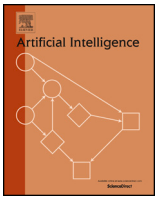
Artificial Intelligence 336 (2024) 104196

Contents lists available at ScienceDirect

Artificial Intelligence

journal homepage: www.elsevier.com/locate/artint



# Manipulation and peer mechanisms: A survey

Matthew Olckers [a,*], Toby Walsh [b]

[a] Macquarie University and the e61 Institute, Australia
[b] School of Computer Science and Engineering, UNSW Sydney, Australia

ARTICLE INFO

*Keywords:*
Peer mechanism
Peer ranking
Peer review
Peer grading
Community-based targeting

ABSTRACT

In peer mechanisms, the competitors for a prize also determine who wins. Each competitor may be asked to rank, grade, or nominate peers for the prize. Since the prize can be valuable, such as financial aid, course grades, or an award at a conference, competitors may be tempted to manipulate the mechanism. We survey approaches to prevent or discourage the manipulation of peer mechanisms. We conclude our survey by identifying several important research challenges.

## 1. Introduction

Imagine that you are competing for a prize at your workplace. The prize is awarded by asking everyone at your work, including you, to nominate who deserves the prize. The person with the most nominations wins. Who should you nominate? If you tell the truth and nominate who you think is most deserving, your nomination could cause you to lose out. Would you be truthful? Do you think your colleagues would be truthful?

As this example illustrates, when the competitors for a prize also determine who wins, the competitors may be tempted to manipulate the outcome. A growing research literature, which we collect under the term "peer mechanisms", aims to prevent or discourage manipulation in these situations. This paper surveys both the theoretical and empirical research on peer mechanisms and offers several research challenges.

The interest in peer mechanisms has been fueled by the variety of high-stakes applications. The prize can take many forms. Closest to the experience of researchers, the prize could be an award at a conference. Further afield, the prize could be grades in a course [79], a time slot to use a telescope [69], aid targeted to people in need [38,3], loans for entrepreneurs [53], a job for freelancers [57], an award for the best soccer player of the year [26], or even the papacy of the Catholic Church [66].

The variety of applications has inspired a variety of models. Some models, known as *peer selection*, assume the mechanism designer wishes to select a single participant or a limited number of participants for the prize. While other models, known as *peer grading*, assume each participant should receive a cardinal score. Besides the mechanism's output, the models can differ in many other dimensions, such as what the participants are asked to report, the type of information participants hold about their peers, and whether the mechanism designer can make payments.

We provide a taxonomy to highlight the differences in each model introduced in the literature, and to identify some common themes. The approaches to prevent manipulation in peer mechanisms can be grouped into one of three categories:

1. *impartial* mechanisms where a participant's report cannot impact their chance of winning the prize,
2. *audits* where the mechanism attempts to detect and punish manipulation,

* Corresponding author.
*E-mail address:* matthew.olckers@mq.edu.au (M. Olckers).

https://doi.org/10.1016/j.artint.2024.104196
Received 4 October 2022; Received in revised form 28 May 2024; Accepted 28 July 2024

Available online 2 August 2024

3. *rewards* for truthful reports.

Although these three approaches are distinct, they are not mutually exclusive. The approaches can be combined. The context will determine which approach or combination of approaches is most suitable.

The bulk of the theoretical research has focused on impartial mechanisms. We delve deeper into these results by focusing on the model of peer selection, where peers nominate each other for a single prize or several identical prizes. Researchers have taken two main approaches: checking if impartiality is compatible with desirable axioms and checking how closely impartial mechanisms can approximate the most desirable outcome (such as awarding the prize to the participants with the most nominations).

The axiomatic and approximation results highlight how flexibility in the number of prizes is crucial for designing impartial peer mechanisms with desirable properties. If the mechanism must always award a fixed number of prizes, discouraging manipulation can lead to undesirable outcomes, such as awarding the prize to a participant who does not receive any nominations. If the mechanism can award more prizes than planned or choose to leave the prize unassigned, manipulation can be discouraged while avoiding many of the undesirable outcomes that stem from a fixed number of prizes. The theoretical results also highlight the importance of randomization. In most cases, mechanisms that use randomization provide better approximation results than mechanisms that use a deterministic rule.

Manipulation in peer mechanisms is not merely a theoretical concern. We survey empirical research that shows that people will try to manipulate peer mechanisms when given the chance. Employees will reduce the peer grades of coworkers when competing for promotion [52], and entrepreneurs bias peer reports in favor of friends and family when competing for loans [53]. Manipulation has also been shown in the lab. Whether experimental participants produce art [16] or label envelopes [30], they are happy to sabotage their peers to increase their own chance of winning a prize.

The empirical research provides several lessons for the theoretical study of peer mechanisms. First, participants often make errors in their peer evaluations. Mechanisms that are robust to errors are more likely to succeed in practice. Second, nepotism is common. Most models assume that participants care only about their own chance of winning a prize and not whether other perhaps related participants win a prize. Third, small amounts of manipulation may be acceptable. Rather than aiming to disincentivize manipulation from all participants, mechanisms could allow some manipulation and still yield good outcomes. Fourth, choosing the optimal manipulation can be difficult for participants. Complexity may be a useful tool to discourage manipulation.

Models of peer mechanisms differ from both the classic approach to mechanism design and the classic approach to social choice. The classic approach to mechanism design is to elicit information from individuals about themselves, such as the amount an individual would be willing to pay for an item at an auction. In peer mechanisms, individuals hold preferences or information about other participants (their peers). The classic approach to social choice is to aggregate the preferences of voters about a set of candidates. The voters and candidates are distinct. In peer mechanisms, the participants are both voters and candidates.

We focus this survey on preventing manipulation in peer mechanisms. We do not include research that focuses on how to aggregate nominations, rankings, or grades. Examples of this line of research include Caragiannis et al. [28,29], which considers how counting methods can aggregate partial rankings, and Wang and Shah [83], which considers how to aggregate grades from individuals that have different standards and ranges. We do not include a recent line of research that studies how to incentivize participants to invite their peers to participate in a mechanism [90]. We restrict our focus to settings where all participants are aware of the mechanism and the prize. We do not include research on peer grading that designs mechanisms to encourage peer graders to exert effort when grading (see Zarkoob et al. [88] for a recent example of this line of work). Our focus is on peer grading mechanisms that prevent graders from improving their own grades or rankings through manipulation.

We include peer prediction mechanisms that have been adapted to evaluating information about people, such as a person's need for financial aid or their entrepreneurial ability. Typically, peer prediction is used for reports about external objects, such as the quality of a product or the forecast of an event. Not to be confused with the "peer" in peer mechanisms, the "peer" in peer prediction refers to the way these mechanisms use reports from multiple participants to incentivize truthful reports without access to ground truth to check the reports. Peer prediction mechanisms make payments to participants that depend on the participant's report and the reports of other participants who evaluate the same target object. We are interested in cases when the target object is information about another participant. See Faltings and Radanovic [45] for a survey of peer prediction mechanisms.

Our survey has some overlap with a recent survey of academic peer review by Shah [75], but our surveys make distinct contributions. Rather than focusing on a single application (academic peer review in the case of Shah [75]), we focus on manipulation in peer mechanisms, which extends to several other applications, such as poverty targeting and peer grading of student assignments. Shah's [75] survey covers some work on manipulation but does not go into the same detail as we do. Also, the models we discuss are often different. The models of peer mechanisms we discuss only correspond to models of academic peer review when each author submits one sole-authored paper and is also available as a reviewer. Authors often submit multiple papers, each paper can have multiple authors, authors may not act as reviewers, and reviewers may not submit papers.

We begin the survey with a motivating example to describe why many peer mechanisms create an opportunity for the participants to manipulate who wins the prize. We then provide a taxonomy of approaches to prevent manipulation and discuss the range of techniques that researchers have proposed. To highlight the two main theoretical approaches of axiomatic and approximation analysis, we focus on the model of peer selection. We survey the empirical studies of peer mechanisms and list key lessons the empirical research provides for theory. We conclude the survey by highlighting several areas in need of further research.

## 2. Motivating example

Suppose a group of people compete for a prize by participating in a peer mechanism. The mechanism determines who wins the prize by asking each participant to nominate one or more peers and awarding the prize to the participant who receives the most nominations. Assume there is only one prize, which cannot be split between multiple participants. In the case of a tie, the prize is awarded uniformly at random between those with the most nominations.

One temptation is for each participant to nominate themselves, so we may want the mechanism to exclude self-nominations.[1] But even when the participants cannot nominate themselves, they may still have opportunities to manipulate who wins.

A simple example with three participants ($a$, $b$, and $c$) demonstrates that they can still manipulate who wins the prize. Suppose that:

- $a$ nominates $b$ and $c$,
- $b$ nominates $c$, and
- $c$ nominates $a$.

Since both $b$ and $c$ have two nominations each, the mechanism awards the prize randomly to $b$ or $c$ with equal probability. Let's consider $c$'s perspective. All else equal, $c$ can ensure he wins the prize by nominating no one or nominating $a$. Even if $c$ believes that $b$ is worthy of the prize, $c$ has a strong incentive to manipulate the outcome to increase his chance of winning.

Whether the participants are asked to nominate, rank, or grade their peers, the same temptation remains. To increase their own chance of winning a prize, participants in a peer mechanism are tempted to manipulate their evaluation of their closest competitors. In the example, $c$'s closest competitor is $b$. By failing to nominate $b$, $c$ increases his chance of winning the prize at $b$'s expense.

## 3. A taxonomy of models

We surveyed the literature for peer mechanisms that address manipulation. Since peer mechanisms are inspired by a variety of applications, there are a variety of different models to describe these applications. In Table 1, we provide a taxonomy of the models we uncovered. We distinguish between different models according to their:

| | |
|---|---|
| Input: | What do the participants need to report about their peers? |
| Output: | What output does the mechanism produce? |
| Information: | What type of information do participants hold about their peers? |

We have some notation within Table 1. We use $m$ for the number of peers each participant is asked to evaluate. In most cases, $m$ is small relative to the number of participants. We use $k$ for the number of winners when the mechanism selects multiple winners.

The table also includes columns to categorize the mechanisms according to:

| | |
|---|---|
| Approach: | Does the mechanism use audits, rewards, or impartiality to discourage manipulation? |
| Technique: | What technique does the mechanism use? |

The taxonomy is useful for several reasons. First, the contributions to peer mechanisms are spread across computer science and economics, and the taxonomy shows which contributions are most closely connected. Second, the taxonomy shows gaps in the literature. For example, the "Approach" column of Table 1 shows that most of the existing work focuses on constructing impartial mechanisms. Less work has focused on using audits or rewards to discourage manipulation.

