Title
Negotiated Fairness: A Multidimensional Framework for Fair Division Beyond Equal Share

ISBN
9798297646070

Author
Liang, Jia-Wei

Publication Date
2025-09-21

Peer reviewed|Thesis/dissertation

Negotiated Fairness:
A Multidimensional Framework for Fair Division Beyond Equal Share

By

JIA-WEI LIANG
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Computer Science

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Hao-Chuan Wang, Chair

Joshua Mccoy

Richard Huskey

Committee in Charge

2025

Abstract

Fair division problems, whether dividing a cake, splitting rent among roommates, or assigning tasks in a team, require participants to agree on what constitutes a "fair share". However, in the presence of self-interest, no individual can perceive absolute fairness or fully align with the expectations of others. This thesis explores the concept of *negotiated fairness*: a dynamic and multidimensional understanding of fairness that allows for differing objectives among participants, while still allowing satisfactory outcomes through communication, transparency, and mutual adjustment.

While classical fair division literature originates in economics, this thesis approaches the problem through an interdisciplinary lens, drawing on mathematics, cognitive science, and human-computer interaction to address three interrelated dimensions of fairness:

1. Material fairness, rooted in legal, logical, and algorithmic systems, provides formal and objective procedures for division. However, such mechanisms often overlook the subjective experience of fairness. Chapter 3 addresses this through the aspects of game-theoretic rules applied to political districting, highlighting both the potential and the limitations of procedural fairness.

2. Cognitive fairness examines how individuals perceive fairness based on their values, biases, and context. Even when formal fairness is achieved, individuals may still feel unsatisfied. Chapter 4 investigates this through experimental studies using EEG and psychological methods to analyze how humans evaluate fairness at a neural and behavioral level.

3. Social fairness, often overlooked, emerges from group negotiation and shared norms. It is the domain where fairness is collectively constructed, and tension between formal procedures and individual perceptions is reconciled. Chapter 5 proposes a transparent and interpretable dashboard system to support fair workload distribution, enabling group-level moderation and feedback.

This thesis contends that addressing only one dimension of fairness is insufficient. By integrating material, cognitive, and social fairness into a unified framework, it offers a new approach to designing division systems that are not only theoretically sound, but also practically effective and socially acceptable. The originality of this work lies in its multi-perspective methodology, combining algorithmic logic, empirical psychology, and interactive system design. It challenges the reductive view that fairness is equality, and provides actionable tools and design principles for more robust and adaptive fair division in real-world contexts.

Contents

List of Figures

List of Tables

Acknowledgments

This thesis would not be possible without the tremendous support of many people around me. Their invaluable contributions, encouragement, and guidance have been the cornerstone of my Ph.D. journey. With deep gratitude, I dedicate this work to each of them.

- My advisor, Prof. Hao-Chuan Wang, who generously welcomed me into his lab after my previous advisor retired. His patience, encouragement, and mentorship guided me into the field of Human-Computer Interaction. His support and guidance helped shape me into the researcher that I am today.
- My previous advisor, Prof. Nina Amenta, who introduced me to the life of a Ph.D. researcher and helped me explore topics beyond traditional computer science. Her guidance opened the door to my fair division research, which ultimately became the focus of this dissertation.
- Members of the dissertation committee: Prof. Richard Huskey and Prof. Joshua McCoy, for their willingness to serve on my committee and for their invaluable feedback throughout this process. Their insights and support were given selflessly and had a lasting impact on my work.
- My collaborators: Ziyu Zhao, Ph.D. student in the Communication Department, who assisted with data collection for the EEG study, and Yi-Hsuan Chiang, currently an M.S. student at Cornell, for her support of collecting user data during her internship at our lab.
- My family: My father, once a Ph.D. student as well, shared his experience and wisdom throughout my journey. My mother provided unwavering emotional support. My elder sister Tina and my younger brother Jack continuously encouraged and believed in me, and their admiration kept me going.
- My lab mates and friends: Jingxian, Jingchao, Xiaoyang and Meredith for welcoming me when I joined the lab and for your guidance. Also, to Claire, Manami, and Anna, whose presence and companionship lightened many long days of research.
- My cat: Caramel, also my steadfast companion, whose quiet presence and unconditional love sustained me through my Ph.D. years.
- Lastly, I am grateful to all participants who participated in my studies and interviews. Your time, thoughts, and contributions made this research possible.

Chapter 1

Introduction

1.1 Motivation

The question of "how to divide" has long intrigued researchers and remains an active area of inquiry. It is a practical topic that we encounter in everyday life and *fairness* is one of the most critical aspects. For example, the simple task of dividing a cake fairly. The cake cutting task appears straightforward; however, it encapsulates deeper challenges in defining and achieving fairness in division, and that is why the topic of fair division, despite its familiarity, does not have a universally accepted solution.

The problem of fair division is accompanied by a number of nontrivial difficulties. For one, there is no single objective definition of fairness, as fairness is inherently a subjective experience and can vary from person to person. For another, the complication arises from the nature of the division. In some situations, it is not only the quantity that matters, but also the quality. For example, a cake might have different toppings or compositions (i.e., what we want to divide can have multiple attributes), making some slices more desirable than others. Beyond such tangible goods, the challenge extends to less easily divisible resources, such as tasks or labor.

Mathematicians have proposed different solutions to the cake cutting problem [3, 147], and one of the most conceptually appealing solution to me is the "divide and choose" protocol [122]. In this protocol, one person divides the cake into two pieces, and the other person selects their preferred piece. This ensures a fair division because the cutter is incentivized to split the cake as evenly as possible to avoid ending up with the less desirable portion, while the chooser is satisfied by having the first choice, thereby preventing envy. What makes this method particularly interesting is that, despite the absence of precise mathematical measurements or complex calculations, a simple procedural rule can lead to an outcome that is perceived as fair to both parties.

Another true story, documented in a decentralized fair division study by Pratt and Zeckhauser [120] shares the similar idea: *When Mary Anna Lee Paine Winsor passed away, her estate included two trunks of silver that had to be divided among her eight grandchildren. Despite avoiding the typical approach of converting heirlooms into impersonal monetary values, a simple auction-inspired mechanism method, produced a result that satisfied all parties.* Each grandchild was given an equal endowment to "bid" on items using probabilistic shares. Their stated preferences served as bids in a second-price auction, where the highest bidder won an item but paid the price of the second highest bid [4]. Although the result did not strictly conform to formal fairness criteria as defined by economists or theorists, all participants felt that the result was fair.

We often think that evenly dividing a resource among n individuals is considered fair; however, in reality, fairness is not always so straightforward. Equal division does not account for individual preferences, values, or the nature of the resource being divided. Moreover, beyond mathematical formulations, there are many human-developed rules and procedures that attempt to address fairness in more practical or context-sensitive ways.

The above raises some crucial questions for me: What does fairness actually mean in practice? Who defines fairness and how can we measure it when perceptions vary between individuals or between majority and minority groups? Is it more important to achieve mathematically defined fairness or to ensure that participants feel satisfied, even with outcomes that might be "unfair" by analytical standards? Are there other methods to enhance human perceived fairness?

I conclude the "fair division problem" into four categories shown in Table 1.1. These categories are based on two key dimensions: (1) whether the resource being divided is desirable (where more is preferred) or undesirable (where less is preferred) and (2) whether the quantity of the resource is the primary concern or whether its quality or variety also plays a significant role. This yields the following four types of fair division problems:

(1) Desirable resource, quantity matters: for example, district cutting, where political actors seek to maximize the number of districts, as more districts often translate into more political power.

(2) Desirable resource, quality/variety matters: for example, the cake cutting problem discussed

earlier, where not all pieces are equal due to differences in toppings or composition.

(3) Undesirable resource, quantity matters: for example, rent division, where individuals aim to minimize their share of the rent payment.

(4) Undesirable resource, quality/variety matters: for example, workload division, where tasks may vary in difficulty or undesirability, and the goal is to minimize the burden, even though the work cannot be divided evenly.

Despite differences in context, the four categories of fair division problems share underlying conceptual structures, allowing for common solution strategies to be applied across them. In this thesis, my focus was more on *desirable resource, quantity matters* and *undesirable resource, quality/variety matters* since these categories are less studied. The thesis proposes a *negotiation-based* framework for fair division that addresses multiple contexts: the complex nationwide redistricting problem, the neuroscientific and psychological exploration of perceived fairness, and group-level workload division. By combining algorithmic models, empirical studies, and interface design, this research aims to bridge the gap between formal fairness and social satisfaction, offering a more nuanced and practical understanding of what it means to divide fairly.

Types of problems	**Same kind of value (quantity matters)**	**Different kinds of value (quality/variety matters)**
Desirable (more is better)	(1) District cutting	(2) Cake cutting
Undesirable (less is better)	(3) Rent division	(4) Work division

Table 1.1: Four types of common fair division problems. Comparison of "same kind of value" and "different kinds of value" across "desirable" and "undesirable".

1.2 Contributions

The traditional notion of fairness, *the outcome fairness*, focuses on studying the result of joint decisions, considering if the distributions of the results are analytically fair or equitable (a more detailed definition of different dimensions of fairness can be found in Section 2.2). However, people started to concern not only about the fairness of the outcomes but also the fairness of the process that determines these outcomes. This is how the notion of *procedural fairness* was brought up. Procedural fairness speaks to the idea of fair processes, and how people's perception of fairness in the procedure (e.g., transparency in information and decision making) is strongly impacted by the quality of their participation experiences and not the end result of these experiences [149]. When procedural fairness is perceived high, people tend to respond more positively to outcome fairness [14, 15].

Another type of fairness, mentioned less but also important, is informational fairness (or interaction justice). It means an individual's expectation on receiving adequate information on the explanation of the process and its outcomes [135]. In a group, it is measured by whether individuals received sufficient information to access a given system.

In this thesis, I build on the insight that no single perspective is sufficient to address the complexity of fair division. I propose a framework that draws from both algorithmic and human-centric traditions to address three critical dimensions of fairness:

- Outcome fairness: ensuring that the results of division are just and acceptable;
- Procedural fairness: designing fair, transparent, and inclusive processes for decision-making;
- Informational fairness: ensuring that participants understand the reasoning behind outcomes and have access to the information necessary to trust the system.

In the following chapters, I present three studies addressing the overarching theme of negotiating fair division. Each chapter explores a different dimension of fairness through distinct methodological lenses and application domains.

Chapter 3: Developing an Algorithm for Fair Redistricting : This chapter focuses on the problem of fair district partitioning. It adopts an algorithmic approach to model and evaluate redistricting fairness, emphasizing outcome fairness through computational mechanisms and game-based simulations.

Chapter 4: Utilizing EEG to Understand Human Perceived Fairness and Social Reciprocity: This chapter uses EEG (electroencephalography) to investigate the subjective perception of fairness. It focuses on procedural fairness, examining how neural responses are shaped by affirmative feedback and reciprocal social behavior.

Chapter 5: Designing a Fairness Work Division Dashboard to Enhance Group Collaboration: This chapter addresses the challenge of workload division in small groups. It takes a human-centered approach, centering on informational fairness by designing an interpretable and transparent dashboard to support equitable task allocation and group negotiation. Finally, Chapter 6 outlines potential directions for future work, including how the proposed multi-dimensional fairness framework could be extended and generalized to broader contexts.

By integrating methods from computer science, cognitive psychology, and human-computer interaction, I show how a more holistic and negotiable model of fairness can be achieved, one that is theoretically grounded, practically usable, and socially acceptable.

1.3 Disclosure

Figure 1.1 and Table 1.2 illustrate the conceptual flow of my approach to solving the fair division problem. It also highlights the conference presentations and publications included in this dissertation. For transparency, this dissertation incorporates first-authored work that has already been published, as well as work that is currently under preparation or planned for future publication.

Type of Fairness	**Description and Chapter Contribution**
Material Fairness	Material fairness, rooted in legal, logical, and algorithmic systems, provides formal and objective procedures for division. However, such mechanisms often overlook the subjective experience of fairness. **arXiv preprint:** Liang, Jia-Wei, and Nina Amenta. *"The Fairness of Redistricting Ghost."* arXiv:2401.07440 (2024) [89].
Cognitive Fairness	Cognitive fairness examines how individuals perceive fairness based on their subjective values, biases, and context. Even when one's fairness criteria is not achieved, individuals may still feel satisfied. **Status:** Planning to submit to CHI 2026 in September 2025.
Social Fairness	Social fairness, often overlooked, emerges from group negotiation and shared norms. It is the domain where fairness is collectively constructed, and tensions between formal procedures and individual perceptions are reconciled. **Publication:** Liang, Jia-Wei, and Hao-Chuan Wang. *"Is It Fair Enough? Supporting Equitable Group Work Assignment with Work Division Dashboard."* In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 2025 [91]. **Publication:** Liang, Jia-Wei, and Hao-Chuan Wang. "Reshaping Group Life: A Transparent and Interpretable Reward Model to Enhance Fairness in Groups." Work-in-Progress Paper. International Conference on Collaboration Technologies and Social Computing. Springer Nature Switzerland, 2023 [90].

Table 1.2: Summary of fairness types, dissertation chapters, and related publications. More detailed definition of each fairness type can be found in Section 2.1.

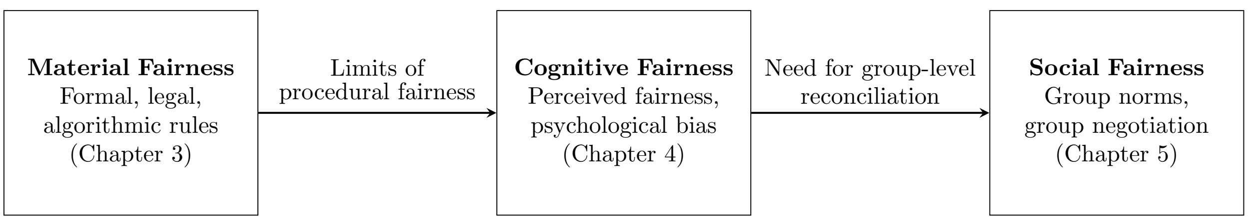


Figure 1.1: Idea flow of fairness concepts in this dissertation.

Chapter 2

Background

2.1 From Material Fairness to Cognitive Fairness to Social Fairness

2.1.1 Material Fairness: Formal, Legal, Algorithmic Rules

The first level of fairness that we normally encounter is *the material fairness.* Material fairness centers on tangible, observable outcomes and emphasizes the need for decisions and actions to be grounded in thorough, unbiased evaluations of all relevant factors. This form of fairness seeks outcomes that are objectively justifiable and equitable for all stakeholders. It is often embedded within legal, logical, and algorithmic systems, where formal and objective procedures guide the distribution of resources and responsibilities [128].

For example, the frequently debated issue in political geography: redistricting. Federal law mandates that the (material) fairness of cutting a district is that districts must be equal in population. It also requires that redistricting efforts do not disenfranchise protected ethnic groups [87]. In addition, there are some commonly adopted redistricting criteria, including preserving community integrity, ensuring population equality, and maintaining geographic compactness [109]. These principles aim to prevent gerrymandering, the deliberate manipulation of district boundaries to favor a particular political party.

Several approaches have been proposed to achieve material fairness in redistricting. One such approach is the concept of partisan symmetry, which holds that redistricting should treat major political parties equally in how votes are translated into seats. A commonly used metric for assessing partisan symmetry is partisan bias, which simulates a hypothetical tied election by adjusting each party's actual vote share relative to a 50% baseline [99]. This method calculates how many more (or fewer) seats a party would receive under a perfectly tied vote. However, a key limitation of partisan bias is that it relies on a theoretical election scenario rather than actual results. To address this issue, the efficiency gap has been introduced as an alternative metric. It measures the difference in wasted votes, and votes that do not contribute to a party's victory between the two major parties, offering a more grounded indicator of partisan advantage [144].

Beyond domestic settings, the principles of fair division extend to fields such as machine learning, where fairness models aim to mitigate data bias, and computer systems, where fair allocation of CPU resources ensures balanced performance in tasks [53]. In all domains, the need for efficient and perceived fair allocation of resources is critical to avoid conflict, inefficiency, and dissatisfaction.

However, such mechanisms often overlook the subjective experience of fairness. For example, experimental work on everyday scenarios, such as rent division, shows that even simple problems can be fraught with complexity. Individuals value different aspects of a room, such as natural light, proximity to the bathroom, or layout, so the division by square footage is often inadequate. Without a structured framework, such negotiations can lead to resentment or perceived injustice [131]. The second level of fairness, *the cognitive fairness*, has shown more importance.

2.1.2 Cognitive Fairness: Perceived Fairness, Psychological Bias

Cognitive fairness examines how individuals perceive fairness based on their values, biases, and context. Even when material fairness is achieved, people may still feel unsatisfied, for example, the classic debate on equality and equity [148]. Equality promotes fairness by reducing discrimination, and equity goes further by acknowledging and addressing the diverse needs from individuals, sometimes even one's differences [77].

Fairness is a subjective experience, and previous research has explored the use of EEG (electroencephalography) that measures the electrical activity of the brain to investigate how individuals perceive fairness. Many of these studies that were designed to measure changes in brain waves during experiments have employed economic games that provide experimental control over objective measures of relative rewards, making them particularly effective for studying fairness in social decision making. One of the most commonly used games is the Ultimatum Game (UG). In the UG, two players are involved, a proposer and a responder. The proposer divides an endowment and offers a portion to the responder, who must then decide whether to accept or reject the offer. If the offer is rejected, both players receive nothing. UG creates a conflict between economic self-interest (maximizing one's own reward) and social motivations such as reciprocity and aversion to inequality. It is a widely used behavioral paradigm in the investigation of fairness decision-making because it effectively measures the adherence of individuals to fairness norms.

Behavioral studies consistently show that individuals are strongly motivated by fairness and reciprocity and are often willing to reward or punish others at considerable personal cost [164]. EEG studies, or neurophysiological evidence, typically focus on medial frontal negativity (MFN), a relatively early negative potential observed at frontocentral sites. The MFN component shows more pronounced negative amplitudes in response to unfair offers compared to fair ones [10]. Other event-related potentials (ERPs) have also been used to investigate the temporal dynamics of brain activity during fairness decision making. For example, feedback-related negativity (FRN) has been shown to be more pronounced in response to negative feedback, indicating unfavorable outcomes [104]. Additionally, P3 (P300) is closely related to the motivational significance of stimuli, with previous studies reporting larger P300 amplitudes elicited by fair offers compared to unfair ones in the context of UG, suggesting its sensitivity to reward processing. Finally, oscillatory brain activity in the delta, theta, and beta bands has also been associated with fairness decision making, further contributing to our understanding of the neural processes underlying fairness judgments [157].

There are also variations of UG designed to explore specific manipulations. For example, researchers may reveal both the outcome and the actions of the other player that led to the outcome, allowing participants to infer the intentions of the proposer. Research has shown that when the intentions of the proposer are perceived as positive, the rejection rate decreases significantly, particularly when the offer is objectively unfair. This suggests that perceiving good intentions can alleviate the social discomfort caused by unsatisfactory outcomes [57]. In another variation, the traditional reward distribution mechanism is modified by adjusting opportunity equity, ensuring that expected payoffs for offers at different fairness levels remain equal, regardless of the number of opportunities presented. This study concludes that neurophysiological evidence supports the idea that people care about both outcome and opportunity equity, with both processes sharing a late-stage integration phase in the brain [88]. However, a limitation of these UG studies is that unfair distributions are often predetermined (e.g., a 2:8 split) based on relative fairness, meaning we can only know the relative response from the participants' perspective, and it is mostly focusing on the financial decision-making driven by fairness norms. In addition, individual differences, such as personal background, should be taken into account [54]. Some studies have introduced a third-party condition to examine violations of the fairness norm beyond the perspective of the first person. One such study investigated the role of individual differences, such as the responder's financial situation, during altruistic punishment decisions. It found that more affluent participants experienced stronger violations of expectations and that FRN amplitudes were more pronounced in those with lower empathy scores. Furthermore, a higher positive affect was associated with more punishment behavior, suggesting that positive emotions play a significant role in the restoration of violating fairness norms [104]. In another study, the presence of an audience, or social observation, supports negative reciprocity in UG, observation increases rejection rates for unfair offers [113].

2.1.3 Social Fairness: Group Norms, Group Negotiation

Cognitive fairness emphasizes individuals' perceptions of fairness, for example, how people personally interpret whether they have been treated justly. However, when fairness must be reconciled at the group level, social fairness becomes essential. Often overlooked, social fairness emerges through group negotiations, shared norms, and collective understanding. It is the domain in which fairness is socially constructed, serving to bridge the gap between formal procedures and individual perceptions by fostering consensus and mutual legitimacy.

The *Fairness Heuristic Theory* indicates that upon joining a social entity, people face a tension between whether to contribute or to withhold personal resources [123]. The decision was made by determining whether a social entity can be trusted, and individuals usually rely on previously formed *fairness*. This is why fairness is important in the group work setting.

The input-process-output model of group work [100] is one of the methods to conceptualize the impact of fairness in a group. It identifies the inputs, outputs, and required processing tasks to transform inputs into outputs. It shows that the success of a group (its outcome) depends upon inputs or resources with which the group has to work, and the interaction among group members. Research on the psychological experience of fairness suggests that it is a basic form of social evaluation that emerges in interactions with group members, and it plays an important role in shaping not only the participation in group life, but also the well-being of individual group members [70]. Unfair treatment can lead to the feeling that one is devalued and excluded from the group [71, 153]. It can also lead to disengagement, which has the potential to negatively affect a group's outcome, including productivity and turnover [66,153]. Moreover, *fairness asymmetry* [79], which indicates that the negative effects of unfairness are substantially stronger than the positive effects of fairness, and *social comparisons* [161], which means that humans not only react to their own situation but also to others' situations with whom we use as references (group members) to compare ourselves with, enlarge these negative feelings.

In a work group, defining, communicating and grounding what is considered fair is important [130]. The relational models also suggest that people's social psychological needs are likely to be satisfied when they interact with others who behave according to the norm of fairness [152].

In the literature, several criteria have been identified to ensure group fairness. A group is considered fair if it follows the following rules: accuracy (provide accurate and valid information), bias suppression (not affected by personal bias), consistency (consistent across people and time), correctness (opportunity to reverse or modify decision), ethically (moral and ethical values), and representative (reflecting the basic concerns of the people affected) [24]. A more succinct criteria will be: independence (the ratio of the work granted to different individuals should be equal), separation (for individual who did hard work, we need to make sure they get similar scores), and sufficiency (is there any individual who did not work hard but get a good score) [23].

Fairness and cooperation are important elements for the effective functioning of groups, organizations, and the society at large. However, achieving mutual cooperation is often challenging if a fairness mechanism is not established in a group, because people recurrently face various resource allocation trade-offs between self and others: the conflict between individual rationality and collective rationality, mentioned more in section 5.2.2.1. Such trade-offs make it tempting for individuals to pursue their own personal interest, which may lead to the breakdown of cooperation [126]. In addition, since fairness is an abstract and potentially sensitive concept that can involve different perspectives and interpretations, and since fairness perceptions are transferable, we should not neglect its important role in a group and need to find ways to support the construction and negotiation of fairness norms in groups.

The distribution of workloads within a group is one of the most contested issues related to fairness. It involves allocating tasks and responsibilities among group members in a way that promotes both efficiency and the achievement of shared goals. When not managed effectively, this process can negatively affect the overall productivity of the group and diminish individual satisfaction.

How to adapt the workload distribution based on the group's dynamics, goals, and member capacities can change over time. Some studies have proposed equal distribution, task-based distribution or skill-based distribution, and so on. However, one thing that has in common is that effective communication and regular feedback are crucial to ensure that the distribution of workload is fair [153].

2.2 Perspectives on Fairness and Its Relationship to Fair Division

Is fairness important? Indeed, it is. Fairness is the metric humans often use to determine what each person deserves. However, due to individual differences in preferences, values, and circumstances, this measurement varies significantly across people, making the pursuit of fairness inherently complex. However, when fairness is effectively achieved, it yields both psychological and practical benefits, enhancing individual satisfaction, motivation, and willingness to engage. For example, in professional sports, fairness underlies the principle of competitive balance. The National Football League (NFL) strives to keep teams evenly matched because close contests are more engaging and ensure that every city has a fair chance to win. This concept, known in law and economics as competitive balance, keeps both players and fans engaged throughout the season [69].

