, aspects such as (i) the transparency of the process, and (ii) what chances individuals have to influence the process are important questions to answer. For distributive fairness, aspects such as (i) what resources are fairness-relevant and distributed to (ii) which groups and (iii) how decisions are made (based on which guiding principles) are important questions to answer.

- Only systems that promote equitable societies can guarantee that these cybernetic systems serve people, and not people serving these systems.
- The implementation of cybernetic systems in practice often fails due to public resistance and equity concerns (Gu et al., 2018).
- Cybernetic systems often support self-coordination of selfish, rational individuals, and to align egoistically-optimal with socially-optimal outcomes; doing so, fairness is the foundation to achieve any form of cooperation of individuals in large populations (Gurney et al., 2021).

Algorithm-driven, automated processes in some form distribute resources to people. As these processes are automated by technology, a discussion of the fairness of cybernetic systems must be a discussion of procedural and distributive fairness (Friedman, 2019; Pereira et al., 2017) (Fig. 1). For example, credit scoring algorithms determine whether a certain person has access to a loan, transportation demand management and congestion pricing determine whether a certain person has access to the road infrastructure of a city, automatic screening algorithms in recruiting determine whether a certain person is invited to a job interview, and user-engagement-maximizing algorithms determine which information is provided to the consumer of a social media platform.

While most discussions in the fairness literature are instrumental to discuss which resources can be considered as fairness-relevant (Nussbaum, 2011; Rawls, 1971; Sen, 2008; Walzer, 1983), which groups shall be compared and what fairness can be considered as conceptually (Goppel et al., 2016), only few answers can be found on how to quantitatively assess fairness in a specific situation based on data, which is crucial for the design and integration into algorithms, that in some form act within their environment using data (Cormen et al., 2022). The quantitative measures proposed in the literature are either too specific to a particular application, suffer from some undesirable characteristics, or are limited to specific ideologies (Jain et al., 1984). What is typically coined *Algorithmic Fairness* for instance, often employs numeric measures to assess biases, and to establish equality between privileged and underprivileged groups (Hellman, 2025); this perspective however is limited to horizontal discussions of fairness, neglecting fairness principles different from equality, and mostly focuses on classification and prediction problems from machine learning rather than resource allocation problems and distributive justice (Beigang, 2022; Kuppler et al., 2022; Scantamburlo et al., 2025).

This work proposes a holistic, integrative, transactional, distributive, procedural, and quantitative fairness framework for dianemetic and diorthotic resource allocation problems based on a rich, interdisciplinary foundation – combining statistical and dispersion metrics with social Welfare functions. This framework distinguishes itself from previous approaches, by its general real-world applicability to cybernetic resource allocation problems, integrative (ideology-agnostic) approach, and its multi-perspective view including teleological and deontological considerations. As such, it enables the quantification of fairness rather than to impose normative, cardinal, comparative or ordinal assessments.

It is the mission of this article (i) to highlight the importance of a quantitative discussion on the fairness of cybernetic systems, (ii) to create a connection between domain-specific literature and the fairness literature, (iii) to enable algorithm design and evaluation based on a holistic, transactional, ideology-agnostic, quantitative fairness framework for a distributive discussion, and (iv) to bridge the gap between algorithmic fairness and distributive justice.

The remainder of this work is organized as follows. Section 2 motivates the societal relevance of fairness, lays the theoretical foundations on fairness (distributive, procedural, and algorithmic fairness), and reviews related works. Section 3 proposes the quantitative fairness framework for distributive justice. Section 4 demonstrates the application of the quantitative fairness framework at the example of a dianemetic and diorthotic case study. Section 5 discusses the properties of the proposed fairness framework, and elaborates further on theoretical contributions. Section 6 concludes this work.

2. Theoretical foundation

Exploring the concept of fairness presents significant challenges, as it is an abstract, philosophical notion deeply rooted in specific social and cultural contexts. Despite over two thousand years of philosophical discourse across human civilizations, a universal consensus on the definition and application of fairness remains elusive. Fairness is widely regarded as a cornerstone of human coexistence and is extensively examined as a multidisciplinary concept across various fields, including philosophy, ethics, biology, sociology, political science, economics,

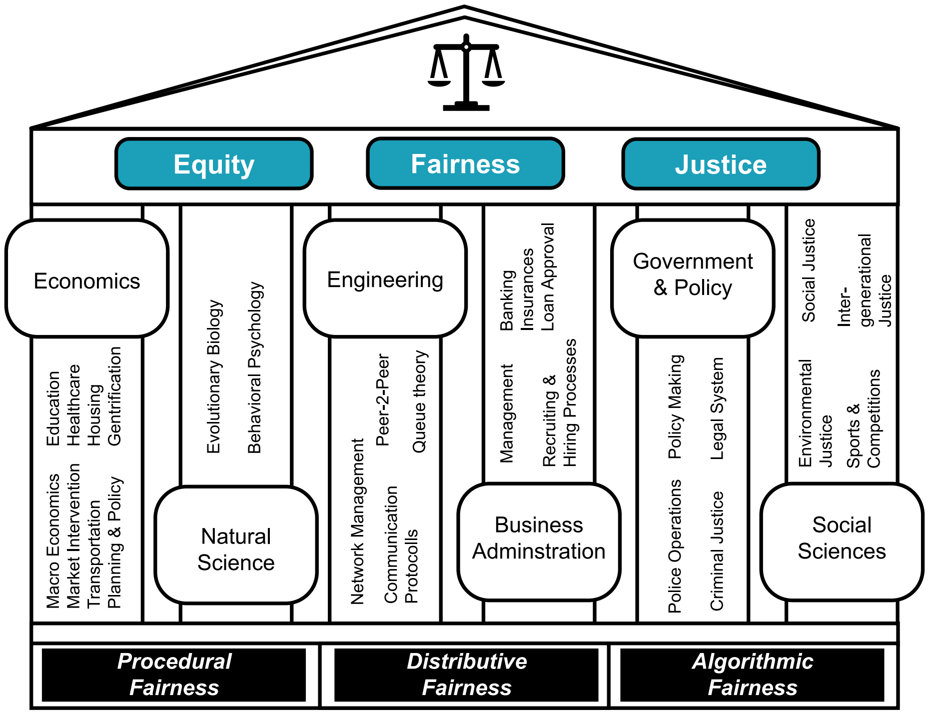


Fig. 2. The Societal Relevance Of Fairness - Interdisciplinary Perspective.
An interdisciplinary review across six societally relevant domains – including economics, natural science, engineering, business administration, government & policy, and social sciences – reveals the pivotal role of fairness as a control variable for the functioning and integrity of complex, socio-technical, cybernetic systems.

and religion studies – which not only motivates but also requires an integrative, interdisciplinary foundation of fairness theories. The terms fairness, equity and justice are used interchangeably in the literature, where equity is often found in the literature of economics and justice in the literature of politics and law (Goppel et al., 2016).

In this section, we start with a broad, interdisciplinary review of fairness to motivate why fairness matters, to emphasize its societal relevance. Then we discourse two important concepts for the framework – distributive and procedural fairness – to lay the theoretical foundations for our framework. Consecutively, we elaborate on algorithmic fairness and related aspects of cybernetic societies (including comparative and procedural concepts of algorithmic fairness, data imbalance, and process obscureness). Finally, we review related works to highlight limitations of previous fairness frameworks, to demarcate, and to motivate the proposed framework.

2.1. Societal relevance of fairness

The discussions of fairness across different, socially relevant domains have in common, that fairness - besides efficiency - is a crucial component when designing systems. Often, a conflict between the individual and the collective perspective causes a egoistic, selfish behavior that leads to an user equilibrium (considered equitable) that is different from a system optimum (considered efficient) (Riehl et al., 2024a, 2024b).

Based on an interdisciplinary review of fairness-relevant topics from six domains (see Appendix A and Fig. 2), we identify the societal relevance of fairness. Fairness constitutes the social and systemic glue that enables cooperation, legitimacy, and stability in any society that allocates limited resources among autonomous actors. Across biological, economic, institutional, and technological systems, fairness functions as a feedback mechanism that regulates behavior and maintains equilibrium between individual incentives and collective welfare. In biological evolution, fairness emerged as a mechanism for reciprocal altruism, sustaining cooperation and group survival. In markets, infrastructures, and networks, fairness governs how benefits and burdens are distributed, balancing efficiency with inclusivity. Within organizations and public institutions, it sustains trust, compliance, and motivation, preventing disintegration through perceived injustice. In the social and ecological context, fairness ensures that prosperity does not erode cohesion, that rules retain legitimacy, and that intergenerational and planetary responsibilities are respected. Fairness not only promotes cooperation among self-interested individuals to achieve better system-level outcomes, but also helps confront public resistance against policy, fosters trust and acceptance of rules and institutions, breaks self-reinforcing cycles of poverty through social mobility, strengthens motivation and inclusion, enhances group performance, and ultimately improves life quality.

As human societies become increasingly cybernetic – governed by data flows, algorithms, and feedback loops – fairness becomes not merely a moral aspiration but an operational requirement: systems perceived as unfair lose trust, trigger adversarial behaviors, and risk destabilization. Quantitative fairness, therefore, represents not only a normative principle but also a control variable for the sustainability, resilience, and ethical integrity of complex socio-technical systems.

2.2. Distributive fairness

Distributive fairness (also distributive justice) tries to find fair ways of allocating resources (goods, opportunities, etc.) across a population as a result of transactions. This can include how benefits but also burdens can be distributed across individuals, reflecting or neglecting individual properties such as wealth, social status, contributions, needs, etc. Hume (1740) requires three conditions to discuss the question of fairness: scarcity of resources, a conflict of interest, and a relative balance of power between the negotiating parties. Important aspects which need clarification in the discussion of distributive fairness include: (i) which resource is distributed (ii) across whom and (iii) how (based on which guiding principle) (Cohen, 1987).

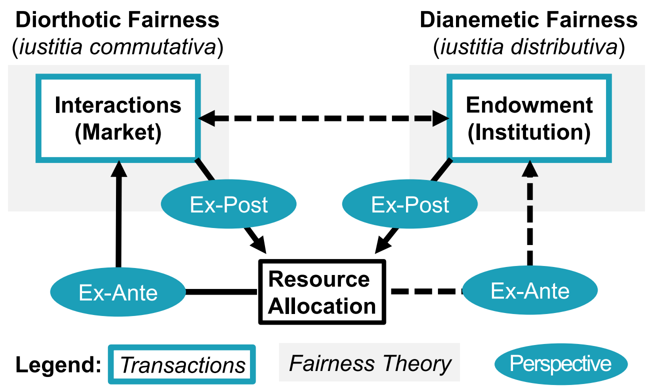


Fig. 3. A Transactional View on Distributive Fairness.
Aristotle distinguishes dianemetic and diorthotic, distributive fairness. Dianemetic fairness is concerned with the fair distribution of resources from top to bottom (endowment) from usually one central decision maker (institution, government) to the population. Diorthotic fairness is concerned with the fair distribution of resources in a decentralized way as a result of transactions between individuals (markets).

Aristotle's Nicomachean Ethics (Wolf, 2002) presents a theory of transactional fairness that distinguishes two types of justice: dianemetic and diorthotic fairness (Fig. 3). Transactions can occur between various parties under different circumstances, including market exchanges, governmental resource allocation, participation in political decision-making, or legal proceedings. Dianemetic fairness focuses on how resources are allocated from an authority to a population, typically involving governmental institutions in cases such as subsidies or societal redistribution. The initial distribution of resources can also be considered a dianemetic endowment. Diorthotic fairness, on the other hand, deals with transactions between individuals or groups within a population, often in the context of market exchanges. Nozick (1974) bridges the gap between these two concepts by arguing that fair, diorthotic transactions can only occur if the initial, dianemetic allocation was already fair.