### 3.1. Inputs

The input into a peer mechanism is the reports from the participants about their peers. Models differ by the detail of the reports and which peers they can report on. In increasing order of detail, the mechanism could ask for a nomination, a ranking, or a grade. A nomination asks the participant to select one or more peers. A ranking asks for a strict order of peers. A grade asks for a cardinal score for each peer.

The appropriate form of peer report can be linked to the level of information each participant holds about their peers. If participants are only able to sort their peers into two groups, such as worthy and unworthy for the prize, nominations are appropriate. If they have more detailed information to order peers but not enough to determine a cardinal score, a ranking is appropriate. Finally, the most detailed information can be modeled by a cardinal score.

[1] As Ng and Sun [70] and Ohseto [72] have shown, excluding self-evaluations can create problems in aggregating peer grades. Suppose participants $a$ and $b$ unanimously grade $a$ higher than $b$, but $a$ uses more generous grades than $b$. Excluding self-evaluations will discard $a$'s generous grade about himself but keep the generous grade he gives to $b$, which may cause $b$ to have a higher aggregated grade than $a$. Ng and Sun [70] provides theoretical results highlighting an incompatibility between unanimity and excluding self-evaluations. Ohseto [72] strengthen Ng and Sun's [70] results to show that if participants can select grades from a finite and large set of real numbers, no aggregation rule excludes self-evaluations and satisfies monotonicity and unanimity.

**Table 1**
Taxonomy of peer mechanisms.

| | Model | | | Mechanism | |
|---|---|---|---|---|---|
| Paper | Input | Output | Information | Approach | Technique |
| Holzman and Moulin [51] | Nominate one peer | Single winner | Subjective | Impartial | Partition |
| Mackenzie [65] | Nominate one peer | Single winner | Subjective | Impartial | Random dictatorship |
| Babichenko et al. [15] | Nominate one peer | Single winner | Subjective | Impartial | Expand possible winners |
| Edelman and Por [44] | Nominate one peer | Single winner | Subjective | Impartial | Random dictatorship |
| Cembrano et al. [34] | Nominate one peer | Single winner | Subjective | Impartial | Permutation |
| Amorós [6] | Nominate one peer | Single winner | Common | Impartial | Fix position |
| Mackenzie [66] | Nominate one peer | At most one winner | Subjective | Impartial | Threshold |
| Bjelde et al. [21] | Nominate one peer | At most two winners | Subjective | Impartial | Permutation |
| Tamura and Ohseto [78], Tamura [77] | Nominate one peer | Top $k$ winners | Subjective | Impartial | Expand possible winners |
| Fischer and Klimm [47,48] | Nominate one or more peers | Single winner | Subjective | Impartial | Permutation |
| Bousquet et al. [24] | Nominate one or more peers | Single winner | Subjective | Impartial | Permutation |
| Babichenko et al. [13] | Nominate one or more peers | Single winner | Subjective | Impartial | Expand possible winners |
| Caragiannis et al. [26] | Nominate one or more peers | Single winner | Subjective | Impartial | Jury |
| Zhang et al. [89] | Nominate one or more peers | Single winner | Subjective | Impartial | Expand possible winners |
| Caragiannis et al. [27] | Nominate one or more peers | Single winner | Subjective | Impartial | Threshold |
| Cembrano et al. [31] | Nominate one of more peers | Single winner | Subjective | Impartial | Threshold |
| Ito et al. [54] | Nominate one or more peers | Single winner | Subjective | Reward | Peer prediction |
| Zhao et al. [91] | Nominate one or more peers | Single winner | Subjective | Impartial | Expand possible winners |
| Alon et al. [5] | Nominate one or more peers | Top $k$ winners | Subjective | Impartial | Partition |
| Cembrano et al. [32] | Nominate one or more peers | Top $k$ winners | Subjective | Impartial | Expand possible winners |
| Wąs et al. [85] | Nominate one or more peers | Grade | Subjective | Impartial | Fix grade |
| Li et al. [63] | Nominate top participant | Ranking | Common | Impartial | Fix position |
| Bao et al. [17] | Nominate network neighbor | Single loser | Ground truth | Audit | Compare to nominee |
| Mattei et al. [68], Lev et al. [61,62] | Rank $m$ peers | Top $k$ winners | Common | Impartial | Threshold |
| Merrifield and Saari [69] | Rank $m$ peers | Ranking | Common | Reward | Reward consensus |
| Xu et al. [87] | Rank $m$ peers | Ranking | Subjective | Impartial | Partition |
| Stelmakh et al. [76] | Rank $m$ peers | Ranking | Ground truth | Audit | Target manipulation |
| Bloch and Olckers [22] | Rank network neighbors | Top $k$ winners | Ground truth | Impartial | Threshold |
| Bloch and Olckers [23] | Rank network neighbors | Ranking | Common | Impartial | Fix position |
| Amorós et al. [8], Amorós [7] | Rank all peers | Ranking | Common | Impartial | Fix position |
| Kahng et al. [56] | Rank all peers | Ranking | Subjective | Impartial | Partition |
| Alcalde-Unzu et al. [4] | Rank all peers | Ranking | Subjective | Impartial | Partition |
| Cembrano et al. [33] | Rank all peers | Ranking | Subjective | Impartial | Fix position |
| Hussam et al. [53] | Rank or grade $m$ peers | Single winner | Ground truth | Reward | Peer prediction |
| Rai [74] | Binary type | Grades | Ground truth | Audit | Target disagreement |
| De Clippel et al. [40] | Relative grade | Cardinal share | Subjective | Impartial | Reduce total reward |
| Kurokawa et al. [58] | Grade $m$ peers | At most $k$ winners | Subjective | Impartial | Expand possible winners |
| Aziz et al. [11,12] | Grade $m$ peers | Top $k$ winners | Subjective | Impartial | Partition |
| Dhull et al. [41] | Grade $m$ peers | Top $k$ winners | Subjective | Impartial | Partition |
| Wang et al. [84] | Grade $m$ peers | Top $k$ winners | Common | Impartial | Partition |
| Chakraborty et al. [36] | Grade $m$ peers | Grades | Ground truth | Audit | Assign audited peers |
| De Alfaro and Shavlovsky [39] | Grade $m$ peers | Grades | Ground truth | Reward | Reward consensus |
| Baumann [20] | Grade network neighbors | Single winner | Ground truth | Audit | Limit misreports |
| Babichenko et al. [14] | Grade network neighbors | Top $k$ winners | Subjective | Impartial | Expand possible winners |
| Cembrano et al. [33] | Grade all peers | Top $k$ winners | Subjective | Impartial | Partition |
| Walsh [82] | Grade all peers | Grades | Ground truth | Reward | Reward consensus |
| Niemeyer and Preusser [71] | Abstract message space | Single winner | Subjective | Impartial | Jury |

To our knowledge, there is little research on the form of peer report that participants prefer. In the context of peer grading, De Alfaro and Shavlovsky [39] reported that:

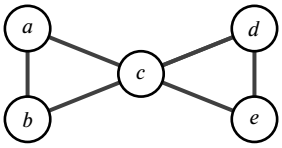
> "Students expressed some uneasiness in ranking their peers, especially as they perceived ranking as a blunt tool, unable to capture the difference between a pair of roughly equivalent submissions, and a pair of submissions, one of which was very good, and the other non-functional."

Further empirical research will be needed to guide theory on the appropriate form of peer reports in peer grading and other contexts.

Although most models use either peer nominations, ranks, or grades, there are some exceptions. Rai [74] uses a model with two participants that can be a binary type (either rich or poor). Each participant reports whether they are poor and whether their peer is poor. However, Rai's [74] model can be thought of as a model of nominations where self-nominations are allowed. Participants can choose to nominate no one, nominate themselves only, their peer only, or both themselves and their peer. Another exception is Niemeyer and Preusser [71], who use an abstract message space. They do not restrict how the participants can communicate with the mechanism designer.

The mechanism must also specify which peers each participant can evaluate. Some models do not restrict which peers each participant can evaluate. Each participant considers all his peers in his reports. Others assume that each participant reports on a fixed number of peers (represented by $m$ in Table 1). When the size of the community is large, reporting on every peer may be too onerous.[2] Distributing peer reports equally among participants is a natural alternative.

Another class of models assumes that the participants are connected by a network, and each participant can only evaluate his network neighbors. For example, in the network shown below, $a$ can evaluate $b$ and $c$, but cannot evaluate $d$ and $e$.




The network can capture situations where participants may learn about their peers through social interactions. The network could represent friendships or co-worker relationships. Social networks often display common structures, such as clustering—if $a$ is friends with $b$ and $c$, then $b$ and $c$ are likely to be friends. A recent line of research, initiated by Baumann [20] and Bloch and Olckers [23] in economics and Babichenko et al. [14] in computer science, investigates how the structure of the network impacts the designer's ability to construct mechanisms that perform well.

The network can also represent who the participants are not allowed to report on. For example, in the context of conference peer review, the network could represent conflicts of interest, such as co-author and student-supervisor relationships [87]. The network of conflicts is the complement or inverse of the network studied by Baumann [20], Bloch and Olckers [23] and Babichenko et al. [14].

Once the network is defined, the model must still define the form of the peer reports. Participants may be asked to nominate [17], rank [23], or grade [20,14] their network neighbors.

In Baumann [20], Bloch and Olckers [23] and Babichenko et al. [14], the network that captures the ability to evaluate peers is unweighted.[3] Either the participant can evaluate a given peer or they cannot. Dhull et al. [41] uses a weighted network where the weight measures expertise to evaluate a given peer. The higher the weight, the more accurate the peer evaluation. In the context of conference peer review, the weight can be thought of as a similarity between two authors' research interests. Dhull et al. [41] study how to assign paper submissions to evaluators to maximize the accuracy of evaluations while ensuring that evaluators' own submissions are not competing with the submissions they are called to evaluate. (This is done using the partition mechanism, which we discuss in more detail in Section 3.4.3.)