Fairness has long been a subject of interest across disciplines, from economics and political science to psychology and computer science. Numerous attempts have been made to convert this inherently subjective notion into formal mathematical definitions for precise measurement. In the field of fair division or the allocation of resources among individuals, researchers have devised algorithms aimed at achieving unbiased results.

Four main criteria have been established as frameworks for evaluating the fairness and efficiency of different outcomes in various scenarios, and they are *Proportional, Equitable, Envy-free* and *Pareto-Optimal (or Pareto-efficiency).* Proportional indicates that each individual thinks their share is at least $1/n$ of the total; equitable indicates a subjective value that each individual thinks their share is the same as everyone else, and they are equally happy; envy-free indicates that each individual has a strategy that can guarantee himself a share that is at least as desirable as any other shares. Last but not least, pareto-optimal indicates the division where there are no other divisions that dominate the share for all individuals, or there is no other feasible agreement that would make at least one agent strictly better off while not making any of the others worse off [20].

Although these formal notions are effective in small, clearly defined settings, real-world scenarios, such as division of labor or collaborative decision-making, are far more complex. In such contexts, resources such as time, effort, and responsibility are difficult to quantify and groups often consist of individuals with diverse and conflicting interests. As a result, researchers have shifted toward human-centered approaches that account for perceived fairness, extending the analytical scope beyond outcomes to include processes and information. This shift has introduced three additional fairness constructs: *outcome fairness, procedural fairness, and information fairness.*

Outcome fairness focuses on whether the final results are perceived as just. However, as people increasingly care about how decisions are made, procedural fairness has gained prominence. It refers to the fairness of the decision-making process itself, such as transparency, consistency, and the opportunity to participate, and often has a stronger impact on satisfaction than actual outcomes [14, 15, 149]. When people feel respected and heard, they are more likely to accept decisions, even unfavorable ones.

People are influenced by procedural fairness because it addresses more symbolic and psychological concerns, such as people's needs for self-esteem, self-identity, and affiliation. In a group, decisions made in the group collaboration process can have a significant impact on individuals in at least two ways. First, fairness in the decision itself, for example, what action to take next as a group, can promote trust and cooperation among group members. Second, when individuals feel their opinions and perspectives are being considered, they are more likely to contribute to the decision-making process and support the final outcome [132].

Informational fairness, a less frequently discussed but equally critical aspect, refers to individuals' access to adequate and honest explanations about processes and decisions [135]. It builds transparency and is especially vital in automated systems. For instance, in algorithmic decision-making (ADM) models used in loan approvals, lack of transparency can lead to bias, exclusion, and mistrust [142]. Enhancing informational fairness can improve public trust and mitigate negative outcomes by clarifying how and why decisions are made [135].

Fairness, or perceived quality of treatment, becomes essential not only for maintaining social harmony but also for improving productivity, creativity, and satisfaction within teams and organizations. Its importance extends beyond formal authority structures to influence dynamics shaped by cultural diversity, behavioral expectations, and cognitive engagement [119, 151]. Ensuring fairness in these dimensions is central to making decisions that are not only optimal, but also widely accepted and respected.

2.3 Two Distinct Approaches of Current Study on Fair Division

Currently, there are two distinct approaches to understanding the problem of fair division: the algorithmic approach and the human-centric approach. The former formalizes fairness using mathematical definitions and computational models, while the latter draws on philosophical, psychological, and social science frameworks to interpret fairness in a human context and increasingly applies these ideas to algorithm design. We use the algorithmic approach to solve material fairness, and the human-centric approach to solve cognitive and social fairness.

Algorithmic fairness emphasizes that the results or decisions should not be unjust, discriminatory, or disparate. However, it is clear that not all causes of fairness or unfairness can be quantified. Algorithms are often effective in small, clearly defined settings, but struggle to scale when applied to larger, more complex social environments. For example, the classic cake-cutting algorithm, though fair in theory, only works with divisible goods and is typically limited to two participants. Moreover, it does not ensure pareto efficiency, that is, it does not guarantee that no participant can be better off without making another worse off [43].

Psychological research further reveals the gap between mathematical models and real human experience. For instance, mathematical solutions do not take into account that human have difficulties articulating what is fair than unfair since notions of unfairness are typically clearer, sharper, and more concrete than notions of fairness [162]. In addition, studies has shown that allocators have a general tendency to overestimate how much the fairness of an allocation procedure will matter to the receivers, which we called *the allocator's illusion*. This usually happens because the differential fairness of allocation procedures is much more salient before an allocation is made than it is afterward; however, the fairness allocation algorithms fails to detect these scenarios [29]. Fair division procedures should consider the psychological determinants of outcome and procedural fairness [68].

In political philosophy, John Rawls' theory of justice proposed that fairness should be rooted in pure procedural justice: a just procedure will lead to a just outcome, regardless of its content. In his cake-cutting example, the person who slices the cake (no matter how many pieces) should receive the last piece, ensuring an incentive to divide fairly. What mattered most to Rawls was an independent, agreed-upon standard and a consistent procedure to implement it [129].

However, while the human-centric approach emphasizes procedural fairness, arguing that fair processes lead to fair outcomes, it is not a panacea. Studies have shown that perfect procedural fairness does not necessarily translate into perceived fairness or satisfaction. Surprisingly, simple and imperfect procedures often produce the most satisfying results. In contrast, overly sophisticated division procedures have been shown to fail in producing outcomes perceived as fair or acceptable. In many cases, participants viewed inefficient but transparent and understandable procedures as the most fair [106].

This paradox reflects a basic human tendency: in social and economic exchanges, individuals expect rewards proportional to their contributions. This notion of equity lies at the heart of outcome fairness in the human-centered approach. Interestingly, algorithmic models such as equity theory also capture this principle, comparing the input-output ratios of individuals across reference groups, and thus offer a more comprehensive analytical tool. These parallels reveal that algorithmic and human-centric approaches are not in conflict but are complementary.

By combining the quantifiable structure of algorithmic models with the experiential sensitivity of human-centered fairness, we can better understand and improve real-world division problems. Metrics and formal models can enhance interpretability and scalability, while psychological and social theories can inform system design to better reflect how fairness is actually perceived. In the following chapters, I present studies that explore fairness through the dimensions of material, cognitive, and social perspectives, employing both algorithmic and human-centric approaches appropriate to each context, and show how are they supportive of one another.

Chapter 3

Material Fairness: Developing an Algorithm for Fair Redistricting

In this chapter, we explore the fairness of a redistricting game introduced by Mixon and Villar, which provides a two-party protocol for dividing a state into electoral districts, without the participation of an impartial independent authority. We analyze the game in an abstract setting that ignores the geographic distribution of voters and assumes that voter preferences are fixed and known. We first show that the minority party can always win at least $p-1$ districts, where p is proportional to the percentage of minority voters, and that when the minority is large they can win more than p districts. We also show that a "cracking" strategy by the majority party limits the number of districts the minority player can win as a function of the size of the minority.

3.1 Overview

In 1974 Montana was the first US state to adopt an independent redistricting commission for the House of Representatives, consisting of two Republicans, two Democrats, and an independent chair. During the most recent redistricting cycle, some of the proposed maps splitting the state into two districts concentrated the Democrats into one competitive district, while in other proposed maps both districts were reliably Republican [37]. The choice was ultimately made by the independent chair, who selected a map with two reliably Republican districts.

Was this fair? In the mathematical sciences, this kind of redistricting question has mainly been addressed by defining fairness properties, and then testing either the map, or the results of an election on the map, against these properties. Especially in the case of a small state like Montana, where rounding is important, different fairness properties lead to different answers.

One of the less-studied ways to define a fair map is to define a fair protocol by which the two parties allocate the voters to districts, and then declare that a map is fair if it is produced (or could be produced) by the protocol. Using a protocol has some advantages over using an independent commission. One is that the objective functions of the parties may be complex, so that a commission has difficulty balancing, or even knowing, all the objectives. More importantly, a small commission's exercise of authority, like the chair's choice in Montana, is itself a cause of dissension and opens the process to accusations of illegitimacy. Independent electoral redistricting commissions in practice do suffer from the appearance or existence of bias, a lack of transparency, and difficulty coming to an agreement [72, 116]. Some have failed to produce maps at all, while others have fallen into legal quagmires. Accepting a protocol as fair may be easier than accepting the judgement of a judge or commission.

While no protocol for redistricting has ever been adopted in practice, neither had independent commissions before 1974. Even if never used for redistricting, a protocol might be useful in evaluating fairness properties. If a redistricting protocol seems fair, it lends validity to properties that its redistricting solutions exhibit. Simulation of a protocol could be used to generate distributions of fair maps, which might be more realistic than the random distributions sometimes used.

Finding an acceptable measure of gerrymandering could be a legal breakthrough as well as a mathematical one. As Supreme Court Justice Kennedy wrote in his pivotal concurring opinion in *Vieth v. Jubelirer*,

> The failings of the many proposed standards for measuring the burden a gerrymander imposes on representational rights make our intervention improper. If workable standards do emerge to measure these burdens, however, courts should be prepared to order relief.

We review the mathematical literature on redistricting protocols in Section 3.2. Several of the protocols proposed so far are inspired by the "I-cut-you-choose" method for cutting a cake into two equally desirable pieces. We focus on a protocol [102] that is instead related to well-known protocols for allocating players to sports teams. The reader may recall the "captains" method from schoolyard games: two team captains are chosen (somehow), and then the captains take turns choosing players. A related mechanism is the "snake draft" used to assign players to teams in professional sports (eg. the NFL draft or Fantasy Football). This turn-taking mechanism is widely familiar, and perceived as fair, at least in the realm of sports.

The analogy between redistricting and choosing sports teams is not perfect. The two parties in redistricting construct a map with several districts, not two teams. Also, there are constraints (differing from state to state) on what constitutes a valid map. Fundamentally, all districts must contain equal numbers of voters. There are also restrictions on the connectivity and shape of the districts, constraints ensuring adequate racial representation, and goals (or soft constraints) such as alignment with municipal or other local boundaries.

Redistricting Ghost: Mixon and Villar's protocol handles many of the issues involved in defining a fair protocol. We will call it Redistricting Ghost, because, as they describe it, the game is inspired by the word game Ghost. In Ghost, two players take turns adding letters to a string, and a player loses if they are the first to spell a word. Each player has to demonstrate, after their turn, that the string they have constructed is a prefix of an English word. In Redistricting Ghost, two players A and B, representing the two parties, take turns assigning a voter to a district (instead of individual voters, the game could also be played with pre-selected equal-sized groups of voters, for example census tracts). On his or her turn, a player places any voter into any district. After their turn, the player must be able to display a complete valid map, meeting whatever legal constraints there are, that extends the current set of partially defined districts. In this way, Redistricting Ghost accommodates realistic constraints on maps and also resembles the mechanisms for picking sports teams.

Our results: Redistricting Ghost may "feel" fair, but what can we say mathematically? Mixon and Villar proved Theorem 1, below, which handles the special case of a perfectly tied electorate in a particular abstract setting. We extend the analysis in the abstract setting to the general case. We find that Redistricting Ghost is reasonably fair from the perspective of the minority party. In particular, let $p = \text{round}(jn/v)$, where v is the total number of voters, n is the number of voters for the minority party, j is the number of districts, and round() is the function that rounds up or down to the nearest integer. We show that the minority party has a strategy that can always win at least $p - 1$ districts. We also consider a "cracking" strategy for the majority party; that is, a strategy that attempts to distribute the minority voters uniformly over all the districts. This classic strategy is essential to gerrymandering, and, when the majority has complete control over redistricting, leads to the majority winning every district. As a strategy in Redistricting Ghost, however, we find that cracking is not very strong. It limits the minority to at most $p - 1$ districts only when the minority is very small. In fact, as far as we know, it might be possible for a large minority to win more than a proportional share of districts.

3.2 Different Methods of Drawing Districts

Redistricting, as an important element of representative democracy, is of great interest in the law, political science and economics, the press, and recently mathematics and computer science. Much of the mathematical literature concerns detecting or quantifying gerrymandering in a given allocation of voters to districts, including metrics such as the efficiency gap [144] and statistical analysis of sets of possible allocations. This work is most closely related to a smaller body of research into ideas for protocols by which political parties can negotiate or collaboratively determine an electoral map.

In analyzing Redistricting Ghost, Mixon and Villar considered an abstract non-geometric setting in which everyone can perfectly predict the party each voter will vote for, and there are no geographic, geometric, racial or other constraints on assigning voters to districts. In this abstract setting they proved

Theorem 1 *(Mixon and Villar) Let j, the number of districts, be even, and let both parties have the same number $n = v/2$ of voters. Then there is a strategy for the second player such that A and B both win exactly $j/2$ districts.*

We extend this analysis to cover parties of different sizes and odd numbers of districts.

They defined a "mirroring" strategy for the player, arbitrarily B, who goes second. B arbitrarily matches each district with another, its "mirror". When A plays an apple to a district, B counters by playing a brick to its mirror district, and similarly when A plays a brick B mirrors it with an apple. In the end A and B win the same number of districts. This mirroring strategy does B no good when he is the minority, however, since A could successfully take a cracking approach. To analyze to games in which the two sides are not perfectly matched, we need a different approach.

Most other research into protocols has addressed cake-cutting mechanisms for redistricting, which generalize the well-know "I cut, you choose" protocol for splitting a cake in two. Like sports drafts, cake-cutting can accommodate differing objective functions for the two players. And like Redistricting Ghost, these protocols can also incorporate map constraints.

Landau et al. [83, 84] proposed an approach in which an independent agent creates a set of nested initial cuts, and then parties A and B follow a protocol to choose one of the cuts, and to assign one side to each of the parties. Each party may then gerrymander their side as they see fit. If the protocol fails, then the sides are assigned randomly. They prove (Theorem 6.1) a *Good Choice Property*: that when the nested cuts are chosen fairly by the independent agent, both parties will achieve a result near the average of the best and worst possible results (based on their individual objective functions) of any map respecting the chosen cut. This mechanism involves an independent agent and randomization, both of which we would like to avoid.

Pegden, Procaccia and Yu cut proposed the *I-cut-you-freeze* protocol, a game in which the two parties switch roles at every turn, with each turn adding one "frozen" district to the map. For example, during her turn A extends the existing set of frozen districts to a complete map, drawing more districts as necessary, and then B chooses one of the newly drawn districts to freeze; the rest of the extended map is discarded. They give a tight bound (Theorem 2.4) on the number of districts that either player can win in the abstract setting, which shows that the number of districts won by either player is close to proportional. The majority player A does a bit better than proportional representation when the minority is small, but (as with Redistricting Ghost) B can always win at least $p-1$ districts. In addition, they show that even in the abstract setting their protocol prevents any designated protected population from being packed into a single district. A drawback is that the party that goes first has a significant advantage when the number of districts is small. In the extreme case of Montana, for example, the player that goes first draws the map.

Recently Ludden et al. [94] proposed a bisection protocol, in which A and B take turns bisecting every remaining large-enough district into two smaller districts, of equal size up to rounding. They give a symmetric optimal strategy for both players, extensive analysis using simulations comparing bisection and I-cut-you-freeze, and some analysis in a semi-geometric (graph-based) setting. In the abstract setting, while they do not completely characterize the maps produced by the bisection protocol they do prove some properties. One result (Lemma 1) is that when j is a power of two, the minority player requires an $\Omega(1/\sqrt{j})$ fraction of the voters to win one district; this suggests that the majority player has a significant advantage. Again, the player who goes first has a significant advantage when the number of districts is small; in Montana, the player who goes first again draws the map. This is not true of Redistricting Ghost, which gives both districts to the majority irrespective of which player goes first.

Tucker-Foltz [150] proposed a more theoretical game in which A is allowed to redistrict as they please, but then B picks a threshold for the election. Any district where the margin of victory does not meet the threshold is assigned randomly to either party with equal probability. He showed that in the Nash equilibria for this game the expected number of districts won by each player differs from proportional representation by at most one; the completely packed map (defined in Section 3.4) is one of the these Nash equilibria, and all equilibria require some packed majority districts, unless the number of voters is equal.

3.3 Definitions and Notation

Following [102], we define Redistricting Ghost on a state with two parties, A and B. The parties use colors (a)pple green and (b)rick red, respectively, and we'll call their voters apples and bricks. Working in the abstract setting, we assume that every voter is consistently an apple or consistently a brick, and that players A and B know which are which. We'll assume there are more apples than bricks, so that B is the minority party.

In Redistricting Ghost, A and B take turns, with B going first. At each move, a player adds a voter to any district which is not yet full. Player A can play either an apple or a brick, and similarly B.

There are j districts, each with positions for $2m+1$ voters, so that the total number of voters $v = j(2m+1)$. Let n be the total number of bricks, and let q be the number of districts won by B at the end of the game. Figure 3.1 shows an example of game play and illustrates the notation.

If a district d contains more apples than bricks, we say A is ahead in d, or that d is an "apple district", and similarly B. If d contains equal numbers of apples and bricks then d is tied (this occurs during the game but not at the end). If there are at least $m+1$ bricks d at the end of the game, we say B wins the district, and similarly A. Let r_b be the number of "remaining" bricks - those that have not yet been assigned to a district - and similarly let r_a count the remaining apples.

A *strategy for* B (or respectively A) is an algorithm describing how B should play at every turn. The strategy for B may reference how A has played in earlier turns, but it does not assume that A plays a particular strategy, and visa versa.

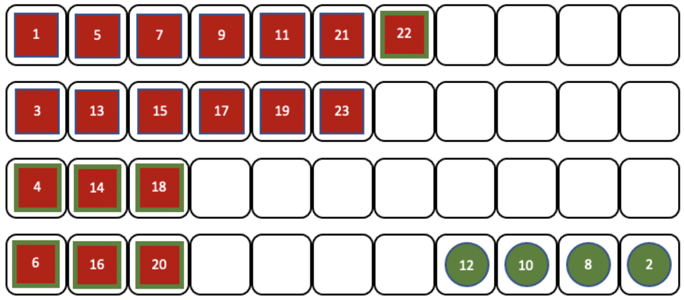


Figure 3.1: A simple example of Redistricting Ghost. An interactive version of the game can be found at `http://ballsandbins.com`. Each row represents a district; so here $j = 4$ and $m = 5$. We draw bricks as red squares and apples as green circles; a brick played by A has a green outline, and an apple played by B has a red outline. The number inside each brick or apple represents the turn at which it was played (B plays odd turns and A plays even). For visual clarity, we place bricks into districts from left to right and apples from right to left, and we sort the rows by the number of bricks they contain, and within that by the number of empty spaces. Here, we see the state of the game after the last brick has been played; at every subsequent turn an apple will be played to an empty position, and the order in which they are played will have no effect on the outcome. So n, the number of bricks, is 19, and q, the number of districts won by B, is 2.

In our analysis, we take j, q and m as given and characterize how many bricks n are necessary or sufficient for B to win q districts. In realistic situations m is orders of magnitude larger than j and q, so we will sometimes give bounds only as a function of m, ignoring additive constants that depend only on j and q, since they make little difference as $m \to \infty$. It seems to make less sense to give bounds that are asymptotic in j, the number of districts; in all the cases we know of j is small (eg. even Mexico has only 300 *distritos electorales*, which in recent cycles have been drawn, in part, by simulated annealing).

One way in which the abstract setting is misleading is that we defined it so that a district is never tied; when a district contains $m + 1$ bricks and m apples, say, we count it as won by B. In reality a nearly-even district could go either way in an election, or could even be exactly tied. We will see later that this artifact produces some skirmishing early in the game, but the overall strategies of the players, when m is large, do not seem to depend on this property. Also, while we do not prove that the results of Redistricting Ghost are indifferent to which player makes the first move, this also does not seem to be an important factor.

3.4 Comparison with Fairness Properties

After analyzing Redistricting Ghost, we will compare its results to other measures of the fairness of a map.

At the end of the game, we define $m+1$ of the bricks in a brick district to be *useful* (they were needed for B to win the district) and the rest of the voters in the district to be *wasted* (so all of the bricks in a district won by A are wasted). We similarly classify apples as useful or wasted. Since m voters in each district are wasted, the total number of wasted votes is always $v * m/(2m+1)$; but one party's votes may be wasted more than the other's. The *efficiency gap* [144] is the difference E in the number of wasted voters for each party as a fraction of the total number of voters. So $0 \leq E \leq 1/2$; and a large efficiency gap - say greater than $1/4$ - is taken as a sign of gerrymandering. In the case of Montana, a reliably Republican map has an efficiency gap near zero, while depending on the result of an election a map containing a swing district might have an efficiency gap of either zero, if the Republicans win, or $1/2$ if the Democrats win. The reliably Republican map seems more fair, then, when judged by the efficiency gap.

Recall that earlier we defined the proportional representation of each party as

$$p = \mathrm{round}(jn/v)$$

where n is the number of bricks, j is the number of districts, and v is the total number of voters. In the abstract setting, there is a deterministic assignment of voters to districts - the "packed map" - that gives each side its proportional representation, as follows. We completely fill as many districts as we can with bricks, completely fill as many districts as we can with apples, and finally construct at most one mixed district containing both apples and bricks. Then p is the number of districts B wins using this allocation. In Montana, a map containing a contested district is similar to the packed map, and thus more likely to provide proportional representation in an election; so judging by proportional representation, a map with a contested district seems more fair.

Since B is the minority party, $p \leq \lfloor j/2 \rfloor$. There is a range of n corresponding to each value of p:

Lemma 2 *For any value of p, we have*

$$p + (2p - 1)m \leq n \leq p + (2p + 1)m$$

Proof: B barely wins p districts in the packed map with $p - 1$ districts packed with bricks and $m + 1$ bricks in the single mixed district. And B will win only p districts when there are p districts packed with bricks and m bricks in the mixed district. Thus

$$(p - 1)(2m + 1) + m + 1 \leq n \leq p(2m + 1) + m$$

Rearranging gives us the bound.

3.5 Strategy for the Minority Player

Let q be the number of districts that B will be able win. Using q, we define a score at any point in the game, which will measure how well B is doing in their quest to win q districts. The score of district d is defined to be $m + 1$ if it contains at least $m + 1$ bricks, and otherwise the score is the number of bricks in d. If there is a set Q of q districts in which B is either ahead or tied, the score of Q is

$$(Q) = \sum_{d \in Q} (d)$$

The score of the game is the maximum score of any choice of Q. When there is no a set of q districts in which B is ahead or tied, we say the score is zero.

A set Q achieving the maximum score is a *maximizing* Q. At the beginning of the game, the score is zero, and, if B succeeds in winning q districts, at the end of the game the score is $q(m+1)$.

Define u to be the minimum score of any district in a maximizing Q.

Lemma 3 *The minimum score u is the same for any maximizing Q, and the number of districts in Q with $(d) = u$ is the same for any maximizing Q.*

Proof: Assume for the purpose of contradiction that some maximizing Q_1 has more districts with minimum $(d) = u$ than some other maximizing Q_2. Then we could replace some district d_1 in Q_1

with $(d) = u$ with a district d_2 from Q_2 with $(d_2) > u$, raising the score of Q_1. This contradicts the assumption that Q_1 is maximizing.

Corollary 4 *Every maximizing Q contains the same number of empty districts.*

Observation 5 *Consider any maximizing district Q. Every district d not in Q either is an apple district or has $(d) \leq u$.*

Let b be the number of moves by B that increase B's score, and let h (for "helping") be the number of moves by A that increase B's score. Both kinds of moves necessarily involve playing a brick. Let w be the number of moves by A that waste a brick, that is, A plays a brick but does not increase B's score.

Lemma 6 *Any strategy by which B can increase the score by at least one at each of his turns will allow B to win q districts if $n \geq 2(q(m+1) - h)$; since $h \geq 0$, this means that B can always win q districts if $n \geq 2q(m+1)$.*

Proof: Define a round to consist of a move by A, followed by a move by B. If the score increases at $q(m+1)$ or more turns, then B will win q districts. We have

$$b \geq w + h$$

since, in every round, B plays a brick and A either helps with a brick, wastes a brick, or plays an apple. We also have

$$n = 1 + (b-1) + h + w = b + h + w \leq 2b$$

since B must play a brick in each round, and A might play a brick in each round. Now assume we have enough bricks, as defined in the statement of the Lemma, so that

$$2b \geq n \geq 2(q(m+1) - h).$$

We simplify this to

$$b + h \geq q(m+1)$$

which implies that B has won q districts.