Goppel et al. (2016) distinguish two types of perspectives on fairness: ex-post, and ex-ante fairness. These perspectives describe what matters for the fairness of a resource allocation; the first one (ex-post) focuses only on the output of a transaction, the second (ex-ante) one focuses only on the input of a transaction. Teleological fairness focuses on outcomes or consequences, while deontological fairness focuses on the adherence to rules, duties, and principles regardless of the outcome. Moreover, a differentiation between horizontal equity (fairness within groups of similar individuals) and vertical equity (fairness between different groups) is possible.

The literature of fairness provides various ideologies and moral guiding principles, that can be summarized by following six principles: equality, proportion, greater-good, difference, equality-of-opportunity, and sufficiency. The equality principle considers an allocation fair, if it ensures equal allocation for all. This principle is the foundation for Egalitarian ideology (Scheffler, 2017). The proportion principle considers an allocation fair, if allocated resources stand in proportion to the status of individuals, where status could refer to social status in a society, economic power, contributions to the system, or needs. This principle is the foundation for Aristocratic (also Aristotelian) ideology (Goppel et al., 2016; Wolf, 2002). The greater-good principle considers an allocation fair, if it maximizes the welfare of the many (society), even if this means the suffering of few (individuals). This principle is the foundation for Utilitarian ideology (Mill, 2016). The difference principle considers an allocation fair, if it achieves the best possible outcome for the least-advantaged (Rawlsian ideology) or the average (Harsanyian ideology) (Harsanyi, 1975; Rawls, 1971). The equality-of-opportunity principle considers an allocation fair, if the initial changes/opportunities of each participating individual to a resource-allocating process were equal upfront. This principle is the foundation for Luck-Egalitarian ideology (Dworkin, 2000). The sufficiency principle considers an allocation fair, if it is guaranteed that each individual receives at least a sufficient minimum. This principle is the foundation for Sufficientarian ideology (Shields, 2012).

The diversity and complexity of fairness motivates to design a holistic, distributive fairness framework that is integrative, and ideology-agnostic, for general applicability to a wide range of problems.

2.3. Procedural fairness

The concept of fairness has evolved throughout history. In ancient philosophy, it was viewed as a personal virtue and a key attribute of divine beings. During the Renaissance, this divine association shifted towards a universal, natural law. The Enlightenment era saw philosophical discussions move away from religious contexts, emphasizing rational thought instead. Contemporary debates on fairness primarily center on political and economic aspects, expanding beyond individual conduct to encompass institutional practices and societal processes (Goppel et al., 2016).

Procedural fairness discusses fairness in processes. These processes usually include processes that resolve disputes and distribute resources, particularly in administrative and legal contexts, but are not limited to those. Other examples could include promotions to higher positions in organizations, hiring for new positions, admission to educational facilities, etc. Procedural fairness intersects with distributive fairness (when distributing resources), and retributive fairness (when punishing mistakes). Two important aspects to procedural fairness are (i) the opportunities of users affecting the process outcome, and (ii) the transparency of the processes and underlying decision-making (Lind & Tyler, 2013). Similarly to distributive fairness, ex-ante and ex-post perspectives on procedural fairness apply.

Three important schools of thought exist in the equality-of-opportunities of processes: formal equality, substantive equality, and pure equality. Formal equality (also meritocracy) often refers to meritocratic provision of opportunities based on performance only, neglecting other discriminatory aspects such as gender, race, or age. Substantive equality refers to equal chances for different groups of people. For example, substantive equality distributes opportunities according to gender and race, rather than merits, to ensure the representation of certain groups. While formal and substantive equality discriminate individuals either based on performances or personal aspects (e.g. gender, race), the pure equality (also known as equality before the law) describes the equal distribution of opportunities to all individuals independent of any personal features (Acemoglu & Wolitzky, 2021; Barnard & Hepple, 2000).

The relevance of procedural fairness, especially in the advent of ever-growing automation of socially relevant resource distribution processes motivates the consideration of this perspective on fairness into the framework.

2.4. Algorithmic fairness

Cybernetic systems use algorithms to control systems, and not only does that involve computer-driven, but also physical-machine-driven, and human-driven processes, such as governments that follow bureaucratic and legal processes (Carver & Scheier, 2012). Algorithms can be used to formalize and automate decision-making processes. An algorithm is a finite sequence of well-defined steps to process inputs and produce outputs (Sedgewick & Wayne, 2011). There are two sources that can be used to define algorithms: First, algorithms can be formalized by humans, that specify inputs, outputs, and process-steps manually to mimic their own decision-making. Second, the advances in computer science, especially in machine learning, allowed computers to derive algorithms directly from data or collected experience (Pessach

& Shmueli, 2023). The fairness of algorithms massively depends on the choice of inputs, outputs, process-steps, and the source of the algorithm. Algorithmic biases can lead to systematic and repeatable errors and discrimination of certain individuals or groups, and cause algorithmic unfairness, which in turn can lead to systematic, self-reinforcing cycles of inequality in a society, which is outlined at the two following examples:

- **Racial Profiling:** Assuming the same criminal rate for white people and people of color – a significantly higher inspection of people of color will lead to more findings of criminal activities performed by people of color. This leads to an increase in findings, and yet to an increased attention of the police to inspect people of color, during police operations (Hurwitz & Peffley, 2005).
- **Hirshleifer-effect and credit scoring:** The Hirshleifer-effect (Hirshleifer, 1978), when applied to algorithmic credit scoring, reveals that more precise group-based information can paradoxically worsen outcomes for society as a whole, even if it seems beneficial for individual firms like banks. Although using better data and algorithms may allow banks to more accurately assess risk, leading them to discriminate against certain groups, this can prevent broader beneficial risk-sharing trades that would be possible under less precise, noisier information. In effect, the efficiency gains for banks and some customers are offset by systemic harm and the exclusion of vulnerable groups, and a lower economic, societal welfare. Better data and sharper algorithms may improve immediate business metrics but often reduce long-term societal welfare by increasing exclusion, amplifying pre-existing biases, and undermining collective economic opportunity (Wang et al., 2024).

According to the *Stanford Encyclopedia of Philosophy*, the term *algorithmic fairness* is used to assess whether machine learning algorithms operate fairly (Hellman, 2025). While metrics of algorithmic success typically center on predictive accuracy or classification performance, fairness demands a more intricate analysis. Notably, outcome-based fairness – measuring group-level disparities in algorithmic outputs – remains insufficient unless accompanied by robust procedural safeguards. Fairness must extend beyond results to encompass the integrity and transparency of the processes that generate them, since these act as preconditions for the legitimacy of algorithmic systems in both technical and social domains. Accordingly, any evaluation of algorithmic fairness should critically consider not only how different groups are treated in aggregate, but also whether the procedures underlying these treatments are themselves fair and explainable. As a consequence, two major components of algorithms are differentiated (Hellman, 2025): (i) algorithmic fairness as a comparative, horizontal aspect, and (ii) as a procedural aspect, related to data imbalance and process obscureness:

1. **Comparative, Horizontal Aspect:** This concept emphasizes how algorithmic decisions produce unequal outcomes across social groups. Fairness is violated when measurable disparities between groups emerge that cannot be justified by relevant data features but instead reflect systematic favoritism or disadvantage encoded in the model's operation. Examples include large language models displaying political leaning (Zhang et al., 2023), chatbots reproducing racist content after online interaction (Neff & Nagy, 2016), facial recognition systems misclassifying darker-skinned individuals as gorillas (Hern, 2018), automated hiring tools discriminating against women (Dastin, 2018), or predictive policing tools disproportionately flagging black individuals (Lum & Isaac, 2016; Richardson et al., 2019).
2. **Procedural Aspect:**
    - **Data Imbalance:** Algorithms with biases often stem from imbalanced data, where certain groups may be underrepresented or misrepresented, leading to unfair outcomes. Addressing these issues involves various strategies, including data preprocessing to balance datasets, in-process adjustments to algorithms during training, and post-process evaluations to ensure equitable outcomes (Pessach & Shmueli, 2023). The challenge of imbalanced data is particularly pronounced in fields like healthcare, where disparities in data collection can exacerbate existing inequalities in treatment and diagnosis. For example, the design and development of medical treatments have historically exhibited a gender bias, often optimizing for male physiology at the expense of female patients (Agyemang et al., 2023).
    - **Process Obscureness:** Increasingly, automated decision-making processes employ deep-learning with neural network models for machine learning, when developing automated decision-making algorithms. These models have an ever-growing number of parameters and thus complexity, which allows them to achieve super-human performance (for example in image recognition and classification) (LeCun et al., 2015; Shinde & Shah, 2018). At the same time, these models are highly problematic, as due to their complexity it is challenging to understand how and based on which specific properties and features from the data, decisions are made (Dwivedi et al., 2023; Holzinger et al., 2020). Transparency and traceability of decision-making processes are a crucial component to procedural fairness. As a consequence, the application of even better performing models is impeded in industry- and governmental applications due to explainability concerns, the often unclear origin of training data and process, the inability to describe guarantees in terms of safety, and equity issues (Xu et al., 2019).

2.5. Related works

In the following we present related works, highlight their limitations to motivate the need for the proposed fairness framework presented in this study.

Casacuberta et al. (2023) propose the concept of welfare-augmented fairness, leveraging scoring profiles, assignment profiles, utility functions, and Social Welfare functions. This concept explicitly differentiates between ex-ante and ex-post fairness, and highlight differences between individual fairness (as defined by Dwork et al., 2012) and perceptions of justice. **Limitations:** While this concept acknowledges the necessity for multiple narratives, their work is limited to Rawlsian, Utilitarian, and Egalitarian ideologies. Moreover, this concept applies to diorthotic resource allocation problems only, neglecting dianemetic problems. Besides, the concept is motivated with a rather over-simplistic fairness definition of the computer science community (individual fairness according to Dwork et al., 2012). Dowrk's individual fairness is based on a Lipschitz-mapping definition, advocating the importance of equal outcome-probability distributions for all individuals in random sampling resource allocation, rather than to judge the shape of the probability distribution itself.

Heidari et al. (2018) present an analytical formulation of Rawlsian fairness for the incorporation of convex loss minimization problems, which is the first to enable a computationally-feasible mechanism for bounding constraints of individual-level inequality. Doing so, the concept establishes a link between fairness in machine learning and cardinal social welfare economics, and trades off prediction accuracy (efficiency), group discrimination, and individual fairness. **Limitations:** While this concept acknowledges the potential of cardinal social welfare economics in the context of algorithmic fairness, it is rather normative (value-based assessment) than quantitative (objective assessment). Furthermore, it focuses only on Rawlsian ideology, neglecting the many aforementioned notions. Finally, while interesting to advance theoretical, mathematical applications, the concept has limited real-world applicability to the complexities of resource allocation problems.

Kuppler et al. (2022) propose the concept of error fairness, that relates to the normative principle of prediction errors to not differ systematically across individuals or groups. Doing so, this concept combines algorithmic fairness (related to prediction tasks) with resource allocating problems (distributive fairness) in a dianemetic context. **Limitations:** While concept acknowledges the need to bridge the gap between algorithmic fairness and distributive fairness, it is limited to dianemetic problems only, and focuses on certain ideologies (Egalitarianism, Prioritarianism, Utilitarianism and Aristotelianism) neglecting other notions of fairness.