### 3.2. Outputs

The output of a peer mechanism describes the form of the prize. Models differ in the number of prizes and if the prizes differ in quality. Inspired by best paper awards at conferences or selecting the most influential user in a social media network, some models assume the output is a single winner. A participant's utility is defined by the probability he wins the prize. Other applications, such as research grants, have inspired models with multiple winners for a prize of equal quality. In Table 1, the number of winners is given by a constant $k$. Finally, peer grading and targeting of aid or loans have inspired models where the mechanism assigns a rank or a grade to each participant, and utility increases in the rank or grade. Similar to a grade, De Clippel et al. [40] defined the output as the share of a divisible prize.

A theme in the theoretical analysis of peer mechanisms (which we will expand on in Section 4) is that flexibility in the output of the mechanism allows the designer to construct mechanisms with desirable properties. For example, suppose a designer who aims to

[2] The burden of grading all peers can be reduced by combining nominations and grades. The mechanism in Cembrano et al. [35] asks each participant to nominate as many peers as they would like and assign grades (or weights) to each nomination. The peers that are not nominated receive a grade of zero.

[3] After the participants grade their peers in Babichenko et al.'s [14] model, the grades can be represented as a weighted network.

award a single prize has the flexibility to award a second prize to a runner-up. If the runner-up is close enough to the winner that the runner-up can divert the prize to himself by changing his peer report, the mechanism can simply award a second prize to the runner-up to discourage the temptation for the runner-up to manipulate who wins.

Nearly every paper listed in Table 1 assumes the prize is desirable. Participants want to be selected or improve their rank or grade. The only exception is a model of criminal networks studied by Bao et al. [17] where the selected participant pays a fine. Rather than selecting a single winner, the mechanism selects a single loser.

### 3.3. Information

Peer mechanisms focus on eliciting the information participants hold about their peers. How is the peer information generated? We divide the models into three categories:

- Subjective information: an opinion about who is worthy of a prize, who should receive a higher rank or a higher grade.
- Common information: all participants agree about who is worthy or who should receive a higher rank or grade.
- Ground truth information: each participant has a value which can be checked through an audit or some other measure.

The types of peer information in the above list are ordered from least to most constrained. Subjective information allows the participants to hold any opinions about their peers. Common information constrains all peers to agree. Ground truth information adds an additional requirement to common information that the information can be checked by the mechanism designer.

A model with common information or ground truth information does not imply that all participants have identical information. Each participant may only hold information about a subset of peers. One participant may know that $a$ is ranked above $b$ while another may know that $b$ is ranked above $c$ but have no knowledge about $a$. Also, the participants may make errors if the common or ground truth information is observed with noise. For example, Lev et al. [61] use the Mallows model to shuffle the ranking that each participant observes. The Mallows model has a dispersion parameter that ranges from perfect information at 0 to no information at 1. As the parameter approaches 0, participants observe a ranking that is concentrated around the true common ranking. At 1, each participant draws a ranking uniformly at random from all possible rankings.

Some participants may make systematically more errors than others. In the context of peer grading, students who have a good understanding of an assignment may grade their peers more accurately than students who struggle to understand the assignment. The mechanism designer may wish to give more weight to graders who got good grades themselves [82]. Grader accuracy or reliability may also be measured directly if the mechanism designer can compare a peer grade to ground truth, such as the grade given by an instructor [36].

Each type of information matches different applications. Social media (such as Twitter) is an example of an application with subjective information. Each user follows peers who they find interesting, but users may disagree on which peers are interesting. The disagreement does not imply that one user is wrong. Peer grading is an example of ground truth information. Each student has a ground truth score according to the marking rubric that could be verified by an expert grader. Common information has similar applications to ground truth, except that the model does not specify how the mechanism designer can verify the information. In the example of targeting aid, participants may agree which peers are most in need of aid, but the designer may not have a method to measure need.

Almost all of the papers in Table 1 assume the mechanism designer has no prior information on the peer reports. Caragiannis et al. [27] is the one exception. Their model assumes that the designer knows the probability that each participant will nominate each peer. The prior information is useful to select a default winner in the case of ties, as ties can create opportunities to manipulate the mechanism.

### 3.4. Mechanisms: approaches and techniques

We provide an overview of the three main approaches (audits, rewards, and impartial mechanisms) used to encourage truthfulness and discourage the manipulation of peer mechanisms. We also highlight the range of techniques used to implement each approach. The three approaches are not mutually exclusive. A single peer mechanism could use all three.

#### 3.4.1. Audits

In some settings, an auditor can check the peer reports. For example, in peer grading of large courses, the instructor can check if students have graded their peers accurately, while in poverty targeting, the grant agency could conduct surveys to measure the poverty level of some households.

If audits can uncover the truth, why not audit everyone? In many applications, including peer grading and poverty targeting, we can assume that audits are more costly than peer reports. The goal is to undertake a limited number of audits to achieve the required performance.

One technique is to *target disagreement* in peer reports [74]. In the poverty targeting setting, if I claim I am poor, but my neighbor says I am rich, the mechanism can audit the disagreement and punish misreports. The mere possibility of an audit can discourage misreporting and may produce a desirable equilibrium where all participants report truthfully, and no audits need to be conducted.

Targeting disagreement is effective when the designer knows that the participants have perfect information about each other. When there is some error or noise in the peer reports, detecting manipulation becomes more difficult. Disagreement no longer implies that

one of the participants is lying. Participants could hold different information about their peers by chance and disagree even when they are both reporting truthfully. For the designer to *target manipulation*, they need to determine if a participant's peer report improves his position by chance or due to manipulation [76].

Audits may also limit the extent to which participants can lie about their peers. When employees compete for a promotion, they may need to support claims about peers and themselves with some evidence. The mechanism can exploit the need for evidence *limits misreports* about themselves and their peers [20]. The best employee can make the highest claim about his own performance, and the most negative peer review he can receive will be better than the most negative peer review any of his participants can receive.

In the constant of peer grading, the mechanism can audit a small number of student assignments and then *assign the audited peers* assignments to all students for grading [36]. Since the students do not know which assignments have been audited, the mechanism discourages manipulation for all assignments.

The mechanism can use peer nominations to decide who to audit. In a model of criminal networks, the mechanism samples one participant and asks that participant to nominate one peer. The *compare to nominee* mechanism audits both the sampled participant and the nominated peer but only fines the one with the higher level of criminal activity. Since audits only provide a signal of criminal activity, the sampled participant minimizes their chance of incurring a fine by nominating the peer who they think has the highest level of criminal activity.

#### 3.4.2. Rewards

To prevent manipulation, the mechanism designer can reward truthful reporting. The rewards can take the form of monetary payments or some other form. In peer grading, the reward might not be money but the student's grade. For instance, the mechanism in Walsh [82] increases a student's grade when they grade other students well.

If the peer reports are entirely subjective, then paying for truthful reports is not particularly applicable. A participant can always claim that their report is their true but subjective opinion. In Table 1, most of the mechanisms that use payments are in settings where there is common information or ground truth. The only exception is Ito et al. [54], where each participant's nomination is subjective. However, the reward component of Ito et al.'s [54] mechanism is used on a part of the model that does have common information—whether both participant $a$ and participant $b$ observe a nomination from $a$ to $b$.

If each peer has common information that all participants agree upon, a natural approach is to *reward consensus*. Multiple peer assessments of the peer's value should converge. A participant who wishes to manipulate the mechanism must consider the cost of diverging from the consensus and reducing their reward. For example, if the prize is the ranking of a grant application, the participants can receive an increase in the ranking as payment for a report that agrees with the consensus [69].

Rewarding consensus can, however, create problems. If a participant believes that the consensus will be biased, he may bias his peer reports to match the consensus. The mechanism must give the participants an incentive to report their true assessment, even if they believe that their true assessment will differ from the consensus.

The field of *peer prediction* uses payments to extract true assessments even though the designer does not have access to the ground truth [45]. In peer prediction, the focus is usually on assessing objects that are unrelated to the peers themselves, such as the quality of a product or a forecast of an event. The peer mechanisms we discuss in this survey involve peers assessing each other.

Peer prediction mechanisms can be adapted to peers assessing each other. For example, Hussam et al. [53] adapt Witkowski and Parkes's [86] peer prediction mechanism to peer reports. The mechanism asks each participant for peer reports and for their belief about what other participants' peer reports will be. If the other participants report truthfully, a participant maximizes their expected payment by truthfully reporting their peer report and their belief of other participants' peer reports.

#### 3.4.3. Impartial mechanisms

The bulk of the research surveyed in Table 1 discourages manipulation by designing an impartial mechanism. A peer mechanism is impartial when a participant cannot influence his chances of receiving a prize or improving his rank or grade.[4]

Although most papers define impartiality so that a mechanism is impartial when every participant cannot change their own probability of receiving a prize or improve their own rank or grade, there are some exceptions. For example, Alcalde-Unzu et al. [4] use a stricter definition of impartiality for the case where the output of the mechanism is a ranking. In Alcalde-Unzu et al.'s [4] definition, a mechanism is impartial when a participant cannot impact their own position in the ranking and who is below and who is above them in the ranking. In contrast, Kahng et al. [56] and Cembrano et al. [33] use the more standard definition of impartiality that each participant cannot change their own position in the ranking.

One technique to achieve impartiality is to *partition* participants into groups [5,51]. For example, the mechanism designer divides participants into two groups, $A$ and $B$. Participants in $B$ pick a winner in $A$, while participants in $A$ pick a winner in $B$. The overall winner is decided with a coin toss. Each individual can only influence the chance a peer outside of his or her group wins the prize, so the mechanism is impartial.

How many participants should be in each group? At one extreme, the *random dictatorship* randomly chooses a single participant for one group and places the rest of the participants in the other group. The single participant (the dictator) decides who wins the prize. *Jury* mechanisms increase the number of participants in the dictatorship group, and these participants act as a jury to decide on the winner among the remaining participants.

[4] The definition of an impartial mechanism is slightly different from a strategy-proof mechanism. In a strategy-proof mechanism, an agent has a weakly dominant strategy to report the truth. As Fischer and Klimm [47,48] point out, impartiality is equivalent to strategy-proofness if the utility of the participant only depends on their chance of winning the prize. Strategy-proof mechanisms are also referred to as dominant-strategy incentive-compatible mechanisms.

Partition mechanisms can have poor performance if all the best participants end up in the same group. For example, suppose a partition mechanism for selecting two winners divides the participants into two groups and selects one winner from the first group and one from the second group. If the top two participants end up in the first group, the mechanism will only select one of them. Aziz et al. [11,12] counter this problem by using more than two groups and deciding on the number of winners in each group based on the grades from peers outside the group.