Next, we define a strategy for the minority player B, in Algorithm 1.

ALGORITHM 1: Strategy for the Minority Player

---

if *no bricks remain* **then**
    Play an apple to any open district;
    break;
end
Select a maxmizing Q including the fewest (possibly zero) tied districts. ;
if *Q contains a non-empty tied district d* **then**
    // Type a move
    Play a brick to d;
    break;
end
else if *Q contains at least one empty district d* **then**
    // Type b move
    Play a brick to d;
    break;
end
else
    // Q contains no tied districts
    // Type c move
    Play a brick to any district d in Q containing $\leq m + 1$ bricks.
end

Our argument that B will be able to increase his score at every round is based on:

Lemma 7 *At the beginning of a round, fix a maximizing Q, and let z be the total number of empty districts (in or outside of Q). Assume that:*

1. Q exists,

2. Q includes brick districts and empty districts, but no non-empty tied districts, and

3. The number of empty districts in Q is at most $\lfloor z/2 \rfloor$.

If B plays the strategy of Algorithm 1, these three conditions will continue to hold at the beginning of the next round.

Proof: Recall that a round is a move by A followed by a move by B. We divide A's possible moves into two categories: A might play to an occupied district, or A might play to an empty district (in or out of Q).

First, assume A plays to an occupied district d. If A plays an apple to an apple district or a brick to a brick district, the conditions still hold. If A plays an apple to a brick district d it might become tied. If $d \notin Q$, the conditions still hold. If $d \in Q$, B then makes a move of type a, restoring

Condition 2. Finally, if A plays a brick to an apple district d it might become tied. If it becomes part of every maximizing Q, again, B makes a move of type a, restoring Condition 2.

Next, we consider the case that A places a brick in an empty district d. Then B makes a move of type b or c. The number of empty districts z decreases by one, and, if Q contained any empty districts before A's move, it is replaced by d, the number of empty districts in Q goes down by one, and Condition 3 still holds.

Finally, assume A places an apple in an empty district d. If there are no empty districts in Q, then Condition 3 still holds. If $d \notin Q$, and there is an empty district d' in Q, B will play a brick to d' (a move of type b), restoring Condition 3. Finally if $d \in Q$, then d drops out of Q, but Condition 3 implies that there is at least one other empty district d' which replaces d in Q, and again B make a move of type b, restoring Condition 3.

Theorem 8 *If $n \geq 2q(m+1)$, playing the strategy in Algorithm 1 ensures that B will win at least q districts.*

Proof: We use induction on the number of rounds. At the beginning of the game, the three conditions of Lemma 7 hold. So as long as B plays using the strategy of Algorithm 1, the three conditions of Lemma 7 will continue to hold at the next round. Following the strategy, as long as there are remaining bricks, B makes a moves of type a, b or c. All of these add a brick to an existing maximizing Q, increasing the score by one. Thus B increases the score in every round, and Lemma 6 ensures that B wins q districts.

3.6 Strategy for the Majority Player

Now we want to show that the majority player A can prevent B from winning q districts when n is too small; so, we will get a lower bound on the number of bricks required for B to win q districts (assuming the given values j, m defining the redistricting problem). In particular, we will show

Theorem 9 *The minority player B can win q districts only if*

$$n \geq f(q) = 2q\left(1 - \frac{q}{j+q}\right)(m+1) - 1$$

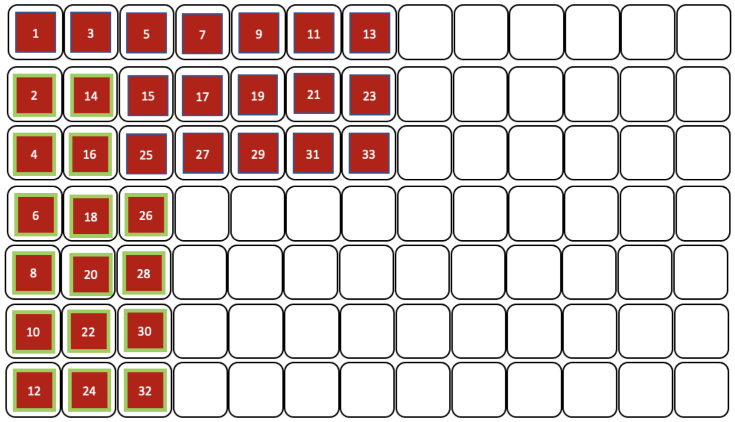


Figure 3.2: Example for the definitions in Theorem 9. Player A fills in columns with bricks, while player B may do anything (we will show two of Brick's strategies in Figure 3.2 and Figure 3.3). In this particular game, B fills in a row with bricks so as to block A from taking more of their balls, and then switches to adding bricks in any of the q rows, which has the most brick after winning a bin. Theorem 9 implies that in this case, A plays their majority strategy. Our lower bound says that if the number of bricks $n < 28$, the B cannot win three districts. Here $n = 33$, and B just barely wins three districts. For $q = 3, j = 7, m = 6$, we get $c = 2$, but A actually manages to fill 3 columns.

We will prove this by giving a strategy for A, which appears in Algorithm 2. In this strategy A uses Theorem 9 to choose the smallest value of q to which they can limit B. Using q, A plays a classic "cracking" strategy in which they ensure that at least

$$c(q) = \left\lfloor \frac{q}{j+q}(m+1) - \frac{1}{j+q} \right\rfloor$$

columns are filled with bricks at the end of the game. The Type a moves (see the algorithm) keep the first $c(q)$ columns free of apples.

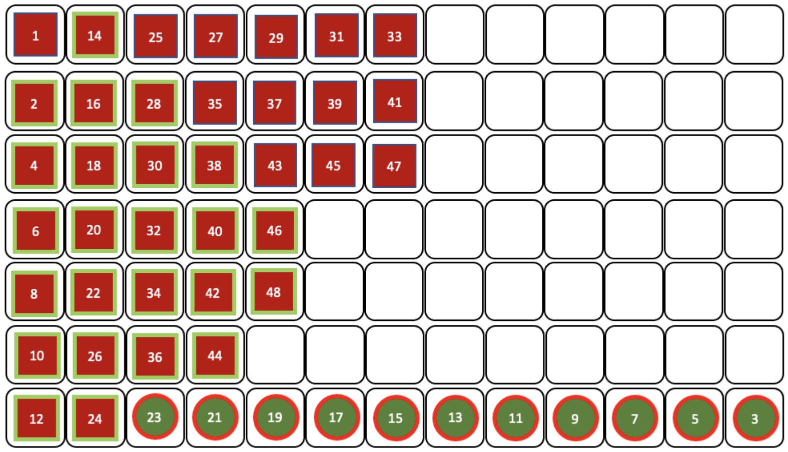


Figure 3.3: Example for the definitions in Theorem 9. Player A fills in columns with bricks, while player B may do anything. In this particular game, B fills in a row with apples so as to block A from adding the final brick to columns 3 and 4, and then switches to adding bricks in the first 3 rows. After this round, all future moves by either player add apples. Using the notation of Theorem 9, we have $j = 7, m = 6, q = 3, c = 2$, and $n = 37$. Theorem 9 implies that in this case, A plays their majority strategy, and here B tries packing apples into a district, forcing A to play into the bottom of column 2 on move 24. This strategy on B's part just makes things worse since B requires $n = 36$ to win three districts.

ALGORITHM 2: Strategy for the Majority Player

Given n, find the largest value of q such that the bound $n < f(q)$.;

if *no bricks remain* **then**

  Play an apple to any open district;

end

else if *there is any open district where the number of open spaces* $< 2(c(q) - r)$, *where* r *is the number of bricks in the district* **then**

  `// Type a move`

  Play a brick to that district ;

end

else

  `// Type b move`

  Let i be the least number of bricks in any open district;

  Play a brick to an open district containing i bricks;

end

This strategy makes no sense unless $f(q)$ is monotone in q, so that a larger q requires a larger n. Fortunately,

Observation 10 *If $p > q$, then $f(p) > f(q)$.*

It also requires the following

Observation 11 *In any round, there is at most one district to which A can make a move of type*

a.

This is because a move of type a is triggered by B placing an apple into a district, reducing the number of open spaces ω to $2(c(q)-r)-1$. The Type a move increases r, restoring $\omega = 2(c(q)-r)$. This gives us

Lemma 12 *If A does not run out of bricks, there will be at least $c(q)$ bricks in each district at the end of the game.*

Now that we have established that the strategy makes sense, let's proceed to

Proof of Theorem 9: Assume at the end of the game that B has won q districts.

The upper-left rectangle of size $q \times (m+1)$ contains only bricks. We claim that at least the first $c = c(q)$ columns also contain only bricks. So assume for the purpose of contradiction that at most $c-1$ columns are filled with bricks.

We observe that every move by A placed a brick into the first c columns; this is always true for moves of type a, and because the first c columns are not full, and cannot contain apples, it will be true of moves of type b as well.

This means that all of the bricks outside of the first c columns - at least $q(m+1-c)$ of them - must have been placed by B. To each of these moves, except possibly the last, A responded by placing a brick into the first c columns. It takes no more than j bricks to fill a column, so it must be that the number of spaces in the first c columns is strictly larger than the number of bricks they contain

$$\begin{aligned} cj &> q(m+1-c)-1 \\ c(j+q) &> q(m+1)+1 \\ c &> \frac{q(m+1)}{j+q} - \frac{1}{j+q} \end{aligned}$$

That is, c is greater than $c(q)$, and we have a contradiction, so it must be the case that the first c columns are filled with bricks.

So the total number of bricks must be

$$
\begin{aligned}
n &\geq q(m+1) + (j-q)c \\
&\geq q(m+1) + (j-q)\left\lfloor \frac{q}{j+q}(m+1) - \frac{1}{j+q} \right\rfloor \\
&\geq q(m+1) + \frac{q(j-q)}{j+q}(m+1) - 1 \\
&\geq q(m+1) + q(m+1) - \frac{2q^2}{j+q}(m+1) - 1 \\
&\geq 2q\left(1 - \frac{q}{j+q}\right)(m+1) - 1
\end{aligned}
$$

3.7 Fairness Relative to Proportional Allocation

Theorem 9 describes the values of n below which B cannot win q districts, and Theorem 8 describes the values of n above which B can always win at least q districts. As a sanity check, we note that the range of n where we do not know either that B can or that B cannot win q districts is

$$2q(m+1) \geq n \geq 2q\left(1 - \frac{q}{j+q}\right)(m+1) - 1$$

and we see that this gap always exists.

Next, we recall that saying that a proportional outcome is for B to win p districts implies that the number of bricks n lies in specific range:

$$(2p-1)m + p \leq n \leq (2p+1)m + p.$$

We can understand how the number of districts that B can win given a minority of size n by examining these breakpoint values of n at which p changes. At the smallest value of n at which the proportional representation is p, Theorem 8 tells us that B can always win at least $p-1$ districts:

$$n = p + (2p-1)m \geq 2(p-1)(m+1)$$

Theorem 9 tells us that B cannot win p districts when p is small:

$$\begin{aligned}
n = p + (2p-1)m &\leq 2p\left(1 - \frac{p}{j+p}\right)(m+1) - 1 \\
2p(m+1) - m - p &\leq 2p(m+1) - \frac{2p^2}{j+p}m - \frac{p}{j+p} - 1 \\
m + p &\geq \frac{2p^2}{j+p}m + \frac{p}{j+p} + 1
\end{aligned}$$

for instance, when $p < \sqrt{j}$.

3.8 Discussion

To visually compare the success of the majority and minority players' strategies, we graph an example for a game of reasonable size in Figure 3.4.

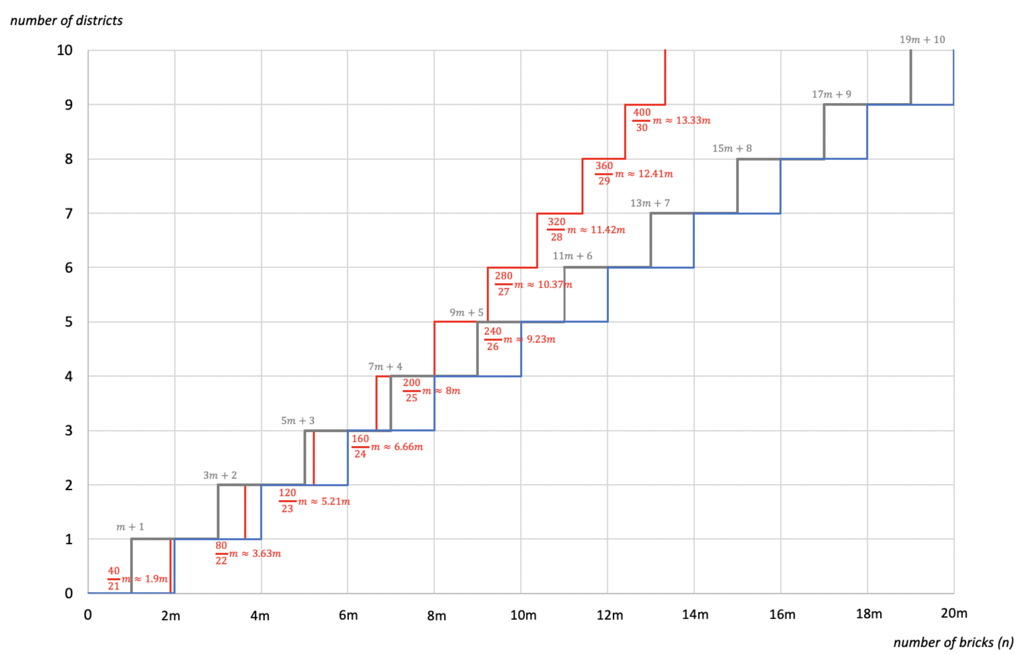


Figure 3.4: The bounds illustrated for the case $j = 20$. The x axis is the number of bricks (voters of the minority party B) as a function of m, the size of a bare majority in a district. The y axis is the number of districts that B can win. The grey plot shows p, the number of districts B wins if they are distributed proportionally. The red plot shows $n = (2q - \frac{2q^2}{j+q})m = \frac{2jq}{j+q}m$; Theorem 9 says that when n is below this value, B cannot win q districts. The blue line shows $n = 2qm$; Theorem 8 says that when n is above this value, B can always win q districts.

In Figure 3.4, we see that the gap between the lower bound (blue line) and upper bound (red line) on the number of minority voters n required to win q districts increases with q. We are working on closing this gap, and we conjecture that improving the strategy for the minority player B would show that B can win more than p (although of course never more than $J/2$) districts for larger values of n. That is, we conjecture that Redistricting Ghost favors the minority player when the minority is large, and the majority player when the minority is small, at least as compared to proportional representation.

It will be important to understand the results of Redistricting Ghost in more realistic contexts. An good first step might be to consider analyze the protocol in a model that simplifies the distribution and geometry with a graph, as in [35]. Because the protocol requires placing individual voters, or small groups of voters such as census blocks, testing it in a fully realistic situation will likely require building bots that play each other.

Redistricting Ghost [102], like I-cut-you-freeze [111] and the bisection protocol [94], does not allow the minority player to achieve proportional representation when the minority is small. The fact that several protocols have the same effect suggests that we might see it as an inherent feature of the redistricting problem: there are many ways to "crack" a small group of minority voters and prevent them from dominating any one district. One could consider this to be fair, or perhaps an unfairness inherent in the idea of electing representatives from geographic districts, even independent of the geographic distribution of voters.

Chapter 4

Cognitive Fairness: Utilizing EEG to Understand Human Perceived Fairness and Social Reciprocity

Understanding how social cues influence fairness perception is essential for decoding human cooperation. Although previous EEG studies have largely focused on the neural consequences of social observation or punitive feedback, the present study explores the effects of positive, observed non-verbal behavior, specifically a thumbs up gesture on fairness evaluation. In this chapter, we use electroencephalography (EEG) to investigate how minimal social feedback modulates behavioral and neural responses during a decision-making task that involves fair and unfair resource distributions.

Our results show that even minimal positive social cues significantly shape perceptions of fairness, with these effects reflected in both behavioral ratings and neural responses, particularly in an established index of evaluative and attentional processing: the P3 component. Crucially, the influence of positive feedback is context-dependent, that is, whether individuals are in a disadvantaged or advantage position. The findings suggest that fairness judgments are not fixed, but are dynamically modulated by relative social status and subtle social signals. This work contributes to a broader understanding of how non-verbal cues shape human judgment and behavior, offering implications for social neuroscience, behavioral economics, and the design of socially interactive systems.

4.1 Overview

Human behavior is profoundly influenced by the social context, which encompasses the environment that shapes the way people perceive and interact with the world around them [125]. This context is shaped by explicit and implicit social norms. Although explicit norms are clearly defined and communicated, implicit norms are learned through the decision-making processes of individuals during social interactions. In the context of group collaboration, there are currently no explicit guidelines for how group work should be organized, particularly with respect to the distribution of group outcomes (rewards). Furthermore, fostering cooperative behavior, where individuals strive for optimal group outcomes, can be challenging due to conflicts between individual and collective interests [82].

Most of the research on decision making has focused on individual decisions, where participants choose solely on their own values and preferences [145,160]. In these studies, participants typically evaluate a set of choices described by different attributes. However, many of our most important decisions occur within interaction with another partner, which is why game theory has been employed to study social decision-making in laboratory settings. Game-based paradigms have the advantage of being easily understood by participants, providing compelling social scenarios, and being relatively simple to adapt for neuroscientific study. One of the most commonly used games to study how human perceive fairness in dividing resources is the Ultimatum Game (UG) [58]. In the UG, two players are involved, a proposer and a responder. The proposer divides an endowment and offers a portion to the responder, who must then decide whether to accept or reject the offer. If the offer is rejected, both players receive nothing. UG creates a conflict between economic self-interest (maximizing one's own reward) and social motivations such as reciprocity and aversion to inequality.

The existing literature has extensively explored the UG; however, a limitation of most UG studies is that the task lacks a clear social context. Typically, the game involves distributing a fixed amount of money between two individuals and the focus is on analyzing how brain activity changes in response to different offers. Although there are many manipulations, for example, by introducing the sanctioning and compensation behavior of a third party [26] or modifying the selection process of the proposer [36], these studies often do not account for the temporal dynamics of social interactions. Most UG research has concentrated on static stimuli, rather than dynamic ones, as the latter can be challenging to examine in psychological research.

Real-life social interactions are inherently dynamic and occur within specific contexts that shape individuals' perception and evaluation processes. Social behaviors play a critical role in everyday functioning, often relying on the interpretation of subtle nonverbal cues to assess others' emotions and intentions. Understanding the dynamics of these interactions is crucial. In this work, we will focus on the dynamic of cooperation in a group. Our aim is to build on previous research by focusing on the fairness aspect of decision making in a more complex context, that is, in *group work settings*. Fairness in group work can be challenging to support, as groups consist of individuals with complex and sometimes conflicting interests, making work division difficult. To our knowledge, no laboratory studies have specifically investigated what constitutes a fair reward distribution among group members.

Human studies often face challenges due to discrepancies between self-reported behavior and actual real-life decisions, such as biases from socially desirable responses, random responses, and demand characteristics. EEG presents a promising alternative as a noninvasive and low-cost method. It has proven to be a reliable tool for clinical research and offers a more direct way to measure brain responses to stimuli related to fairness. EEG originates in the field of brain-computer interface (BCI) technology, which shares the goal of enhancing communication between humans and computers. However, EEG specifically focuses on direct communication between the brain and the computer. Initially applied in the medical field, EEG has since gained widespread acceptance in cognitive neuroscience research [56]. As a non-invasive biometric tool, EEG records brain activity using electrical signals from the scalp. These signals, generated by neural activity within milliseconds, follow unique neural pathways that are difficult to forge according to physiological theories [166]. In addition, compared to other BCI technologies (e.g., fMRI or fNIR), EEG is faster, more affordable, and provides accessible information on brain function. We plan to design a laboratory study that reflects the dynamics of group work in real life and incorporates EEG technology, with a focus on understanding the neural mechanisms underlying fairness.

We seek to explore the role of positive "observed nonverbal behaviors", rather than punitive responses from one's partner. Our findings contribute to a growing body of evidence showing that minimal social feedback can profoundly influence perception of fairness. These effects are not only behavioral, but are reflected in neural markers such as the P3 component, which indexes evaluative processing and attention allocation. Understanding how and when such cues shape human judgment has important implications not only for neuroscience but also for social psychology, behavioral economics, and human-computer interaction, particularly in designing systems that aim to simulate or mediate social behavior. In addition, we highlight the context-dependent nature of social modulation. The same feedback (a thumbs up) has different psychological effects depending on whether a participant is in a relatively advantaged or disadvantaged position. When outperforming a partner, a thumbs up can increase feelings of entitlement, making fair offers seem less generous. In contrast, when underperforming, the same gesture can be interpreted as an unexpected affirmation, leading to a greater appreciation of fairness. This supports the idea that fairness judgments are not static but dynamically constructed based on relative social status, perceived intent, and emotional framing. The novelty of our study lies in its design of interactive scenarios involving partners to investigate this phenomenon.

4.2 Fairness in Group and Its Measurements

4.2.1 Fair Workload Division: Fairness Norms in Group Work

The compliance of the fairness norm is heterogeneous, covering different types of behavior, such as reliable fair players (voluntary compliants), individuals who comply to avoid sanctions or punishment (sanction-based compliers), and those who are reliably unfair (non-compliers) [54]. However, there is a natural tendancy towards fairness, which is unavoidable because people tend to dislike and punish unfair behaviors in social interactions. In UG for example, although the responder can only maximize their payoffs by accepting all offers, acceptance rates are substantially lower for unequal than equal offers, finding that the responders typically reject offers that constitute only 20% of the total sum [18].

Furthermore, although the fairness norm is often defined as a shared expectation, it is also shaped by the values and beliefs that guide group members' behavior. Research has shown that proposers often aim to *appear fair* rather than *to be fair*. When there is a chance that their offer will not be seen by the responder, they are more inclined to make unfair allocations. This suggests that implicit social norms, such as the desire to maintain a positive reputation, can override explicit task instructions or fairness principles [113].

The current discussion on the fairness norms regarding fair division is surrounded by the four main criteria, which are *Proportional, Equitable, Envy-free* and *Pareto-Optimal (or Pareto-efficiency)* [20]. More detailed explanation can be found in Section 2.2. Proportional indicates that each individual thinks their share is at least $1/n$ of the total; equitable indicates a subjective value that each individual thinks their share is the same as everyone else, and they are equally happy; envy-free indicates that each individual has a strategy that can guarantee himself a share that is at least as desirable as any other shares. Last but not least, pareto-optimal indicates the division where there are no other divisions that dominate the share for all individuals, or there is no other feasible agreement that would make at least one individual strictly better off while not making any of the others worse off.

These frameworks disregard the procedure for making these decisions, paying little attention to the preferences, thoughts, and feelings of the individuals involved [153]. The aim of this work is to understand fairness norms in a group setting, specifically how humans perceive fairness and what factors can enhance it.

4.2.2 The Power of Affirmative Feedback

Researchers have long explored the dynamics of proposer-responder interactions, particularly in the context of unfair decisions. After such decisions are made, responders often feel more vulnerable and consequently display reduced trust toward the unethical decision-maker. In response, the transgressor may express remorse, offering an explanation for the unfair decision. Apologies, especially non-perfunctory ones, tend to be more frequent as the severity of the social predicament increases [134]. Moreover, the perceived sincerity of the apology significantly affects its effectiveness in altering behavior within a UG and in shaping the responder's reaction to the unfair outcome [139]. Social accounts also play a pivotal role in trust judgments, with denials leading to lower trustworthiness ratings compared to apologies. In fact, denials tend to exacerbate the situation, while responders have a strong desire for social information, particularly in response to unfair offers [33]. In addition to expressing remorse, promises have been shown to facilitate cooperation [137]. A promise to change behavior can significantly accelerate the trust recovery process, though prior deception can undermine the promise's effectiveness in rebuilding trust.

