Table of Contents

ACM SIGecom Exchanges, Vol. 24, No. 1, June 2026

Editors-in-Chief: Irene Lo and Sam Taggart

Communications Team: Yang Cai, Kira Goldner, and Jinzhao Wu

ACM Staff: Irene Frawley

Editors' Introduction

NICK ARNOSTI
University of Minnesota
and
SAM TAGGART
Oberlin College

With EC 2026 just around the corner, we are happy to add to the festivities with the Summer 2026 issue of the SIGecom Exchanges. This issue includes an update from the SIGecom executive committee, and a recap of the 2026 SIGecom Winter Meeting, written by Sara Fish, Vrinda Pareek, and Rui-Jie Yew. The technical part of the issue starts with a letter highlighting an award-winning paper from EC 2025. Next, we have four (!) excellent surveys across different topics in EconCS. The issue concludes with an annotated reading list. We detail the technical contributions below.

Our opening letter by Peng Shi describes three of his recent papers. Each of these appeared in EC, and one was chosen as the Exemplary Applied Modeling Paper in 2025. These papers study recommendation systems in two-sided marketplaces where providers vary in quality and capacity, and can choose their own prices. Although each provider's capacity is limited, a central message from these papers is that the platform need not explicitly account for this constraint: providers' will choose prices that distribute demand. These papers study three simple ranking strategies (ranking by estimated consumer surplus, by providers' willingness to pay for attention, and by a weighted combination of these terms) and show that these strategies can perform well even when the platform has only limited information.

Our first survey, by Paul Dütting, Michal Feldman, and Inbal Talgam-Cohen, offers an introduction to algorithmic contract theory. Contract theory is a pillar of classic microeconomic theory, and in the last decade or so, the EconCS community has uncovered a rich, complementary algorithmic view of the subject. Their survey provides a great first treatment of this fundamental topic.

In our next survey, Daniel Halpern, Evi Micha, Ariel Procaccia, Benjamin Schiffer, Itai Shapira, and Shirley Zhang show how tools from social choice theory can be applied to the alignment process for large language models. Intuitively, tuning processes like reinforcement learning from human feedback are typically treated as a mapping from human annotation (i.e. comparison) of LLM outputs to a global utility function over outputs. This utility function is then used to change the base model's output distribution. The process of producing the utility function is voting in disguise: the survey describes how to use existing theory, as well as new theoretical issues arising from LLM alignment specifically.

Our third survey is by Moses Charikar, Prasanna Ramakrishnan, and Kangning Wang, who provide an introduction to modern voting theory. While classical

---

Author's address: arnosti@umn.edu, staggart@oberlin.edu.

theory is filled with impossibility results, recent work has introduced new desiderata, leading to positive results and new voting rules. This survey shows how two game-theoretic concepts – maximal lotteries and stable lotteries – have been integral to proving new results on the *metric distortion* problem (where voters have cardinal utilities but express only ordinal rankings) and the *committee selection* problem (where each modestly-sized coalition of voters should be assured adequate representation on the committee).

Our final survey comes from Martino Bernasconi and Federico Fusco, who provide an overview of recent work on online learning in bilateral trade, complementing their EC 2025 tutorial on the same topic. Bilateral trade is a classic problem in mechanism design, and the Bayesian version is famously subject to the impossibility result of Myerson and Satterthwaite. This survey avoids this impossibility by considering a (non-Bayesian) online learning setting. The authors do a great job providing intuition for the key technical ideas set bilateral trade apart from other online pricing problems.

We close with an annotated reading list by Amin Rahimian and Yuxin Liu on Differential Privacy. This includes three classic papers and seven modern papers (all from the last four years), which connect differential privacy to platform design, revenue management, recommendation systems, and more.

Looking ahead to future issues, we are pleased to announce an **open problem initiative**, which will be supported jointly by the Exchanges staff and SIGecom communications team. The idea will be to circulate writeups of important unsolved problems in EconCS, both through the Exchanges and in an online repository. The initiative is still in the very early stages: expect further discussion at the EC 2026 business meeting and more formal solicitation over the SIGecom listserv. In the meantime, start thinking whether you have something to contribute!