One participant can also be part of many different partitions. In the *permutation* mechanism introduced by Fischer and Klimm [47,48], the participants are placed in random order, and the mechanism only counts nominations from peers that are before the participant in the order. To decide on the winner, the mechanism starts with the first participant as the candidate winner and moves through the order to update the candidate winner. A participant $p$ becomes the candidate winner if he is above the current candidate $c$ in the order, and the number of nominations $p$ receives from peers before $p$ in the order (excluding $c$) is greater than or equal to the number of nominations $c$ receives from peers before $c$ in the order. The winner is the candidate winner after moving through all participants in the order. The mechanism is impartial because each participant can only influence if a peer wins when the participant is no longer able to win.

Researchers often assume the designer must award a fixed number of prizes, which creates a strong incentive for participants to misreport. For example, if the mechanism must award a single prize, the participant in second place has a strong incentive to share a negative review of the participant in first place.

If the designer has some flexibility in awarding the prize, impartiality can be achieved without resorting to a partition. One approach is to *expand the set of possible winners* to include participants that could win if they changed their peer reports [78,58]. If the designer is constrained to choose at most $k$ winners, she can randomly pick $k$ from the expanded set of possible winners. But, the probability that each participant is selected cannot depend on their peer report. One solution is to have the option of not selecting any winners to remove the participant's incentive to reduce the number of possible winners [58]. If the designer has no constraints on the number of winners, another approach is to choose an exogenous *threshold* and only award prizes to participants who exceed the threshold [68]. By tweaking the mechanism, the designer can award $k$ prizes in expectation.

Some flexibility in the number of winners also helps to generalize the permutation mechanism described above. The permutation mechanism selects one winner. If the designer aims to select two winners, Bjelde et al. [21] show that the permutation mechanism can be adapted if the mechanism is allowed to select one winner in a certain case. The permutation mechanism is run both forward and backward, and one winner is selected on each run. If the same participant is selected in both the forward and backward run, the mechanism only selects one winner.

If the prize is divisible, the designer can *reduce the total reward* to achieve impartiality. Suppose the designer would like to share the prize according to the participants' reports when the participants have consensus. The designer can discourage deviations from consensus by reducing the share of the prize awarded to participants that disagree and increasing the share of a default participant [40].

Most of the research uses a strict definition of impartiality—a participant cannot influence whether he receives the prize no matter what his peers report. We could relax this definition so that the participant cannot influence whether he receives the prize provided his peers report truthfully. Rather than looking for strategy-proof or dominant-strategy incentive-compatible mechanisms, we look for ex post incentive-compatible mechanisms. Ex post incentive compatibility is achieved if each participant has no incentive to lie when all other participants report truthfully. Several papers, including Amorós et al. [8], Amorós [6,7], Li et al. [63], Bloch and Olckers [23] and Babichenko et al. [15], use more relaxed definitions of impartiality and incentive-compatibility to construct mechanisms with good performance.

If the mechanism only needs to meet ex post incentive compatibility, the designer can construct what we refer to as *fix position* mechanisms. Provided all other participants report the truth, the participant's position in the output is fixed, and he cannot change his probability of winning the prize. Consider the following example of a fix position mechanism. Suppose three participants, $a$, $b$, and $c$, are asked to rank their peers to win a prize, and the true order is $a$ first, $b$ second, and $c$ last. If all participants report truthfully, do they have an incentive to deviate from this equilibrium? Consider $b$'s perspective when both $a$ and $c$ have reported truthfully that $a \succ b$ and $b \succ c$. No matter what $b$ reports, the designer can still use $a$ and $c$'s reports to fix $b$ in second place. At equilibrium, $b$'s report does not influence his chance of winning the prize.

The fix position mechanism applies when the mechanism's output is a ranking or selection of participants. When the mechanism outputs a grade, and the participants care only about their grade (and not their relative position), a simple way to achieve impartiality is to *fix the grade* of each participant to depend only on other participants' peer reports [85]. Participants can change other participants' grades but not their own.

We conclude this section with the disclaimer that impartiality, alone, does not guarantee a good outcome. If each participant cares only whether they receive a prize and the mechanism is impartial, each participant will be indifferent between reporting their true knowledge or preferences about their peers and any other report. This indifference gives rise to multiple equilibria, some of which may have undesirable outcomes.[5] For example, suppose an impartial mechanism selects a winner who would not receive any nominations under the participants true preferences. The participants have no incentive to change their report even though their change may cause a more desirable outcome. Mechanisms need to satisfy other properties in addition to impartiality to guarantee desirable outcomes. In the next section, we discuss the other properties that impartial mechanisms can satisfy.

[5] Some models, such as Amorós et al. [8], assume that participants prefer to report truthfully if their report does not influence their own chances of winning a prize. This assumption avoids many undesirable equilibria.

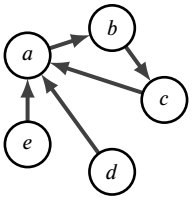
## 4. Theoretical results

The theoretical results on peer mechanisms illuminate the limits and potential of peer mechanisms. Researchers have taken two main approaches: proving which mechanisms (if any) satisfy a set of axioms and showing how certain impartial mechanisms can approximate the optimal truthful outcome.

To highlight the differences between the two approaches, we focus on peer selection, where each participant can nominate peers to receive a prize. As Table 1 shows, the model of peer selection is popular. Nominations are used as the input of the mechanism in 21 of the 37 papers on impartial peer mechanisms.

The axiomatic study of peer selection was initiated by Holzman and Moulin [51] while the approximation approach was initiated by Alon et al. [5]. In the following two subsections, we describe results from each of these two seminal works and the papers that built upon them.

The nominations can be represented as a directed graph. An edge from a participant to a peer shows that the participant nominates that peer. In the graph below, $a$ nominates $b$ and $b$ nominates $c$. Participant $a$ is nominated by $c$, $d$, and $e$ so $a$ receives the most nominations.




### 4.1. Axioms

Holzman and Moulin [51] use a model where each participant can nominate one peer, and there is a single prize. Self-nominations are not allowed. In this model, Holzman and Moulin [51] show that impartial peer mechanisms fail to satisfy two weak axioms simultaneously:

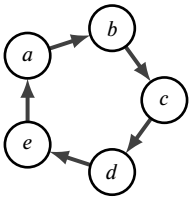
*positive unanimity:* A participant always wins if he is nominated by everyone else.
*negative unanimity:* The winner gets at least one nomination.

**Theorem 1** *(51). There exists no nomination rule that satisfies impartiality, positive unanimity, and negative unanimity*

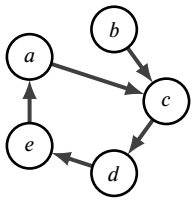
The strong impossibility result raises the question of whether similar results apply in different (perhaps more complex) models. If we allow the mechanism to award more than one prize, the impossibility no longer holds. Tamura and Ohseto [78] show that a peer mechanism with single nominations and more than one prize can simultaneously satisfy impartiality, positive unanimity, and negative unanimity, provided there are at least four participants.

To prove the result, Tamura and Ohseto [78] defines a mechanism called "plurality with runners-up". The participant with the most nominations wins a prize. If there is a tie for the most nominations, all the tied participants win a prize. Each runner-up also wins if and only if he nominates the single participant with the most nominations who only wins by one point.

The downside of plurality with runners-up is that every participant could win a prize, and, in this situation, a single participant could reduce the number of winners from everyone to just himself and one other by changing his nomination. Consider the example below; there is a cycle of nominations. Since every participant receives one nomination, plurality with runners-up awards a prize to everyone.




Suppose $a$ switched his nomination from $b$ to $c$ (as shown below). Since $c$ receives the most nominations, he would still win a prize. Participants $a$, $b$, $d$, and $e$ are all one nomination behind $c$, but only $a$ would continue to win a prize. Recall that plurality with runners-up only awards a prize to a runner-up if he nominates the single participant with the most nominations who only wins by one nomination.




Tamura and Ohseto [78] show that it is possible to adjust plurality with runners-up to have at most two winners while simultaneously satisfying impartiality, positive unanimity, and negative unanimity. However, the adjustment requires that in the case of a tie for the most nominations, the participant who is earlier in a pre-defined order always wins. Thus, participants who are earlier in the order have an advantage over participants who are later in the order.

Favoring certain applicants who happen to be earlier in an order may be an undesirable property. To address this challenge, Tamura [77] search for impartial peer mechanisms that satisfy the following axioms:

*Symmetry:* The determination of the winners is independent of the order of the participants.
*Anonymity:* An exchange of nominations between two participants does not affect whether any other participant wins.
*Monotonicity:* Receiving an additional nomination cannot cause a winner to lose his prize.

**Theorem 2** *(77). Plurality with runners-up is the only minimal nomination rule satisfying impartiality, symmetry, anonymity, and monotonicity.*

The definition of *minimal* is that the nomination rule must have the smallest set of winners while still satisfying the four axioms. For example, a rule that always gives a prize to every participant would satisfy impartiality, symmetry, anonymity, and monotonicity but would not be minimal since plurality with runners-up can satisfy the same axioms and choose a strictly smaller number of winners for at least one profile of nominations.

Anonymity is a desirable property because it gives the participants privacy when they report their nominations. The winner can be determined even if the participants complete their nominations anonymously. Unfortunately, anonymity is a difficult property for impartial mechanisms to satisfy.

If we return to the single winner setting of Holzman and Moulin [51], the only impartial nomination rules that satisfy anonymity give the prize to the same default participant—no matter the profile of nominations. However, the result only applies to deterministic mechanisms. If the mechanism designer is willing to use randomization, Mackenzie [65] shows that:

**Theorem 3** *(65). An impartial nomination rule satisfies anonymity and negative unanimity if and only if it is a uniform random dictatorship.*

The uniform random dictatorship picks a single participant uniformly at random, and this participant picks the winner. If the nominations are placed anonymously in a box, the uniform random dictatorship can be implemented by randomly selecting one of the nominations.[6]

Besides randomization, are there other features of the model that can allow the designer to break away from the incompatibility between impartiality and anonymity? Holzman and Moulin's [51] model assumes that a winner must be selected—the prize cannot remain unassigned. If the prize can remain unassigned, a threshold mechanism can satisfy impartiality, anonymity, and other desirable axioms, such as positive unanimity and monotonicity. The threshold mechanism assigns the prize to the participant who receives at least a threshold number of nominations and leaves the prize unassigned otherwise. The threshold must be greater than half the number of nominations to ensure that there is, at most, a single winner. Mackenzie [66] shows that the threshold mechanism is the only deterministic impartial nomination rule that satisfies anonymity, monotonicity, positive unanimity, and candidate neutrality (all participants are treated symmetrically when they are considered as possible winners).