Scantamburlo et al. (2025) highlight that algorithmic fairness is not enough to achieve fair decision-making. Beyond the scope of algorithmic (prediction) fairness, they explicitly differentiate decision fairness (causing consequences to human lifes through resource allocation). They propose a conceptual framework that combines "prediction-modeler" with "decision-makers" and offer a new perspective shifting the focus from an abstract concept of algorithmic fairness to the concrete context-dependent nature of algorithmic decision-making. offer a new perspective shifting the focus from an abstract concept of algorithmic fairness to the concrete context-dependent nature of algorithmic decision-making. **Limitations:** While highlighting the need for more than just algorithmic fairness discussions, their work lacks practical implications and specific, applicable, quantitative measures due to their rather theoretical focus.

Speicher et al. (2018) introduce the concept of algorithmic unfairness through inequality indices. The concept highlights the need for satisfactory, useful, and quantitative measures of fairness, to inform ordinal, comparative assessments, such as how to decide which of two given algorithms is more fair (unfair). **Limitations:** While highlighting the need for more quantitative discussions, the concept is limited to the Egalitarian ideology of fairness, and focuses mostly on algorithmic fairness discussed at prediction problems, rather than resource allocating problems.

To summarize, previous concepts in the context of fairness either (i) focused purely on algorithmic fairness rather than distributive, decision-making (resource allocating) fairness, or (ii) neglected the complexity of multiple fairness ideologies, (iii) discussed either dianemetic or diorthotic problems, (iv) were rather theoretical and normative in character, or (v) simply lacked a real-world practicability. These limitations of previous works highlight the need for (i) a holistic fairness framework that combines algorithmic with distributive fairness in the context of cybernetic societies, which is (ii) integrative (multiple fairness ideologies) and (iii) transactional (both dianemetic and diorthotic) in nature, and enables (iv) quantitative rather than normative discussions for real-world applicability.

3. Quantitative fairness framework

The quantitative fairness framework aims to provide a holistic toolset to quantitatively assess the fairness of resource allocations in the context of cybernetic societies, and hence make equity considerations accessible to resource allocating, algorithmic design in general. This framework distinguishes itself from previous approaches, by its general applicability, integrative (ideology-agnostic) and holistic (algorithmic and distributive) approach, and its multi-perspective view (teleological and deontological considerations). Rather than advocating a specific ideology, the framework enables an integrative analysis that combines different guiding principles and allows for systematic comparison.

The proposed quantitative fairness framework consists of a three step approach, and combines statistic dispersion metrics of distributions with econometric social welfare functions, as outlined in Fig. 4. Each of the three steps, and its foundations in the literature are outlined in the following.

3.1. Step 1: Modeling as resource allocating process

The first steps consists of modeling the situation to be assessed. This includes identifying the fairness-relevant resource, describing the situation as a process with inputs and outputs, and identifying relevant giving and receiving entities.

What? – Fairness-Relevant Resource: To identify the fairness-relevant resource of interest that needs to be distributed, three important concepts are useful to determine the fairness-relevance of a specific resource: primary goods, distributive spheres of justice, and capabilities. Rawls (1971) contends that primary goods, which are universally valuable and impact the well-being of all individuals regardless of their personal preferences, are crucial considerations in determining fairness. Walzer (1983) argues that goods with a distinct social meaning need to be distributed in dedicated, fair, distributive spheres, contrary to normal goods. The capability approach advocates to discuss resources that determine capabilities (freedom of choice through many opportunities) rather than functionings (actual outcomes) (Nussbaum, 2011; Sen, 2008). In practice, such fairness-relevant resources could be money, time, accessibility, opportunities for instance.

How? – Process-Relevant Aspects: Resource allocation – within this framework – is modeled as a process, in order to be applicable to any Nicomachean (diorthotic and dianemetic) discussion and any fairness ideology (including teleological and deontological). A resource-allocating, decision-making process starts with an input x (the initial situation), and an output y (the situation afterwards), where x and y might be resources (depending on the context), and y is the fairness relevant resource of interest. Furthermore, for certain Utilitarian ideologies, one might consider the utility u that receiving entities create out of those allocated resources y, with a utility-function f: $u = f(y)$. In the notation, x, y, and u represent vectors of length n, where each field of that vector represents the amount of individual i. Arguably, any decision-making system in some form uses information about an initial situation to make an informed decision, and the outcome of any decision is related in some form to the distribution of resources, be it useful goods, opportunities to be selected, or burdens in retributive contexts. Depending on the specific context and procedural-fairness aspects, further questions shall be answered during the captivation of process-relevant aspects:

- Are individuals able to influence the outcome of the process (by contributions x)?
- What information do individuals have available in beforehand?
- How transparent is the process itself?
- How often is the resource allocation taking place ? (once or regularly)
- Does the outcome distribution y affect the initial distribution x of a consecutive resource allocation process?

Who? - Participating Entities: To identify fairness-relevant entities, such as individuals, organizations, or groups of individuals and/or organizations, one needs to identify which of those entities are participating in the resource-allocating process, either as giving or receiving entity (Goppel et al., 2016).

3.2. Step 2: Transactional characteristics of the process

The second steps consists of determining the transactional characteristics of the resource allocation process, including: whether it is a dianemetic or diorthotic process, possible allocation space, and relevant ideological metrics.

Whether the process is dianemetic or diorthotic in nature depends on the relative size of giving and receiving entities; governmental endowments are typically dianemetic, while transactions of many entities of similar size in a market-like setup is typically considered diorthotic (Wolf, 2002).

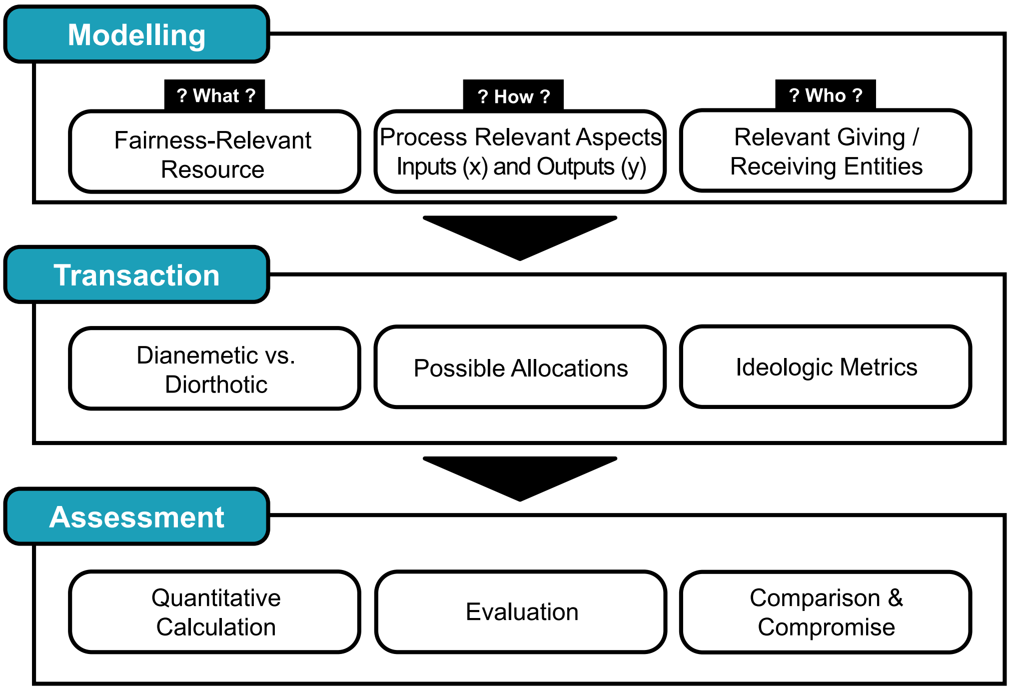


Fig. 4. Quantitative Fairness Framework – Three Step Approach.
The proposed quantitative fairness framework consists of a three step approach combining statistic dispersion metrics of distributions with econometric social welfare functions.

Table 1
Dianemetic fairness metrics.
Statistical dispersion and economic concentration metrics are used to assess the dianemetic equity of distributions with n values x_i representing inputs (contributions), and y_i representing outputs (receiving) to a resource allocating process.

Principle	Perspective	Metric	Optimization goal
Difference	Ex-Post	Average or Minimum of y_i	Maximize
Equality	Ex-Post	Dispersion of y_i	Minimize
Equality-of-opportunity	Ex-Ante	Dispersion of x_i	Minimize
Greater-good	Ex-Post	Sum of u_i	Maximize
Proportion	Both	Dispersion of ratio y_i/x_i	Minimize
Sufficiency	Ex-Post	Threshold-share of y_i	Maximize

The possible allocation space might be continuous or discrete, and bounded by the scarcity of the given resource and further constraints. For instance, the sum of distributed resources across entities $\sum_i y_i$ cannot be greater than the total resources available Y. Furthermore, it might be that the this sum is even smaller due to losses l_i of receiving entities:

$$\sum_i y_i \times l_i \leq Y \quad , \quad l_i \in [0,1] \tag{1}$$

We acknowledge the challenge of a unifying definition and guiding principle for fairness as a panacea (Casacuberta et al., 2023; Kuppler et al., 2022), and therefore advocate the use of an integrative framework, that has the capacity to reflect multiple, rather than to focus on dedicated ideologies. As a result, we provide a framework that is able to reflect following guiding principles: equality, proportion, greater-good, difference, equality-of-opportunity, and sufficiency. As not all of these ideologies combine the ex-post and ex-ante perspective on fairness, some will require input and (or) output information. Fairness is a question of distribution – according to the principles of distributive fairness (Cohen, 1987). Therefore, we advocate the use of statistic and dispersion metrics (see Table 1) to quantify dianemetic fairness, and the use of dedicated social welfare functions (see Table 2) for diorthotic fairness, in line with previous works (Casacuberta et al., 2023; Heidari et al., 2018).

For the dianemetic context, further notes should be taken to account to quantify fairness in the sense of a specific ideologies:

- Difference principle: the goal is the maximization of the average (Harsanyian fairness) or minimum of any y (Rawlsian fairness).
- Sufficiency principle: the goal is the maximization of the share of all individuals for which the outcome exceeds a certain threshold T ($y_i > T$).
- Utilitarian interpretation of difference, equality, proportion, and sufficiency principle could involve u instead of y as well. Similarly, the greater-good principle could also involve y instead of u.

For the diorthotic context, further notes should be taken to account to quantify fairness in the sense of a specific ideologies:

- Harsanyian ideology: based on the difference principle, could be reflected by the average of x as Bergson–Samuelson Welfare function.
- Sen and Foster Welfare function: not only concerned with equality, but also with average of y. This welfare functions not only capture relative distribution but also absolute distribution of resources. For example, this implies that if a slightly more unequal distribution of resources could increase the average resource received, this would be preferred over perfect equality. This could make sense in the context of resources, that deteriorate by splitting them into many small pieces.

Table 2
Diorthotic fairness metrics.
Social welfare functions are used to assess the diorthotic equity of distributions with n values x_i representing inputs (contributions), and y_i representing outputs (receiving) to a resource-allocating process.

Principle	Perspective	Welfare function
Difference	Ex-Post	Rawlsian welfare function
Equality	Ex-Post	Sen and Foster welfare function
Equality-of-opportunity	Ex-Ante	Dispersion as Leontief–Lerner function
Greater-good	Ex-Post	Benthamite welfare function
Proportion	Both	None or Dispersion of ratios y_i/x_i
Sufficiency	Ex-Post	Threshold-share of y_i

- Equality-of-opportunity principle: could employ dispersion metrics as Leontief–Lerner functions (based on the inputs x_i). Contrary to the equality principle, only the relative distribution would play a role here, as the decision-making process is assumed to affect the output y only, but not to affect the input x.
- Proportion principle: could use the dispersion of ratios as welfare function. The proclaimer of this principle (Aristotle) argued that any market activity and resulting prices and outcomes are fair per se, if these are the result of free decisions of individuals.