Thanks to all the authors in this issue for their insights and hard work. Further thanks to communications chair Yang Cai, technical lead Jinzhao Wu, and social media chair Kira Goldner. Their help publishing this issue is greatly appreciated. Please continue to volunteer letters, surveys, annotated reading lists or position papers. We hope you enjoy this issue.

Letter from the SIGecom Executive Committee

MICHAL FELDMAN (chair)
Tel Aviv University
and
FEDERICO ECHENIQUE (vice-chair)
UC Berkeley
and
BRENDAN LUCIER (secretary-treasurer)
Microsoft Research New England

As we enter our third year serving as the SIGecom Executive Committee, we continue to be inspired by the remarkable growth, breadth, and energy of the Economics and Computation community. Over the past several years, SIGecom has expanded substantially in size, visibility, and intellectual diversity, and we feel fortunate to help support this evolving community during such an exciting period. Our goals remain to strengthen the foundations of the SIG while continuing to foster interdisciplinary exchange, broaden participation, recognize impactful research, and build structures that help the community thrive as it scales.

At the same time, our community is entering a period of rapid change driven by the growing role of AI in research and scientific practice. These developments raise important questions for our field: Substantive questions about the role of AI in subjects that we study (markets, allocation mechanisms, decision-making, algorithms, etc.), as well as methodological questions. How AI should be used in paper writing and reviewing, how conferences and journals should adapt their processes, how to preserve fairness and trust in evaluation, and more broadly, how to ensure that our research culture continues to encourage creativity, rigor, and genuine scientific contribution. We believe that the economics and computation community is uniquely well positioned to help address many of these challenges. Questions surrounding incentives, strategic behavior, information, marketplaces, and mechanism design are at the core of our field, and they are becoming increasingly central in the AI era. As a community, we hope not only to adapt thoughtfully to these changes, but also to help shape the broader scientific and societal conversation around them.

Our flagship event, the ACM Conference on Economics and Computation (EC), continues to grow in size and diversity.

EC'25 was hosted at Stanford University this past summer, with 927 submissions and 206 accepted papers. The conference featured outstanding work across theory, empirics, and applications. We are grateful to General Chair Ilya Segal, PC Chairs Itai Ashlagi and Aaron Roth, and the many organizers and volunteers who helped make EC'25 a success. The conference was accompanied by a robust workshop program, a special session on highlights beyond EC, along with a virtual preview week that featured the annual mentoring workshop and many high-quality tutorials.

The continued expansion of the conference is tremendously exciting and reflects

Authors' addresses: mfeldman@tauex.tau.ac.il, fede@econ.berkeley.edu,brlucier@microsoft.com

the vitality of the field, but it also raises important questions about how to preserve the quality and character of EC as the community grows. In particular, the increasing submission volume continues to put pressure on both the reviewing process and the conference structure itself. Together with conference organizers, the PC chairs, the advisory committee, and the entire community, we are actively exploring different approaches for addressing these challenges while maintaining the openness, rigor, and sense of community that make EC special. As always, we very much welcome feedback and ideas from the community as we continue thinking about the future evolution of EC.

We look forward to EC'26 this year, to be hosted in Rome, Italy. EC'26 has brought with it a lot of change and exciting challenges that we look forward to sharing with you. One important change is that we are moving to a 5 day conference format. This allows us to increase the conference size in line with increased interest, submissions, and topic diversity. Overall, EC'26 had 1115 submitted papers from which 290 were accepted.