The results on anonymity we have discussed thus far use nominations as the input to the mechanism. Do other inputs, such as ranks or grades, also display a tension between impartiality and anonymity? Alcalde-Unzu et al. [4] show that when participants are asked to rank all their peers and the mechanism outputs a ranking, impartiality and anonymity are incompatible. Similar to Holzman and Moulin's [51] result for nominations, the only mechanisms satisfying impartiality and anonymity for ranking output the same default ranking—ignoring the participants' reports.

### 4.2. Approximation

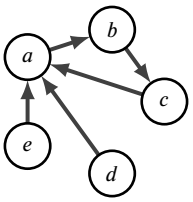
If everyone reported truthfully, the ideal mechanism would be to award the prize to the participant who receives the most nominations. Unfortunately, this ideal mechanism is not impartial. For example, suppose $a$ and $b$ both receive the same number of

[6] Anonymity and negative unanimity is not the only way to characterize the uniform random dictatorship. Edelman and Por [44] show that the rule can also be characterized by other axioms.

nominations, and one of the nominations $b$ receives is from $a$. By simply changing his nomination to any other peer, $a$ can reduce the number of nominations $b$ receives and put himself in the top position.

Can we design a mechanism that is close to the ideal but still retains impartiality? A line of research initiated by Alon et al. [5] seeks to answer this question by designing mechanisms that are impartial and closely approximate the ideal of awarding the prize to the participant with the most nominations. Alon et al. [5] defined the *approximation ratio* as the number of nominations the selected participant receives divided by the number of nominations received by the participant with the most nominations. The closer this ratio is to 1, the better the mechanism performs.

To better understand the approximation ratio, consider the example shown below. Participant $a$ receives 3 nominations, which is the most nominations received by any participant. Participants $b$ and $c$ both receive 1 nomination and $d$ and $e$ receive zero nominations. If the mechanism selected participant $b$, the approximation ratio is $\frac{1}{3}$.




If the designer is restricted to deterministic mechanisms, the results are disappointing. Alon et al. [5] find that no deterministic mechanism can provide a finite approximation ratio. The result mirrors Holzman and Moulin's [51] result that an impartial deterministic mechanism may assign the prize to a participant who does not receive any nominations.

Randomization provides more encouraging results. Alon et al. [5] show that the partition mechanism that divides the participants into two equal-size groups, only counts nominations across groups, and randomly picks which group the winner is selected from, provides an approximation ratio of $\frac{1}{4}$ in expectation. Since only the between-group nominations are counted, each nomination is counted with probability $\frac{1}{2}$. And the winner is selected from a given group with probability $\frac{1}{2}$.

In the partition mechanism with two groups, the mechanism may ignore many of the nominations. To attain better approximation ratios, Fischer and Klimm [47,48] tackle this problem by increasing the number of partitions and allowing participants to be part of many different partitions. Since each participant is part of many different partitions, they call their mechanism the "permutation mechanism". The permutation mechanism achieves an approximation ratio of $\frac{1}{2}$, which turns out to be the best possible approximation ratio.

Consider the example shown below on the left. The two participants, $a$ and $b$, both nominate each other. Without loss of generality, let's focus on participant $a$. To achieve impartiality, the probability that $a$ wins must be the same whether they nominate $b$ or abstain (the situation shown below on the right). Thus, the probability $a$ wins must be $\frac{1}{2}$ when they abstain and $b$ nominates them. But now $b$ must win with probability $\frac{1}{2}$ even though they do not receive any nominations. This situation shows that any mechanism cannot be more than $\frac{1}{2}$ optimal.


The above example relies on the case of two participants who are allowed to abstain. Several papers have circumvented the ceiling of $\frac{1}{2}$ by ruling out this case with conditions on the profile of nominations. If participants cannot abstain, the permutation mechanism achieves an approximation ratio of at least $\frac{7}{12}$ [47,48]. If each participant submits exactly one nomination, the permutation mechanism has an approximation ratio of $\frac{2}{3}$ [34]. If the participant with the most nominations receives at least a threshold number of nominations, Bousquet et al. [24] design a "slicing mechanism" that has an approximation ratio close to one. The slicing mechanism first samples some participants to decide how the remaining participants should be partitioned and the order in which the partitions should be considered. The slicing mechanism adds a sampling step to the techniques used in partition and permutation mechanisms. The nearly optimal approximation ratio of the slicing mechanism relies on placing conditions on the nominations—the input to the mechanism. We can also consider how conditions on the output of the mechanism impact the approximation ratio.

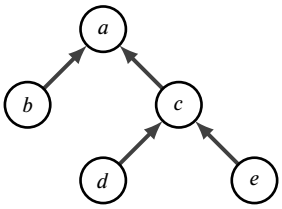
Similar to how flexibility in the prizes provides mechanisms with better axiomatic properties, flexibility in the prizes also allows for better approximation ratios. If the mechanism targets $k$ winners, Bjelde et al. [21] show that allowing for the mechanism to select fewer than $k$ winners in some situations can allow for improved approximation ratios. In comparison to the case where the mechanism must select exactly two winners, allowing the mechanism to sometimes select fewer winners can improve the approximation ratio from $\frac{7}{12}$ to $\frac{2}{3}$.

The approximation results described above all focus on the approximation ratio. Why should the mechanism designer target the ratio and not some other metric? Caragiannis et al. [26] propose additive approximation. Instead of the ratio, the designer targets the difference between the winner and the participant with the most nominations. For example, suppose the winner receives 5 nominations, but the participant with the most nominations receives 8 nominations. The approximation ratio is $\frac{5}{8}$ whereas the difference is 3. In the case of single nominations and no abstentions, a randomized partition mechanism provides an additive approximation of $\mathcal{O}(\sqrt{n})$ [26]. A deterministic threshold mechanism can also achieve an additive approximation of $\mathcal{O}(\sqrt{n})$ [31].

The study of the additive approximation ratio by Caragiannis et al. [26] and Cembrano et al. [31] assumes the mechanism can select at most one winner. If the mechanism is allowed to select many winners, the additive approximation ratio must be defined differently. Cembrano et al. [32] propose that the number of nominations for the participant with the most nominations should be compared to the selected participant with the least nominations. For example, suppose the most popular participant received 8 nominations, and the mechanism selected two participants, one with 7 nominations and another with 5 nominations. The min-additive approximation proposed by Cembrano et al. [32] is $8 - 5 = 3$. The plurality with runners-up mechanism proposed by Tamura and Ohseto [78] is 1-min-additive, which turns out to be the best possible min-additive approximation Cembrano et al. [32].

The approximation results described above all focus on the participant with the most nominations. In social networks, the mechanism designer may wish to target the most influential user rather than the user who is the most popular. In social networks, participant $a$ nominating participant $b$ can be thought of as user $a$ following user $b$. If we consider the graph of nominations, the approximation results discussed above target the participant with the maximum in-degree, but other measures of centrality may more accurately capture influence. For example, Babichenko et al. [13] define an influence measure using the expected number of paths that will end at a given participant when starting randomly at any participant.

Similar to the way participants can manipulate which participant receives the most nominations, participants can also manipulate which participant has the highest influence measure. Consider the example shown below. Participant $a$ and $c$ both have two nominations (or "follows" in the language of social media), but $a$ is more influential than $c$ because he can influence $d$ and $e$ via his influence on $c$. Suppose the mechanism selects $a$. If $c$ chooses to abstain and remove his nomination, $c$ would become the most influential participant. Participant $c$ can change the outcome by changing his nomination.




Babichenko et al. [13] design mechanisms that are impartial and approximate the ideal of selecting the most influential participant. When the nomination graph is a tree or a forest, an impartial mechanism with a constant approximation ratio to the maximum influence exists. Babichenko et al. [15] provide tighter bounds on the approximation for forests, Zhang et al. [89] improve the bounds for directed acyclic graphs when participants can only manipulate by hiding nominations, and Zhao et al. [91] design a mechanism that achieves the upper bound shown in Zhang et al. [89].

## 5. Empirical evidence

In this section, we discuss the empirical evaluation of peer mechanisms. After providing an overview of the studies and discussing evidence of manipulation, we highlight several key lessons the empirical studies provide for theory.

Table 2 lists research studies that test peer mechanisms in practice. We focus on studies where the participants providing the peer reports are also eligible for the prize. Many fascinating studies that ask a third party to report on the participants are excluded.[7]

We separate studies into three settings: field experiments, lab experiments, and observational studies. The field experiments introduce experimental treatments in real-life settings. For example, Hussam et al. [53] assigned business grants to entrepreneurs in India based on peer rankings of profitability. The lab experiments invite participants into a controlled laboratory setting and study the impact of experimental treatments. Lab experiments may be conducted in physical locations or online. Carpenter et al. [30] recruit student participants to complete tasks in a laboratory environment while Kotturi et al. [57] conduct experiments using the Amazon Mechanical Turk online crowdsourcing platform. The line between field and lab experiments can be unclear. As Chakraborty et al. [36] point out, their lab experiment studies peer grading in a classroom setting so could be considered a field experiment. The final setting is observational studies, where the researchers did not introduce any experimental treatments.

As peer mechanisms can have many different applications, the studies listed in Table 2 have been conducted in many different contexts. Many studies focus on the context of government transfers, subsidies, and loans. Policymakers have recognized that local community members may have superior information compared to the central government on which community members are in most

[7] For example, Maitra et al. [67] study mechanisms where local traders and political representatives nominate farmers to receive loans. We do not include Maitra et al.'s [67] study in Table 2 because the local traders and political representatives are not eligible to receive loans.