Information we may gather during social decision making includes aspects such as whether we are in person with someone and how they interact with us [5]. In behavioral economics, research has emphasized the importance of intention, with actions perceived as fair when the underlying intention is seen as kind [124], and vice versa [76]. Recent studies have increasingly focused on the concept of social interactions. The UG has demonstrated altruistic punishment, in which individuals punish others at personal cost, even without immediate material benefit. However, further research has shown that unrelated individuals often seek to ensure fairness toward others in daily life, with the expectation of receiving fair treatment and cooperation in return. This behavior is referred to as reciprocal affirmation.

Reciprocal affirmation, the process of creating a balanced give-and-take of validation and support where both parties feel respected and valued, has been well supported by previous research in therapeutic communities (TCs). Studies have shown that when residents of these communities receive affirmation from peers (but not from staff), they tend to increase their prosocial behavior [39]. Trust between residents also grows from the perception that community members care sincerely for each other [8]. This highlights the importance of peer connections and the affirmation cycle within a community.

A similar concept, *reciprocal expertise affirmation*, has been observed in group work settings. In groups with higher levels of reciprocal expertise affirmation, where members confirm each other's abilities, participants are more motivated to contribute their expertise to the team task, ultimately enhancing team performance [42, 55]. Other studies have supported this finding from different perspectives, particularly in the context of dynamic social interactions where human intentions play a key role. For example, third-party interests that do not have an interest often view intentional social harm as more severe than those that are unintentional [104]. People's reactions are heavily influenced by their perceptions of others' intentions [57]. In addition, people tend to accept unfair offers proposed by a computer rather than a human agent [95]. In this work, our aim is to manipulate how social affirmation would influence individuals' perceptions of fairness in a group setting. We will investigate whether receiving a thumbs up from a partner affects participants' perceptions.

4.3 The Present Study

4.3.1 Research Questions

While prior research has examined the effects of affirmative feedback, our study aims to delve more deeply into how such interventions influence individuals' responses to reward distribution. Specifically, we investigate both behavioral outcomes, as measured by Likert scale ratings **(RQ1)**, and neural correlates, particularly ERP responses **(RQ2)**. Additionally, we seek to determine whether the participant's relative performance, whether outperforming or underperforming their partner, modulates these effects **(RQ3)**.

4.3.2 Study Design

The study consists of 8 rounds of games, with participants facing 50 reward conditions in total. There are two context conditions: fair and unfair. Participants were randomly assigned 25 fair and 25 unfair conditions, with a randomized order of assignment. They were instructed to make decisions based on their initial impressions, without overthinking the process. In fair conditions, participants are presented with a reward distribution based on a fair algorithm that is proportional to their performance. In unfair conditions, participants are presented with a reward distribution based on the same fair algorithm, but with a multiple of 20% in the allocation, resulting in a distribution that is highly disproportionate to their performance. The choice of a 20% is informed by the literature on UG, where studies have consistently found that proposals that offer only 20% of the endowment are rejected approximately half the time [19].

Within these two conditions, we manipulate an additional factor: reciprocal affirmation. On the game summary page, we will display a message from the partner. Participants will either see a thumbs up message from their partner or no message at all (reciprocal affirmation). We plan to run a 2x2 within-subjects design as shown in Table 4.1 for the fair and unfair conditions, as well as receiving reciprocal affirmation or not. In summary, we will run 8 trials of the game (blocks) for each participant. Each block will consist of 50 rounds of reward distribution to improve the signal-to-noise ratio of the EEG and ensure a sufficient number of valid trials. In total, there will be 400 tests, and participants will take a break after each block to minimize fatigue. We make sure the number of trials has reached the signal-to-noise level [11].

4.3.2.1 Proportional v.s. Penalization Algorithm

We define our fair and unfair algorithms as proportional and penalization, respectively. For fairness, we adopt a proportional algorithm, meaning that an individual's reward should be directly aligned with their contribution to the group. Research has shown that using team-based merit pay to recognize individual contributions helps to create a better balance between group and individual

	Fair Offer	**Unfair Offer**
W/ thumbs up	Fair, W/ thumbs up	Unfair, W/ thumbs up
W/O thumbs up	Fair, W/O thumbs up	Unfair, W/O thumbs up

Table 4.1: Four experiment conditions. Our goal is to manipulate whether reciprocal affirmation or social comparison influences individuals' perceptions of fairness in a group setting. Specifically, we examine whether receiving a thumbs up from a partner constitutes reciprocal affirmation and whether revealing participants' game performance as easier or more difficult than their partner's serves as social comparison.

goal achievement [67]. In contrast, for unfairness, we apply a penalization rule, where the reward an individual should receive proportionally is reduced by 20%. The following are the formulas for our algorithm:

Measurement

$$\text{Proportional} = \frac{\text{Individual Score}}{\text{Group Score}}$$

$$\text{Penalization} = \frac{\text{Individual Score}}{\text{Group Score}} * 20\%$$

$$\text{Group Score} = \text{Individual Score} + \text{Partner Score}$$

4.3.3 Experiment Process

We design a scenario in which participants are told that we have developed a fairness algorithm and need their help to improve it by providing feedback. To begin, participants will play a Tetris game for one minute, which serves as a proxy to complete a task. During the game, participants will be able to see how their partner performs on a side-by-side screen. However, using the Wizard-of-Oz method, participants are led to believe they are playing with an online partner, though they are actually interacting with a pre-programmed computer setup designed by us. The purpose of using a pre-set computer program is to ensure that each game is unique and consistent. Initially, we invited 10 players to participate and recorded their performance (mean score = 67). The game is designed with a range of possible and reasonable scores to ensure variability and fairness in the results. The computer partner's scores in the 8 rounds are as follows: $40, 0, 20, 300, 10, 110, 30,$

and 130. During this playing time, both players must act quickly under time pressure, with their actions potentially influencing each other's success. The side-by-side setup allows players to track their partner's performance.

After the game, participants will be presented with an option to send a thumbs up to their partner, with the choice to either "send" or "skip". A game summary page will then display the score and number of lines cleared by each player. Depending on the study condition, a system message will indicate whether they receive a thumbs up message from the partner, or no message at all. Before revealing the reward distribution based on the fairness algorithm, participants will focus on a fixation cross to direct their attention to a specific area of the screen. Finally, after the reward distribution is presented, participants will rate the fairness of the algorithm using a 5-point Likert scale, with 1 indicating "unfair" and 5 indicating "most fair". A detailed experiment process can be found in Figure 4.1.

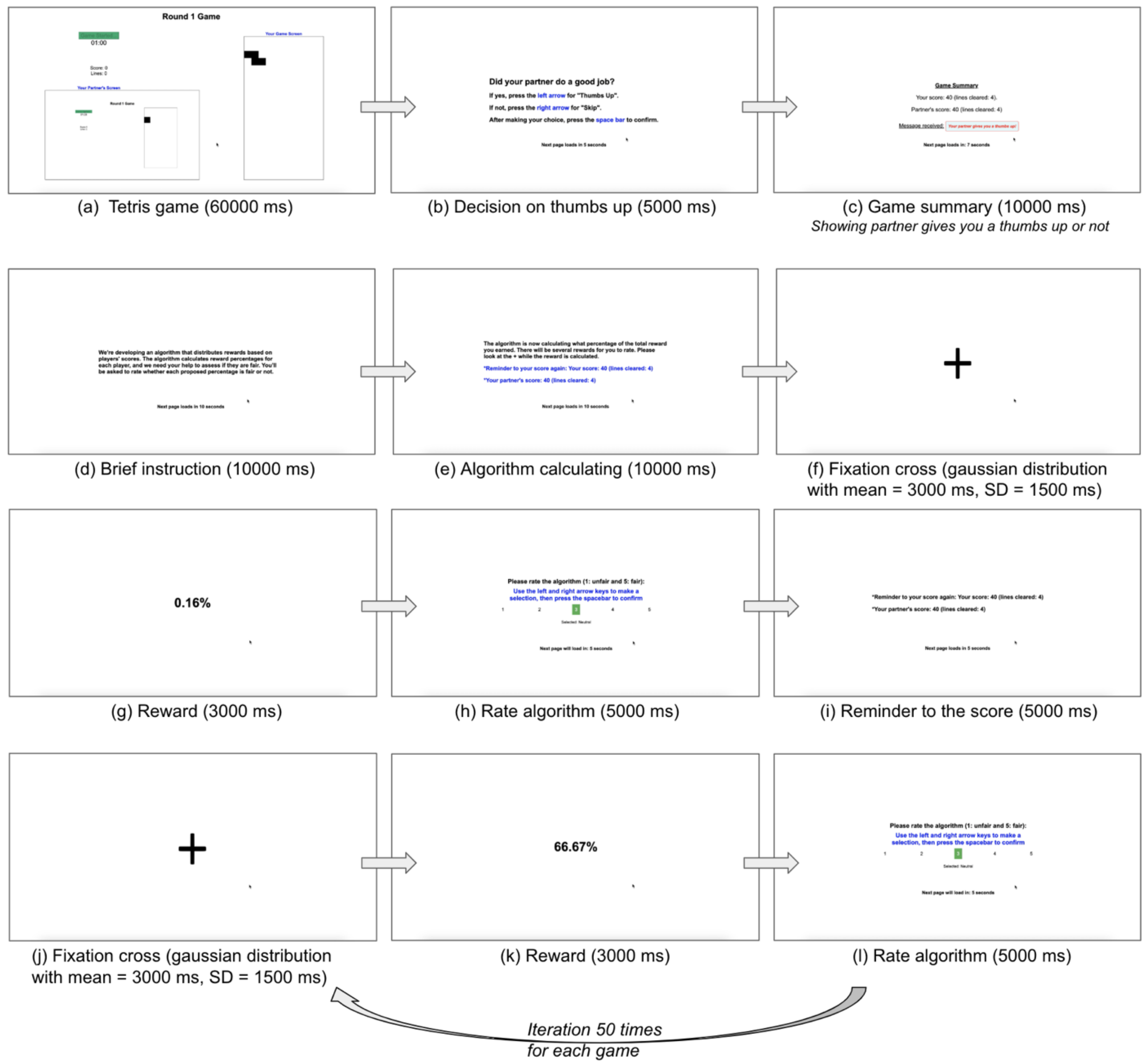


Figure 4.1: Experiment process. (a) A 1-minute Tetris game with the partner's playing screen displayed side-by-side. Duration: 60 seconds. (b) An options page that asks participants whether they want to send a thumbs up to their partner. Duration: 5 seconds. (c) A game summary page showing both players' scores. Depending on the study conditions, a system message will either display a thumbs up message (or not) from the partner, or show nothing. Duration: 10 seconds. (d) A brief introduction of our study goal, "we are developing an algorithm that distribute rewards based on players' scores. We need your help to assess if they are fair". Duration: 10 seconds. (e) A loading page with the message, "The algorithm is calculating the percentage of the total reward you earned. Please focus on the fixation cross while the reward is calculated". Duration: 10 seconds. (f) Fixation cross. Duration: average 3 seconds. (g) The calculated reward. Duration: 5 seconds. (h) An algorithm rating page regarding fairness, presented as a 5-point Likert scale. Duration: 5 seconds. (i), (j), (k), (l) are just iterations.

4.3.4 Participants

41 healthy, right-handed participants, aged 19-30 years, with no history of neurological disorders or mental health conditions, participated in the study. All participants had normal or corrected-to-normal vision. They voluntarily enrolled and provided their written informed consent. The study was approved by the Institutional Review Board (IRB). During the experimental session, the participants were seated in front of a computer screen inside an acoustically shielded EEG cabin. Before participating in the study, participants were asked to arrive on time to avoid meeting their partner. In reality; however, they are partnering with a computer.

4.3.5 EEG Recordings and Offline Processing

EEG data were obtained at a sampling rate of 2000 Hz, using a 32-channel electrode cap (MBT O smarting pro 32 EEG cap), and we have made sure that the impedance of all electrodes was below 5Ω. Data were processed using EEGLAB and ERPLAB toolbox and analyzed in MATLAB. We followed the study procedure [46] to ensure the quality of the data collected. Furthermore, following Luck's ERP data analysis standards and procedure [93], we remove 6 participants for whom more than 25% of the trials contained artifacts while performing EEG analysis. We also remove 1 participant's data for not being able to earn points in a time-constraint Tetris game (4 out of 8 rounds are zero point). In total, we are analyzing EEG signals from 34 participants.

First, the EEG data were imported to EEGLAB, which involved temporary re-referencing to Cz, followed by downsampling to 256 Hz, low-pass FIR filtering at 30 Hz, high-pass FIR filtering at 0.1 Hz, and epoch each trial using a window of 200 ms pretrial onset to 800 ms after trial onset. Second, noisy EEG channels were manually rejected using SEM (2 SD), and noisy epochs were removed using the criteria of $\pm 100uw$ for all channels. Third, independent component analysis (ICA) was used to decompose the EEG channel data and remove components that capture eye movement and other artifacts based on statistical and visual confirmation. After artifact component removal, the EEG data were projected back from the ICA with artifacts. We then use correction to minimize the noise produced by blinks, and rejection to eliminate epochs that contain large artifacts that are not easily corrected. Lastly, the data were re-referenced to an average reference and the reference selected during data import to EEG (Cz) was added back to the data. The rejected channels were at that time interpolated.

4.4 Result

As we aim to understand whether independent variables *thumbs up vs. no thumbs up* or *fair vs. unfair offers* affect dependent variables *fairness perception ERP responses* in group work, we apply mixed model ANOVAs. This approach accounts for the repeated measures structure of the design and the potential interdependencies among observations introduced by group collaboration.

4.4.1 User's Inputs From the 5-points Likert Scale

As shown in Figure 4.2, participants rated fair offers more favorably than unfair offers in both thumbs up and no thumbs up conditions. This indicates that the fairness manipulation was effective, as the distinction between fair and unfair offers remained substantial regardless of positive feedback. Interestingly, although the presence of a thumbs up appeared to have a minimal impact on perceived fairness at the descriptive level, statistical analysis revealed a highly significant main effect of fairness on the Likert scale's ratings ($p < 0.001$), as well as a significant interaction between fairness and the thumbs up condition ($p = 0.0338$). This suggests that the influence of fairness on participant ratings depended, to some extent, on whether a thumbs up was shown. The main effect of the thumbs up itself was marginal ($p = 0.0592$), indicating a potential trend that did not reach conventional significance. These findings answered **RQ1**, but motivate a deeper investigation into how reciprocal affirmation (such as a thumbs up) might modulate responses at the neural level. To explore this, we turn to participants' EEG data.

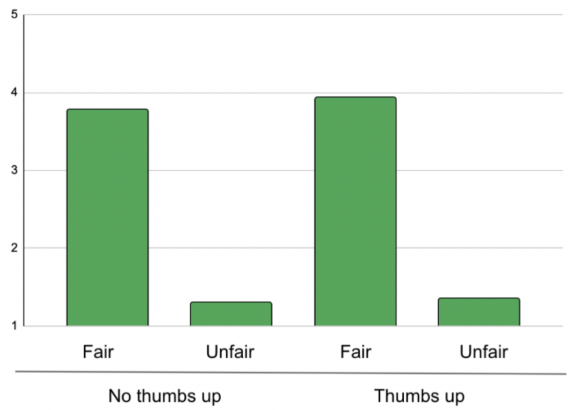


Figure 4.2: Participants rated fair offers significantly higher than unfair offers, regardless of whether the partner gives a thumbs up. To better understand this positive effect, we aim to examine participants' EEG data.

4.4.2 Participant's Grand Average ERP at Pz

In this work, we focus on the Pz channel within the 250–500 ms time window to investigate the P3 (P300) component, as this region and time frame are where P3 is typically strongest and most relevant for tasks involving evaluation processing. Following Picton's guidelines [114], we also examined three key scalp locations: Fz, Cz, and Pz, where the P3 component is generally prominent. Accordingly, we conducted a Region of Interest (ROI) analysis across these electrodes. The results shown in Figure 4.3 revealed a consistent trend in all three sites, closely matching the pattern observed at Pz alone (see Figure 4.4a). This not only reinforces the robustness of our findings, but also supports our focus on the Pz channel, which is widely recognized in the literature as the primary site for observing the P3 component [114].

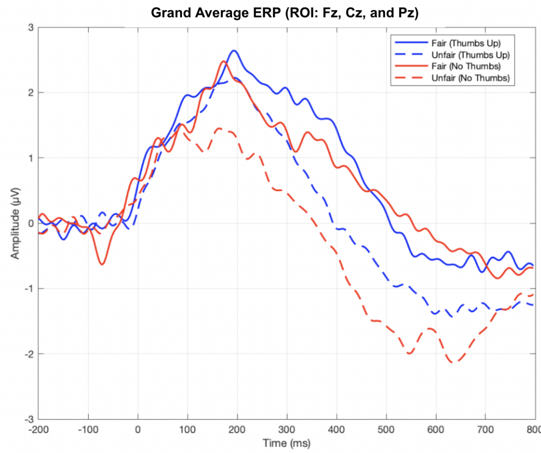


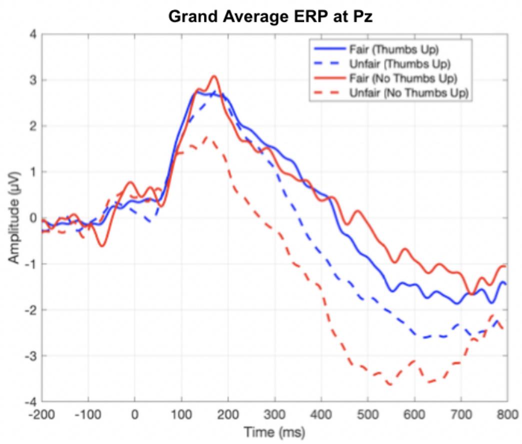
Figure 4.3: Grand average ERP across Fz, Cz, and Pz channels. The Region of Interest (ROI) analysis revealed a consistent pattern across all three channels, mirroring the trend observed in the Pz channel alone.

The classic component P3 is usually elicited by infrequent or unexpected stimuli and is associated with attentional and cognitive processing [114]. A high amplitude (larger P3) could mean that the brain processes the outcome as more relevant to motivation or more attentional (how relevant or important the outcome is). In our study, we are particularly interested in examining whether positive feedback (e.g., a thumbs up) modulates the P3 component, especially in the context of perceived fairness. To address this, we place specific emphasis on analyzing the Pz channel and present results centered on this region.

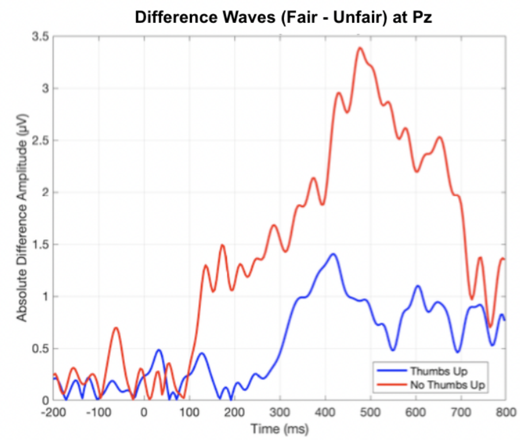
4.4.2.1 Positive Feedback Changes the Size of Fairness Effect

As shown in Figure 4.4a, from 250 to 500 ms, fair offers (solid lines) are generally associated with slightly higher P3 amplitudes than unfair offers (dashed lines), particularly in the thumbs up condition. This pattern aligns with previous research suggesting increased cognitive or evaluative engagement with norm-congruent (i.e., fair) outcomes. Furthermore, conditions involving thumbs up feedback (blue lines) elicited more positive P3 amplitudes compared to those without feedback (red lines), especially in the 200–350 ms time window. This finding indicates that positive social feedback can improve the motivational significance or cognitive evaluation of the outcomes ($F[1, 33] = 5.39, p < 0.05$).

The most pronounced divergence appears between the Unfair + No Thumbs Up condition (dashed red) and the Fair + Thumbs up condition (solid blue), with the former showing a markedly reduced and more negative P3 response. Even within unfair trials, the presence of thumbs up feedback (dashed blue) appears to increase P3 positivity relative to the no feedback counterpart, suggesting a potential buffering effect of positive feedback in the context of unfair treatment in 300 to 400 ms ($F[1, 33] = 4.04, p < 0.05$). Figure 4.4b shows that in the thumbs up condition, participants exhibit reduced neural responses, particularly in the P3 amplitude. This suggests that positive feedback may attenuate the neural impact of outcome evaluation.



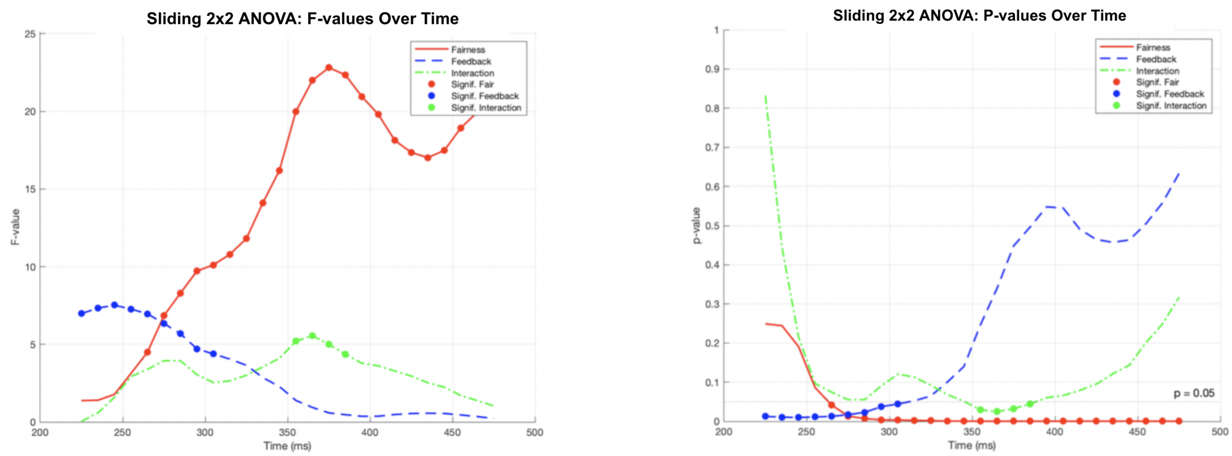
(a) Average P300



(b) Difference in P300

Figure 4.4: Comparison of Pz waveforms across conditions. In (a), we can see that in the thumbs up condition, the solid blue (fair) is slightly more positive than dashed blue (unfair); however, in the no thumbs up condition, the solid red (fair) is much more positive than dashed red (unfair). In (b) shows the difference waves. This suggests that the thumbs up feedback from the partner reduced the difference between fair and unfair outcomes. This is a crossover interaction, indicating that the fairness effect is stronger without thumbs up, weaker with no thumbs up.

We set a threshold for statistical significance at an F-value of 5 and a P-value of 0.05. As shown in Figures 4.5a and 4.5b, fairness consistently reaches significance throughout the 200–500 ms time window. Positive feedback shows its strongest significance between 200–300 ms, while the interaction effect between fairness and feedback is most significant between 350–400 ms. These time-specific effects suggest that fairness is processed more broadly over time, whereas feedback and interaction effects emerge in more temporally specific windows, but still show effects in the overall time range.



(a) F-value over the time

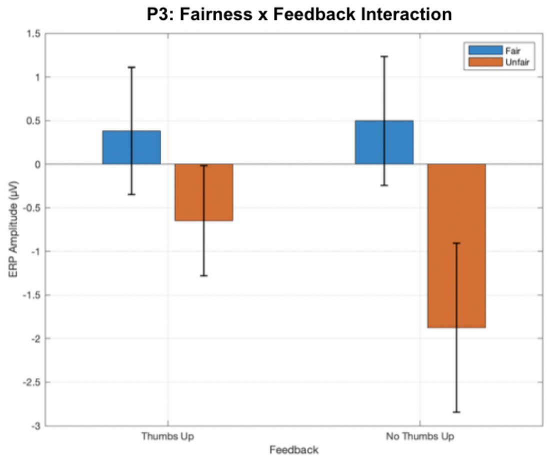
(b) P-value over the time

Figure 4.5: Sliding 2x2 ANOVA Result: F-value and P-value over time.