There have also been several changes to the reviewing process, spearheaded by our PC Chairs Alex Teytelboym and Matt Weinberg. First, they welcomed hundreds of new reviewers to EC. This lowered the reviewing load, improved the paper-expertise matching, and brought many new excellent researchers into the EC community, but also required more hands-on structure for the review process. Additionally, they implemented a much narrower and focused author response period, and also asked for more structured engagement from SPCs. Unsurprisingly, AI has also been a point of change in the reviewing process. All submissions were offered a chance to get AI feedback from Refine.Ink, while the PC chairs took exploratory steps to understand how future years might integrate AI effectively into the review process. They were also strict about AI disclosures for both authors and reviewers, and were careful to ensure every paper had truly sufficient expertise/discussion in case of concerns surrounding inappropriate AI usage. This year, EC has added 5 AEA journals (AER, AER:Insights, AEJ: Micro/Applied/EconPolicy) and Quantitative Economics to the forward-to-journal program, bringing the total up to 23 journals.

This is also the first year that the ACM Open program goes fully live for all ACM conferences, including EC. We recognize that this is a meaningful change to the publication process that may introduce new costs for some authors. The SIG is committed to easing this transition and has set aside additional funds to help cover fees this year. We encourage anyone with questions or concerns to reach out.

We are grateful to General Chair Stefano Leonardi, PC Chairs Alex Teytelboym and Matt Weinberg, and the many organizers and volunteers, without whom EC would not be possible.

The SIGecom Winter Meetings have continued to grow into a vibrant forum for focused discussions. The 2026 meeting, co-organized by Rachel Cummings and Inbal Talgam-Cohen, focused on the *interplay of Algorithmic Game Theory, AI, and Law.* Through a combination of invited speakers, contributed talks, and a fireside chat, it explored emergent research topics in the intersection of CS and law during a time when the impact of AI is making the influence of algorithms greater than ever.

We are working with the ACM to build a historical repository for the SIG: ar-

Fig. 1. Left: The SIGecom Executive Committee hard at work. Right: EC'25 plenary session.

tifacts and documents related to the history of SIGecom from its founding to the present day. To this end, **we need your help!** If you have any historical documents about the SIGecom and/or EC, or perspectives on the history of our community, please reach out – we would love to hear from you.

We are pleased to announce a new SIGecom committee to help coordinate nominations from our community to the Communication of the ACM (CACM) Research Highlights section. The committee will be led by Robert Kleinberg and Amy Greenwald, who will work together with the EC PC chairs and the broader community to identify especially strong papers that may be suitable for RH nomination. We encourage members of the community to reach out to them with suggestions for papers that they believe would be exciting and broadly impactful candidates for CACM Research Highlights.

The SIGecom Communications Committee was recently formed to support and expand SIGecom's public-facing communication channels and community resources. The committee consists of Yang Cai (chair), Kira Goldner (social media lead), and Jinzhao Wu (technical lead). Its current activities include maintaining SIGecom webpages; uploading and organizing EC-related videos on YouTube; managing SIGecom's social media presence; and providing technical support for SIGecom's online infrastructure. The committee is also developing new community-facing initiatives, including a centralized repository of EC-related courses and a possible open-problems initiative. The course repository would collect links to publicly available course materials across the field, while the open-problems initiative would make research questions, background, references, and possible directions more visible and accessible to the community. As a crucial part of the SIG's communication efforts, we also want thank the *SIGecom Exchanges* Editors-in-Chief, Sam Taggart and Nick Arnosti, for all of their hard work.

We'd like to take this opportunity to reinforce our ongoing call for *Special Initiative* proposals. SIGecom has annual discretionary funding available to support community-building efforts such as mentoring programs, inclusion efforts, and other pilot events. Past initiatives have included the annual EC Mentoring Workshop, EC childcare support funding, and the SIGecom Winter Meetings. We encourage members to propose creative new initiatives that further SIGecom's mission.

Finally, we're thankful for the many volunteers who contribute to SIGecom's

success in both visible and behind-the-scenes roles. From SIG award committees, to conference and workshop organizers, to the many editors and reviewers in economics and computation: the SIG could not function without you! As always, we invite members of our community to get involved. Whether you're interested in organizing an event, submitting a nomination for one of our awards, launching a special initiative, or serving in a formal role, we welcome your participation.

We once again thank you for the opportunity to serve.