There is a variety of dispersion metrics in dianemetic contexts, and a compilation of the most common dispersion metrics used in statistics and econometrics is summarized in Table B.3 of the Appendix. These dispersion metrics strongly correlate; the choice of a specific metric can be decided context-specifically, but does not make a significant difference in practice.

Similarly, there is a variety of social welfare functions in diorthotic contexts, and a compilation of the most common social welfare functions is summarized in Table B.4. Leontief–Lerner welfare functions describe welfare as a function of available resources. Bergson–Samuelson welfare functions describe welfare as a function of individual utilities. Capability-approach welfare functions, describe welfare as a function of incomes. The isoelastic welfare function is a general formalization of the Bergson–Samuelson welfare function, with the possibility to weight each individuals utility with α_i; for $\rho = 0$ it becomes the Benthamite welfare function in the sense of utilitarian fairness; for $\rho = +\infty$ it becomes the Rawlsian welfare function in the sense of the Rawlsian fairness; for $\rho \to 1$ it becomes the Bernoulli–Nash welfare function; for any other ρ it is an intermediate welfare function between the extremes of utilitarian and Rawlsian fairness. Sen and Foster (Sen, 2008) apply dispersion metrics such as the Gini-Coefficient G and the Theil-Index T_T.

3.3. Step 3: Quantitative assessment of outcome scenarios

Given the resource allocation modeling and the transactional formulation of the process, the third step comprises quantitative calculation, evaluation, and comparison & compromise. The quantitative calculation results in the assignment ideological fairness metric values to the possible allocation space. This enables the evaluation of the possible allocation space in terms of different fairness guiding principles and ideologies. The evaluation indicates fairness-optimal allocations from different perspectives, allows for the ranking of possible allocations, as well as it allows to quantitatively assess how far away different allocations are from each other. Finally, the comparison & compromise sub-step reflects on how different the recommendations of various fairness ideologies actually are, and what resource allocation satisfies the different fairness definitions overall as a good compromise.

4. Case study: Fair, dianemetic & diorthotic resource allocation

In this section we want to showcase the usage of the proposed, quantitative fairness framework at two exemplary case studies relating to well-studied resource allocation problems from the literature:

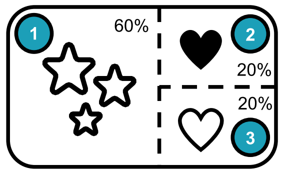
- The fair cake-cutting problem (Brams & Taylor, 1996; Moulin, 2004) is used to illustrate dianemetic fairness. The problem involves dividing a (partially-divisible), heterogeneous resource among multiple agents, and can be applied to land division, property partitioning, inheritance, partnership settlement, fair allocation in multi-agent systems (e.g. of computing resources), voting theory, market design, and natural resource sharing (e.g. emission rights).
- The daily work of two fishermen at the pond (Hardin, 1968; Ostrom, 1990; Ostrom et al., 1994) is used to illustrate diorthotic fairness. The problem refers to sharing, common-pool (public good), renewable resources, and can be applied to sustainable resource management (e.g. forests, water, fishing, hunting), cooperative game theory, Pareto efficiency, tragedy of the commons, and bargaining theory.

4.1. Cake-cutting problem & dianemetic fairness

The fair cake-cutting (Brams & Taylor, 1996) is a typical division problem discussed in economics and dianemetic, distributive justice. There are different types of problems, including whether only the size of the cake piece matters (homogeneous goods) or also other features such as the toppings (heterogeneous goods), whether the cake can be cut everywhere (fully-divisible) or only to specific discrete pieces (partially-divisible), and whether the times of cutting creates waste of no use for anyone (non-lossy vs. lossy division).

The fair cake-cutting problem is a metaphor for various dianemetic, distributive fairness contexts:

- Imagine a manager that must distribute projects to staff; staff members can usually be staffed fully on one project only, and personal preferences and interests will affect how much staff members can learn from or are willing to invest into a project. Usually, it is the manager's decision, and therefore a dianemetic context. This could be an example for a heterogeneous, partially-divisible cake-cutting problem.
- Imagine a government that tries to stimulate its stagnating economy during a recession with subsidies, tax-reductions, or cash subsidies amongst its population. How should different parts of the population (or companies) benefit from this support programme, e.g. it could make sense to support family households more than single households. This could be an example for a homogeneous, fully-divisible cake-cutting problem.
- Imagine a traffic signal controller at an intersection. It distributes delays to movement phases (respectively green time for passage). Every time there is a transition from one green movement phase to another, there is a short amount of time where both phases are red for security. Transitioning is important so that queues and waiting times do not get too long, but too frequent transitions will cause wasted time, where no vehicle can pass the intersection. How should the green time be divided to different movement phases? This could be an example for a lossy, fully-divisible cake-cutting problem.



The Individuals

Individual	Contribution	Preferences: Shape ☆	Preferences: Shape ♡	Preferences: Flavour ○	Preferences: Flavour ●
A	0.9	0.1	0.1	0.0	0.1
B	0.1	-0.1	0.1	0.1	0.2

The Quantitative Analysis

	Resource Allocation: A gets	Resource Allocation: B gets	Situation for A: X	Y	U	Situation for B: X	Y	U	Quantitative Fairness Assessment: Difference	Equality	Equality-Of-Opportunity	Greater-Good	Proportion	Sufficiency
1	1 2 3		0.9	1.0	1.4	0.1	0.0	0.0	0.00	0.99	0.56	1.40	**1.09**	0.50
2	1 2	3	0.9	0.8	1.1	0.1	0.2	0.4	0.40	0.49	0.56	1.50	1.96	0.50
3	2 3	1	0.9	0.4	0.7	0.1	0.6	0.6	0.60	**0.07**	0.56	1.30	3.69	**1.00**
4	1 3	2	0.9	0.8	1.0	0.1	0.2	0.5	0.50	0.35	0.56	1.50	2.75	**1.00**
5	1	2 3	0.9	0.6	0.7	0.1	0.4	0.9	**0.70**	0.14	0.56	**1.60**	5.81	**1.00**
6	2	1 3	0.9	0.2	0.4	0.1	0.8	1.0	0.40	0.42	0.56	1.40	6.76	0.50
7	3	1 2	0.9	0.2	0.3	0.1	0.8	1.1	0.30	0.57	0.56	1.40	7.54	0.50
8		1 2 3	0.9	0.0	0.0	0.1	1.0	1.5	0.00	1.07	0.56	1.50	10.61	0.50

Fig. 5. Case Study: Fair Cake-Cutting Problem & Dianemetic Fairness.
A given cake that can be cut into three heterogeneous pieces only shall be distributed across two individuals, that differ in their contributions to paying or making the cake, and their preferences on the toppings. Eight allocations of the three pieces to the two agents are possible. Different guiding principles on fairness result in different, fairness-optimal recommendations for how to distribute the cakes.

(1) Modeling: For this case study, let us consider a rectangular cake with chocolate toppings as displayed in Fig. 5. The toppings have two features: shape (stars and hearts), and flavor (white and dark chocolate). The cake can only be cut into three pieces. This cake must be distributed to two individuals A and B, that differ in their contributions to making or paying the cake, and preferences on topping features. They have in common, that they value the amount of cake similarly. The individual's utility function f for the received pieces of cake is the sum of three components:

- amount of cake (full cake equals one unit of utility)
- utility points according to topping shape
- utility points according to topping flavor

(2) Transaction: There are eight different ways (possible allocation space) to distribute the three pieces amongst the two individuals (assuming that no piece is wasted). The quantitative, dianemetic fairness measures enable the objective analysis of this problem, and to fairness-optimally allocate the resource amongst the individuals. Following measures are used for the different guiding principles: minimum utility (difference principle), standard deviation of utility (equality principle), standard deviation of contributions (equality-of-opportunity principle), sum of utilities (greater-good principle), standard deviation of utility-contribution ratios (proportion principle), threshold share of utility, with 0.50 utility as a sufficient minimum utility (sufficiency principle).

Let us discuss one of the eight possible allocations to outline the calculations. In this case study, we leave the contributions constant, meaning A contributes 0.9 X to the cake, and B contributes 0.1 X to the cake. Assume A gets piece 1 and 3, and B gets piece 2 (allocation scenario 4), then A receives 0.8 Y, B receives 0.2 Y. The utility the individuals experience based on the cake they get depends on the amount Y and the toppings. A experiences a utility of U= $1.0 = 0.8 + 2 \times 0.1$, as A receives 0.8 amount of cake, and additionally two toppings in heart and star form which both add 0.1 up to the utility; as A is interested in dark chocolate only but pieces 1 and 3 are of white chocolate, no additional utility can be generated for A. B experiences a utility of U= $0.5 = 0.2 + 0.2 + 0.1$, as B receives 0,2 amount of cake, and additionally one topping in heart form which adds 0.1 up to the utility, and the topping is made of dark chocolate which adds 0.2 up to the utility.

(3) Assessment: The quantitative fairness assessment of the different resource allocation scenarios reveals different levels of fairness depending on the different guiding principles. Following the difference principle, scenario 5 achieves the highest (best) worst case with a utility of 0.7. Following the equality principle, scenario 3 achieves the lowest inequality/dispersion (standard deviation) in utilities. A discussion of the equality-of-opportunity in this case study might not be relevant, as we did not provide details on the reasons for contributions. Following the greater-good principle, scenario 5 achieves the highest total utility. Following the proportion principle, scenario 1 achieves the highest alignment with contributions. Following the sufficiency principle, scenarios 3-5 provide a sufficient minimum experienced utility for both individuals.

These differing ratings for fairness by the different guiding principles highlight the conflicting differences in understanding of what fairness is. Scenario 5 for instance, achieves the best situation in total, and even for the poorest, while it is twice as unequal when compared to scenario 3. From a purely merits based perspective, A should receive everything as in scenario 1, as A contributes almost everything for buying the cake.

4.2. Fishermen & diorthotic fairness

Let us discuss diorthotic fairness at the example of the daily work of two fishermen (Ostrom, 1990), as shown in Fig. 6. Assume there is a village with two fishermen, that go to the same pond for fishing every day. The fishermen differ in their capabilities due to experience and catching methods they use, but also in their working attitude and willingness to work. Should each fisherman keep what they catched, or

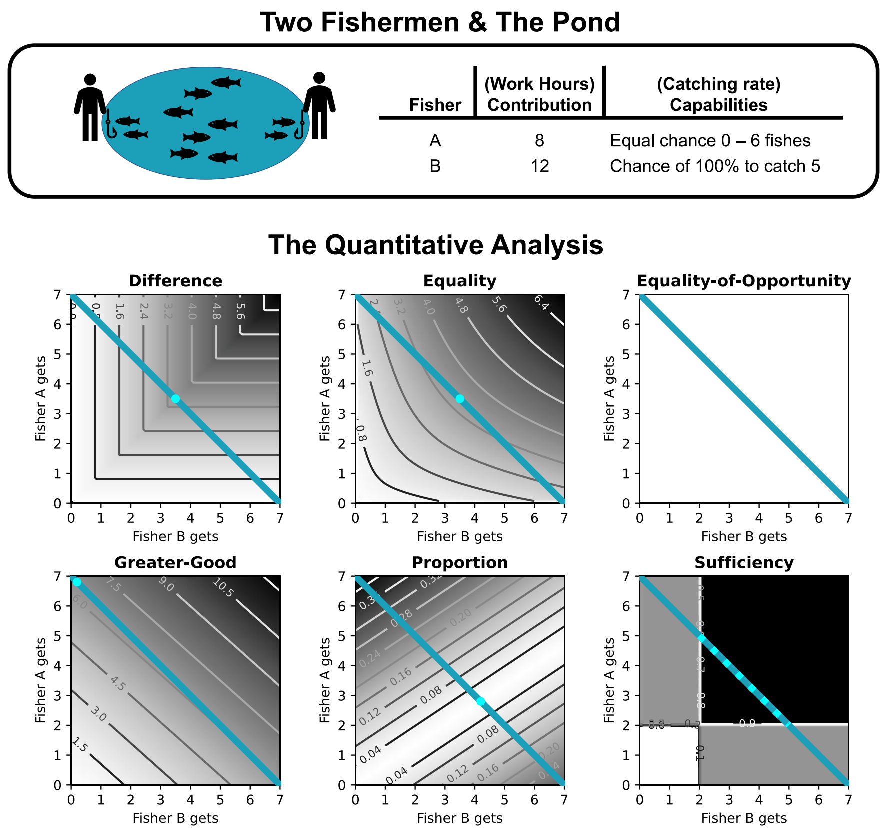


Fig. 6. Case Study: Fishermen & Diorthotic Fairness.
Two fishermen go fishing at a pond every day, and at this specific day they fish seven fishes together. They differ in the working hours and catching rates due to different fishing techniques. Their outcome is stochastic. How should they distribute the total fish catch per day? Assuming fish is divisible, there is a continuous allocation space of the fish ranging from zero to seven fishes for each fisher. The Pareto-efficient frontier (blue line) displays this solution space. Different guiding principles on fairness result in different heatmaps and social welfare functions (contour plots of the heatmaps), with different recommendations on how to distribute the daily fish catch.

should the village agree on some form of legislation to distribute the daily catch "fairly" to both fishermen?