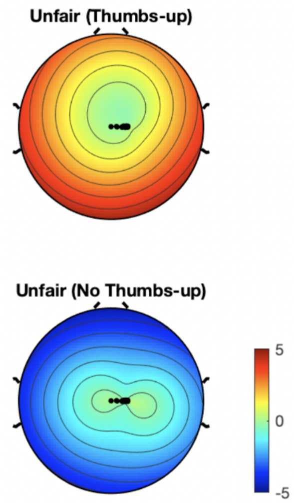
4.4.2.2 Thumbs Up Dampens or Reverses the Typical ERP Responses to Unfairness

As shown in Figure 4.6a, there is an interaction effect between fairness and feedback. Specifically, participants' neural sensitivity to fairness appears diminished when they receive a thumbs up from their partner, particularly in response to unfair offers.

In addition, the scalp topography plots in Figure 4.6b provide a spatial representation of brain activity in response to fair versus unfair offers, under thumbs up versus no thumbs up conditions, during the time window of 250–500 ms. The ERP response to unfair offers is more positive when a positive feedback (thumbs up) is present (top panel). In contrast, unfair offers without feedback elicit more negative ERP responses (bottom panel). Section 4.4.2.1 and 4.4.2.2 answer **RQ2**.



(a) The interaction effect of fairness and feedback. Within the thumbs up condition, the ERP response to fair and unfair is less different.



(b) The ERP topographical maps. Unfair offer evoke more positively when thumbs up is present.

Figure 4.6: The interaction effect and topographical map in the time range of 250 to 500 ms.

4.4.3 When Participant Outperform or Underperform Their Partner

To answer **RQ3**, we define a score gap of 50 points or more (corresponding to a difference of 5 or more rows) as a large score difference. In our analysis, we aim to examine whether such score disparities influence ERP responses, particularly in relation to fairness and feedback conditions.

4.4.3.1 Participant's Score is Much Higher Than Their Partner

When an individual outperforms their partner by a large margin, we would expect a reduced response of P3 to proportional (fair) results, consistent with previous findings that the amplitude of P3 decreases for expected or unexpected events [118]. However, we found out that the positive feedback in this condition can change one's fairness expectations, that is, receiving a thumbs up can lead participants to feel more confident or worthy.

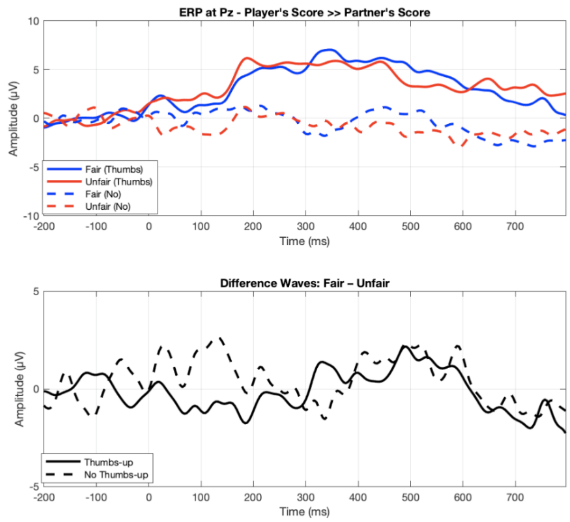
When we look at Figure 4.7a (top panel), the P3 response remains relatively strong, even resembling the pattern observed in the Unfair + Thumbs Up condition. It indicates that participants still engage evaluatively with fairness, despite their advantageous position, and are surprised by the proportional distribution. In this context, proportional offers might be perceived as less generous or rewarding, given their elevated self-assessment.

Similarly to the findings of van der Veen and Sahibdin [154], our results suggest that both fairness and subjective expectations influence neural responses. In their study, fairness and the size of the offer affected acceptance rates, and only unfair high offers (i.e., where the initial stake was large) elicited a medial frontal negativity (MFN), while unfair low offers did not. Although the stakes in their monetary paradigm are not directly comparable to the points earned in our design, the underlying principle remains relevant: when individuals mentally calculate that they deserve more (based on context or performance), their neural responses to perceived unfairness are amplified. In our case, this is reflected in an enhanced component of P3, suggesting a stronger emotional or evaluative engagement when fairness expectations are violated under conditions of perceived merit.

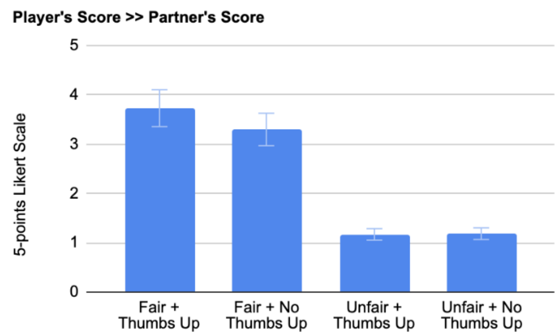
In the Unfair + Thumbs Up condition, the larger P3 amplitude is understandable, as participants received positive social feedback but were immediately confronted with an unfavorable distribution decision from the same partner. This incongruence between feedback and outcome can increase cognitive and emotional processing, reflected in the increased P3 response.

In contrast, in the No Thumbs Up condition, the lack of positive feedback can already act as a prediction error early in the trial, reducing the significance of subsequent fairness-related processing. For No Thumbs Up + Fair trials, the fair offer may be seen as acceptable but unsurprising, resulting in a modest P3. For No Thumbs Up + Unfair trials, the combination of negative social feedback and an unfair outcome could be perceived as a violation of merit-based fairness expectations, possibly eliciting frustration or disengagement. This is reflected in a reduced amplitude of P3, consistent with previous studies linking lower P3 to emotional stress or motivational withdrawal [63], which also aligns with the findings of the UG, where responders are typically sensitive to unfairness and often reject unfair offers, even at personal cost, to punish proposers for violating fairness norms [107].

In Figure 4.7a (bottom panel), there is no clear visual gap between the Thumbs Up and No Thumbs Up conditions. This contrasts with the grand average findings, which suggest that when the score gap between the participant and their partner is too large, whether the participant scores higher or lower, the effect of the thumbs up feedback diminishes or becomes ineffective. However, statistical analysis reveals that feedback has a significant main effect on neural responses from 300-350 ms, $F(1, 5) = 44.89, p < 0.05$, and the interaction between feedback. These results again support the conclusion that positive feedback (a thumbs up) can influence the perception of fairness of participants, even when the visual difference is not immediately apparent in the ERP waveforms. The results from the 5-point Likert scale in Figure 4.7b showed consistently high ratings for fair offers and low ratings for unfair offers, as expected.



(a) The neural responses when player's score is much higher than partner's score



(b) 5-points likert scale result

Figure 4.7: (a) Both solid lines, red and blue, indicate that even in an advantageous position, participants continue to engage evaluatively with fairness, including when they receive a fair (proportional) offer. In this context, proportional offers may be perceived as less generous or less rewarding due to an elevated sense of self-worth or entitlement. This mismatch between expectations and actual outcomes may lead to a sense of surprise, which in turn drives continued cognitive engagement, as reflected in the sustained neural responses. There is no clear visual gap between the Thumbs-Up and No Thumbs-Up conditions. (b) The 5-point Likert scale shows expected high ratings for fair offers and low ratings for unfair offers.

4.4.3.2 Participant's Score is Much Lower Than Their Partner

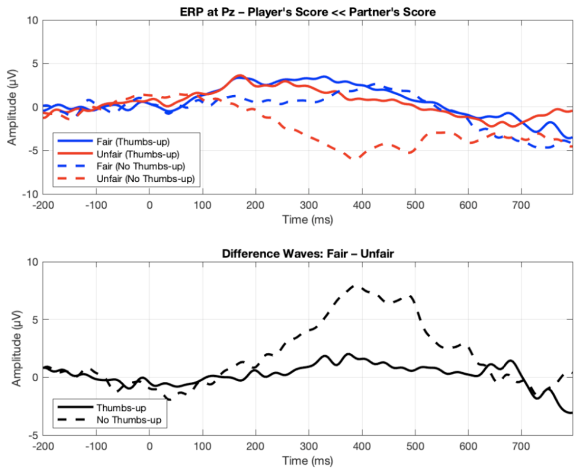
Contrary to the findings in Section 4.4.3.1, when an individual underperforms relative to their partner by a large margin, we would expect a reduced P3 response to penalizing (unfair) outcomes, likely due to lowered expectations or acceptance of disadvantage. As shown in Figure 4.8a, this pattern holds for the Unfair + No Thumbs Up condition. However, when a thumbs up is given in this context, the P3 amplitude increases, reflecting a violation of expectation or surprise. This suggests that positive feedback from a higher scoring partner followed by an unfair offer creates a mismatch that heightens neural processing, consistent with the role of P3 in signaling unexpected or incongruent outcomes.

We also observe that, for fair offers, there is little difference in neural responses between the Thumbs Up and No Thumbs Up conditions. This may be because participants typically accept that performing worse than their partner justifies receiving a less favorable outcome. Even when the penalization is relatively harsh, as shown in Figure 4.8b, the participants appear to accept the outcome without a higher neural response. However, when they receive a fair offer despite underperforming, participants may experience a sense of gratitude or inclusion, which could influence their emotional and cognitive engagement.

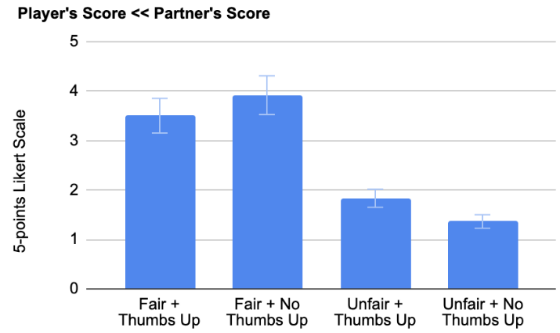
In Figure 4.7a (bottom), there is a clear visual gap between the Thumbs Up and No Thumbs Up conditions. Statistical analysis confirms a significant interaction effect between fairness and feedback on neural responses from 300-400 ms, $F(1, 5) = 9.75, p = 0.03$. Consistent with the grand average results, this suggests that positive feedback can reduce the neural distinction between fair and unfair outcomes, possibly by altering fairness expectations or buffering the negative impact of unfair offers.

As shown in Figure 4.8b, the results of the 5-point Likert scale suggest that when the participants' scores are significantly lower than those of their partner, they may not expect to receive positive feedback. In the Fair + No Thumbs Up condition, participants may be surprised that despite their poorer performance and the lack of explicit positive feedback, their partner still provided a fair result. This unexpected fairness may evoke a sense of gratitude, leading to the highest fairness ratings among all conditions.

In contrast, the Unfair + No Thumbs Up condition aligns more closely with expectations, receiving the lowest ratings within this group. However, even in this case, the scores are slightly higher compared to the condition in which the participant's score was much higher than their partner's. This suggests that participants may be more forgiving of unfair treatment when they perceive themselves to be in a disadvantaged position, possibly due to lower expectations or greater empathy.



(a) The neural responses when player's score is much higher than partner's score



(b) 5-points likert scale result

Figure 4.8: (a) In a disadvantageous position, participants may anticipate receiving no thumbs-up or an unfair offer. As a result, any deviation from this expectation, such as receiving fair treatment or affirmative feedback can elicit surprise and heightened cognitive processing. The neural effect of fairness is amplified when affirmative feedback is present. The presence of a thumbs-up may accentuate the brain's differential response to fair versus unfair treatment. (b) The 5-point Likert scale suggest that when the participants' scores are significantly lower than their partner, they may not expect to receive affirmative feedback compared to when they score significantly higher.

4.5 Discussion

Humans rely significantly on cooperation, and much of the research on fairness revolves around financial decision-making influenced by fairness norms. Previous studies have developed theories such as inequality aversion [44] and altruistic punishment [12], both of which suggest that individuals have a strong preference for fairness and tend to oppose situations where resources are distributed unfairly. In particular, individuals who engage in altruistic punishment (often referred to as strong reciprocator) are willing to incur personal costs to punish unfair behavior or non-cooperation, even when these actions do not result in any material gain [47].

The same concept extends to economics, where most models traditionally assume that individuals act based on self-interest. However, research has shown that a substantial portion of people are also motivated by considerations of fairness. Fairness is a key determinant in shaping human behavior and influences many individuals' decisions. Economic studies have examined concepts such as social comparison and loss aversion, revealing that even at personal cost, people are willing to punish unfair behavior [48]. In conclusion, (1) humans care about being treated fairly and about treating others fairly; (2) humans are willing to resist unfair practices, even at a personal cost; and (3) humans have systematic implicit rules that define which actions by firms are considered unfair [75, 163].

4.5.1 How Does a Partner's Thumbs Up or No Thumbs Up Affect Fairness Perception?

Nonverbal cues, such as thumbs up or lack thereof, play a powerful, yet often underappreciated, role in shaping human perceptions of fairness. These minimal social signals function as rapid feedback mechanisms that trigger evaluative processes that influence how individuals interpret social outcomes. In our study, although the reward distribution was determined algorithmically, participants responded as if the results were socially mediated, especially when a thumbs up was present. This finding underscores the deep-rooted human sensitivity to social signals, even in contexts where explicit social interaction is minimal or artificial.

A thumbs up gesture typically communicates approval, inclusion, or validation. From a psychological perspective, it acts as a social reward, engaging systems involved in reinforcement learning and social cognition. These systems, particularly those involving the anterior cingulate cortex (ACC) and ventromedial prefrontal cortex (vmPFC), are known to assess the salience and value of social outcomes [141]. When participants receive a fair or even unfair distribution after a thumbs up, it can alter the way they interpret the outcome, either enhancing perceived fairness or buffering the negative impact of an unfavorable result.

Furthermore, this feedback operates within the framework of expectancy violation theory, where deviations from anticipated social behavior generate greater attention and neural engagement. For example, a thumbs up following poor performance or after an unfair offer can produce a stronger neural response (e.g., increased P3 amplitude) precisely because it violates expectations, prompting a re-evaluation of the partner's intention and fairness norms. This aligns with theories of social information processing, where individuals continuously assess others' actions to guide future social behavior.

Unfair or unethical decisions usually erode trust and trigger cognitive mechanisms to seek explanations or social accounts [33]. In many social contexts, people rely on these accounts, either verbal or non-verbal to determine whether an action was intentionally unfair or circumstantial. In our findings, the thumbs up can be seen as a proxy for such a social account. Although the reward allocation was computer generated, participants seemed to treat the feedback as socially meaningful, suggesting that the attribution processes are automatically engaged in the presence of social signals.

4.5.2 Social Feedback Not Only Enhances Outcome Evaluation But Also Modulates Fairness Expectations Based on Relative Status

Our findings highlight the nuanced role of social feedback, specifically, a simple thumbs up message, in shaping the fairness evaluations of individuals and neural responses. As demonstrated in Figures 4.8a and consistent with the grand average ERP results in 4.4a, positive social feedback generally elicits larger P3 amplitudes, particularly in response to fair offers. This aligns with previous research suggesting that positive social signals improve evaluative processing by increasing attentional engagement or reward sensitivity.

However, our data extend these insights by demonstrating that the effects of social feedback are context-dependent, particularly with respect to the relative performance status of the participant. When participants were in a relatively advantageous position (outperformed their partner), receiving a thumbs up appeared to amplify their perceived entitlement. In these cases, even proportionally fair offers might be viewed as insufficient or unremarkable, leading to a diminished fairness response. This suggests that elevated self-assessment can distort fairness expectations such that equitable treatment may not be experienced as generous or rewarding.

In contrast, when participants were in a disadvantaged position, the presence of a thumbs up appeared to have a buffering or enhancing effect. It may have been interpreted as a gesture of inclusion or social approval, which in turn increased evaluative engagement and increased the amplitude of P3. In particular, in the absence of feedback, participants in this disadvantaged condition showed reduced neural engagement and seemed more accepting of penalizing outcomes. This pattern suggests that lack of social affirmation can reduce motivation to contest unfair treatment, possibly reflecting internalized expectations of low status or value.

These findings underscore the power of social moderation. Positive feedback does not simply enhance the evaluation of the outcome; it also dynamically modulates fairness expectations based on the individual's perceived status within the interaction. This interaction between feedback, fairness, and relative performance suggests a complex cognitive-emotional mechanism underlying social decision making. Importantly, these neural effects occur even in the absence of real monetary stakes, demonstrating the profound influence of minimal social signals on perception of fairness.

Future research could build on these results by exploring how other forms of social feedback affect fairness judgments, especially in competitive or hierarchical contexts. It would also be valuable to investigate whether these effects persist in real-world scenarios involving material consequences, such as monetary exchanges or resource allocation, where fairness considerations are more directly tied to self-interest.

4.5.3 How People Perceived Fairness in a More Complex Environment: Different Roles and Types of Feedback

People often evaluate fairness not only on the basis of the outcome of an interaction but also on the perceived credibility or authority of the person involved. If a well-liked or trusted performer (high-score performer) gives a thumbs up, the action may be seen as more legitimate and fair. Conversely, if the performer is seen as a loner or untrustworthy, the same gesture can be interpreted as unfair or questionable. This highlights that fairness perception is not solely outcome-based, but also context-dependent on who is delivering the feedback. This result is consistent with the findings presented in Section 4.4.3.2.

An intriguing question to explore is how participants rate fairness in decision-making contexts. Specifically, does a high performer exhibit more generosity? Studies suggest that participants' decision making in the UG may be influenced by different cognitive processes. For example, rejecting ultimatum offers is often associated with intuitive heuristic-based thinking, while accepting offers tends to reflect more deliberate analytical decision making [17].

Moreover, fairness perceptions in the UG are complex. Both fair and unfair punishers coexist within the game, and punishment decisions are not easily categorized as purely prosocial or anti-social. Instead, they often involve a mixture of both motivations, indicating that fairness-related decisions in the UG are nuanced and context-dependent [13].

Although this study focused on the effects of reciprocal positive feedback (i.e., affirmation from a partner), future research could explore conditions where feedback is either nonreciprocal or nonaffirming. For example, what happens when a participant receives neutral or even negative feedback following a fair or unfair offer? Investigating these variations could help unravel whether the observed neural and behavioral effects are driven specifically by reciprocity, affirmation, or broader social dynamics such as perceived fairness, trust, or rejection.

4.6 Conclusion

Traditional ultimatum game studies often focus on fixed roles, typically examining the proposer–responder dynamic, rather than exploring how individuals respond within a shared team context. In such studies, participants are usually pre-assigned as either responders or allocators, which limits the investigation of more fluid or reciprocal interactions. In contrast, our study aims to examine how individuals react to the distribution of rewards within a collaborative setting, where roles and expectations are more ambiguous. We found that even a minimal nonverbal cue (a thumbs up) can significantly influence both behavioral and neural responses. Importantly, our results suggest that fairness perception is context-dependent; the same outcome can be interpreted differently depending on the individual's relative status or situational framing.

Previous research has shown that cooperative behavior in social decision-making is highly sensitive to contextual factors. However, the specific ways in which dynamic social cues modulate cooperation and their corresponding neural underpinnings remain insufficiently understood. Given the growing interest in promoting ethical behavior within organizational environments, we argue that it is equally important to investigate how individuals respond to perceived unfairness or unethical decisions. Our findings underscore the value of integrating behavioral measures with EEG data to capture a more comprehensive view of the cognitive and affective processes that drive decision-making in socially complex situations.

Chapter 5

Social Fairness: Designing a Fairness Dashboard to Enhance Group Collaboration

Uneven work division and assignment can demotivate group members as no matter how much workers contribute individually, the entire group can end up with sharing the same under-performing outcome, rendering individual contributions invisible and unaccountable. Although fair division of outcomes or costs has been studied extensively as an optimization or economic problem, there is limited understanding of how group members perceive work division in collaborative settings and how to better support fair negotiations that would encourage taking individual differences and equity in consideration. In this chapter, we propose supporting fair work assignments in groups with a group support interface, designed as a work division dashboard, that displays information on (1) parity of division: system recommendation of how to equally divide work (which may not be equitable), and (2) individual accountability in group work: visualization providing the awareness of how individual members' contributions added up as group outcome. We recruit 60 participants to play a *Group Tetris game* and negotiate work division using the supporting interface, offering understandings of how offering information of equal division and individual accountability affect fairness perceptions, group productivity, and the adoption of more equitable strategies in work assignment.

5.1 Overview

Human work is often bonded with groups. Many of the previous studies have looked at how group work may benefit individuals through aggregation and synergy [21, 81]. For example, by complementing the insufficient resources and capabilities of individuals, stimulating interpersonal discussions and inspiring new ideas beyond what separate individuals can think of, as well as all different types of tasks where working in groups can shine [101]. However, more and more studies have found that although people tend to believe in the additional values and effectiveness offered by group collaboration, the actual productivity gain of collaboration could be limited, depending on collaborative processes and contexts [34]. The opposite phenomenon of productivity loss or discoordination can lead to *groups performing less well than sum of their members' base contributions* [60, 105, 146]. Simply working in groups does not necessarily lead to increased performance.

In technological designs to support group work, practitioners have been primarily interested in improving the *performance* aspect of group outcomes. Past work examining issues around productivity gain or loss in work groups has also focused only on circumstances in which individuals will and will not exert great effort [65]. Various group awareness tool (GAT) designs have been proposed and studied, such as using numeric charts to show real-time participation [38] or visualizing language use in group conversations [86]. However, an aspect that is often overlooked in group collaboration studies is *workload assignment*. We believe that this is due to the inherent challenges in designing such studies. First, there is the lack of appropriate study paradigms to translate real-world group workload assignment scenarios into a controllable experimental setting. Second, it remains unclear what are the appropriate ways to measure and operationalize associated factors, such as accountability and fairness perception, to effectively capture the nuances of dynamics in group workload assignments.

To support fair assignment of work in groups, we understand the challenges of proceeding the research and look at theories across management science, behavioral economics, and psychology. We then focus on two common challenges in dividing and assigning work to group members. First, there is a tendency that individuals' motivation and group's goal for work are in conflict, known as *the social trap* in economy [117]. Second, there is no feedback that shows how one's contributions can lead to individual and group rewards. We propose to develop a fairness-oriented work division dashboard to support group discussion and decision making about work assignment for sequential collaboration tasks, specifically additive and compensatory tasks (see a more detailed explanation in Section 5.2.1). The dashboard displays information on (1) how to achieve equal division of work: system recommendation of how to equally divide the work (which may not be equitable), and (2) group awareness of individually accountable contributions: visualization showing how individual members' contributions added up as group outcome. Since the amount of effort individuals contribute to a group is multi-determined, and we should study the combination of multiaffecting factors, such as task difficulty, one's capability, and background difference all combined. We recruit 60 participants to play a group Tetris game that resembles real-life group work using different versions of the dashboard, offering novel understandings of how system recommendation and accountability affect fairness perceptions, group productivity, and the adoption of more equitable strategies in work assignment. This work contributes new understandings to support the equitable division and assignment of work that have been critical but understudied in the past [6]. We highlight key theories across studies that are relevant to designing fair and collaborative systems. In addition, we offer information on how the dashboard design should be structured, identifying the essential elements to include.

5.2 Types of Group Work and the Challenges of Working in a Group

5.2.1 Fairness and Different Types of Tasks in Group Work

Fairness in group work could be challenging to support, as groups consist of individuals with complex and sometimes conflicting interests, making work division difficult. Recent research has started to adopt a human-centered angle to navigate this issue, focusing on how fairness is perceived by people, which adds another layer of complexity to the inquiry of fairness on top of mathematical and algorithmic analyses [78]. Two key concepts of human-centered fairness have emerged: outcome fairness and procedural fairness.