SIGecom Winter Highlights 2026

SARA FISH
Harvard University
and
VRINDA PAREEK
Columbia Law School
and
RUI-JIE YEW
Simons Institute for the Theory of Computing, UC Berkeley
Brown University

Sara Fish is a PhD Candidate in Applied Mathematics at Harvard University and Graduate Fellow at the Harvard Kempner Institute. Her research spans diverse topics at the intersection of EconCS and artificial intelligence, including algorithmic collusion, AI-enhanced democratic processes, and AI evaluations. Previously, she earned her B.S. in Mathematics from Caltech.

Vrinda Pareek holds an LL.M. degree from Columbia Law School. Her coursework focused on technology law, including copyright law and the regulation of artificial intelligence. She is an India-qualified lawyer, where she earned her B.A. and LL.B. degrees.

Rui-Jie Yew is a Law & Society Fellow at the Simons Institute for the Theory of Computing and a PhD candidate in computer science at Brown University. She conducts research on AI policy and incentives, alignment, and avoision in regulation. She holds an S.M. from the Massachusetts Institute of Technology and a B.A. from Scripps College.

The sixth annual ACM SIGecom Winter Meeting took place on 25 February, 2026. Organized by Rachel Cummings and Inbal Talgam-Cohen, the meeting explored topics at the confluence of Artificial Intelligence, Law, and Algorithmic Game Theory. We share some highlights from the 2026 Winter Meeting. Thanks to the organizers and all the participants!

The meeting kicked off with opening remarks from **Michal Feldman**. As part of these remarks, Feldman discussed topics of AI safety, AI regulation, and incentive alignment towards the prevention of harm and provided an overview of the meeting's structure. This year's meeting took the form of a virtual workshop and included talks, panels, and fireside chats with experts. These segments included presentations on motivating tne field of CS&LAW, discussion on published work, as well as discussion on future research ideas directions.

1. AT THE INTERSECTION OF COMPUTER SCIENCE AND LAW (DANNY WEITZNER)

As part of the Introductory Talks, we hear from Daniel Weitzner as the presenter and Kobbi Nissim as the discussant about what it means to do research at the intersection of Computer Science and Law.

Daniel Weitzner is a founding director of the MIT Internet Policy Research Initiative and is a Principal Research Scientist at MIT's CSAIL. He leads technically grounded research to inform critical internet public policy and data governance

research. He has worked at the National Telecommunications Information Agency, the Center for Democracy and Technology, the Electronic Frontier Foundation, and more. As part of this talk series, Weitzner discusses what it means to do research at the intersection of CS&LAW

Kobbi Nissim is a Professor of Computer Science at Georgetown University and an Affiliate Professor at Georgetown Law. Before that, he was a professor of computer science at Ben-Gurion University, and he spent several years at Harvard University in the Center for Research and Computation and Society. His work establishes rigorous practices for privacy in computation, including being one of the inventors of differential privacy. He also works at the intersection between privacy law and policy.

This talk is based on *At the Intersection of Computer Science and Law* [Feigenbaum and Weitzner 2026], a paper that he and **Joan Feingenbaum** submitted to ACM CS&LAW. In the paper, they lay down the motivations for launching the conference, which is now roughly in its fifth year. In this talk, there are two main problems that Weitzner presents as motivation for the studying of CS and law together:

(1) There are gaps in addressing the legal and policy challenges information systems have presented us.

(2) On the other hand, there is a lack of understanding in the building of governable systems or when systems are governable.

These are problems in both technology policy and system design that Weitzner has worked on and that a number of colleagues in CS&LAW have worked on that are still not particularly well-addressed. He discusses some research that answers to some of these gaps.

1.1 Starting from Law

This class of research starts from how can you tell if system architecture satisfies legal requirements. Some examples of this class of work include [Cohen and Nissim 2020], which defines the formal property of "predicate singling out" in understanding the notion of singling out in the GDPR. Another work includes [Meding and Sorge 2025] shows that a number of editing features will blur lines between editing and manipulation.