This example is a metaphor for various diorthotic fairness contexts:

- Imagine a nation's labor economy. People have different capabilities to earn an income, different chances depending on the economic situation of markets, exports, industries, and also different levels of discipline and perseverance. A government could apply an income taxation to redistribute wealth amongst the individuals of the labor force. This could create an overall healthier economy, as workers with bad luck or health in one season can still survive based on the solidarity support of others. However, this could also create a demotivation of the top performers, as higher achievements mean higher contributions to the others, that might not work as hard as them.
- Imagine a nation's public health system. People have different chances to get sick based on genetic and socio-economic dispositions that they cant affect, but also do they make life choices that lead to different health-affecting life-styles (e.g. smoking). Should the government enforce a mandatory health insurance system, so that everyone pays into the system, even when not using it? The solidarity could lead to less extreme poverty, as often severe diseases, accompanied by job loss and divorce, lead to homelessness. However, many diseases could be prevented by a healthier life style, and a governmentally-enforced safety net could incentivize a free-riding behavior, in which individuals have even less incentivizes to take care of them, as they can get cured on the costs of others. If such a public health system and insurance would be implemented, how should people pay for it? Based on their income, based on their lifestyle, or even based on their genetic and socio-economic dispositions?

(1) Modeling: For this case study, let us consider two fishermen, *A* and *B*. Fisherman *A* works 8 h per day, and has an equal chance to find 0 to 6 fishes per day. This means, on average, *A* catches 3 fishes per working day, and 0.375 fishes per hour. Another fisherman *B* works 12 h per day, and due to his experience and catching method, catches exactly 5 fishes per day. This means, on average *B* catches 5 fishes per working day, and 0.417 fishes per hour.

(2) Transaction: Let us assume one day, *A* catches 2 fishes, and *B* catches 5 fishes, so there are 7 fishes in total to be distributed. As fish is a perishable good, it does not make sense to not allocate all fishes, so the sum of all fishes of *A* and *B* (5+2=7) forms the Pareto-efficient frontier. Moreover, each fisherman has the same utility for the amount of fish received, however when walking back home, fisherman *A* looses around 5%, and fisherman *B* looses around 15%.

Fish can be considered as a divisible good, one could assume continuous distribution based on weight. This means there is an infinitely large solution space for allocations along the Pareto-efficient frontier. The quantitative, diorthotic fairness measures enable to objectively analyze this problem and to fairness-optimally allocate the resources. Following measures are used for the different guiding principles: Rawlsian welfare function (difference principle), Foster welfare function (equity principle), standard deviation of working time (equality-of-opportunity principle), sum of all utilities (greater-good principle), standard deviation of received fish-working time ratios (proportion principle), threshold share of received fish, with 2 units as a sufficient minimum (sufficiency principle).

(3) Assessment: The quantitative fairness assessment of the different resource allocation scenarios reveals different levels of fairness depending on the different guiding principle of fairness. Following the difference principle, it would be most fair to equally distribute the catched fish, meaning both fishermen get 3.5 units of fish. Following the equality principle, it would be most fair to equally distribute as well. A discussion of the equality-of-opportunity in this case study might not be relevant, as we did not provide details on the reasons for working times. It might be, that A is older and cannot work that long any more as B does. In this case, one would need to discuss ways to account for that. Following the greater-good principle all fish should be given to A, as A can bring most of fish home where it can actually be cooked, while giving fish to B would generate slightly more waste. If the greater-good principle would be applied to resources y rather than utilities u, then the greater-good principle would be indifferent on any allocation along the Pareto-efficient frontier. Following the proportion principle, A should receive 2.8 and B should receive 4.2 units of fish, according to their contributed working time. One could argue, that not working time, but actual contribution matters as Aristotle argues; in this case each fisherman should keep as much as they catched without sharing. The sufficiency principle would be indifferent to all allocations along the Pareto-efficient front between 2 to 5 fishes for A resp. B.

The different social welfare functions exhibit different shapes. Difference and equality principle share the same optima and similar gradients. The greater-good principle is almost parallel, and the proportion principle is even almost orthogonal to the Pareto-efficient front. The sufficiency principle rather distributes the solution space into distinct areas.

5. Discussion

In this section, we discuss the properties of the proposed framework to highlight its theoretical contribution, advocate compromise over optimize, and provide some further notes on the complexities of the equality-of-opportunities guiding principle in the context of the case studies.

5.1. Quantitative framework properties

The proposed quantitative fairness framework comprises following properties, making it a unique, valuable, and powerful contribution to the literature on quantitative fairness assessments:

- **Holistic:** The framework combines distributive, procedural, and algorithmic fairness.
- **Integrative:** The framework integrates multiple fairness ideologies and guiding principles rather than to advocate a specific one.
- **Transactional:** The framework enables a fairness discussion of both diorthotic and dianemetic resource allocation processes.
- **Multi-Perspective:** The framework enables both a teleological and deontological fairness discussion due to its quantitative nature. Either can it be used to judge the outcomes of a resource allocating process (teleological) or can it be used to inform designing better processes (deontological).
- **Quantitative:** The framework not only contributes normative definitions, or does it enable cardinal, ordinal, and comparative rankings of algorithms and allocations, but does it also enable to quantitatively assess how much more fair (unfair) a certain process or algorithm is (e.g., compared to a certain optimum).

5.2. Compromise over optimize

The proposed quantitative framework purposely does not advocate one over another guiding principle. Deviating from the fairness-optimal allocations from one guiding principle by just a bit, can achieve significant improvements from the perspective of another guiding principle. A compromise solution might be derived based on the preferences and weights the decision maker provides to the different guiding principles. Alternatively, rankings of the possible allocation space (based on the different guiding principles) could be created, and then aggregated to a combined ranking for a final decision-making.

Contrary to previous, rather transcendental, qualitative and mostly normative, philosophic works, our framework can be used to analyze situations from a more transitive, and quantitative perspective. Not only can two situations be compared to decide which one is more fair (transitive), but also can the framework be used to assess how much more fair it is (quantitative). This enables a more systematic discussion of the deviation from strict optima, and encourages an inclusive discussion that allows for the combination of different goals, including different fairness and efficiency definitions. Similar to quantitative definitions of efficiency, quantitative definitions of fairness can be used as a goal metric for the design and multi-goal optimization of algorithms or the outcomes of these algorithms.

5.3. Notes on equality-of-opportunities guiding principle

The equality-of-opportunities guiding principle is related to the chance aspect of procedural fairness and a guiding principle for the distributive fairness. In the case studies we excluded a discussion, as further assumptions must be taken. Besides, the question that needs to be answered for a discussion from this guiding principle's perspective includes how the resource allocation can affect the opportunities and chances individuals have.

If there is a clear relationship between the allocated resources in one cycle and the chances of individuals in the consecutive cycle, then equality-of-opportunity is clearly relevant to discussing the allocation of resources. For instance, one could assume that having more money at the beginning of a market place's opening will allow to generate more trading profits, which will then enable even more chances and opportunities at the beginning of the following day. This could be considered as a feedback loop, and therefore redistribution of shares of profits that are due to luck rather than capabilities or efforts, might be considered as fair, when they provide more chances to everyone else.

If there is no clear relationship between the allocated resources and the chances, a different discussion is necessary (e.g. how much fish you get as fisherman in the case study wont affect how much fish you can get on the next day). One could rather focus on which decision criteria an algorithm uses, or how it weights different inputs to the decision-making. Doing so, one could aim for inputs that reflect more equal opportunities for individuals to participate and actually affect the outcome.

Rather than focusing on pure contribution of making or buying the cake in the dianemetic case study above, one could try to adjust and normalize the inputs for capability (how much pocket money do you have available to pay for the cake) or ability (how much knowledge and tools do you have to make a cake). Rather than focusing on pure contribution of fish, or working time in the diorthotic case study above, one could try to adjust and normalize these inputs for capability (catching rate of fish) or ability (age, gender, size), to better reflect the pure willingness to work.

6. Conclusion

This work set out to propose an useful, holistic, quantitative, transactional, distributive fairness framework, which enables the systematic design of socially-feasible decision-making systems for resource allocation in the context of equitable cybernetic societies. After the review of distributive, procedural, and algorithmic fairness theories from relevant literature and domains, we identified limitations of existing frameworks and motivate the proposition of a new kind of framework.

The proposed quantitative fairness framework offers measures for dianemetic and diorthotic fairness discussions based on statistic metrics, dispersion metrics, and social welfare functions. Two case studies on fair cake-cutting and fishermen demonstrate the usefulness and flexibility of the proposed framework. A discussion highlights the properties and novelty of the proposed framework, and advocates compromise over optimize.

Future work could focus not only on situational quantification of fairness at a specific time, but to include a temporal component for repeated settings. For example, forms of aggregation over many iterations of the same algorithm could be part of investigation. A useful way could be the probabilistic, stochastic discussion of the distributive effects of algorithms.

CRediT authorship contribution statement

Kevin Riehl: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anastasios Kouvelas:** Writing – review & editing, Supervision. **Michail A. Makridis:** Writing – review & editing, Supervision.

Funding

This work has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101203465 (project "FEDORA"), and from the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract No. REF-1131-52301.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. An interdisciplinary review of fairness

Based on an interdisciplinary review of six domains (economics, natural science, engineering, business administration, government & policy, social sciences), we identify that fairness:

- acts as the social and systemic glue enabling cooperation, legitimacy, and stability;
- functions as a feedback mechanism balancing individual incentives and collective welfare;
- evolved biologically as reciprocal altruism, sustaining cooperation and group survival;
- regulates distribution of benefits and burdens in markets, infrastructures, and networks;
- sustains trust, compliance, and motivation within organizations and institutions;
- supports social cohesion, inclusion, and upward mobility through fair opportunities;
- maintains legitimacy of governance and policy by reducing perceived injustice;
- extends to intergenerational and ecological justice, ensuring sustainability; and
- becomes an operational requirement in cybernetic societies governed by algorithms and feedback loops.

