

Entitlement Theory and Algorithmic Fairness Measures

Edward Speer
California Institute of Technology
espeer@caltech.edu

Llama-2-7b
Meta Platforms, Inc.

January, 2025

Abstract

This paper explores the relationship between entitlement justice and measures of algorithmic fairness. ...

1 Introduction

The rise of algorithmic decision making in the public sector has caused significant public concern. As algorithms increasingly make decisions that affect individuals' lives, from determining creditworthiness to predicting criminal recidivism, the public has grown fearful of their potential to perpetuate existing social inequalities. A 2018 study showed that 58% of Americans believe that algorithms will always have some level of bias [Smith \(2018\)](#), and as we know from the famed COMPAS case, these fears are not unfounded [Angwin et al. \(2016\)](#).

In response to these concerns, researchers have developed two broad and increasingly vast bodies of work. The first, which we will refer to as *algorithmic fairness*, focuses on developing statistical and computational tools to ensure that algorithms do not discriminate against protected groups. The second, referred to as *algorithmic accountability*, focuses on explainability and interpretability — developing tools to help users understand and interpret the decisions made by algorithms. The former area of research is what we will focus on in this paper.

The field of algorithmic fairness is often conceptualized as the application of the philosophical notion of distributive justice to algorithmic decision making. At first glance, this seems like a natural fit. The goal of distributive justice is to ensure that the allocation of the benefits and burdens of society are distributed fairly among its members, and the goal of algorithmic fairness measures are to ensure that the allocation of decisions by algorithms complies with some notion of fairness. However, recent work questions this analogy [Hertweck et al. \(2024\)](#), analytically showing that the extent to which algorithmic fairness measures can be seen as a form of distributive justice is quite limited, and isolated to egalitarian concepts of justice [Kuppler et al. \(2021\)](#).

In this paper, we propose a new direction for research that incorporates a previously overlooked distributive justice concept: entitlement justice. Entitlement theory, which roots justice in the idea of respecting individuals' property rights, offers a more nuanced and context-sensitive understanding of algorithmic fairness. We argue that by incorporating entitlement justice into the

design of algorithmic fairness measures, we can create a more robust framework for evaluating algorithmic decisions. When this framework is applied to the broader sociotechnical systems in which algorithms are embedded, we can better understand the social implications of algorithmic decision making and develop more effective strategies for mitigating their negative effects.

The rest of this paper is organized as follows. In Section 2, we provide an overview of the existing literature on algorithmic fairness and distributive justice. We draw on the formalism from Kuppler et al. (2021) and Corbett-Davies et al. (2023) to create a unified model for understanding algorithmic fairness and distributive justice consistently with each other. In Section 3, we introduce the concept of entitlement justice and discuss its historical development. We confront the traditional objections to entitlement theory and show how they can be overcome in the context of algorithmic decision making. In Section 4, we propose a new framework for understanding algorithmic fairness through the lens of entitlement justice. We analyze the implications of this framework for existing algorithmic fairness measures and show an example of how it can be applied to a real-world case study. Finally, in Section 5, we conclude with a discussion of the broader implications of our work and suggest directions for future research.

2 Background

3 Entitlement Justice

4 Entitlement Fairness

5 Conclusion

Aknowledgements

References

- Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine bias: There’s software used across the country to predict future criminals. and it’s biased against blacks. *ProPublica*, May 2016. URL <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Accessed: 2025-01-12.
- Sam Corbett-Davies, Johann D. Gaebler, Hamed Nilforoshan, Ravi Shroff, and Sharad Goel. The measure and mismeasure of fairness, 2023. URL <https://arxiv.org/abs/1808.00023>.
- Corinna Hertweck, Christoph Heitz, and Michele Loi. What’s distributive justice got to do with it? rethinking algorithmic fairness from the perspective of approximate justice, 2024. URL <https://arxiv.org/abs/2407.12488>.
- Matthias Kuppler, Christoph Kern, Ruben L. Bach, and Frauke Kreuter. Distributive justice and fairness metrics in automated decision-making: How much overlap is there?, 2021. URL <https://arxiv.org/abs/2105.01441>.

Aaron Smith. Public attitudes toward computer algorithms, November 2018. URL <https://www.pewresearch.org/internet/2018/11/16/public-attitudes-toward-computer-algorithms/>. Accessed: 2025-01-12.