1.2 Starting from CS

This class of research starts from observations about properties of systems which may give rise to behaviors or functions that require some regulation/control. One example here includes the question of how courts should apply principles to modern encryption systems. Here, [Cohen et al. 2022] provides formal definitions that could help courts determine when a specific requirement to decrypt communication would run afoul of this constitutional protection. Another example of research that falls into this category is [Lee et al. 2024] which asks the "right questions" in considering where liability ought to be assigned in AI production under copyright law.

1.3 Broader Examples

Weitzner also discusses broader examples of research that may not fit neatly in either of the above two categories. For example, [Chan et al. 2025] tackles the question of balancing privacy and accountability in misconduct settlements and shows how use of cryptographic techniques like zero-knowledge proofs and private computation can help strike balance between privacy and accountability.

1.4 Intersection Rules

Finally, Weitzner discusses what it means to conduct "scholarship that makes a contribution in the intersection, as opposed to the union, of computer science and law": (1) **takes both disciplines seriously**, (2) **could be "publishable" contribution in both or either field**, (3) **but publishable value could be solely the intersection**.

Going forward, he hopes this area of research will help policymakers apply more rigorous system applications and conversely that it will help system designers build more governable systems. He hopes researchers continue to study the fit between law and CS with an eye toward effective laws and efficient systems.

1.5 Response: Scope and Mission of CS&Law: (Kobbi Nissim)

Nissim starts by motivating the urgency involved in answering questions at the intersection of the two fields. There are a growing number of decisions of legal consequence are made in sociotechnical systems and there are questions surrounding the scale of having to review the decisions made by sociotechnical systems. And, even if only a small fraction of the decisions made in sociotechnical systems required human review, they would quickly overwhelm our judiciary or administrative systems. When he first got started in this field, he noticed the two fields (CS and law) having: (1) **same words, different meaning**: PII, composition of datasets and (2) **different reasoning**: "and" meaning as a logical connector, **different/conflicting values and goals**: CS theory values precision while law values flexibility.

Nissim concludes with hope regarding the identification of paradigms for bridging between the two disciplines, while respecting their differing approaches and values. He looks forward to seeing the paper which does not satisfy intersection rule, but which does belong to CS&LAW.

2. CS&LAW MEETS ALGORITHMIC GAME THEORY (KATRINA LIGETT)

Katrina Ligett is a professor of computer science at the Hebrew University, a director of the Federal Center for the Study of Rationality, and a visiting associate at Caltech. She's one of the founders of the field of CS and law and led one of the first workshops at the Simon Institute back in 2019, which brought together experts from theoretical computer science, law, economics, and the social sciences more broadly. In this talk, Ligett catalogs questions that may be of interest to the triangle of CS, Law, and Econ. How do we walk into this simplex of disciplines? What are the questions and tools this can raise for us? Ligett starts from the less developed CS-law edge in the simplex and considers the value that a lens of algorithmic game theory (AGT) can bring.

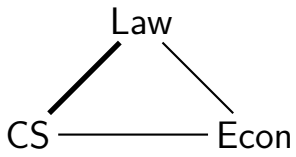


Fig. 1. Simplex of law, computer science, and economics with the CS-law edge highlighted.

Combing through the topics as part of the call for papers at the ACM CS&LAW conference [Feigenbaum and Weitzner 2026] with AGT "glasses", she identifies the following themes: **(1) Incentives**, **(2) Information**, **(3) Contracts**, and **(4) Attribution**. She presents the following research questions related to these themes:

2.1 How can AI and humans collaborate?

This body of questions concerns settings where you have different parties interacting. Here, *asymmetry* is a recurring theme and there are a variety of different asymmetries. Some examples include big tech companies and those regulating them, companies vs those using them, etc. These challenges are amenable to people with a toolkit (e.g., in economics and AGT) to model these asymmetries. An example of work concerning this topic is [Collina et al. 2026].