**Table 2**
Empirical evidence of peer mechanisms.

| Paper | Setting | Context | Input | Participants | Sample Size |
|---|---|---|---|---|---|
| Alatas et al. [2] | Field | Government aid programs | Influence beneficiary lists | Residents | 3998 |
| Alatas et al. [1] | Field | Government cash transfers | Rank 8 households | Residents | 5633 |
| Hussam et al. [53] | Field | Business grants | Rank 5 entrepreneurs | Entrepreneurs | 1345 |
| Huang et al. [52] | Field | Employee promotion | Grade peers | Employees | 432 |
| Trachtman et al. [80] | Field | Cash transfers | Rank 10 households | Residents | 300 |
| Dupas et al. [42] | Field | Poverty targeting | Rank neighbors | Residents | 507 |
| Carpenter et al. [30] | Lab | Worker compensation | Grade 7 peers | Students | 224 |
| Balietti et al. [16] | Lab | Art exhibition | Grade 3 peers | Students | 144 |
| Chakraborty et al. [36] | Lab | Peer grading | Grade 5 peers | Students | 69 |
| Leibbrandt et al. [60] | Lab | Neutral | Grade 3 peers | Students | 200 |
| Kotturi et al. [57] | Lab | Freelance job applications | 50 pairwise comparisons | MTurk workers | 320 |
| Stelmakh et al. [76] | Lab | Neutral | Rank 4 peers | Students | 55 |
| Bao et al. [17] | Lab | Crime | Nominate 1 peer | Students | 300 |
| Piech et al. [73] | Observational | Peer grading | Grade 4 peers | Students | 3600 |
| Basurto et al. [19] | Observational | Government subsidies | Nominate households | Residents | 1559 |
| Vera-Cossio [81] | Observational | Government loans | Assess loan applications | Residents | 710 |

need of aid or will make the most productive use of a loan. Mechanisms that rely on the local community to target aid (often called "community-based targeting") are a popular means to decide which community members should receive aid.

Another popular study context is peer grading of student assignments. As online delivery has enabled instructors to scale their courses to thousands of students, the instructor does not have the time to grade all the assignments. Peer grading offers a scalable solution for grading in massive open online courses (MOOCs). Although there are many studies on different facets of peer grading, we focus on studies that highlight how the manipulation of grades can be prevented.

Outside of the targeting of government programs and peer grading of student assignments, empirical studies of peer mechanisms have focused on a diverse range of contexts. Huang et al. [52], Carpenter et al. [30], and Kotturi et al. [57] focus on labor markets: peer evaluations can influence which employee is promoted, set salary bonuses, or decide which freelancer is hired. Balietti et al. [16] use peer grading to assess artworks. Bao et al. [17] use a lab experiment to show how allowing criminal suspects to nominate another suspect can reduce the overall crime level. Hussam et al. [53] use peer evaluations to determine which entrepreneurs will create the largest return from a business grant or loan.

We also note that Spliddit, a popular online tool for using fair division algorithms, has implemented De Clippel et al.'s [40] peer mechanism to divide credit for a joint project among the members of the team [50]. Empirical studies of Spliddit have focused on other functionality, such as the rent sharing algorithms [49], and the peer mechanism part of the tool has not been the main focus. Spliddit's peer mechanism is a small part of Lee and Baykal's [59] user study that focuses primarily on the algorithms for assigning chores. The study did not provide insights on preventing the manipulation of peer mechanisms.

The empirical studies listed in Table 2 use the full range of inputs discussed in the taxonomy in Section 3: nominations, rankings, and grades. Except for Kotturi et al. [57], most studies asked participants to evaluate a small set of peers for nominations, rankings, and grades. Evaluating a large number of peers is likely to be tedious or cognitively demanding. However it is unclear how many peers people can evaluate before accuracy begins to erode. Participants may also find it easier to use nominations than ranking or grading as the number of peers increases.

In the final two columns of Table 2, we list the type of study participants and sample size. The sample sizes range from large countrywide studies to small lab experiments.

### *5.1. Evidence of manipulation*

Although peer mechanisms can create opportunities to manipulate who wins the prize, do participants take these opportunities? The following examples show that they do.

- In the context of employee promotion, Huang et al. [52] shows that employees will reduce peer grades for coworkers eligible for the same promotion. If their coworker was not eligible and therefore not a direct competitor, employees tended to inflate their peer grade of the coworker.
- In a field experiment with entrepreneurs in India, Hussam et al. [53] show that peer reports decrease substantially in accuracy when the reports influence the chance of receiving a business grant.
- During a lab experiment conducted by Carpenter et al. [30], participants were asked to print, seal, and address letters to a list of recipients—complete with handwritten addresses. Each participant was then asked to count the number of letters and rate the quality of the work of each of their seven peers in the experiment. When the peer reports determined who received a bonus, the

participants under-counted the number of letters and reduced quality ratings. Participants relied more on the subtle manipulation of reducing the quality rating than the more obvious manipulation of under-counting.

- During a lab experiment framed as an art competition, participants gave lower peer review scores to direct competitors than to other peers when the prize was split among the winners [16]. When all winners received the same prize and the number of winners was unlimited, participants gave similar scores to direct competitors and other peers.

As these examples show, participants in peer mechanisms do tend to take opportunities to manipulate peer mechanisms in their favor. The examples highlight one type of manipulation—downgrading competitors—but manipulation can take other forms, such as collusion and nepotism.

Notorious cases in academic peer review provide examples of collusion [46,64]. One researcher gives another researcher a positive review on their paper in exchange for a positive review in return. The collusion can be more complex than two researchers exchanging positive reviews. A collusion ring may form where researcher a larger group of researchers exchange reviews.[8] For example, in a collusion ring with three researchers, $a$ writes a positive review about $b$, then $b$ writes a positive review about $c$, and the ring closes with $c$ writing a positive review about $a$.

For an example of nepotism, entrepreneurs in Hussam et al.'s [53] study tended to increase the rank of friends and family members—especially when the peer rankings influenced the chance of receiving a business grant.

### 5.2. Lessons for theory

The empirical research provides several key lessons for the theoretical analysis of peer mechanisms. We recognize that theoretical models should not be too complex, so we encourage researchers to choose carefully which of these lessons, if any, to include in their models.

#### 5.2.1. Participants make errors in their peer evaluations

Across different contexts, participants make errors in their peer evaluations. In peer grading of student assignments, some students consistently inflate or deflate the grades they give peers, and students differ in the reliability of their grades [73]. In the context of poverty targeting, community members often report that they don't know the ranking of two fellow community members [1].

If the mechanism designer can assume that participants do not make errors, she can design a mechanism that punishes differences in peer evaluations. If two participants share different evaluations of a given peer, one must be lying. The chance of errors makes such mechanisms difficult to implement. The designer does not know whether the participant is lying or making an honest error. Any mechanism that punishes differences or rewards consensus must consider the chance of errors.

In some contexts, participants may hold very little accurate information about their peers. For example, Dupas et al. [42] asked survey respondents in Abidjan, Côte D'Ivoire to rank neighbors from poorest to richest. The ranking contained many cycles and had a weak correlation with other measures of wealth. An open problem is to design a mechanism that can adjust to the level of peer information participants hold. Perhaps the mechanism could award more prizes or larger prizes when participants provide more accurate peer evaluations.

A lack of consensus in peer evaluations is a clear sign that the participants have made errors. But, the participants may make errors even when they agree. Trachtman et al. [80] found that although residents agreed on peer rankings of need, the rankings reflected long-term attributes, and the residents could not identify which of their fellow residents were in immediate need. If the mechanism designer aimed to collect a ranking of immediate need, the consensus peer rankings would not reflect this aim.

#### 5.2.2. Nepotism is common

The empirical research shows several examples where participants favor family members and friends in peer evaluations. Entrepreneurs gave higher ranks to family members and friends when asked to rank fellow entrepreneurs according to profitability [53]. Village chiefs were more likely to give food subsidies to family members [19]. A committee assigning local business loans was more likely to give loans to community members who were socially connected to the committee [81].

Most models of peer mechanisms consider that participants treat all peers equally. The empirical research shows that we cannot ignore the social context of peer mechanisms. Participants often favor their friends and family. Just as researchers have developed many creative approaches to discourage selfish manipulation, we also need creative approaches to discourage or detect nepotism.

#### 5.2.3. Small amounts of manipulation can be acceptable

The same studies that have shown evidence of nepotism also emphasize that peer mechanisms can still be the best option even if they are susceptible to manipulation [2,19]. A common alternative to a peer mechanism decides on winners based on participants' applications sent to an external committee. This more centralized approach may be more costly and less effective than a peer mechanism—even when participants have manipulated the outcome of the peer mechanism.

Most models of peer mechanisms try to find equilibria where all participants report the truth.[9] Allowing for small amounts of manipulation may allow researchers to take new approaches to designing peer mechanisms.

[8] Jecmen et al. [55] provide evidence that collusion rings are difficult to detect.

[9] Baumann [20] is one exception. The mechanism uses an equilibrium where participants misreport their peer evaluations.

#### 5.2.4. Choosing optimal manipulation can be difficult

During a lab experiment conducted by Stelmakh et al. [76], participants were encouraged to manipulate their peer rankings. The experiment showed that many participants struggled to employ the optimal manipulation. Participants often chose to reverse their reported peer ranking, which did not help to move themselves up the final ranking. Reversing the ranking could make the participant worse off if the participant was near the bottom of the ranking by boosting the score of their closest competitors (peers also near the bottom of the ranking).

The optimal manipulation put peers closest to the participant at the bottom of the ranking and peers furthest away at the top. With some nudging, most participants employed the optimal manipulation after a few rounds of practice playing against truthful bots.

A layer of complexity was added when playing against participants who could also manipulate their peer reports. Each participant needed to reason about the types of manipulations their peers might employ. The optimal manipulation depends on their peers' actions.

Peer mechanisms could take advantage of the complexity of finding optimal manipulations. Participants may prefer to report truthfully if they are uncertain about the optimal manipulation or cannot calculate the optimal manipulation. Conitzer and Walsh [37] survey how computational complexity can be a barrier to manipulation in voting—similar insights may apply to peer mechanisms.

## 6. Research challenges

Based on our reading of the theoretical and empirical research on peer mechanisms, we highlight several important research challenges.

### 6.1. Collusion

If peers can communicate, they can collude. "I will give you a positive review if you give me a positive review." Peer mechanisms must prevent manipulation by groups as well as by individuals.

Most existing peer mechanisms can be manipulated by groups. One partial exception is the partition approach. Reviewers placed within the same group cannot collude because their reviews only impact peers outside of their group. Unfortunately, the partition approach cannot prevent collusion between participants in different groups.

Preventing all forms of collusion is likely impossible. For example, in the model of nominating one or more peers for a fixed number of winners, Alon et al. [5] prove the impossibility of designing a group-strategy-proof peer mechanism with good performance. Even if we cannot design peer mechanisms immune to collusion, the challenge of discouraging and detecting collusion is still important.

At a minimum, the mechanism should not encourage collusion.[10] Consider a simple peer mechanism for poverty targeting—give aid to a person if he claims he is poor and his neighbor agrees. The mechanism is impartial but provides a strong incentive to collude. Suppose the claimant is rich. He could claim he is poor and pay his neighbor to agree by giving his neighbor a portion of the aid.