Outcome fairness focuses on the distribution of results, assessing whether outcomes are fair, which aligns with the principle of equality, the idea that everyone should receive the same share of resources or rewards. However, fairness also involves the process through which these outcomes are determined. This leads to procedural fairness, or equity, which emphasizes the fairness of the decision-making process, particularly when individual differences are considered. Procedural fairness highlights the importance of the process itself in shaping perceptions of fairness, with factors such as transparency and inclusiveness playing a key role. Research suggests that when procedural fairness is perceived as high, individuals are more likely to view outcomes as fair, regardless of the result [155]. This underscores the impact of process-oriented fairness on outcome perceptions [14,15]. Our goal is to address the challenges of understanding group fairness, focusing on procedural fairness, and integrating both fairness of the process and equity of the results to promote a balanced approach.

Different types of tasks significantly influence group collaboration and how we assess group performance. Steiner identified five types of tasks, each with distinct combinatorial rules to determine the outcomes [143]. *Disjunctive tasks* depend on the best performing member, whereas *conjunctive tasks* are limited by the weakest member. *Additive tasks* sum individual contributions, meaning that every member's effort impacts the final result. In *compensatory tasks*, the group estimates a collective outcome by averaging individual contributions. Lastly, *complementary tasks* involve breaking down work into subtasks that are assigned based on each member's expertise. In our study, we focus specifically on additive and complementary tasks because these types of tasks often raise important questions about fairness in task division. The debate over how to divide work evenly and justly is particularly relevant for these task types, whereas other task structures, such as disjunctive or compensatory tasks, typically do not involve the same level of scrutiny regarding fairness since the outcomes are more directly determined by individual performance or averaged contributions, making the division of work inherently clearer.

5.2.2 Challenges in Ensuring Fair Group Collaboration

5.2.2.1 Group Collaboration Margin When Individuals' and Group's Goals Are Different

Recent studies have highlighted a limit to productivity gains in groups, and many groups did not function as expected [59]. In these groups, a *competitive* environment often emerges, where individuals prioritize their personal goals over collective goals, creating a dynamic similar to *social comparison*. When individual goals are overemphasized in group work, turning collaboration into competition, it can disrupt the overall functioning of the group. Fairness in collaborative work, including how members perceive fairness, is crucial to realigning and integrating individual and group goals. Research suggests that ensuring fairness can prevent the fragmentation of group efforts and foster more cohesive collaboration [167]. Concepts like *the prisoner's dilemma* and *the tragedy of the commons* vividly illustrate how shared resources and collective goods can be jeopardized when individual and group goals are misaligned. In these game-theoretic scenarios, each individual makes decisions based on their own rationality, but these decisions often lead to a collectively undesirable outcome [64, 115].

To mitigate this, it is essential to understand the type of information group members need in order to assess the interdependencies between their actions and the group's outcomes. Providing this insight can help reset expectations and reconnect individual goals with the broader objectives of the group, thus reducing competition and fostering more effective collaboration.

5.2.2.2 No Visible Contribution-Reward Model Available to Support Group's Self-Regulation

Individual group members often struggle to determine how work should be best divided and assigned to achieve group goals [22, 96]. In some cases, they rely on unverified or erroneous heuristics, such as assigning more tasks to high-performing members and excluding lower-performing individuals from contributing. Although these heuristics may seem intuitive, they could potentially harm the productivity of the group and, more importantly, undermine the fairness and motivation of members by denying their opportunities to participate [103]. Despite the well-documented impact of such practices, effective design solutions remain scarce. In this work, we propose implementation of the *social translucence concept* [45] to create a visible contribution-reward model accessible to all members. It will be a system recommendation tool that tracks each member's contribution and aligns the corresponding rewards based with those contributions.

5.2.3 Realigning Individuals and Group with Group Awareness and System Recommendation Tools

Few studies focus on enhancing group fairness by linking individual goals to group goals or visualizing individual contributions to group performance. However, research on group awareness tools has explored visualizing cognitive states (e.g., understanding of a topic) [156], behavioral traces (e.g., number of posts one contributes) [110], and social interactions (e.g., communication friendliness) [80]. These tools promote transparency and help prevent misunderstandings about the roles of each member and reduce issues such as social loafing, free-riding, and under-participation [27].

Various group awareness tool (GAT) designs have been proposed and studied, such as using numeric charts to show real-time participation [38, 74], using tag cloud visualization to highlight often-mentioned concepts by collaborative learners [112], or providing a visual summary of group work characteristics and activities [51]. Although group awareness allows people in a group to be aware of what everyone else is doing [40], the emphasis can still be on how "individual contributions" compare, which promotes the competitive thinking of how one obtains a larger share of outcomes by working hard [50, 73]. That is, the GAT designs, if not properly constructed, can still encourage individual members to compete with one another, especially when the system rewards individuals with high performance and imposes penalties on low performers. Competition is often considered a means to improve productivity and efficiency in individual work. However, group work also relies on group processes that provide mutual support and synergies [52]. There is a need to design GAT to motivate individuals to realign individual and group goals, consider diversity in individuals' needs and constraints, and collectively achieve equitable assignments to accomplish as a group [7].

In addition, there can be direct interventions, such as system-generated recommendations, to regulate group members' behaviors, such as assigning roles, giving evaluation (assessments) feedback or applying incentives and penalties to direct one's attention and behaviors in ways desirable to the group [32, 98, 127]. System recommendations have been commonly developed and used in productivity tasks. For example, studies have investigated *intelligent task routing*, where a system can locate and assign tasks that are suitable to individual workers in a community to ensure that optimal work allocation is achieved considering the needs of the community, the characteristics of the task and the expertise of individual workers [30, 31]. These recommendation tools are shown to regulate and increase members' contributions; however, they tend to focus on individual productivity and efficiency, but do not consider other social and perceptual constructs such as fairness in the assignment of work.

5.3 The Present Study

Our study aims to explore the design space for developing a work division dashboard that displays information to promote a more equitable work division and assignment among members of the group work. During a group's discussion of how to assign work to its memberships, we consider the effects of (1) the indirect intervention of providing the group with a shared awareness of how individual contributions add up to the group's shared outcomes, and (2) the direct intervention of providing equality-based recommendation to assign work to individual members of the group in a game-based study of sequential collaborative gameplay with three-person groups.

5.3.1 Our Proposed Workload Division Dashboard

To explore how system recommendations and group awareness affect fairness and productivity, we design four versions of a workload division dashboard to aid group discussions on task allocation (see Section 5.3.5 for study design). All versions share a core feature that allows players to specify their preferred workload, which is operationalized as *play time* in this study, using a scroll bar. Versions differ in the availability of two types of supportive information displayed: system recommendation, and/or group awareness indicators. These variations help assess how either design factor jointly influence the group's ability to negotiate a fair and productive task distribution.

As shown in Figure 5.1, the dashboard has several regions. Region (a) displays system recommendations, suggesting how much time a player should contribute (e.g., "You should play xxx seconds in the second round"). The individual bar chart in region (c) and the group aggregation chart in region (d) show the recommended contributions based on an equal fairness model (see Section 5.3.2). Region (b) includes a scroll bar for players to specify their preferred play time, initially set to 120 seconds for equal distribution. Scrolling left reduces play time, while scrolling right increases it. The label "The time you selected: xxx seconds" shows the exact chosen time. Region (d) displays group performance, allowing players to see how their individual efforts contribute to the group's overall score, including projected points if they follow the system's recommendation.

5.3.2 The Contribution-Reward Model

The system recommendation is based on a contribution-reward model, which links individual contributions to corresponding rewards. This model can be represented in a two-dimensional space, where individual efforts (e.g., time allocated to players) are mapped to their outcomes (e.g., individual scores). It follows the concept of the *expected contribution method* [98], which forms the basis for various methods used to adjust for differences in contributions [165], and the goal is to adjust the group's overall grade by incorporating a deviation from the expected contribution.

In this contribution-reward model, the diagonal line in the two-dimensional space represents

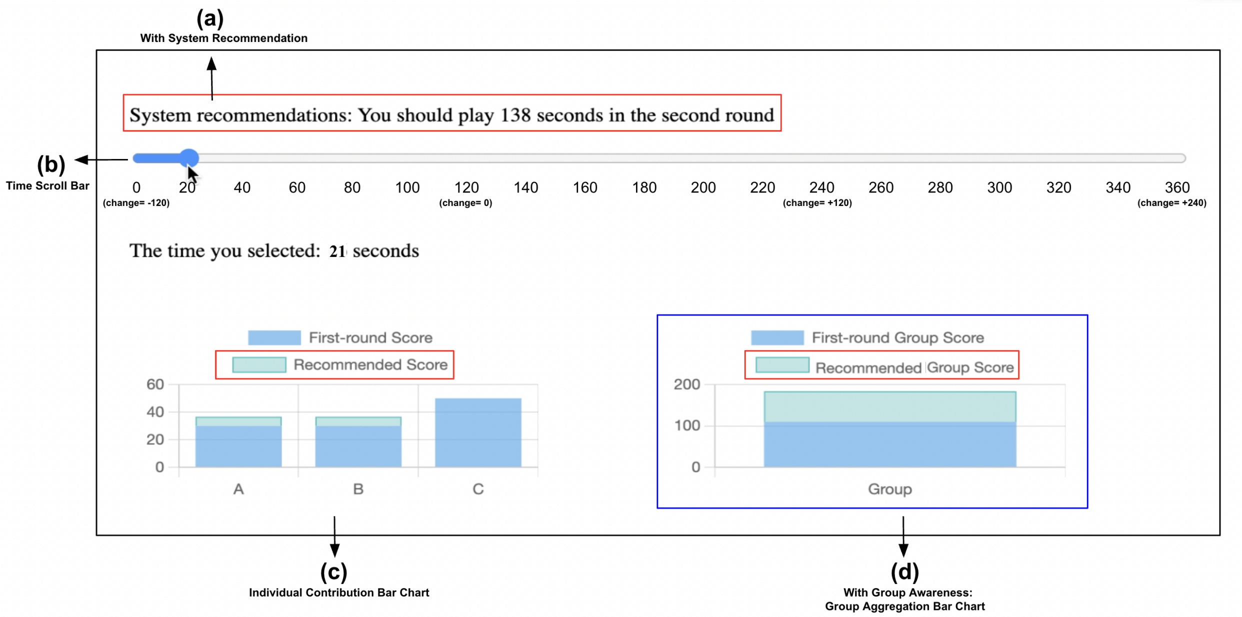


Figure 5.1: Workload division dashboard design. (a) shows system recommendation, (b) is the time scroll bar, and below it shows how much time you choose, (c) shows individual contribution bar chart, and lastly, (d) shows group aggregation bar chart, which resembles group awareness. This is the most comprehensive design in our study, and other designs for different study conditions follow the same logic, with one or two components taken away.

equal contributions and equally shared rewards. Deviations from this line show an under- or over-contribution relative to the group reward. The model recommends future contributions based on past performance. For example, if Player A earns 10 points, Player B earns 20 points, and Player C earns 30 points, the system will suggest that Player A contributes at least 10 more points in the second round to align with the group's total score of 60 points, with the goal of an equal distribution of contributions. Additionally, the model ensures accountability by requiring all members to contribute to the full 6 minutes of play time; if one player contributes less, the others must play more.

5.3.3 The Group Tetris Game

To simulate real-world work division processes within groups, we design a modified version of the classic Tetris game, referred to as *The Group Tetris Game*. Our aim is to use specific settings of the game to reflect important areas of consideration, such as how members divide workload in real group work. See Figure 5.2 for a screenshot of the game.

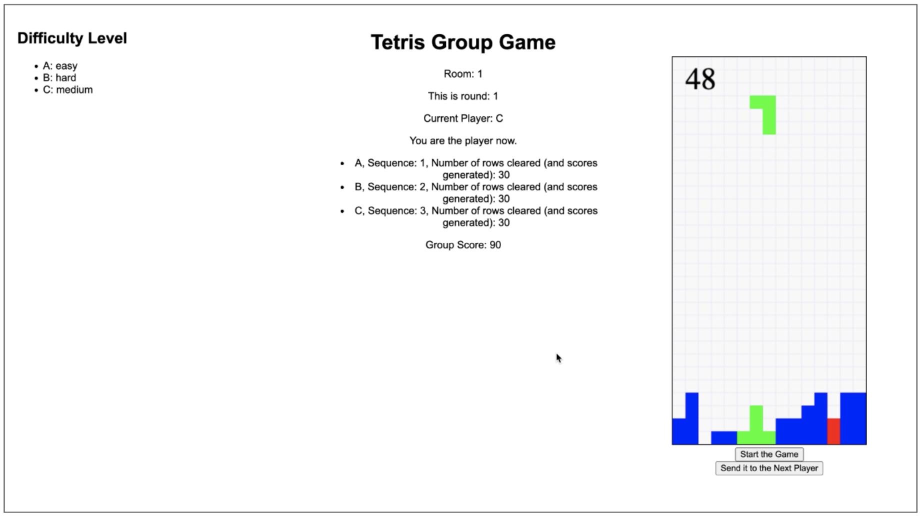


Figure 5.2: The Group Tetris game platform. The panel on the left shows the difficulty level assigned to workers; the middle section is the playing sequence and individual participants' playing scores, which increase 10 everytime when a solid row is cleared; the panel on the right is the canvas for playing the Tetris game, which also displays the remaining time (in seconds) of each member's turn.

We selected Tetris for this study due to its widespread familiarity and its balance of simplicity and challenge [1]. The game rule does not require prior training, making it accessible to participants without introducing a steep learning curve. Tetris, however, can be challenging to play as the game progresses, effectively captures aspects of cognitive work that demand attention and decision making in a timely manner.

In the Group Tetris game, three players participate in a 12-minute game split into two rounds (each round is 6-minute). In the first round, each player plays for 2 minutes. In the second round, players can adjust their play times, but the total still sums to 6 minutes. The players earn 10 points for each cleared row, and the group's total score is the sum of individual scores. Participants receive a \$10 Amazon gift card for participation, with an additional \$10 prize for the winning group to ensure full engagement.

At the start of the game, players are randomly assigned an order (first, second, or third) and a difficulty level (easy, medium, or difficult), both of which remain fixed. The difficulty level is based on the shapes of the tetrominoes one gets. Tetrominoes with complex shapes (e.g., more edges) are commonly considered more difficult to play. The game follows a sequential turn-based structure: The first player plays for 2 minutes, followed by the second player, and then the third player. During this time, members can observe their each others gameplay on their screens. The game *does not reset* between turns and the players build on each other's progress, reflecting the additive nature of group work. After the first round, individual and group scores are displayed, and players enter a negotiation phase to discuss and decide on their respective workload assignments for the second round. At this stage, players are presented with four different visualization layouts. During this phase, players face the dilemma between pursuing individual interests and cooperating to maximize the group score, mirroring the real-world group dynamics.

5.3.4 Research Questions

While prior work in the literature has initially explored the effects of group awareness and system recommendation tools on collaborative work in different contexts, the present study aims to further examine (1) the joint effects of the two types of intervention designs in collaborative work and (2) the effects of the interventions on not just productivity, but also perceived fairness and specific strategies for work assignments discussed and adopted by groups.

Specifically, we aim to investigate how the experimental manipulation of the system recommendation and group awareness features in the proposed dashboard will impact two key outcomes: **individuals' fairness perceptions (RQ1)** and the **group's productivity (RQ2)**. Given that groups are free to discuss and negotiate work assignments, which can be guided by different work allocation strategies that emphasize values such as equity versus equality, we are also interested in understanding how these interventions influence the **strategies** that groups adopt for dividing and assigning tasks **(RQ3)**. Finally, recognizing that engagement in open communication and negotiation is a critical component of procedural fairness, we seek to explore how different dashboard designs affect the **intensity of communication** within the group **(RQ4)**.

5.3.5 Study Design

As shown in Figure 5.3, during the negotiation stage, different versions of *workload distribution dashboard* are provided to the groups to support their discussion. We conducted a 2 by 2 mixed experiment that manipulates the availability of system recommendation (with or without visibility of system suggested points contribution or not - based on the contribution-reward model described in 5.3.2) as a between-subject factor, and the availability of group awareness display (with or without visibility of the connection of one's individual performance to the group or not) as a within-subject factor. The order of which group awareness condition was presented to groups was counterbalanced in prevention of ordering effects. All of our four different experiment conditions is summarized in Table 5.1. In the following paragraphs, for simplicity, we will sometimes call the system recommendation *recommendation*, and the group awareness *awareness*.

	W/ awareness	**W/O awareness**
W/ recommendation	W/ recommendation, W/ awareness	W/ recommendation, W/O awareness
W/O recommendation	W/O recommendation, W/ awareness	W/O recommendation, W/O awareness

Table 5.1: Four experiment conditions. We run a between-subject design on with or without recommendation, and a within-subject design on with or without awareness. These conditions are counterbalanced to prevent ordering effects.

Figure 5.3: Full study procedure for the Group Tetris Game. Each trial consists of two rounds of gameplay, where all players play for the same amount of time in the first round. After the first round, participants enter a negotiation stage, where they review their performance and discuss the allocation of individual playtime for the second round. Four different versions of the dashboard are designed in this stage to study the effects of system recommendations and group awareness on players' decision-making.

5.3.6 Participants and Procedure

We recruit 60 participants from campus recruitment flyers to form 20 three-person gaming groups. All participants completed a pre-experiment survey regarding demographics and prior experience of working in a group. The result shows that 53.2% of the participants were female and 46. 8% were male. Their age ranged from 18 to 35 years, and they were students and company workers, all of whom had prior experience working in groups.

Participants were randomly assigned to one of the two conditions between subjects: with versus without system recommendation, and participated in two trials of sequential collaboration. In each collaboration trial, the group will have access to the group awareness display or not during workload negotiation. As mentioned, the order in which group awareness condition is present first is counterbalanced to account for the possible ordering effect in within-subject design. We ensure that the participants do not know each other. The study was conducted online through the Zoom video conferencing system (with audio support) in an effort to ensure the consistency of the study protocols under different conditions. We allow participants to speak because their conversational data, particularly during the negotiation stage, is one of the key focuses of our group collaboration analysis.

5.3.7 Measure and Analysis

As we want to evaluate how the dashboard affects the fairness perception and group productivity of group members, we gather different sources of data, including *fairness survey* and *behavioral log* throughout the game trials. The Fairness In Work (FIW) scale is developed by the authors to assess how group members perceived fairness in collaborative work. The scale consists of five questions to cover three aspects of fairness perceptions: self-assessment of individual contribution; treatment from group members; and the connection between individual performance and group performance (see the following paragraphs for detailed questions). The questionnaire items used 5-point Likert scales anchored by (1) "strongly disagree" and (5) "strongly agree", such that higher scores indicate higher perceived fairness in group work. While behavioral logs are extracted from participants' gaming behaviors, which is the group score improvements between the two rounds of game.

A fairness score is calculated by averaging the responses to the five questions (items) of the fairness survey (Cronbach's alpha =0.72). Whereas, productivity is the improvement of game performance after negotiation, which is calculated by subtracting second round group score from the score of the first round of game. We reference Steiner's models for inferring relationships between group size and potential group productivity [143].

The Fairness In Work (FIW) Scale

1. I feel the amount of time I played is reasonable considering the sequence and difficulty level.
2. Given the time, I feel I deserve the amount of individual score I received.
3. I feel I was not being taken advantage of.
4. I feel the amount of score that I contribute is reasonable compared to the final group score.
5. I feel my group members did their allocated tasks properly.

Measurement

The following formulas are for fairness and productivity calculation:

$$\text{Fairness} = \frac{\text{sum of individual's responses on the fairness items}}{\text{number of items}}$$

$$\text{Productivity} = \text{second round group score} - \text{first round group score}$$

5.3.8 Identifying Work Division Strategies in Negotiations

By analyzing the conversational scripts during the negotiation stage of the 20 groups, we categorize each group into one of the four categories according to their common strategy used to determine the division and assignment of the work for the second round of game play: *performance group*, *equality group*, *equity group*, and *others.* A performance group indicates that group decisions were made based on the performance or capability of group members in the first round (e.g., who is the strongest player). An equality group indicates that group decisions were made simply by dividing the playing time into three halves. An equity group indicates that the group decisions were made based on individual differences, for example: computer-assigned sequence or difficulty level, and lastly, there are undefined others, which indicates there is not much discussion or group members just simply chose to follow system recommendation for convenience. The table below (Table 5.2) is a summary of the four types of negotiation strategies.

Strategy	**Definition**	**Example from Transcripts**
Performance	Based on players' first-round performance and capability	- *Stronger player should play more to gain more points in the second round.* - *I gained the least score previously, can I play less in the next round?*
Equality	Share time equally or consciously say everyone's portion should be equal	- *I think the way to earn more points is that everyone plays carefully and not make mistakes. I think we should all play 120 seconds and, by focusing, everyone should be able to gain reasonable points.*
Equity	Take into consideration individual differences, for example: the player's assigned difficulty level or sequence	- *I think the first player should play the longest since he can set a good foundation for the following players.* - *Since easy games encounter more straight bars, we should let the player who gets an easy game play longer for better second-round score.*
Others	Simply follow system recommendation, no talk or cannot be defined	- *Tell me what you think, and I will follow.* - *How about let's just follow the system's recommendation?*

Table 5.2: Summary of the four types of negotiation strategies.

Since the negotiation script for each group is composed of several strategies brought up by different members, we follow the *turn-taking model* from Sacks, Schegloff and Jefferson, who have described how conversationalists construct *turns or changes* at speaking and how they allocate them in a systematic way [133]. We segment each group's negotiation data into several units, and each unit has one specific strategy type (mentioned by any of the players). The entire scripts can be composed of multiple units, depending on the time length of their negotiation, or the number of turns members made. A group's final label is determined by two rules: (1) we first follow the majority strategy's unit count, if there are identical numbers and we cannot rank, then (2) we follow the player's suggestion that all members agreed upon. We pre-process the transcripts to anonymize the participants. The following example shows that there were 5 units, and this group's work division strategy was labeled as an *equity* group since equity has the majority unit count (5 units with equity: 3, performance: 1, equality: 1).

> Player A: Since player B got the easy difficulty level, she should play more for the second round group score to be higher **(unit 1: equity)**.
> Player B: I don't think so, the important thing is what kind of blocks we get. Easy will have a lot of long bars and squares which are not great to clear **(unit 2: equity)**.
> Player A: But you and Player C's scores are higher, you two should play more to get the group score to be higher **(unit 3: performance)**.
> Player B: 10 points is very minor. The score everyone gets is very similar. We should play equal time. **(unit 4: equality)** Or how about I play a bit more since I get easy **(unit 5: equity)**, and you two can share the remaining time equally? Let's just use the first round as a trial, we can discuss time division again in the next test.

Two independent coders perform the coding task. Inter-coder reliability based on 10% of the data was satisfactory (Cohen's Kappa=0.8). The mean of the total negotiation units per negotiation episode is 3.68 ($SD = 1.53$), with performance accountable of 1.43; equality 1.05; equity 1.15, and others 0.05.

5.4 Results

As we aim to understand whether independent variables *with or without system recommendation* or *with or without group awareness* affect dependent variables such as *fairness perception* or *productivity* in group work, we apply mixed model ANOVAs to account for possible local interdependency between data points caused by group collaboration and repeated measures in the study design. Two linear mixed models (as shown in Figures 5.4 left and right, respectively) were created to model the variance of group members' *fairness perception* and groups' *productivity* scores. For the linear model to model the fairness score, individual participant nested in the group, and group was included as random effects. System recommendation, group awareness, and the interaction between the two variables are included as fixed main effects in the model. Whereas, for group productivity, of which the score is calculated from group score improvements between the two rounds of game at the group level, group is included was a random effect, while system recommendation, group awareness, and the interaction term of the two variables are the fixed effects. We conduct the mixed-model regression analysis using SAS JMP 18.

5.4.1 Fairness Perception

By using fairness score as the dependent variable of a linear mixed model, we want to answer **RQ1**. There is a main effect of system recommendation, $F[1, 18] = 10.81, p < 0.001$, as well as a main effect of group awareness on fairness score, $F[1, 58] = 26.29, p < 0.001$. From Figure 5.4 (Left), we can see that with system recommendations, the scores are higher than without recommendation; and with group awareness, the scores are higher than without awareness. Perceived fairness is the highest when both recommendation and awareness are available ($mean = 4.47, SD = 3.8$). In contrast, without recommendation and without awareness has the lowest fairness score ($mean = 3.60, SD = 3.08$). The result shows that recommendation and awareness are both effective intervention, and it would be ideal to combine both for fairness enhancement.