2.2 What *can* AI produce? How can we incentivize richness in the output of AI?

These types of questions relate to learning theory and results from learning theory about the limitations of the ability to recognize a language from samples is a computationally hard problem. [Kleinberg and Mullainathan 2024] show that it is possible to generate from a language given access to samples from that language. But there are also interesting challenges that are sort of implicitly raised here. Just because you can generate more samples from a language doesn't mean you're actually representing the full sort of breadth of that language. So, how do you potentially get better representation over the sort of the limits of getting richness in the outputs of AI?

2.3 How to incentivize (human) creative work?

This set of questions concerns the certification of creative work. A world where we can generate content cheaply and at scale raises a whole host of questions about how would you recognize creative work, how could you certify creative work, what is authorship, what rights should authors have in their work, how do you protect the rights of authors in their work. The existing infrastructure that we have for incentivizing human creative work, e.g., the copyright/patent family of tools on the regulatory legal side, seem to be insufficient to recognize and promote the things that we (potentially) want to promote in today's world. But there are a huge number of questions here, and some of them very quickly start to look like AGT questions: questions about what credit belongs to previous works, how do you identify intellectual contributions as people are building on each other's ideas. There are various initial works that chip off various pieces of these questions while echoing things that are familiar to us, like Shapley values.

2.4 How to balance freedom of expression with protections?

Now that we can produce lots of content very cheaply now with AI, we can also produce a lot of false, extremely convincing, personalized speech. Ligett asks the question here of: if we value meaningful, positive, constructive, useful, correct information and discourse, how can we promote that? What are the mechanisms that can be put in place? And again, there are questions about incentives, there are questions about recognizing this, there are questions about how do we build the systems that can promote this and protect this and ensure that there's room for expression that maybe we still value.

2.5 How can we ensure that AI companies' incentives align with societal objectives?

Stochasticity in modern machine learning and company incentives can make naively auditing AI safety difficult. There are also problems surrounding visibility. Most individual interactions with AI are not directly made public. Example works concerning this topic include [Saig et al. 2024; Singh et al. 2026; Hadfield-Menell and Hadfield 2019].

2.6 Who/what is responsible for what?

This surrounds responsibility for speech/responsibility for actions. When something goes right/wrong who gets the credit/is held liable? The tools that we have to hold entities liable are somewhat limited. A related question here is: how can LLMs avoid violating copyright and attribution/credit/holding actors liable.

2.7 How to track/allocate trust in information?

Now, we are flooded with fake information, and so we are going to need some new infrastructure to help us understand how to get access to information we value. Some infrastructure we are missing is infrastructure that allows us to build networks of trust to make sense of claims/events/facts.

2.8 When algorithms are used for pricing: what's harmful, what's helpful, and what's legitimate?

We're going to see and we're already seeing personalization of prices. We have to decide what are the limits, what's acceptable, what's unacceptable. And then it raises all sorts of questions about how can we enforce, how can we detect, how can we ensure, how do we build incentives. There are a lot of concerns around the ability of AI, again, to be extremely personalized and extremely convincing, but also potentially to trigger our biases and take advantage of us in various ways. AI is also going to be in a position to potentially have access to much more information than individual humans, and potentially even individual humans who have their weaker AIs working on their behalf. So we have to be concerned about taking advantage of information deficits that consumers might have. We have to ask ourselves questions about when is personalization beneficial for social welfare, for fairness, for efficiency? What do we want to accept? What do we not want to accept? There are algorithmic and definitional questions here, as well as questions concerning the translation of sort of more abstract concepts that might come from ethics of the law or policy and trying to figure out what do they mean for us at a

technical level and how can we realize them. Example of works concerning this topic and algorithmic price discrimination in particular include [Bar-Gill et al. 2023].

2.9 How to define/detect/prevent algorithmic collusion?

There are really interesting questions around as these algorithms are often operating in wider and wider ranges of decision-making, we have a lot of potential for them to collude. They're going to be interacting with each other. They're going to be interacting with complex systems. And so there are a lot of questions about how do we define, how do we detect, how do we prevent algorithmic collusion? What does this look like? In particular, one tension between CS and law is that of intent in collusion. Example works concerning this topic include [Hartline et al. 2024a] [Fish et al. 2024].