As Rai [74] shows, the pressure to collude can be alleviated by limiting the aid budget. If two neighbors both claim to be poor, they each receive half the budget, whereas if they claim that one of the neighbors is rich, the poor neighbor receives the full budget. The disadvantage is that poor claimants with rich neighbors receive less aid than poor claimants with poor neighbors.

Besides early work by Rai [74] on poverty targeting, we know little about how to discourage collusion in peer mechanisms, in which settings collusion is most likely, or how to detect collusion. We encourage work on these open questions.

### 6.2. Nepotism

Much research on peer mechanisms starts with the assumption that people only care about their own chance of winning the prize. If it's not me, then I don't care who wins.

Academic peer review provides a counterexample to the assumption. Reviewers are often asked to list conflicts of interest—to avoid biases towards colleagues, coauthors, and students. The conflicts of interest are usually public knowledge and can be modeled as a conflict graph. The mechanism designer can prevent manipulation by choosing a partition that respects the conflict graph [87].

What if the conflicts of interest are not public knowledge? As discussed above, empirical research shows many examples of nepotism. Participants often favor family and friends. The mechanism designer may not observe these social connections. Can the mechanism discourage nepotism without observing the conflicts of interest? Hussam et al. [53] show that using peer prediction mechanisms to pay for accuracy can discourage nepotism. Are there other approaches that can discourage nepotism?

### 6.3. Punishments not prizes

Most contexts that motivate peer mechanisms, such as peer grading and awarding research grants, involve awarding a prize. The participants want to be selected. However, peer mechanisms can also be used for assigning chores or punishments [17]. In this case, the participants want to avoid being selected.

Is the problem of assigning prizes equivalent to the problem of assigning punishments? Or are there differences that allow for different types of mechanisms? In fair division, the allocation of chores is somewhat different to the allocation of goods (e.g. [9]). We expect peer mechanisms for punishments to behave somewhat differently to peer mechanisms for prizes.

[10] For example, in designing algorithms to assign wine producers to certify fellow producers, Barrot et al. [18] recognize that if two producers are assigned to review each other, they may be tempted to collude. The algorithms include a constraint that two producers cannot review each other.

### 6.4. Machine learning methods

We might view this not as the problem of designing a peer mechanism with good axiomatic properties but as a machine learning task, with a focus that shifts to accuracy and error. We might also view the problem of identifying manipulation (and colluding peers) as a prediction problem suitable for machine learning. For example, reinforcement learning has been proposed to prevent collusion in online e-commerce platforms [25]. And similar to how deep learning has been used to design optimal auctions [43], we might be able to employ machine learning methods to learn peer mechanisms with good properties.

### 6.5. Mechanisms that can respect constraints

We might have constraints on the winners. For example, we might want a gender-balanced group of winners. One solution to this problem is to have the women vote on the male winners and the men on the female winners. However, this may not be very satisfactory if the women know more about each other and the men similarly. How then can we adapt the peer mechanisms discussed so far to deal with additional constraints like this? Such constraints have proved useful for capturing real-world issues in other areas of social choice (e.g. diversity constraints in school choice mechanisms [10]), and we may be able to adapt ideas from these domains to peer mechanisms.

Peer mechanisms that respect constraints can have unintended consequences that create new challenges. A lab experiment studied how a gender quota impacted groups of two men and two women completing tasks for payment [60]. In one version, performance was measured by peer review, and only the top two performing participants in each group received higher pay. Participants could manipulate the mechanism by under-reporting the number of tasks their peers completed. When a gender quota was in place (where at least one woman received higher pay), peers were more likely to under-report the performance of women. The difference was driven by women being more likely to sabotage women in the presence of a gender quota while men sabotaged women and men equally. The gender quota had the unanticipated consequence of intensifying competition between women and focusing women's manipulation of peer review on other women.

## 7. Conclusion

Manipulation is a very real problem in peer mechanisms where a group is selecting one or more of the group to win a prize, receive a ranking, or be given a grade. This survey identified three broad approaches to prevent such manipulation: mechanisms designed to be impartial so that a participant cannot impact their outcome, audits to detect and punish manipulation, and rewards for truthful reports. Empirical evidence of manipulation in practice suggests several outstanding research challenges, such as dealing with collusion between participants as well as various forms of nepotism. Despite the considerable body of research in this area, there remain many significant obstacles to be overcome in the design of peer mechanisms to address a range of issues met in the real world.

## CRediT authorship contribution statement

**Matthew Olckers:** Conceptualization, Investigation, Project administration, Writing – original draft, Writing – review & editing. **Toby Walsh:** Supervision.

## Declaration of competing interest

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

## Data availability

No data was used for the research described in the article.

## Acknowledgements

This research was supported under Australian Research Council's Laureate Fellowship (project number FL200100204). We thank Felix Fischer, Andrew Mackenzie, Herve Moulin, Axel Niemeyer, Shinji Ohseto, and Nihar Shah for their helpful suggestions. We also thank the anonymous referees for the detailed comments that helped to improve the paper.

## References

[1] V. Alatas, A. Banerjee, A.G. Chandrasekhar, R. Hanna, B.A. Olken, Network structure and the aggregation of information: theory and evidence from Indonesia, Am. Econ. Rev. 106 (2016) 1663–1704.
[2] V. Alatas, A. Banerjee, R. Hanna, B.A. Olken, R. Purnamasari, M. Wai-Poi, Does elite capture matter? Local elites and targeted welfare programs in Indonesia, in: AEA Papers and Proceedings, 2019, pp. 334–339.
[3] V. Alatas, A. Banerjee, R. Hanna, B.A. Olken, J. Tobias, Targeting the poor: evidence from a field experiment in Indonesia, Am. Econ. Rev. 102 (2012) 1206–1240.
[4] J. Alcalde-Unzu, D. Berga, R. Gjorgjiev, Impartial social rankings: Some impossibilities, Working Paper SSRN 4068178, 2022.