Thus, quantitative fairness is not only a normative ideal but a control variable for the sustainability, resilience, and ethical integrity of complex socio-technical systems.

Natural science

Evolutionary biology & behavioral psychology:
Fairness in biological systems often manifests as cooperative behaviors. From an evolutionary biology perspective, sensitivity to fairness is a behavioral trait that evolved in social animals, whose survival relies on cooperation and group dynamics. This perception of fairness can be observed in various social animals, including Capuchin monkeys, vampire bats, and humans (Brosnan & de Waal, 2014). At the individual level, the subjective perception of fairness plays a crucial role in both physical health and psychological well-being. The brain of social animals, particularly the insula, reacts to perceived unfairness with a sense of disgust. Interestingly, this reaction is more pronounced when individuals feel they are unfairly disadvantaged compared to when they are unfairly advantaged (Jackson et al., 2006). Furthermore, perceptions of fairness significantly influence how individuals form relationships and interact with one another (Cremer et al., 2010). At the societal level, fairness plays a crucial role in fostering robust communities, encouraging adherence to social norms, and promoting cooperation. The perception of fairness is shaped by shared, group-specific norms and cultural values. Moreover, fairness serves to enhance social cohesion, mitigate conflicts, bolster group identity and trust, and ultimately leads to improved group work outcomes (Cremer et al., 2010; Hitti et al., 2011). The evolution of fairness in biological systems is closely tied to the concept of reciprocal altruism. Organisms that engage in fair exchanges are more likely to maintain beneficial relationships over time, increasing their chances of survival and reproduction. This has led to the development of various strategies and mechanisms in nature to enforce fairness, such as partner choice and reputation systems in some species. Often, fairness reflects a balance between contribution and reward that is sustainable and beneficial for the species or ecosystem as a whole (Schino & Aureli, 2009).

Engineering

Network management:
Human societies use networks for the transportation and distribution of resources, e.g. energy networks, telecommunication networks, supply chain networks, rail- and road networks, computation networks, and social networks. Fairness in network management is a critical aspect of ensuring that resources are distributed equitably among users and applications. This principle is essential for maintaining the efficiency, reliability, user satisfaction, and motivation to contribute to a network. Discussions of fairness in networks include allocated resources, but also quality of service (e.g. delays or bandwidth in the context of internet networks). Achieving fairness in network management often involves trade-offs between efficiency and equity. For instance, while it is desirable to allocate resources fairly, it is also important to ensure that the network operates efficiently and maximizes overall throughput. This balance can be challenging, particularly in heterogeneous networks where users have varying requirements and capabilities. Utility optimization methods are often used to address this issue, where the goal is to maximize the aggregate utility of all users while ensuring fair resource allocation. Various fairness models, such as max–min fairness, proportional fairness, α-fairness, and the Jain metric have been developed to address these allocation challenges (Bonald et al., 2006; Jain et al., 1984).

Economics

Macro economics:

Economics studies the production, consumption and distribution (allocation) of resources in human societies. Fairness and economics are deeply intertwined concepts that have significant implications for market structures, competition, policies and taxation. Governmental intervention into markets is often justified with restoring competition that is affected by monopolies, public goods, and externalities. Minimum wages, wealth redistribution and income taxation play an important role in fairness-promoting policies. Behavioral economics has provided new insights into how people form judgments about what is fair. Factors like framing, social norms, and perceptions of intentions all influence whether a given outcome is seen as fair or unfair, which has important implications on public support for policies. Welfare economics is a branch of economics that studies social welfare, and evaluates the overall well-being of a society, where fairness is an important component to most definitions of welfare. Especially welfare-economists have shaped the fairness literature of the recent (Alesina & Angeletos, 2005; Feldman, 2018; Fleurbaey, 2008).

Micro economics & fairness of market prices:

When can markets and market prices be considered as fair? The fairness of markets was closely linked with the philosophical, political and economic discussions of fairness. Aristotle argues, a diorthotic transaction is fair, when the exchanged resources are of equal value. Albertus Magnus and Thomas Aquinas introduced the term of a fair price for transactions at monetary markets. The fair price primarily reflects the efforts for the generation of the resource, but can also include marginal profits of traders. The school of Salamanca puts the term of a fair price equal to the market price, assuming efficient, ideal markets. Smith (1776) developed the theory of the invisible hand which claims, that any selfish, egoistic behavior and any price in transactions is fair, as free markets lead to societal optima as a result. Besides market failure and arbitrage, purposeful phenomena such as price differentiation, dynamic pricing, price discrimination, and personalized pricing are heavily discussed in the literature. The fairness of markets is therefore closely linked to the fairness of prices (Kahneman et al., 1986).

Education:

Education systems are essential for societal development and individual growth. As a catalyst for social mobility, education equips individuals with knowledge and skills that can improve job prospects and socioeconomic status (Brown, 2017). Early childhood education is especially key to promoting equity. Fairness ensures every child has the chance to reach their full potential. Because children enter school with diverse backgrounds and abilities, equal access to quality teachers, textbooks, and learning environments is crucial. Additionally, students with special needs, including disabilities, may require more resources. Governments play a vital role in providing public education to address disparities that affect disadvantaged children (Tharp, 2018).

Housing & gentrification:

Gentrification is a complex urban phenomenon involving housing, economic development, and social equity. It can improve neighborhoods through better infrastructure and economic activity but also raises concerns about fairness and displacement of long-time residents. It typically involves wealthier residents moving into lower-income areas, causing rising property values, rents, and changes in local culture. This often displaces long-term residents, impacting social cohesion and raising questions of housing fairness and social justice (Krings & Schusler, 2020). Housing markets and rental prices are frequently regulated to address discrimination, racial segregation, economic disparities, and challenges faced by marginalized communities, adding complexity to the issue (Von Hoffman, 2000).

Healthcare:

Access to medical services is vital for sustaining healthy and worth-living lives, and for keeping up productivity of the workforce. In many societies, access to healthcare and quality of treatment outcomes depends upon socio-economic contexts. At the same time, disease and sickness have been identified as one of the most important reasons for people turning into poverty (and even homelessness in extreme cases) (Jamison, 2018). Fairness-relevant discussions in this context include health insurance systems on the societal level, and various ethical questions on the individual level, such as prioritization of patients in the context of emergency rooms and organ transplants (Daniels et al., 1996; Ding et al., 2019).

Transportation planning & policy:

Transportation infrastructure design determines how effectively people and goods can move around. Fairness arises as an important theme in the field of transportation. Accessible, affordable, safe, inclusive, and barrier-free transportation are the key aspects of discussion. The planning of transportation infrastructure, such as roads, railways, and public transport are important to enable an equitable access for as many as possible, with positive outcomes for the economy and life-quality. Often, transportation infrastructure faces a demand which is higher than its supply, resulting in congestion, and long waiting queues; policies for traffic demand management such as congestion pricing are part of highly controversial debates. Especially fairness and equity-concerns are the major impediments for the real-world implementation of transportation policies (Gu et al., 2018; Martens, 2016).

Business administration

Management:

Fairness in management is essential for effective leadership and organizational success. It involves equitable treatment, transparent decision-making, and inclusive work environments. Prioritizing fairness fosters trust, morale, and productivity. Key aspects include consistent treatment, transparency, equal growth opportunities, conflict resolution, and fair compensation. By emphasizing these, managers create workplaces where employees feel valued and respected, benefiting both individuals and the organization's long-term success (Simons & Roberson, 2003).

Recruiting & hiring processes:

Ensuring fairness in recruitment and selection not only promotes diversity and inclusion, but also enhances the reputation of the organization as an equitable employer, increases the chance to find the most skilled workers for a position, and improves organizational efficiency. Procedural fairness, transparency on the recruiting process, and bias awareness play an important role. To account for diversity and inclusion, both the applicant pool diversity and hiring outcome diversity are common measures to quantify fairness in this context. The increased use of automated decision-making allow for the consideration of larger number of applications, but also harbor the threat for technological biases, that need to be considered carefully (van den Broek et al., 2020).

Banking & loan approval:

Access to financial services for all individuals is key to inclusion and customer participation. Historically, marginalized groups have faced systemic barriers such as discriminatory lending, biased credit scoring, and opaque loan approvals. Addressing these issues is essential for equity and public trust. With the rise of machine learning in credit decisions, algorithmic fairness has become critical, as algorithms may reinforce existing biases. Various fairness metrics and bias mitigation strategies aim to prevent discrimination against marginalized groups. At the same time, financial institutions must manage lending risk by assessing creditworthiness without disproportionately affecting certain groups. Balancing fair access and risk management remains a complex challenge needing innovative, ongoing solutions (Lee & Floridi, 2021).

Table B.3
Dispersion metrics.
Statistical dispersion and economic concentration metrics are used to assess the equality and thus dispersion of distributions. The formulas cover the most commonly used measures and are denoted to determine the equality of a distribution of n values x_i, where $\bar{x}$ represents the average over all x_i.

Metric	Reference	Formula
Atkinson-Index	Atkinson et al. (1970)	$A_\epsilon = \begin{cases} 1 - \frac{1}{\bar{x}}(\frac{1}{n}\sum_{i=1}^{n} x_i^{1-\epsilon})^{1/(1-\epsilon)} \\ for\ 0 \leq \epsilon < 1 \\ 1 - \frac{1}{\bar{x}}(\Pi_{i=1}^{n} x_i)^{1/N} \\ for\ 0 \leq \epsilon < 1 \\ 1 - \frac{1}{\bar{x}} \min_i(x_i) \\ for\ \epsilon = +\infty \end{cases}$
Gini-Coefficient	Dorfman (1979)	$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} \lVert x_i - x_j \rVert}{2n\sum_{i=1}^{n} x_i}$
Herfindahl-Index	Herfindahl (1950)	$HH_n = \frac{HH - \frac{1}{n}}{1 - \frac{1}{n}}, HH = \sum_{i=1}^{n}(\frac{x_i}{\sum_{j=1}^{n} x_j})^2$
Hirschmann-Index/ Hoover-Index	Hoover (1936)	$H = \frac{1}{2}\frac{\sum_i^n \lVert x_i - \bar{x} \rVert}{\sum_i^n x_i}$
Palma-Index	Palma (2011)	$P = \frac{\int_{0\%}^{40\%} L(x)dx}{\int_{90\%}^{100\%} L(x)dx}$
Standard Deviation		$\sigma = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
Theil-Index T	Theil (1965)	$T_T = \frac{1}{n}\sum_{i=1}^{n} \frac{x_i}{\bar{x}} ln(\frac{x_i}{\bar{x}})$
Theil-Index L	Theil (1965)	$T_L = \frac{1}{n}\sum_{i=1}^{n} ln(\frac{\bar{x}}{x_i})$

Table B.4
Welfare functions.
Social welfare functions aim to define the social welfare W of resource allocations, where x describes the input to the resource allocation process (e.g. financial wealth), y describes the output of the resource allocation process (e.g. allocated resources), and u describes the utility of the allocated resources to the individuals. There are three types of welfare functions: Leontief–Lerner, Bergson–Samuelson, and Capability-approach functions.