2.10 Incentives: How to nudge the legal system (and the world) to adopt beneficial technologies?

So some examples are things like zero-knowledge proofs [Bitan et al. 2022], where we have a lot of technical capabilities that can do a lot of good, and that can improve a lot of pretty broken systems that we have. These questions surround filling that last gap between theory making things work in the real world.

2.11 In what sense does the law learn?

Common law has this concept of precedent where we sort of look to decisions that have been made in the past, and these help inform basically what the rules are today. But this can be framed as a really sort of interesting type of a learning problem where there's a decision to be made about whether to bring a case, and that decision encompasses a bunch of considerations, including possibly sort of the cost of bringing the case, the likelihood of succeeding. But if there is an interesting case that is never brought, then we never get precedent, and in some sense the law doesn't learn. Sometimes part of the incentive to bring a case is actually to set precedent. And so there's been some work on sort of this learning kind of algorithmic take, which has a kind of law and economics vibe in the sense that sort of trying to model the system that is the law [Hartline et al. 2022][Dutz et al. 2025]. Ligett also points out the computational and algorithmic flavor to this learning problem.

3. TALIA GILLIS

Talia Gillis is a professor of law at Columbia Law School, and studies the law and economics of consumer markets. She is interested in household financial behavior and how consumer welfare is shaped by technological changes. She is a recipient of the 2022 junior faculty grant and the Richard Paul Reichman Center for Business Law and Public Policy Grant. She completed her SJD degree at Harvard and a PhD in economics and clerked for Deputy Chief Justice Mer at the Supreme Court of Israel.

Gillis addressed law's (and lawyers') discomfort with ambiguity, and posited that regulatory ambiguity may, in fact, be an important strategic tool. Gillis examined the complex relationship between legal standards and algorithmic systems, focusing on the persistent problem of regulatory ambiguity. While she initially sought

to eliminate vagueness in laws governing consumer finance and anti-price discrimination, her views evolved through her engagements with policymakers. Now she thinks of this imprecision as a strategic tool used by regulators. Gillis argued that this flexibility in legal standards or regulatory approaches prevents firms from "discrimination hacking" or gaming fixed regulatory targets, and it allows regulatory oversight to adapt to rapidly evolving technological changes. By framing legal design as a mechanism design question, Gillis suggested that the optimal level of precision desirable in regulation depends on the specific strategic environment and the incentives of the actors involved. Ultimately, her work highlighted the tension between the desire for clear compliance rules and the necessity of maintaining enforcement discretion to prevent substantive harms in automated markets. Gillis encouraged participants to think of ambiguity in legal standards not as 'sloppy' drafting, but as a deliberate decision to preserve flexibility, deter gaming by regulated actors, or maintain enforcement discretion.

4. PANEL ON GENERATIVE AI, AGENTIC AI, LAW, AND CS

The 2026 Winter Meeting featured a dynamic panel discussion on GenAI, Agentic AI, Law, and CS, featuring four panelists with rich experience across economics, law, and computer science.

Vince Conitzer is a Professor of Computer Science at Carnegie Mellon University, where he directs the Foundations of Cooperative AI Lab (FOCAL). He is also Head of Technical AI Engagement at the Institute for Ethics in AI, and Professor of Computer Science and Philosophy, at the University of Oxford. He began his PhD in 2001 under Thomas Sandholm and has been an active member of both the EC and AI research communities throughout his career.

Peter Henderson is an assistant professor at Princeton University, with appointments in the Department of Computer Science and the School of Public and International Affairs, as well as the Center for Information Technology Policy. He received his J.D. from Stanford Law School and PhD in computer science from Stanford University.

Zoë Hitzig is a Junior Fellow at the Harvard Society of Fellows. She received her PhD in economics from Harvard in 2023. Between 2024 and 2026 she was Research Scientist at OpenAI, working on their safety research team and also working on their economic research team.