5.4.2 Group's Productivity

We include group productivity as the dependent variable in a linear mixed model at the group-level to answer **RQ2**. Contrary to the results on fairness perception, we can see in Figure 5.4 (Right) that there is a significant interaction effect between the availability of system recommendations and group awareness on group productivity, $F[1, 18.1] = 4.07, p = 0.059$. With recommendation and without awareness reaches the highest productivity improvement ($mean = 39, SD = 21.5$), and without recommendation and without awareness results in the lowest improvement on productivity ($mean = 12, SD = 9.24$). With or without awareness overall does not have significant impact on productivity, $F[1, 17.8] = 0.006, p = n.s.$, while with or without recommendation has significant effect $F[1, 17.8] = 2.55, p < 0.05$. The result shows that the availability of system recommendation as an external support matters for productivity, especially there is no group awareness available; however, if group have awareness indicators, with or without recommendation does not have too much impact on group's productivity.

Compared with the previous result on fairness perception, we note that recommendation and awareness factors can be utilized as support for different goals. For fairness goals, internal regulatory mechanisms like group awareness indicators are essential, while at the same time also receives visible improvement in productivity. For productivity goals, both recommendation and awareness interventions could be potentially useful, but recommendation can risk lowest productivity improvement when there's not awareness of how individual contributions connect to group, and also risk low fairness perception.

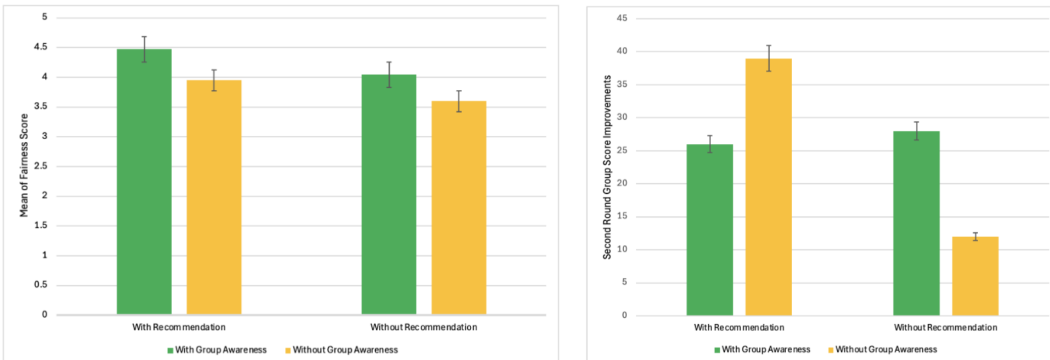


Figure 5.4: (Left) fairness perception; (Right) productivity improvement across conditions with/without system recommendation and with/without group awareness indicators. Means and standard errors were estimated by linear mixed models.

5.4.3 How Recommendation and Awareness Tool Affect Work Division Strategies in Groups

We now examine the negotiation strategies that participants used under different conditions that could affect fairness and productivity. The raw frequencies of the strategies mentioned during the negotiation in percentages are shown in Figure 5.5.

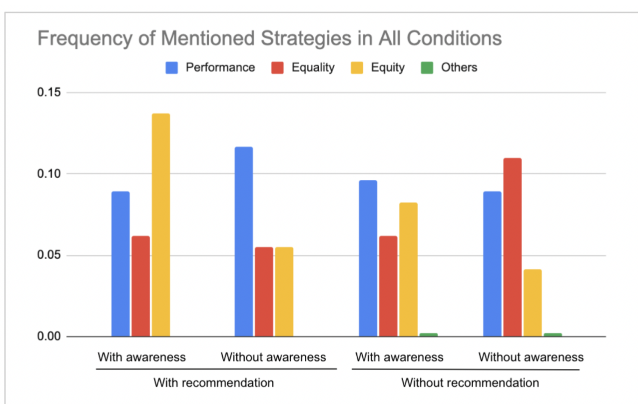


Figure 5.5: Frequencies of mentioned strategies (in percentages) across different conditions.

5.4.3.1 Group Awareness Enhances Equity in Group Work

Based on the frequencies of strategies mentioned as shown in Figure 5.5, we could answer **RQ3** regarding how different interventions affect negotiation of work. We noted that both with awareness conditions created higher equity groups (count=96) than those without awareness (count=42). Specific linear mixed models (Figure 5.6) are created with respect to the groups' *number of equity strategy mentioned* during the negotiation. As we set the number of equity strategies mentioned as the dependent variable, group awareness has a significant impact on creating equity groups, $F[1, 18] = 10.41, p < 0.05$; while system recommendation has no impact $F[1, 18] = 2.88, p = n.s.$.

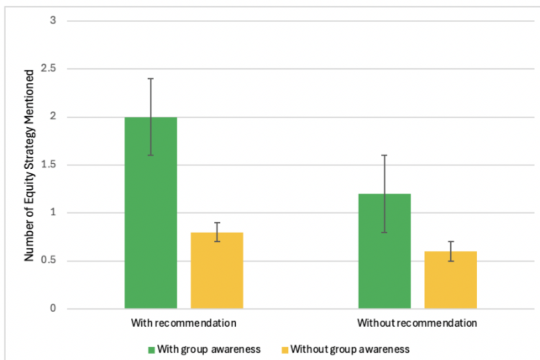


Figure 5.6: Equity strategy count per group with recommendation (or not) and with group awareness (or not) conditions. Means and standard errors were estimated by linear mixed model.

5.4.3.2 System Recommendations Encourage Further Negotiation

Although system recommendation does not have a significant impact on the creation of equity or equality groups; however, with recommendation, people tend to negotiate more ($F[1, 18] = 4.84, p < 0.05$; see Figure 5.7). In conjunction with correlation analysis, longer negotiation times are highly correlated with a higher negotiation unit, $p < 0.001$, and the negotiation unit is highly correlated with mention of equity count, $p < 0.001$. Combined with 5.4.3.1, a mix of awareness and recommendation tools will provide the most equity groups, which leads to a higher fairness perception score. This is aligned with our findings in 5.4.1, and further answer **RQ4**.

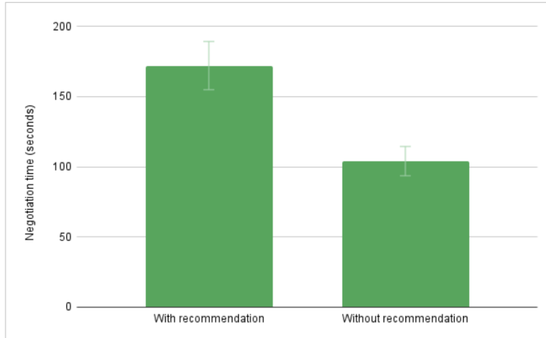


Figure 5.7: Negotiation time is significantly longer in with-recommendation conditions.

5.5 Discussion

5.5.1 Visualizing Group Ability: Using Self-Regulation to Bridge Actual and Potential Performance

As highlighted in the related work section, much of the existing research on group collaboration has focused primarily on measuring actual group performance and outcomes. Fewer studies have explored the *potential* of a group: the unrealized performance a group could possibly achieve, which is also mentioned as *group ability* in most papers [62, 136]. Although many articles emphasize the importance of group ability, understanding the internal mechanisms and surrounding contexts of it can provide deeper insights into group dynamics, task allocation, and member formation [121]. To our knowledge, there has been limited exploration in this regard, especially in the context of fair group collaboration. Group collaboration involves both intrapersonal and interpersonal actions through which individuals transform resources into decisions. However, many non-productive behaviors, triggered by frustration, competing motivations, or inadequate understanding, are often overlooked previously. In addition, the nuanced nature of collaboration matters. Even in additive task scenarios, where group performance is the sum of individual contributions, it can be challenging to ensure that all members align and maximize their contributions. Understanding the gap between the actual performance of a group and its potential is crucial to improving group dynamics and motivating sustainable team performance [143].

In our dashboard design, which includes system recommendations, we observe significant improvements in group productivity, the difference in scores between the first and second rounds of the game. Although the recommendation of the system serves as a self-regulatory tool in our study, by clearly presenting this difference, the visualization of the gap between actual and potential performance aligns with *the goal setting theory*. When individuals are presented with achievable goals based on their potential, they are more likely to set higher goals and strive to achieve them [92]. The self-regulation visualization not only helps distinguish what groups currently accomplish and what they have the ability to accomplish, but also improves the shared understanding of member behaviors. Ultimately, by highlighting the potential of a group, this approach can increase motivation, drive performance, and foster greater group productivity.

5.5.2 Designing for Engagement: Leveraging Leadership Facilitation and Transparency to Combat Social Loafing

As Latane, Williams, and Harkins introduced the concept of social loafing, the tendency of individuals to exert less effort when working in a group compared to when working alone, it has been recognized as one of the main factors that affect productivity [85], which also pose threats on fairness perceptions. Many studies have explored ways to mitigate this effect, including strategies to increase personal accountability [108], enhance group cohesiveness [159], clarify task meaning [140], and implement system tools that actively observe and track individual contributions [41]. However, the characteristics of groups have evolved over time [61]. Traditionally, work groups in organizations had a designated leader. Today, many groups are self-managing, with members sharing leadership responsibilities and collectively monitoring and managing their own activities [97]. Although previous research suggested that social loafing could be prevented with an external regulatory authority, this approach no longer fully applies to modern group structures. In the current landscape, designing tools that encourage active participation of all members of the group is crucial.

Our dashboard does not require a designated leader; instead, all members have the autonomy to take actions they believe are best for the group. It represents the key element of *leadership facilitation* [158], empowering team members to make decisions, resolve conflicts, and assume greater responsibility. Importantly, there are no restrictions on members' participation in the system; the tool will serve solely as a facilitator of collaboration, rather than a prescriptive authority, making it more likely that members will actively participate. This active participation helps counteract social loafing, as individuals are not passive recipients of tasks, but are instead empowered to shape their participation. In addition, social transparency can act as a preventive measure against social loafing by making individual contributions visible and holding members accountable, as in our design of group awareness indicators. Transparency, which is strongly correlated with effective communication, suggests that a tool aimed at reducing social loafing should also enhance group communication. By providing real-time feedback on individual efforts and fostering an open environment for discussion, such a tool can create an atmosphere where each member feels responsible for the group's success and motivated to contribute fully. The manipulation of group awareness in the dashboard also proved to be effective in influencing negotiation strategies, promoting a larger number of equity groups, and improving individuals' perceptions of fairness.

5.5.3 The Dilemma of Focus in a Collaboration Group

We begin by investigating how different dashboard designs can enhance fairness perception in the division of the group workload. We then examine the impact of these designs on performance, specifically analyzing improvements in scores. Our findings reveal an intriguing insight: fairness and productivity are not necessarily correlated, and also are not necessarily at odd. A fairer group does not always lead to lower productivity, and vice versa. This raises an important question. Should we still think to prioritize just one of the two outcomes in group processes without considering one another?

This answer is not straightforward, as both fairness and productivity are essential to effective, healthy group dynamics, yet they may not always align. Fairness is essential to fostering trust and cohesion within a group, promoting greater cooperation and long-term collaboration. However, productivity, measured by tangible results, such as performance scores, directly reflects a group's ability to achieve its goals and meet deadlines. It is crucial to consider the broader context: a sole focus on productivity could overlook the interpersonal factors that contribute to a healthy and sustainable work environment. Imbalances in workload division can lead to frustration, burnout, or disengagement, which in the long run could harm productivity.

Thus, the choice is not an either-or proposition, but a matter of finding a balance between the two, such as through the design and use of supportive and regulatory tools, as illustrated in our study. Future research could explore how to further fine-tune this balance, potentially through dynamic dashboard designs that both enhance fairness and adapt to varying productivity needs across different group compositions and contexts (e.g., when members have distinct skill levels and abilities to contribute). Ultimately, we need to consider not just the nature of the task and other short-term and long-term objectives of the organization, but also emerging opportunities introduced by algorithmic and data-driven tools for co-optimizing both fairness and productivity goals.

5.6 Conclusion and Future Work

We examine key factors that influence group fairness and performance, such as individual differences, the balance between individual and group rationality, and the role of self-regulated reward systems. Our findings indicate that equity-based groups, where individual backgrounds and abilities are considered, generally have the highest perceived fairness. However, this approach does not always translate into the highest productivity. Additionally, we develop a workload division dashboard that differs from traditional group awareness tools, which typically focus on tracking individual contributions. Our dashboard encourages members to set collective goals and evaluate both positive and negative group interactions. This collaborative approach helps balance fairness with productivity, optimizing group dynamics for individual and collective success. For future work, we would like to incorporate real-time feedback and dynamic adjustments of the workload distribution. As group dynamics changes (e.g., if some members dominate or suffer), the system should provide timely feedback to adjust the workload distribution accordingly. This ensures that fairness and sustainable productivity is maintained throughout the collaboration process, even as individual contributions of group members fluctuate over time.

Chapter 6

Lesson Learned and Future Work

Chapter 3, Chapter 4, and Chapter 5 in this thesis, spanning from redistricting, psychological perception to collaborative task division indicates the multidimensional nature of fairness. Each chapter also corresponds to a distinct dimension: material fairness (formal rules-based systems), cognitive fairness (how individuals perceive and internalize fairness), and social fairness (how fairness emerges and stabilizes through group negotiation and moderation). By addressing these dimensions alongside with each other, I highlight the limitations of single-perspective solutions and present a comprehensive framework for fair division that is both theoretically grounded and practically applicable. The novelty of this work lies not only in connecting these disparate dimensions, but also in demonstrating how fairness can be systematically understood, measured, and improved across domains. Rather than seeking a universe solution, I use algorithmic and human-centered approaches to advocate flexible, interpretable, and context-aware mechanisms, where algorithms, perceptions, and social dynamics should coexist.

6.1 Practical Usability of Fair Division

The dissertation lays the foundation of understanding fair division through the lenses of algorithmic, psychological, and social fairness. However, there are several promising directions remain for future exploration. First, a key limitation in the current studies is the focus on relatively bounded domains and group sizes. Future research could investigate how the proposed frameworks scale in more complex, high-stakes environments, such as large-scale online platforms, public policy debates, or workplace resource negotiations. Second, incorporating longitudinal studies would help examine how perceptions of fairness shift over time and whether interventions like feedback or transparency tools lead to lasting changes in behavior or trust. Lastly, advancing the research into real-world applications, such as participatory budgeting, classroom tools, or civic technology, offers the opportunity to co-design with users and validate fairness models in ecologically valid settings. These directions aim to deepen the impact of this dissertation by moving beyond theoretical fairness to build systems that are not only equitable but also negotiable, interpretable, and socially robust.

6.2 Potential and Implication of Human-AI Collaboration Group

With recent advances in pattern recognition, predictive modeling, and decision-making, AI is increasingly being positioned as a cognitive tool to increase human capabilities and reduce labor costs. It enables people to accomplish more with fewer resources [16]. Cognitive tools that offload or extend human cognition are not new; even analog tools such as written notes or visual charts have long helped workers organize, remember, and reason effectively [25,28]. What sets AI apart is its adaptive learning ability, which enables it to handle more complex cognitive tasks such as synthesizing information, generating personalized recommendations, and making high-stakes decisions in domains such as finance or medicine.

This evolution raises important questions. Although AI promises to improve productivity, it also poses challenges to human agency in the workplace, fueling anxieties about job displacement or automation-driven inequality [9,49]. Discussions around responsible AI design have focused mainly on balancing automation with human control and on preparing workers through up-skilling and reskilling efforts [2,138].

However, an increasing number of people start to view AI not only as a tool but as a collaborative actor in team-based work. When AI is introduced as a partner that shares responsibilities with human teammates, the questions of coordination, communication, and fairness become crucial. Workers must negotiate not only tasks and timelines, but also perceived abilities, limitations, and contributions of human and AI collaborators. In such Human-AI teaming environments, fair work allocation is a key concern. Unfairness, whether real or perceived, can lead to disengagement, distrust, and poor team performance, similar to human-only teams. It is important to understand how fairness in work allocation is perceived and conceptualized by human workers and can be systematically further supported.

For future research, I propose expanding this work by exploring how fairness is perceived, negotiated, and operationalized in Human-AI collaboration. Specifically, it remains unclear how human workers coordinate and share workloads with AI teammates and how they evaluate fairness in such settings. For example, how do they interpret the role of AI, tool, peer, or something in between? What expectations do they have for equitable contribution and accountability? This research could help uncover how fairness norms in workload division adapt (or fail to adapt) when AI enters the team, and how systems can better support transparent, interpretable, and socially acceptable Human-AI workload negotiation. Additionally, it would pioneer the understanding of fairness in hybrid teams and offer concrete design directions for the future of collaborative AI.

Chapter 7

Conclusion

The dissertation contributes to the development of a unified framework that integrates material, cognitive, and social fairness (extending beyond traditional notions of outcome and procedural fairness). It seeks to deepen the understanding of how fairness is perceived by humans and introduces a whole new concept: *negotiated fairness*.

Negotiated fairness encompasses multiple dimensions of fairness and combines both algorithmic and human-centered approaches. It allows individuals with different objectives to reach mutually satisfactory outcomes through communication and transparency. It also provides a more holistic understanding of fairness, recognizing that no single perspective is sufficient on its own. My work offers a novel approach to designing fair division systems that are not only in theory but also practically effective and socially acceptable. Although I highlight outcome, procedural, and informational fairness as distinct aspects, future research could develop integrated models similar to the concept of negotiated fairness that dynamically balance the trade-offs between them. Such models could leverage adaptive or hybrid systems that respond to evolving user feedback and group dynamics, with particular potential in the context of collaborative environments between humans and AI.

Bibliography

[1] Dan Ackerman. *The Tetris effect: the game that hypnotized the world.* Public Affairs, 2016.

[2] Sayed Fayaz Ahmad, Heesup Han, Muhammad Mansoor Alam, Mohd Khairul Rehmat, Muhammad Irshad, Marcelo Arrano-Munoz, and Antonio Ariza-Montes. Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications*, 10(311), 2023.

[3] A.K. Austin. Sharing a cake. *The Mathematical Gazette*, 66(437):212–215, 1982.

[4] Lawrence Ausubel and Paul Milgrom. The lovely but lonely vickrey auction. *Combinatorial Auction*, 17, 01 2006.

[5] Daniel Balliet, Joshua M. Tybur, and Paul A. M. Van Lange. Functional interdependence theory: An evolutionary account of social situations. *Personality and Social Psychology Review*, 21(4):361–388, 2017.

[6] Wendy L. Bedwell-Torres, Eduardo Salas, Gregory Funke, and Benjamin Knott. Team workload: A multilevel perspective. *Organizational Psychology Review*, 4:99–123, 04 2013.

[7] Beersma Beersma, John. R. Hollenbeck, Stephen E. Humphrey, Henry Moon, Donald. E. Conlon, and Daniel R. Ilgen. Cooperation, competition, and team performance: Toward a contingency approach. *Academy of Management Journal*, 46(5):572–590, 2003.

[8] David C. Bell. Connection in therapeutic communities. *The International journal of the addictions*, 29(4):525–543, 1994.

[9] Sudip Bhattacharya. Artificial intelligence, human intelligence, and the future of public health. *AIMS Public Health*, 9(4):644–650, 2022.

[10] Maarten A S Boksem and David De Cremer. Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Social neuroscience*, 5(1):118–128, 2010.

[11] Megan A. Boudewyn, Steven J. Luck, Jaclyn L. Farrens, and Emily S. Kappenman. How many trials does it take to get a significant erp effect? it depends. *Psychophysiology*, 55(6):S285–S300, 2018.

[12] Robert Boyd, Herbert Gintis, Samuel Bowles, and Peter J. Richerson. The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences*, 100(6):3531–3535, 2003.

[13] Pablo Branas-Garza, Antonio M. Espin, Filippos Exadaktylos, and Benedikt Herrmann. Fair and unfair punishers coexist in the ultimatum game. *Scientific reports*, 4(6025), 2014.

[14] Joel Brockner. Making sense of procedural fairness: How high procedural fairness can reduce or heighten the influence of outcome favorability. *The Academy of Management Review*, 27(1):58–76, 2002.

[15] Joel Brockner, Ya-Ru Chen, Elizabeth A. Mannix, Kwok Leung, and Daniel P. Skarlicki. Culture and procedural fairness: When the effects of what you do depend on how you do it. *Administrative Science Quarterly*, (1):138–159, 2000.

[16] Alison Burke. Group work: How to use groups effectively. *The Journal of Effective Teaching*, 11(2):87–95, 2011.

[17] Dustin P. Calvillo and Jessica N. Burgeno. Cognitive reflection predicts the acceptance of unfair ultimatum game offers. *Jugment and Decision Making*, 10(4), 2015.

[18] Colin Camerer. Behavioral game theory experiment in strategic interaction. *Journal of Socio-economics*, 32, 12 2003.

[19] Colin Camerer and Richard H. Thaler. Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives*, 9(2):209–219, 1995.

[20] Ioannis Caragiannis, Christos Kaklamanis, Panagiotis Kanellopoulos, and Maria Kyropoulou. The efficiency of fair division. *Proceedings of the 5th International Workshop on Internet and Network Economics (WINE '09)*, (LNCS 5929):475–482, 2009.

[21] Priyanka B. Carr and Gregory M. Walton. Cues of working together fuel intrinsic motivation. *Journal of Experimental Social Psychology*, 53:169–184, 2014.

[22] Heather Caruso and Anita Woolley. Harnessing the power of emergent interdependence to promote diverse team collaboration. *Research on Managing Groups and Teams*, 11:245–266, 2008.

[23] Alessandro Castelnovo, Riccardo Crupi, Greta Greco, Daniele Regoli, Ilaria G. Penco, and Andrea C. Cosentini. A clarification of the nuances in the fairness metrics landscape. *Scientific Reports*, 2022.

[24] David Chan. Perception of fairness. *Research Collection School of Social Sciences, Paper 2796*, 2011.

[25] Ken Chapman and Stuart Van Auken. Creating positive group project experiences: An examination of the role of the instructor on students' perceptions of group projects. *Journal of Marketing Education*, 23:117–127, 08 2001.

[26] Alex K. Chavez and Cristina Bicchieri. Third-party sanctioning and compensation behavior: Findings from the ultimatum game. *Journal of Economic Psychology*, 39:268–277, 2013.

[27] Dengkang Chen, Yi Zhang, and Yuru Lin. Group awareness and group awareness tools in computer-supported collaborative learning: A literature review. In *2022 International Symposium on Educational Technology (ISET)*, pages 18–22. IEEE, 2022.

[28] Andy Clark and David Chalmers. The extended mind. *Analysis*, 58(1):7–19, 1998.

[29] Gus Cooney, Daniel T. Gilbert, and Timithy D. Wilson. When fairness matters less than we expect. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 113(40):11168–11171, 2016.

[30] Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. Using intelligent task routing and contribution review to help communities build artifacts of lasting value. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, page 1037–1046. Association for Computing Machinery, 2006.

[31] Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. Suggestbot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th International Conference on Intelligent User Interfaces*, IUI '07, page 32–41. Association for Computing Machinery, 2007.

[32] W. Martin Davies. Groupwork as a form of assessment: common problems and recommended solutions. *Higher Education*, 58:563–584, 2009.

[33] David De Cremer, Eric Van Dijk, and Madan M. Pillutla. Explaining unfair offers in ultimatum games and their effects on trust: An experimental approach. *Business Ethics Quarterly*, 20(1):107–126, 2010.

[34] Frank R. de Wit, Lindred L. Greer, and Karen A. Jehn. The paradox of intragroup conflict: a meta-analysis. *The Journal of applied psychology*, 97(2):360–390, 2012.

[35] Daryl DeFord, Moon Duchin, and Justin Solomon. Recombination: A family of markov chains for redistricting. *arXiv preprint arXiv:1911.05725*, 31, 2019.

[36] Elif E. Demiral and Johanna Mollerstrom. The entitlement effect in the ultimatum game – does it even exist? *Journal of Economic Behavior & Organization*, 175:341–352, 2020.

[37] Eric Dietrich. How montana got its new congressional map. *The Montana Free Press*, 2021.