[5] N. Alon, F. Fischer, A. Procaccia, M. Tennenholtz, Sum of us: strategyproof selection from the selectors, in: Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge, 2011, pp. 101–110.
[6] P. Amorós, A natural mechanism to choose the deserving winner when the jury is made up of all contestants, Econ. Lett. 110 (2011) 241–244.
[7] P. Amorós, Implementing optimal scholarship assignments via backward induction, Math. Soc. Sci. 125 (2023) 1–10.
[8] P. Amorós, L.C. Corchón, B. Moreno, The scholarship assignment problem, Games Econ. Behav. 38 (2002) 1–18.
[9] H. Aziz, I. Caragiannis, A. Igarashi, T. Walsh, Fair allocation of indivisible goods and chores, in: Proceedings of 28th International Joint Conference on Artificial Intelligence, 2019, pp. 53–59.
[10] H. Aziz, S. Gaspers, Z. Sun, T. Walsh, From matching with diversity constraints to matching with regional quotas, in: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, 2019, pp. 377–385.
[11] H. Aziz, O. Lev, N. Mattei, J. Rosenschein, T. Walsh, Strategyproof peer selection: mechanisms, analyses, and experiments, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2016, pp. 390–396.
[12] H. Aziz, O. Lev, N. Mattei, J.S. Rosenschein, T. Walsh, Strategyproof peer selection using randomization, partitioning, and apportionment, Artif. Intell. 275 (2019) 295–309.
[13] Y. Babichenko, O. Dean, M. Tennenholtz, Incentive-compatible diffusion, in: Proceedings of the 2018 World Wide Web Conference, 2018, pp. 1379–1388.
[14] Y. Babichenko, O. Dean, M. Tennenholtz, Incentive-compatible classification, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2020, pp. 7055–7062.
[15] Y. Babichenko, O. Dean, M. Tennenholtz, Incentive-compatible selection mechanisms for forests, in: Proceedings of the 21st ACM Conference on Economics and Computation, 2020, pp. 111–131.
[16] S. Balietti, R.L. Goldstone, D. Helbing, Peer review and competition in the art exhibition game, Proc. Natl. Acad. Sci. 113 (2016) 8414–8419.
[17] Z. Bao, L. Gangadharan, C.M. Leister, Deterrence using peer information, Working Paper SSRN 3725400, 2021.
[18] N. Barrot, S. Lemeilleur, N. Paget, A. Saffidine, Peer reviewing in participatory guarantee systems: modelisation and algorithmic aspects, in: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, 2020, pp. 114–122.
[19] M.P. Basurto, P. Dupas, J. Robinson, Decentralization and efficiency of subsidy targeting: evidence from chiefs in rural Malawi, J. Public Econ. 185 (2020) 104047.
[20] L. Baumann, Robust implementation with peer mechanisms and evidence, Working Paper, 2023.
[21] A. Bjelde, F. Fischer, M. Klimm, Impartial selection and the power of up to two choices, ACM Trans. Econ. Comput. 5 (2017) 1–20.
[22] F. Bloch, M. Olckers, Friend-based ranking in practice, in: AEA Papers and Proceedings, 2021, pp. 567–571.
[23] F. Bloch, M. Olckers, Friend-based ranking, Am. Econ. J. Microecon. 14 (2022) 176–214.
[24] N. Bousquet, S. Norin, A. Vetta, A near-optimal mechanism for impartial selection, in: International Conference on Web and Internet Economics, Springer, 2014, pp. 133–146.
[25] G. Brero, N. Lepore, E. Mibuari, D.C. Parkes, Learning to mitigate AI collusion on economic platforms, ArXiv preprint arXiv:2202.07106, 2022.
[26] I. Caragiannis, G. Christodoulou, N. Protopapas, Impartial selection with additive approximation guarantees, in: International Symposium on Algorithmic Game Theory, Springer, 2019, pp. 269–283.
[27] I. Caragiannis, G. Christodoulou, N. Protopapas, Impartial selection with prior information, ArXiv preprint arXiv:2102.09002, 2021.
[28] I. Caragiannis, G.A. Krimpas, A.A. Voudouris, How effective can simple ordinal peer grading be?, in: Proceedings of the 2016 ACM Conference on Economics and Computation, 2016, pp. 323–340.
[29] I. Caragiannis, G.A. Krimpas, A.A. Voudouris, How effective can simple ordinal peer grading be?, ACM Trans. Econ. Comput. 8 (2020) 1–37.
[30] J. Carpenter, P.H. Matthews, J. Schirm, Tournaments and office politics: evidence from a real effort experiment, Am. Econ. Rev. 100 (2010) 504–517.
[31] J. Cembrano, F. Fischer, D. Hannon, M. Klimm, Impartial selection with additive guarantees via iterated deletion, ArXiv preprint arXiv:2205.08979, 2022.
[32] J. Cembrano, F. Fischer, M. Klimm, Optimal impartial correspondences, in: International Conference on Web and Internet Economics, 2022, pp. 187–203.
[33] J. Cembrano, F. Fischer, M. Klimm, Impartial rank aggregation, ArXiv preprint arXiv:2310.13141, 2023.
[34] J. Cembrano, F. Fischer, M. Klimm, Improved bounds for single-nomination impartial selection, ArXiv preprint arXiv:2305.09998, 2023.
[35] J. Cembrano, S.M. Griesbach, M.J. Stahlberg, Deterministic impartial selection with weights, ArXiv preprint arXiv:2310.14991, 2023.
[36] A. Chakraborty, J. Jindal, S. Nath, Removing bias and incentivizing precision in peer-grading, J. Artif. Intell. Res. 79 (2024) 1001–1046.
[37] V. Conitzer, T. Walsh, Barriers to manipulation in voting, in: F. Brandt, V. Conitzer, U. Endriss, J. Lang, A.D. Procaccia (Eds.), Handbook of Computational Social Choice, Cambridge University Press, 2016, pp. 127–145.
[38] J. Conning, M. Kevane, Community-based targeting mechanisms for social safety nets: a critical review, World Dev. 30 (2002) 375–394.
[39] L. De Alfaro, M. Shavlovsky, Crowdgrader: a tool for crowdsourcing the evaluation of homework assignments, in: Proceedings of the 45th ACM Technical Symposium on Computer Science Education, 2014, pp. 415–420.
[40] G. De Clippel, H. Moulin, N. Tideman, Impartial division of a dollar, J. Econ. Theory 139 (2008) 176–191.
[41] K. Dhull, S. Jecmen, P. Kothari, N.B. Shah, Strategyproofing peer assessment via partitioning: the price in terms of evaluators' expertise, in: Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, 2022, pp. 53–63.
[42] P. Dupas, M. Fafchamps, D. Houeix, Measuring relative poverty through peer rankings: Evidence from Côte d'Ivoire, NBER Working Paper 29911, 2022.
[43] P. Dütting, Z. Feng, H. Narasimhan, D. Parkes, S.S. Ravindranath, Optimal auctions through deep learning, in: International Conference on Machine Learning, 2019, pp. 1706–1715.
[44] P.H. Edelman, A. Por, A new axiomatic approach to the impartial nomination problem, Games Econ. Behav. 130 (2021) 443–451.
[45] B. Faltings, G. Radanovic, Game theory for data science: eliciting truthful information, Synth. Lect. Artif. Intell. Mach. Learn. 11 (2017) 1–151.
[46] C. Ferguson, A. Marcus, I. Oransky, The peer-review scam, Nature 515 (2014) 480.
[47] F. Fischer, M. Klimm, Optimal impartial selection, in: Proceedings of the Fifteenth ACM Conference on Economics and Computation, 2014, pp. 803–820.
[48] F. Fischer, M. Klimm, Optimal impartial selection, SIAM J. Comput. 44 (2015) 1263–1285.
[49] Y. Gal, M. Mash, A.D. Procaccia, Y. Zick, Which is the fairest (rent division) of them all?, J. ACM 64 (2017) 1–22.
[50] J. Goldman, A.D. Procaccia, Spliddit: unleashing fair division algorithms, ACM SIGecom Exch. 13 (2015) 41–46.
[51] R. Holzman, H. Moulin, Impartial nominations for a prize, Econometrica 81 (2013) 173–196.
[52] Y. Huang, M. Shum, X. Wu, J.Z. Xiao, Discovery of bias and strategic behavior in crowdsourced performance assessment, ArXiv preprint arXiv:1908.01718, 2019.
[53] R. Hussam, N. Rigol, B.N. Roth, Targeting high ability entrepreneurs using community information: mechanism design in the field, Am. Econ. Rev. 112 (2022) 861–898.
[54] K. Ito, S. Ohsawa, H. Tanaka, Information diffusion enhanced by multi-task peer prediction, in: Proceedings of the 20th International Conference on Information Integration and Web-Based Applications & Services, 2018, pp. 96–104.
[55] S. Jecmen, N.B. Shah, F. Fang, L. Akoglu, On the detection of reviewer-author collusion rings from paper bidding, ArXiv preprint arXiv:2402.07860, 2024.
[56] A. Kahng, Y. Kotturi, C. Kulkarni, D. Kurokawa, A.D. Procaccia, Ranking wily people who rank each other, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2018, pp. 1087–1094.
[57] Y. Kotturi, A. Kahng, A. Procaccia, C. Kulkarni, Hirepeer: impartial peer-assessed hiring at scale in expert crowdsourcing markets, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2020, pp. 2577–2584.
[58] D. Kurokawa, O. Lev, J. Morgenstern, A.D. Procaccia, Impartial peer review, in: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015, pp. 582–588.

[59] M.K. Lee, S. Baykal, Algorithmic mediation in group decisions: fairness perceptions of algorithmically mediated vs. discussion-based social division, in: Proceedings of the 2017 Acm Conference on Computer Supported Cooperative Work and Social Computing, 2017, pp. 1035–1048.
[60] A. Leibbrandt, L.C. Wang, C. Foo, Gender quotas, competitions, and peer review: experimental evidence on the backlash against women, Manag. Sci. 64 (2018) 3501–3516.
[61] O. Lev, N. Mattei, P. Turrini, S. Zhydkov, Peer selection with noisy assessments, ArXiv preprint arXiv:2107.10121, 2021.
[62] O. Lev, N. Mattei, P. Turrini, S. Zhydkov, Peernomination: a novel peer selection algorithm to handle strategic and noisy assessments, Artif. Intell. 316 (2023) 103843.
[63] Z. Li, L. Zhang, Z. Fang, J. Li, A two-stage mechanism for ordinal peer assessment, in: International Symposium on Algorithmic Game Theory, Springer, 2018, pp. 176–188.
[64] M.L. Littman, Collusion rings threaten the integrity of computer science research, Commun. ACM 64 (2021) 43–44.
[65] A. Mackenzie, Symmetry and impartial lotteries, Games Econ. Behav. 94 (2015) 15–28.
[66] A. Mackenzie, An axiomatic analysis of the papal conclave, Econ. Theory 69 (2020) 713–743.
[67] P. Maitra, S. Mitra, D. Mookherjee, S. Visaria, Decentralized targeting of agricultural credit programs: Private versus political intermediaries, NBER Working Paper 26730, 2020.
[68] N. Mattei, P. Turrini, S. Zhydkov, Peernomination: relaxing exactness for increased accuracy in peer selection, in: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020, 2020, pp. 393–399.
[69] M.R. Merrifield, D.G. Saari, Telescope time without tears: a distributed approach to peer review, Astron. Geophys. 50 (2009) 4–16.
[70] Y.K. Ng, G.Z. Sun, Exclusion of self evaluations in peer ratings: an impossibility and some proposals, Soc. Choice Welf. 20 (2003) 443–456.
[71] A. Niemeyer, J. Preusser, Simple allocation with correlated types, Working Paper, 2022.
[72] S. Ohseto, Exclusion of self evaluations in peer ratings: monotonicity versus unanimity on finitely restricted domains, Soc. Choice Welf. 38 (2012) 109–119.
[73] C. Piech, J. Huang, Z. Chen, C.B. Do, A.Y. Ng, D. Koller, Tuned models of peer assessment in moocs, in: Proceedings of the 6th International Conference on Educational Data Mining, 2013, pp. 153–160.
[74] A.S. Rai, Targeting the poor using community information, J. Dev. Econ. 69 (2002) 71–83.
[75] N.B. Shah, Challenges, experiments, and computational solutions in peer review, Commun. ACM 65 (2022) 76–87.
[76] I. Stelmakh, N.B. Shah, A. Singh, Catch me if I can: detecting strategic behaviour in peer assessment, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2021, pp. 4794–4802.
[77] S. Tamura, Characterizing minimal impartial rules for awarding prizes, Games Econ. Behav. 95 (2016) 41–46.
[78] S. Tamura, S. Ohseto, Impartial nomination correspondences, Soc. Choice Welf. 43 (2014) 47–54.
[79] K. Topping, Peer assessment between students in colleges and universities, Rev. Educ. Res. 68 (1998) 249–276.
[80] C. Trachtman, Y.H. Permana, G.A. Sahadewo, How much do our neighbors really know? The limits of community-based targeting, Working Paper, 2021.
[81] D. Vera-Cossio, Targeting credit through community members, J. Eur. Econ. Assoc. 20 (2022) 778–821.
[82] T. Walsh, The peerrank method for peer assessment, in: Proceedings of the Twenty-First European Conference on Artificial Intelligence, 2014, pp. 909–914.
[83] J. Wang, N.B. Shah, Your 2 is my 1, your 3 is my 9: handling arbitrary miscalibrations in ratings, in: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, 2019, pp. 864–872.
[84] Y. Wang, H. Fang, C. Cheng, Q. Jin, Tsp: truthful grading-based strategyproof peer selection for moocs, in: 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), IEEE, 2018, pp. 679–684.
[85] T. Wąs, T. Rahwan, O. Skibski, Random walk decay centrality, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2019, pp. 2197–2204.
[86] J. Witkowski, D.C. Parkes, A robust Bayesian truth serum for small populations, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2012, pp. 1492–1498.
[87] Y. Xu, H. Zhao, X. Shi, N.B. Shah, On strategyproof conference peer review, in: Proceedings of the 28th International Joint Conference on Artificial Intelligence, AAAI Press, 2019, pp. 616–622.
[88] H. Zarkoob, G. d'Eon, L. Podina, K. Leyton-Brown, Better peer grading through Bayesian inference, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2023, pp. 6137–6144.
[89] X. Zhang, Y. Zhang, D. Zhao, Incentive compatible mechanism for influential agent selection, in: International Symposium on Algorithmic Game Theory, Springer, 2021, pp. 79–93.
[90] D. Zhao, Mechanism design powered by social interactions, in: Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems, 2021, pp. 63–67.
[91] Y. Zhao, Y. Zhang, D. Zhao, Incentive-compatible selection for one or two influentials, in: IJCAI International Joint Conference on Artificial Intelligence, 2023, pp. 2931–2938.