Function title	Formalization
Leontief–Lerner	$W = F(x_1, \dots, x_n)$
Bergson–Samuelson	$W = F(u_1, \dots, u_n)$
Isoelastic	$W = \frac{1}{1-\rho}\sum_{i=1}^{n} \alpha_i u_i^{1-\rho}$
Benthamite	$W = \sum_{i=1}^{n} u_i$
Rawlsian	$W = \min_{i=1}^{n} u_i$
Bernoulli–Nash	$W = \Pi_{i=1}^{n} \alpha_i u_i$
Capability-approach	$W = F(y_1, \dots, y_n)$
Sen	$W = \bar{y}(1 - G)$
Foster	$W = \bar{y}e^{-T_T}$

Social sciences

Social justice:

Social justice involves the fair distribution of resources, opportunities, and privileges within society, based on equality, human rights, and collective responsibility. Its goal is to address systemic inequalities and promote inclusivity regardless of background. It focuses on rectifying disparities in economic status, race, gender, and disability rights. Achieving social justice often requires challenging power structures through policy reform, activism, education, and awareness. Movements for social justice have advanced civil and workers' rights throughout history (Harvey, 2010). In the digital age, social media has become a vital tool for organizing and amplifying marginalized voices but also raises issues like digital divides and online harassment (Eubanks, 2012).

Environmental & intergenerational justice:

A particular form of social justice examines the relationship between humans and their ecological environment, exploring how justice between generations can be addressed amid demographic changes and climate change. Environmental justice seeks equal protection from environmental and health hazards for all people, regardless of race, color, national origin, or income, emphasizing how marginalized communities often face greater risks (Mohai et al., 2009). Intergenerational justice broadens this concept, stressing the responsibility of current generations to preserve the environment for those to come and recognizing the lasting impacts of today's actions (Barry, 1997).

Sports & competitions:

Fairness is a key principle in sports, ensuring competitions are meaningful and equitable. It centers on equal opportunity based on skill, effort, and merit, rather than external advantages. Sports organizations design rules and structures to give all competitors a fair chance to succeed. Adherence to these rules, including anti-doping measures, maintains integrity and prevents unfair advantages. Beyond rule-following, fairness embraces sportsmanship, respect, and ethical behavior, fostering a positive sports culture. Moreover, fairness sustains public trust and interest, enhancing the credibility and appeal of sports events (Loland, 2010).

Government & policy

Policy making:

Fairness is vital in policy development and implementation across governance. Policies should be created through transparent processes involving public input and stakeholder consultation. Implementation must be consistent, with clear explanations for decisions. Policymakers need to consider the impacts on different groups and aim for just outcomes, sometimes requiring targeted support for vulnerable populations. Policies should promote equality of opportunity, recognizing that strict equality of outcomes is not always feasible. Anti-discrimination measures help ensure fair access across various social lines. Evidence-based policymaking supports fairness by relying on objective data, though data must be scrutinized to avoid reinforcing inequities. Perceptions of fairness vary culturally and evolve, so ongoing dialogue with diverse communities is essential (Gilley, 2017).

Legal system & criminal justice:

Fairness in the legal system is fundamental to the rule of law, ensuring justice is applied equitably and impartially. It is key to maintaining public trust. Procedural fairness includes the right to a fair hearing and an impartial decision-maker. In criminal cases, it protects the accused's rights, such as the presumption of innocence, legal representation, and a fair trial, helping prevent miscarriages of justice. Equality before the law means all individuals are subject to the same laws and entitled to equal protection, without discrimination based on race, gender, ethnicity, or socioeconomic status. Achieving this requires ongoing efforts to address systemic biases and ensure fairness for all (Berk et al., 2021; Hurwitz & Peffley, 2005).

Police operations:

Fairness is crucial in policing, affecting both public perception and the internal functioning of police organizations. Procedural fairness – treating people with dignity, explaining actions, and ensuring neutrality – enhances police legitimacy, which builds public trust and cooperation, even when outcomes are unfavorable. A study of New York City residents found procedural fairness key to police legitimacy and compliance with the law (Sunshine & Tyler, 2003). Racial profiling, which uses race or ethnicity in decisions to stop or search individuals, highlights the tension between fairness and effectiveness. While some argue it aids crime prevention, racial profiling raises concerns about discrimination, civil rights, and loss of public trust. It judges groups rather than individuals, potentially creating a self-fulfilling prophecy where increased scrutiny leads to higher arrest rates, justifying further profiling (Hurwitz & Peffley, 2005).

Appendix B. Tables

See Tables B.3 and B.4.

References

Abbott, J., McKiernan, K., & McNulty, S. (2024). Technocracy for the people? The impact of government-imposed democratic innovations on governance and citizen well-being. *Comparative Political Studies*, *57*(2), 187–220. http://dx.doi.org/10.1177/00104140231178725.

Acemoglu, D., & Wolitzky, A. (2021). A theory of equality before the law. *The Economic Journal*, *131*(636), 1429–1465. http://dx.doi.org/10.1093/ej/ueaa116.

Agyemang, M., Andreae, D. A., & McComb, C. (2023). Uncovering potential bias in engineering design: a comparative review of bias research in medicine. *Design Science*, *9*, Article e17. http://dx.doi.org/10.1017/dsj.2023.17.

Alesina, A., & Angeletos, G. M. (2005). Fairness and redistribution. *American Economic Review*, *95*(4), 960–980. http://dx.doi.org/10.1257/0002828054825655.

Ashby, W. R. (1956). *An introduction to cybernetics*. London: Chapman & Hall Ltd., ISBN: 1-61427-765-6.

Atkinson, A. B., et al. (1970). On the measurement of inequality. *Journal of Economic Theory*, *2*(3), 244–263. http://dx.doi.org/10.2307/1924845.

Barnard, C., & Hepple, B. (2000). Substantive equality. *The Cambridge Law Journal*, *59*(3), 562–585. http://dx.doi.org/10.1017/S0008197300000246.

Barry, B. (1997). Sustainability and intergenerational justice. *Theoria*, *44*(89), 43–64. http://dx.doi.org/10.3167/004058197783593443.

Beigang, F. (2022). On the advantages of distinguishing between predictive and allocative fairness in algorithmic decision-making. *Minds and Machines*, *32*(4), 655–682. http://dx.doi.org/10.1007/s11023-022-09615-9.

Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, *50*(1), 3–44. http://dx.doi.org/10.1177/0049124118782533.

Bonald, T., Massoulié, L., Proutiere, A., & Virtamo, J. (2006). A queueing analysis of max-min fairness, proportional fairness and balanced fairness. *Queueing Systems*, *53*, 65–84. http://dx.doi.org/10.1007/s11134-006-7587-7.

Brams, S. J., & Taylor, A. D. (1996). *Fair division: From cake-cutting to dispute resolution*. Cambridge University Press, ISBN: 978-0521556446, http://dx.doi.org/10.1017/CBO9780511598975.

van den Broek, E., Sergeeva, A., & Huysman, M. (2020). Hiring algorithms: An ethnography of fairness in practice. In *40th international conference on information systems* (pp. 1–9). Association for Information Systems, ISBN: 0-9966831-9-4.

Brosnan, S. F., & de Waal, F. B. (2014). Evolution of responses to (un) fairness. *Science*, *346*(6207), Article 1251776. http://dx.doi.org/10.1126/science.1251776.

Brown, P. (2017). Education, opportunity and the prospects for social mobility. In *Education and social mobility* (pp. 60–82). Routledge, http://dx.doi.org/10.4324/9781315651972-11.

Carver, C. S., & Scheier, M. F. (2012). Cybernetic control processes and the self-regulation of behavior. *The Oxford Handbook of Human Motivation*, 28–42. http://dx.doi.org/10.1093/oxfordhb/9780195399820.001.0001,

Casacuberta, S., Robinson, I., & Wagaman, C. (2023). Augmenting fairness with welfare: A framework for algorithmic justice. In *European Workshop on Algorithmic Fairness*.

Cohen, R. L. (1987). Distributive justice: Theory and research. *Social Justice Research*, *1*(1), 19–40. http://dx.doi.org/10.1007/BF01049382.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2022). *Introduction to algorithms*. MIT Press, ISBN: 0-262-04630-X.

Daniels, N., Light, D., & Caplan, R. L. (1996). *Benchmarks of fairness for health care reform*. USA: Oxford University Press, ISBN: 0-19-510237-1, http://dx.doi.org/10.1093/oso/9780195102376.003.0003.

Dastin, J. (2018). Insight - amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, URL: https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/.

De Cremer, D., Brockner, J., Fishman, A., van Dijke, M., van Olffen, W., & Mayer, D. M. (2010). When do procedural fairness and outcome fairness interact to influence employees' work attitudes and behaviors? The moderating effect of uncertainty. *Journal of Applied Psychology*, *95*(2), 291. http://dx.doi.org/10.1037/a0017866.

Ding, Y., Park, E., Nagarajan, M., & Grafstein, E. (2019). Patient prioritization in emergency department triage systems: An empirical study of the Canadian triage and acuity scale (CTAS). *Manufacturing & Service Operations Management*, *21*(4), 723–741. http://dx.doi.org/10.1287/msom.2018.0719.

Dorfman, R. (1979). A formula for the gini coefficient. *The Review of Economics and Statistics*, 146–149. http://dx.doi.org/10.2307/1924845.

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., et al. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys*, *55*(9), 1–33. http://dx.doi.org/10.1145/3561048.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference* (pp. 214–226). http://dx.doi.org/10.1145/2090236.2090255.

Dworkin, R. (2000). *Sovereign virtue: the theory and practice of equality*. Harvard University Press, ISBN: 0-674-00219-9, URL: http://www.jstor.org/stable/j.ctv1c3pd0r.

Eubanks, V. (2012). *Digital dead end: Fighting for social justice in the information age*. MIT Press, ISBN: 978-0-262-01498-4, http://dx.doi.org/10.7551/mitpress/8073.001.0001.

Feldman, A. M. (2018). Welfare economics. In *The new palgrave dictionary of economics* (pp. 14536–14549). Springer, http://dx.doi.org/10.1057/978-1-349-95189-5_1417.

Fleurbaey, M. (2008). *Fairness, responsibility and welfare*. Oxford University Press, ISBN: 978-0191808579, http://dx.doi.org/10.1093/acprof:osobl/9780199215911.001.0001.

Friedman, J. (2019). *Power without knowledge: a critique of technocracy*. Oxford University Press, ISBN: 0-19-087717-0, http://dx.doi.org/10.1093/oso/9780190877170.001.0001.

Gilley, B. (2017). Technocracy and democracy as spheres of justice in public policy. *Policy Sciences*, *50*, 9–22. http://dx.doi.org/10.1007/s11077-016-9260-2.

Goppel, A., Mieth, C., & Neuhäuser, C. (2016). *Handbuch gerechtigkeit*. Springer, ISBN: 978-3-476-02463-3, http://dx.doi.org/10.1007/978-3-476-05345-9.

Gu, Z., Liu, Z., Cheng, Q., & Saberi, M. (2018). Congestion pricing practices and public acceptance: A review of evidence. *Case Studies on Transport Policy*, *6*(1), 94–101. http://dx.doi.org/10.1016/j.cstp.2018.01.004.

Gurney, G. G., Mangubhai, S., Fox, M., Kim, M. K., & Agrawal, A. (2021). Equity in environmental governance: perceived fairness of distributional justice principles in marine co-management. *Environmental Science & Policy*, *124*, 23–32. http://dx.doi.org/10.1016/j.envsci.2021.05.022.

Hardin, G. (1968). The tragedy of the commons. *Science*, *162*(3859), 1243–1248. http://dx.doi.org/10.1126/science.162.3859.1243.

Harsanyi, J. C. (1975). Can the maximin principle serve as a basis for morality? A critique of John Rawls's theory. *American Political Science Review*, *69*(2), 594–606. http://dx.doi.org/10.2307/1959090.

Harvey, D. (2010). *Social justice and the city*: *Vol. 1*, University of Georgia Press, ISBN: 978-0-8203-3604-6.

Heidari, H., Ferrari, C., Gummadi, K., & Krause, A. (2018). Fairness behind a veil of ignorance: A welfare analysis for automated decision making. *Advances in Neural Information Processing Systems*, *31*.