[38] Joan Morris DiMicco, Anna Pandolfo, and Walter Bender. Influencing group participation with a shared display. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, CSCW '04, page 614–623. Association for Computing Machinery, 2004.

[39] Nathan J. Doogan and Keith Warren. A network of helping: generalized reciprocity and cooperative behavior in response to peer and staff affirmations and corrections among therapeutic community residents. *Addiction Research & Theory*, 25(3):243–250, 2017.

[40] Paul Dourish and Victoria Bellotti. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work*, CSCW '92, page 107–114. Association for Computing Machinery, 1992.

[41] Marilyn A. Dyrud. Group projects and peer review. *Business Communication Quarterly*, 64(4):106–112, 2001.

[42] Amy C. Edmondson, Kate Roloff, and Lucy H. MacPhail. Collaboration across knowledge boundaries within diverse teams: Reciprocal expertise affirmation as an enabling condition. *In Exploring Positive Identities and Organizations: Building a Theoretical and Research Foundation, Psychology Press*, page 311–332, 2009.

[43] Ulle Endriss. Lecture notes on fair division. *arXiv:1806.04234*, 2018.

[44] Thomas F. Epper, Ernst Fehr, Claus Thustrup Kreiner, Søren Leth-Petersen, Isabel Skak Olufsen, and Peer Ebbesen Skov. Inequality aversion predicts support for public and private redistribution. *Proceedings of the National Academy of Sciences*, 121(39), 2024.

[45] Thomas Erickson and Wendy A. Kellogg. Social translucence: an approach to designing systems that support social processes. *ACM Transaction on Computer Human Interaction*, 7(1):59–83, 2000.

[46] Jaclyn Farrens, Aaron Matthew Simmons, Steven Luck, and Emily Kappenman. Electroencephalogram (eeg) recording protocol for cognitive and affective human neuroscience research. *Protocol Exchange*, 2019.

[47] Ernst Fehr, Urs Fischbacher, and Simon Gächter. Strong reciprocity, human cooperation, and the enforcement of social norms. *Human nature*, 13(1):1–25, 2002.

[48] Ernst Fehr and Klaus M. Schmidt. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868, 1999.

[49] Susan Brown Feichtner and Elaine Actis Davis. Why some groups fail: a survey of students' experiences with learning groups. *Organizational Behavior Teaching Review*, 9(4):58–73, 1984.

[50] Martin R. Fellenz. Toward fairness in assessing student groupwork: A protocol for peer evaluation of individual contributions. *Journal of Management Education*, 30:570–591, 2006.

[51] Gweon Gahgene. Providing insight into group process. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, page 2645–2648. Association for Computing Machinery, 2008.

[52] Huseyin Gençer. Group dynamics and behaviour. *Universal Journal of Educational*, 7(1):223–229, 2019.

[53] Ali Ghodsi, Matei Zaharia, Benjamin Hindman, Andy Konwinski, Scott Shenker, and Ion Stoica. Dominant resource fairness: Fair allocation of multiple resource types. In *8th USENIX Symposium on Networked Systems Design and Implementation (NSDI 11)*. USENIX Association, 2011.

[54] Lorena R. R. Gianotti, Kyle Nash, Thomas Baumgartner, Franziska M. Dahinden, and Daria Knoch. Neural signatures of different behavioral types in fairness norm compliance. *Scientific Reports*, 8, 2018.

[55] Hanneke Grutterink, Gerben S. Van der Vegt, Eric Molleman, and Karen A. Jehn. Reciprocal expertise affirmation and shared expertise perceptions in work teams: Their implications for coordinated action and team performance. *Applied Psychology*, 62:359–381, 2012.

[56] Qiong Gui, Maria V. Ruiz-blondet, Sarah Laszlo, and Zhanpeng Jin. A survey on brain biometrics. *ACM computing Surveys*, 51:1–38, 2019.

[57] Qing guo Ma, Liang Meng, Zhexiao Zhang, Qing Xu, Yue Wang, and Qiang Shen. You did not mean it: Perceived good intentions alleviate sense of unfairness. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 96 3:183–190, 2015.

[58] Werner Güth, Rolf Schmittberger, and Bernd Schwarze. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4):367–388, 1982.

[59] J. Richard Hackman, editor. *Groups that work (and those that don't)*. Jossey-Bass, 1990.

[60] J. Richard Hackman. Why teams don't work. *Leader to Leader*, 1998(7):24–31, 1998.

[61] J. Richard Hackman and Namcy Katz. Group behavior and performance. *Handbook of social psychology*, pages 1208–1251, 2010.

[62] J. Richard Hackman and Charles G. Morris. Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. *Advances in Experimental Social Psychology*, 8:45–99, 1975.

[63] Greg Hajcak, Annmarie MacNamara, and Doreen Olvet. Event-related potentials, emotion, and emotion regulation: An integrative review. *Developmental neuropsychology*, 35:129–55, 02 2010.

[64] Garrett Hardin. The tragedy of the commons. *Science*, 162(3859):1243–1248, 1968.

[65] S. G. Harkins and R. E. Petty. Effects of task difficulty and task uniqueness on social loafing. *Journal of Personality and Social Psychology*, 43(6):1214–1229, 1982.

[66] James K. Harter, Frank L. Schmidt, and Theodore L. Hayes. Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: a meta-analysis. *Journal of Applied Psychology*, 87(2):268–279, 2002.

[67] Robert L. Heneman and Courtney Von Hippel. Balancing group and individual rewards: Rewarding individual contributions to the team. *Compensation & Benefits Review*, 27(4):63–68, 1995.

[68] Dorothea K. Herreiner and Clemens Puppe. Inequality aversion and efficiency with ordinal and cardinal social preferences—an experimental study. *Journal of Economic Behavior & Organization*, 76(2), 2010.

[69] Brad R. Humphreys. Alternative measures of competitive balance in sports leagues. *Journal of Sports Economics*, 3(2), 2002.

[70] Yuen J. Huo and Kevin R. Binning. Why the psychological experience of respect matters in group life: An integrative account. *Social and Personality Psychology Compass 2*, 2(4):1570–1585, 2008.

[71] Yuen J. Huo, Kevin R. Binning, and Ludwin E. Molina. The interplay between fairness and the experience of respect: Implications for group life. *Research on managing groups and teams: Fairness and groups*, 13:95–120, 2010.

[72] David Imamura. The rise and fall of redistricting commissions: Lessons from the 2020 redistricting cycle. *Human Rights Magazine*, 48(1), 2022.

[73] Georgeta Ion, Anna Díaz-Vicario, and Cristina Mercader. Making steps towards improved fairness in group work assessment: The role of students' self- and peer-assessment. *Active Learning in Higher Education*, 0:570–591, 2023.

[74] Jeroen Janssen, Gijsbert Erkens, Gellof Kanselaar, and Jos Jaspers. Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Comput. Educ.*, 49(4):1037–1065, dec 2007.

[75] Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler. Fairness and the assumptions of economics. *The Journal of Business*, 59(4):285–300, 1986.

[76] Laura Kaltwasser, Andrea Hildebrandt, Oliver Wilhelm, and Werner Sommer. Behavioral and neuronal determinants of negative reciprocity in the ultimatum game. *Social Cognitive and Affective Neuroscience*, 11(10):1608–1617, 2016.

[77] Helen Kennedy, Hannah Ditchfield, Susan Oman, Jo Bates, Itzelle Medina Perea, Monika Fratczak, and Mark Taylor. How people connect fairness and equity when they talk about data uses. *Big Data & Society*, 11(4), 2024.

[78] Christoph Kern, Frederic Gerdon, Ruben L. Bach, Florian Keusch, and Frauke Kreuter. Humans versus machines: Who is perceived to decide fairer? experimental evidence on attitudes toward automated decision-making. *Patterns*, 3(10), 2022.

[79] Maryam Kouchaki, Isaac H. Smith, and Ekaterina Netchaeva. Not all fairness is created equal: Fairness perceptions of group vs. individual decision makers. *Organization Science*, 26(5):1301–1315, 2015.

[80] Karel Kreijns and Paul Kirschner. Group awareness widgets for enhancing social interaction in computer-supported collaborative learning environments: design and implementation. In *32nd Annual Frontiers in Education*, volume 1. IEEE, 2002.

[81] Marjan Laal and Seyed Mohammad Ghodsi. Benefits of collaborative learning. *Procedia - Social and Behavioral Sciences*, 31:486–490, 2012. World Conference on Learning, Teaching & Administration - 2011.

[82] Daniel Ladley, Ian Wilkinson, and Louise Young. The impact of individual versus group rewards on work group performance and cooperation: A computational social science approach. *Journal of Business Research*, 68(11):2412–2425, 2015.

[83] Zeph Landau and Francis Edward Su. Fair division and redistricting. *AMS Special Sessions on the Mathematics of Decisions, Elections, and Games*, pages 17-36, 2010.

[84] Zeph Landau and Reid I Yershov. A fair division solution to the problem of redistricting. *Social Choice and Welfare*, 32(3):479-492, 2009.

[85] Bibb Latané, Kipling D. Williams, and Stephen G. Harkins. Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, 37:822–832, 1979.

[86] Gilly Leshed, Dan Cosley, Jeffrey T. Hancock, and Geri Gay. Visualizing language use in team conversations: designing through theory, experiments, and iterations. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, page 4567–4582. Association for Computing Machinery, 2010.

[87] Justin Levitt. Where are the lines drawn? `https://redistricting.lls.edu/redistricting-101/where-are-the-lines-drawn/`, 2020.

[88] Qi Li, Ya Zheng, Jing Xiao, Kesong Hu, and Zhong Yang. Neural mechanisms of fairness decision-making: An eeg comparative study on opportunity equity and outcome equity. *NeuroImage*, 305, 2024.

[89] Jia-Wei Liang and Nina Amenta. The fairness of redistricting ghost. *arXiv*, 2024.

[90] Jia-Wei Liang and Hao-Chuan Wang. Reshaping group life: A transparent and interpretable reward model to enhance fairness in groups. In *Collaboration Technologies and Social Computing: 29th International Conference, CollabTech 2023, Osaka, Japan, August 29–September 1, 2023, Proceedings*, page 209–216. Springer-Verlag, 2023.

[91] Jia-Wei Liang and Hao-Chuan Wang. Is it fair enough? supporting equitable group work assignment with work division dashboard. *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, pages 2480–2490, 2025.

[92] Edwin A. Locke and Gary P. Latham. Building a theory by induction: The example of goal setting theory. *Organizational Psychology Review*, 10(3-4):223–239, 2002.

[93] Steven J. Luck. *Applied Event-Related Potential Data Analysis*. LibreTexts, 2022.

[94] Ian G Ludden, Rahul Swamy, Douglas M King, and Sheldon H Jacobson. A bisection protocol for political redistricting. *INFORMS Journal on Optimization*, 2022.

[95] Yi Luo, Tingting Wu, Lucas S Broster, Chunliang Feng, Dandan Zhang, Ruolei Gu, and Yue-Jia Luo. The temporal course of the influence of anxiety on fairness considerations. *Psychophysiology*, 51(9):834–842, 2014.

[96] L. H. MacPhail, K. S. Roloff, and A. C. Edmondson. Collaboration across knowledge boundaries within diverse teams: Reciprocal expertise affirmation as an enabling condition. *Exploring positive identities and organizations: Building a theoretical and research foundation*, page 319–340, 2009.

[97] Charles C. Manz. Self-leadership: Toward an expanded theory of self-influence processes in organizations. *The Academy of Management Review*, 11(3):585–600, 1986.

[98] Michael Wyvill Mark Lejk and Stephen Farrow. A survey of methods of deriving individual grades from group assessments. *Assessment & Evaluation in Higher Education*, 21(3):267–280, 1996.

[99] Eric McGhee. Measuring partisan bias in single-member district electoral systems. *Legislative Studies Quarterly*, 39(1):55–85, 2014.

[100] Joseph McGrath. *Social Psychology: A Brief Introduction*. Holt, Rinehart & Winston, 1964.

[101] Joseph E McGrath. *Groups: Interaction and Performance*. Prentice-Hall, 1984.

[102] Dustin G Mixon and Soledad Villar. Utility Ghost: Gamified redistricting with partisan symmetry. *arXiv:1812.07377*, 2018.

[103] Agustin Molina, Carolina Moliner, Vicente Martínez-Tur, Russell Cropanzano, and Jose Maria Peiro. Validating justice climate and peer justice in a real work setting. *Journal of Work and Organizational Psychology*, 32(3):191–205, 2016.

[104] Hendrik Mothes, Soren Enge, and Alexander Strobel. The interplay between feedback-related negativity and individual differences in altruistic punishment: An eeg study. *Cognitive, Affective, & Behavioral Neuroscience*, 16:276–288, 2016.

[105] Goran Murić, Andres Abeliuk, Kristina Lerman, and Emilio Ferrara. Collaboration drives individual productivity. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 2019.

[106] Dupuis R. Nicolas and Gosselin Frederic. The simpler, the better: A new challenge for fair-division theory. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 2011.

[107] Martin A Nowak, Karen M Page, and Karl Sigmund. Fairness versus reason in the ultimatum game. *Science*, 289(5485):1773–1775, 2000.

[108] Jay F. Nunamaker, Bruce A. Reinig, and Robert O. Briggs. Principles for effective virtual teamwork. *Communications ACM*, 52(4):113–117, 2009.

[109] National Conference of State Legislatures. Redistricting criteria. `https://www.ncsl.org/elections-and-campaigns/redistricting-criteria`, 2025.

[110] S. Ollesch, L.and Heimbuch and D. Bodemer. Improving learning and writing outcomes: Influence of cognitive and behavioral group awareness tools in wikis. *International Journal of Computer-Supported Collaborative Learning*, 16(2):225–259, 2021.

[111] Wesley Pegden, Ariel D Procaccia, and Dingli Yu. A partisan districting protocol with provably nonpartisan outcomes. *arXiv:1710.08781v1*, 2017.

[112] Yu Peng, Yanyan Li, You Su, Kailiang Chen, and Shiyan Jiang. Effects of group awareness tools on students' engagement, performance, and perceptions in online collaborative writing: Intergroup information matters. *The Internet and Higher Education*, 53:100845, 2022.

[113] Jutta Peterburs, Rolf Voegler, Roman Liepelt, Anna Schulz, Saskia Wilhelm, Sebastian Ocklenburg, and Thomas Straube. Processing of fair and unfair offers in the ultimatum game under social observation. *Scientific Reports*, 7(44062), 2017.

[114] Terence W. Picton. The p300 wave of the human event-related potential. *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society*, 9(4):456–479, 1992.

[115] Christopher Pieper. Decision theory and game theory. In *Encyclopedia of Violence, Peace, & Conflict*, pages 258–267. Academic Press, third edition edition, 2022.

[116] Olga Pierce and Jeff Larson. How democrats fooled california's redistricting commission. *ProPublica, Dec*, 21, 2011.

[117] John R. Platt. Social traps. *American Psychologist*, 28(8):641–651, 1973.

[118] John Polich. Updating p300: An integrative theory of p3a and p3b. *Clinical Neurophysiology*, 118(10):2128–2148, 2007.

[119] Irene Poort, Ellen Jansen, and Adriaan Hofman. Does the group matter? effects of trust, cultural diversity, and group formation on engagement in group work in higher education. *Higher Education Research & Development*, 41(2):511–526, 2022.

[120] John W. Pratt and Richard J. Zeckhauser. The fair and efficient division of the winsor family silver. *Management Science*, 36(11):1293–1301, 1990.

[121] R. D. Pritchard, S. D. Jones, P. L. Roth, K. K. Stuebing, and S. E. Ekeberg. Effects of group feedback, goal setting, and incentives on organizational productivity. *Journal of Applied Psychology*, 73(2):337–358, 1988.

[122] Ariel D Procaccia. Cake cutting: not just child's play. *Communications of the ACM*, 56(7):78–87, 2013.

[123] Devon Proudfoot and E. Allan Lind. Fairness heuristic theory, the uncertainty management model, and fairness at work. In *The Oxford Handbook of Justice in the Workplace*. Oxford University Press, 07 2015.

[124] Matthew Rabin. Incorporating fairness into game theory and economics. *The American Economic Review*, 83(5):1281–1302, 1993.

[125] Annisa Rahma and Wantini Wantini. Human behavior in social context. *Indonesian Impression Journal*, 3:411–417, 06 2024.

[126] David G Rand and Martin A Nowak. Human cooperation. *Trends in cognitive sciences*, 17(8):413–425, 2013.

[127] Amirhossein Rasooli, Jim Turner, Tunde Varga-Atkins, Edd Pitt, Shaghayegh Asgari, and Will Moindrot. Students' perceptions of fairness in groupwork assessment: validity evidence for peer assessment fairness instrument. *Assessment & Evaluation in Higher Education*, pages 1–16, 2024.

[128] John Rawls. Distributive justice: Some addenda. *Natural Law Forum*, 1968.

[129] John Rawls. *A Theory of Justice: Original Edition*. Harvard University Press, 1971.

[130] Binns Reuben. On the apparent conflict between individual and group fairness. *Conference on Fairness, Accountability, and Transparency*, 25:115–191, 2020.

[131] Siobhan Roberts. To divide the rent, start with a triangle. *The New York Times*, 2014. `https://www.nytimes.com/2014/04/29/science/to-divide-the-rent-start-with-a-triangle.html`.

[132] Carla Anne Roos, Tom Postmes, and Namkje Koudenburg. Feeling heard: Operationalizing a key concept for social relations. *PLoS One*, (11), 2023.

[133] Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. A simplest systematics for the organization of turn taking for conversation. In *Studies in the Organization of Conversational Interaction*, pages 7–55. Academic Press, 1978.

[134] Barry R. Schlenker and Bruce W. Darby. The use of apologies in social predicaments. *Social Psychology Quarterly*, 44(3):271–278, 1981.

[135] Jakob Schoeffer, Niklas Kuehl, and Yvette Machowski. "there is not enough information": On the effects of explanations on perceptions of informational fairness and trustworthiness in automated decision-making. *ACM Conference on Fairness, Accountability, and Transparency*, 2022.

[136] Stefan Schulz-Hardt and Felix C. Brodbeck. Group performance and leadership. *Introduction to Social Psychology*, 2012.

[137] Maurice E. Schweitzer, John C. Hershey, and Eric T. Bradlow. Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes*, 101(1):1–19, 2006.

[138] Ben Shneiderman. Human-centered artificial intelligence: Three fresh ideas. *AIS Transactions on Human-Computer Interaction*, 12(3):109–124, 2020.

[139] Daniel P. Skarlicki, Robert Folger, and Julie Gee. When social accounts backfire: The exacerbating effects of a polite message or an apology on reactions to an unfair outcome. *Journal of Applied Social Psychology*, 34(2):322–341, 2006.

[140] Brian N Smith, Natalie A. Kerr, Michael J. Markus, and Mark F. Stasson. Individual differences in social loafing: Need for cognition as a motivator in collective performance. *Group Dynamics: Theory, Research, and Practice*, 5:150–158, 2001.

[141] David V. Smith, John A. Clithero, Sarah E. Boltuck, and Scott A. Huettel. Functional connectivity with ventromedial prefrontal cortex reflects subjective value for social rewards. *Social cognitive and affective neuroscience*, 9(12):2017–2025, 2014.

[142] Christopher Starke, Janine Baleis, Birte Keller, and Frank Marcinkowski. Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society*, 9(2), 2022.

[143] Ivan D. Steiner. Models for inferring relationships between group size and potential group productivity. *Behavioral science*, 11(4):273–83, 1966.

[144] Nicholas Stephanopoulos and Eric McGhee. Partisan gerrymandering and the efficiency gap. *Public Law and Legal Theory Working Paper*, 493, 2014.

[145] Paul G Straub and J Keith Murnighan. An experimental investigation of ultimatum games: Information, fairness, expectations, and lowest acceptable offers. *Journal of Economic Behavior & Organization*, 27(3):345–364, 1995.

[146] Wolfgang Stroebe and Michael Diehl. Why groups are less effective than their members: On productivity losses in idea-generating groups. *European Review of Social Psychology*, 5(1):271–303, 1994.

[147] Walter Stromquist. How to cut a cake fairly. *The American Mathematical Monthly*, 87(8):640–644, 1980.

[148] David T. Takeuchi, Tiziana C. Dearing, Melissa W. Bartholomew, and Ruth G. McRoy. Equality and equity: Expanding opportunities to remedy disadvantage. *Generations: Journal of the American Society on Aging*, 42(2):13–19, 2018.

[149] The Justice Collaboratory, Yale Law School. Procedural justice. `https://law.yale.edu/justice-collaboratory/procedural-justice`, n.d. Accessed: August 11, 2025.

[150] Jamie Tucker-Foltz. A cut-and-choose mechanism to prevent gerrymandering. *arXiv:1802.08351*, 2019.

[151] Tom R. Tyler and Yuen J. Huo. *Trust in the Law: Encouraging Public Cooperation with the Police and Courts Through*. Russell Sage Foundation, 2002.

[152] Tom R. Tyler and E. Allen Lind. A relational model of authority in groups. *Advances in experimental social psychology*, 25:115–191, 1992.

[153] Alarith Uhde, Nadine Schlicker, Dieter P. Wallach, and Marc Hassenzahl. Fairness and decision-making in collaborative shift scheduling systems. *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[154] Sahibdin Priya P. Van der Veen, Frederik M. Dissociation between medial frontal negativity and cardiac responses in the ultimatum game: Effects of offer size and fairness. *Cognitive, Affective, & Behavioral Neuroscience*, 11:516–525, 2011.

[155] Neil Vidmar. The origins and consequences of procedural fairness. *Law Social Inquiry*, 15(4):877–892, 1990.

[156] Axi Wang, Wang Minhong Yu, Shengquan, and Ling Chen. Effects of a visualization-based group awareness tool on in-service teachers' interaction behaviors and performance in a lesson study. *Interactive Learning Environments*, 27(5–6):670–84, 2019.

[157] Yiwen Wang, Zhen Zhang, Liying Bai, Chongde Lin, Roman Osinsky, and Johannes Hewig. Ingroup/outgroup membership modulates fairness consideration: neural signatures from erps and eeg oscillations. *Scientific Reports*, 7, 2017.

[158] Richard G. Weaver and John D. Farrell. *Managers as facilitators : a practical guide to getting work done in a changing workplace.* Berrett-Koehler Publishers, 1997.

[159] Barbara A. Wech, Kevin W. Mossholder, Robert P. Steel, and Nathan Bennett. Does work group cohesiveness affect individuals' performance and organizational commitment?: A cross-level examination. *Small Group Research*, 29(4):472–494, 1998.

[160] Jordan Wells and David G Rand. Strategic self-interest can explain seemingly fair offers in the ultimatum game. *Social Science Research Network (SSRN)*, 2013.

[161] Judith B. White, Ellen J. Langer, Leeat Yariv, and John C. Welch. Frequent social comparisons and destructive emotions and behaviors: The dark side of social comparisons. *A Journal of Adult Development*, 13:36–44, 2006.

[162] Lan Xia, Kent B. Monroe, and Jennifer L. Cox. The price is unfair! a conceptual framework of price fairness perceptions. *Journal of Marketing*, 68(4):1–15, 2004.

[163] Toshio Yamagishi, Yutaka Horita, Nobuhiro Mifune, Hashimoto Hirofumi, Yang Li, Mizuho Shinada, Arisa Miura, Inukai Keigo, Haruto Takagishi, and Dora Simunovic. Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proceedings of the National Academy of Sciences*, 109(50):20364–20368, 2012.

[164] Keith J. Yoder and Jean Decety. Me first: Neural representations of fairness during three-party interactions. *Neuropsychologia*, 147:107576, 2020.

[165] Bo Zhang and Matthew W. Ohland. How to assign individualized scores on a group project: An empirical evaluation. *Applied Measurement in Education*, 22(3), 2009.

[166] Shuai Zhang, Lei Sun, Xiuqing Mao, Cuiyun Hu, and Peiyuan Liu. Review on eeg-based authentication technology. *Computational Intelligence and Neuroscience*, 2021.

[167] Yan Zhang and Chi-Yue Chiu. Goal commitment and alignment of personal goals predict group identification only when the goals are shared. *Group Processes & Intergroup Relations*, 15(3):425–437, 2012.