Hellman, D. (2025). Algorithmic fairness. In E. N. Zalta, & U. Nodelman (Eds.), *The stanford encyclopedia of philosophy* (Fall 2025 ed.). (pp. 1–100). Metaphysics Research Lab, Stanford University, [ISSN: 1095-5054].

Herfindahl, O. C. (1950). *Concentration in the steel industry* (Ph.D. thesis), Columbia University.

Hern, A. (2018). Google's solution to accidental algorithmic racism: ban gorillas. *The Guardian*, URL: https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people.

Hirshleifer, J. (1978). The private and social value of information and the reward to inventive activity. In *Uncertainty in economics* (pp. 541–556). Elsevier, http://dx.doi.org/10.1016/B978-0-12-214850-7.50038-3.

Hitti, A., Mulvey, K. L., & Killen, M. (2011). Social exclusion and culture: The role of group norms, group identity and fairness. *Anales de Psicología*, *27*(3), 587–599. http://dx.doi.org/10.6018/analesps.

Holzinger, A., Saranti, A., Molnar, C., Biecek, P., & Samek, W. (2020). Explainable AI methods-a brief overview. In *International workshop on extending explainable AI beyond deep models and classifiers* (pp. 13–38). Springer, http://dx.doi.org/10.1007/978-3-031-04083-2_2.

Hoover, E. M. (1936). The measurement of industrial localization. *The Review of Economic Statistics*, 162–171. http://dx.doi.org/10.2307/1927875.

Hossain, M. S., Muhammad, G., & Guizani, N. (2020). Explainable AI and mass surveillance system-based healthcare framework to combat COVID-I9 like pandemics. *IEEE Network*, *34*(4), 126–132. http://dx.doi.org/10.1109/MNET.011.2000458.

Hume, D. (1740). A treatise of human nature: Being an attempt to introduce the experimental method of reasoning into moral subjects. Self Published Manuscript.

Hurwitz, J., & Peffley, M. (2005). Explaining the great racial divide: Perceptions of fairness in the US criminal justice system. *The Journal of Politics*, *67*(3), 762–783. http://dx.doi.org/10.1111/j.1468-2508.2005.00338.x.

Jackson, B., Kubzansky, L. D., & Wright, R. J. (2006). Linking perceived unfairness to physical health: The perceived unfairness model. *Review of General Psychology*, *10*(1), 21–40. http://dx.doi.org/10.1037/1089-2680.10.1.21.

Jain, R. K., Chiu, D. M. W., Hawe, W. R., et al. (1984). A quantitative measure of fairness and discrimination. *Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA*, *21*, 1. http://dx.doi.org/10.48550/arXiv.cs/9809099.

Jamison, D. T. (2018). Disease control priorities: improving health and reducing poverty. *The Lancet*, *391*(10125), e11–e14. http://dx.doi.org/10.1016/S0140-6736(15)60097-6.

Kahneman, D., Knetsch, J. L., Thaler, R., et al. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review*, *76*(4), 728–741. http://dx.doi.org/10.1515/9781400829118-011.

Krings, A., & Schusler, T. M. (2020). Equity in sustainable development: Community responses to environmental gentrification. *International Journal of Social Welfare*, *29*(4), 321–334. http://dx.doi.org/10.1111/ijsw.12425.

Kuppler, M., Kern, C., Bach, R. L., & Kreuter, F. (2022). From fair predictions to just decisions? conceptualizing algorithmic fairness and distributive justice in the context of data-driven decision-making. *Frontiers in Sociology*, *7*, Article 883999. http://dx.doi.org/10.3389/fsoc.2022.883999/full.

Larsson, O. L. (2022). Technocracy, governmentality, and post-structuralism. In *Technocracy and the epistemology of human behavior* (pp. 103–123). Routledge, ISBN: 1-003-32838-5, http://dx.doi.org/10.4324/9781003328384.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. http://dx.doi.org/10.1038/nature14539.

Lee, M. S. A., & Floridi, L. (2021). Algorithmic fairness in mortgage lending: from absolute conditions to relational trade-offs. *Minds and Machines*, *31*(1), 165–191. http://dx.doi.org/10.1007/s11023-020-09529-4.

Lind, E. A., & Tyler, T. R. (2013). *The social psychology of procedural justice*. Springer Science & Business Media, ISBN: 0-306-42726-5, http://dx.doi.org/10.1007/978-1-4899-2115-4.

Loland, S. (2010). Fairness in sport. *The Ethics of Sports. A Reader*, 116–124. http://dx.doi.org/10.4324/9780367766924-RESS179-1.

Lum, K., & Isaac, W. (2016). To predict and serve? *Significance*, *13*(5), 14–19. http://dx.doi.org/10.1111/j.1740-9713.2016.00960.x.

Martens, K. (2016). *Transport justice: Designing fair transportation systems*. Routledge, ISBN: 0-415-63832-1.

Mill, J. S. (2016). Utilitarianism. In *Seven masterpieces of philosophy* (pp. 329–375). Routledge, http://dx.doi.org/10.4324/9781315508818-7.

Mohai, P., Pellow, D., & Roberts, J. T. (2009). Environmental justice. *Annual Review of Environment and Resources*, *34*(1), 405–430. http://dx.doi.org/10.1146/annurev-environ-082508-094348.

Moulin, H. (2004). *Fair division and collective welfare*. MIT Press, ISBN: 978-0262633116.

Neff, G., & Nagy, P. (2016). Talking to bots: Symbiotic agency and the case of tay. *International Journal of Communication*, *10*, 4915–4931.

Nozick, R. (1974). *Anarchy, state, and utopia*. John Wiley & Sons, ISBN: 978-0-631-19780-5.

Nussbaum, M. C. (2011). *Creating capabilities: The human development approach*. Harvard University Press, ISBN: 978-0-674-07235-0, http://dx.doi.org/10.2307/j.ctt2jbt31, 10.2307/j.ctt2jbt31.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press, ISBN: 978-0511807763, http://dx.doi.org/10.1017/CBO9780511807763.

Ostrom, E., Gardner, R., & Walker, J. (1994). *Rules, games, and common-pool resources*. University of Michigan Press, ISBN: 0-472-09546-3.

Palma, J. G. (2011). Homogeneous middles vs. heterogeneous tails, and the end of the 'inverted-U': It's all about the share of the rich. *Development and Change*, *42*(1), 87–153. http://dx.doi.org/10.1111/j.1467-7660.2011.01694.x.

Pereira, R. H., Schwanen, T., & Banister, D. (2017). Distributive justice and equity in transportation. *Transport Reviews*, *37*(2), 170–191. http://dx.doi.org/10.1080/01441647.2016.1257660.

Pessach, D., & Shmueli, E. (2023). Algorithmic fairness. In *Machine learning for data science handbook: data mining and knowledge discovery handbook* (pp. 867–886). Springer, http://dx.doi.org/10.1007/978-3-031-24628-9_37.

Rawls, J. (1971). *A theory of justice*. Harvard University Press, http://dx.doi.org/10.2307/j.ctvjf9z6v.

Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review*, *94*.

Riehl, K., Kouvelas, A., & Makridis, M. A. (2024a). Karma economies for sustainable urban mobility–a fair approach to public good value pricing. *Npj Sustainable Mobility and Transport*, *1*(1), 14. http://dx.doi.org/10.1038/s44333-024-00014-4.

Riehl, K., Kouvelas, A., & Makridis, M. A. (2024b). Resource allocation with karma mechanisms—A review. *Economies*, *12*(8), 211. http://dx.doi.org/10.3390/economies12080211.

Scantamburlo, T., Baumann, J., & Heitz, C. (2025). On prediction-modelers and decision-makers: why fairness requires more than a fair prediction model. *Ai & Society*, *40*(2), 353–369. http://dx.doi.org/10.1007/s00146-024-01886-3.

Scheffler, S. (2017). What is egalitarianism? In *John rawls* (pp. 309–344). Routledge, http://dx.doi.org/10.4324/9781315251431-12.

Schino, G., & Aureli, F. (2009). Reciprocal altruism in primates: partner choice, cognition, and emotions. *Advances in the Study of Behavior*, *39*, 45–69. http://dx.doi.org/10.1016/S0065-3454(09)39002-6.

Sedgewick, R., & Wayne, K. (2011). *Algorithms*. Addison-wesley professional, ISBN: 978-0-321-57351-3.

Sen, A. (2008). The idea of justice. *Journal of Human Development*, *9*(3), 331–342. http://dx.doi.org/10.1080/14649880802236540.

Shields, L. (2012). The prospects for sufficientarianism. *Utilitas*, *24*(1), 101–117. http://dx.doi.org/10.1017/S0953820811000392.

Shinde, P. P., & Shah, S. (2018). A review of machine learning and deep learning applications. In *2018 fourth international conference on computing communication control and automation* (pp. 1–6). IEEE, http://dx.doi.org/10.1109/ICCUBEA.2018.8697857.

Simons, T., & Roberson, Q. (2003). Why managers should care about fairness: The effects of aggregate justice perceptions on organizational outcomes. *Journal of Applied Psychology*, *88*(3), 432. http://dx.doi.org/10.1037/0021-9010.88.3.432.

Smith, A. (1776). *The wealth of nations*. Self published manuscript.

Speicher, T., Heidari, H., Grgic-Hlaca, N., Gummadi, K. P., Singla, A., Weller, A., & Zafar, M. B. (2018). A unified approach to quantifying algorithmic unfairness: Measuring individual & group unfairness via inequality indices. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2239–2248). http://dx.doi.org/10.1145/3219819.3220046.

Sunshine, J., & Tyler, T. R. (2003). The role of procedural justice and legitimacy in shaping public support for policing. *Law & Society Review*, *37*(3), 513–547. http://dx.doi.org/10.1111/1540-5893.3703002.

Tharp, R. (2018). *Teaching transformed: Achieving excellence, fairness, inclusion, and harmony*. Routledge, http://dx.doi.org/10.4324/9780429496943.

Theil, H. (1965). The information approach to demand analysis. *Econometrica*, *33*(1), 67–87. http://dx.doi.org/10.2307/1911889.

Von Hoffman, A. (2000). A study in contradictions: The origins and legacy of the housing act of 1949. *Housing Policy Debate*, *11*(2), 299–326. http://dx.doi.org/10.1080/10511482.2000.9521370.

Walzer, M. (1983). *Spheres of justice: a defense of pluralism and equality*. New York: Basic Books, ISBN: 0-465-08189-4.

Wang, X., Wu, Y. C., Ji, X., & Fu, H. (2024). Algorithmic discrimination: examining its types and regulatory measures with emphasis on US legal practices. *Frontiers in Artificial Intelligence*, *7*, Article 1320277. http://dx.doi.org/10.3389/frai.2024.1320277.

Wolf, U. (2002). *Aristoteles"" Nikomachische Ethik''*. Wissenschaftliche Buchgesellschaft Darmstadt, ISBN: 978-3-499-55651-7.

Xu, M., David, J. M., Kim, S. H., et al. (2018). The fourth industrial revolution: Opportunities and challenges. *International Journal of Financial Research*, *9*(2), 90–95. http://dx.doi.org/10.5430/ijfr.v9n2p.

Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., & Zhu, J. (2019). Explainable AI: A brief survey on history, research areas, approaches and challenges. *Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part II 8*, 563–574. http://dx.doi.org/10.1007/978-3-030-32236-6_51.

Zhang, J., Bao, K., Zhang, Y., Wang, W., Feng, F., & He, X. (2023). Is chatgpt fair for recommendation? evaluating fairness in large language model recommendation. In *Proceedings of the 17th ACM conference on recommender systems* (pp. 993–999). http://dx.doi.org/10.1145/3604915.3608860.