Georgios Piliouras is a senior staff research scientist at Google DeepMind, where he leads the Game Theory team, and Associate Professor in the Engineering Systems and Design Pillar at the Singapore University of Technology and Design (SUTD).

Below are selected questions and panelist responses, edited for clarity and brevity.

What do you see as the most pressing research questions in generative and agentic AI that can be addressed by the EconCS and CS&Law communities?

Georgios Piliouras: Sometimes we use syntactically similar languages to talk about rather different types of problems. For example, contracts. In CS, we talk about algorithmic contract theory, and we have some very well-designed problems in that space, on which we have made reasonable progress. But also, there are the

types of contracts that lawyers develop in practice. These two things are rather different. I would be interested to see if there are possible connections to make in this space.

Zoë Hitzig: I think that one of the most important questions for this community to be thinking about is data governance and property rights for data. Right now, the conversation about who owns what data and how different kinds of data get used in large AI systems has been co-opted by, first of all, the big labs who have an incentive to talk about it in a certain way, but also to some degree by lawyers who have one main apparatus for thinking about property rights for data, which is copyright. But as economists, we understand copyright from first principles: you want some mechanism like copyright so that people are incentivized to produce valuable stuff. So I think one of the most important and understudied questions right now is: what are the incentives that people have to produce the data that goes into large machine learning systems? And how do different mechanisms and institutions change the landscape and incentives that people had to produce that data in the first place? I think there are ways to talk about this both on a macro scale and on a micro scale.

Vince Conitzer: In my lab, we're thinking a lot about the interactions between AI systems, especially from a strategic perspective. We also think a lot about the ways in which AI agents are different from human agents and what that allows you to do. For example, an AI agent can be run in a simulation to see what it will do—you can't quite do that with a human. That gives different ways to achieve trust and cooperation. Finally, related to Zoë's points on incentives for content production, a very difficult—yet often legally significant—question is: can we understand why a language model produced any given output?

Peter Henderson: There are infinitely many things to work on right now because of how fast the field is moving. But I'll give three examples of things I consider high priority. First, the problem of specifying natural language rules and standards for models to follow. You can think about this from an economics perspective (e.g. contracts), from a legal perspective (e.g. statutes), or from a computer science perspective (e.g. reward specification). I think this is an overlooked problem, but one where all three fields can come together to make progress. Second, how do we reshape the law when it comes to copyright and fair use? This will require both a technical understanding of the models and how they're trained, as well as economic questions of impacts on the market and the value of the labor. Third, what are the right policies and legal tools by which to tackle potential labor market impacts? This again requires all three fields to come together to understand: How well are the models actually performing, so we can better project what might happen? What are the potential macroeconomic and microeconomic effects? And then, what are the legal and policy options that we have that we can actually intervene at a reasonable timescale?

As we approach AGI—that is, AI that can match or surpass human capabilities across virtually all cognitive tasks—there is a tension between utilizing this emerging technology, and protecting human jobs and expertise. What do you see as our way through this?

Vince Conitzer: When we should keep humans in the loop is a difficult question.

I think that it varies across different professions. You could make a good argument in law, that maybe there's a special reason to have humans still be part of it. More broadly, if AI doesn't advance too fast in the near future, maybe one of the main problems is disruption of hiring pipelines. We need people at the senior level who really understand their profession, but they won't have a path to get there if the entry-level work is done by AI. So maybe that's another reason to keep people in the process—to understand what the processes are.

Zoë Hitzig: Hayek's key insight in *"The Use of Knowledge in Society"* was that we need people to have decision-making power on the fringes. We can think about this in terms of two theses. First, things need to be decentralized because people have tacit, immovable knowledge in their heads or because of where they are physically. Second, even if we could move all the information to some central source, that source wouldn't be able to process all of it. As Eric Brynjolfsson and I have argued [Brynjolfsson and Hitzig 2025], these two premises are eroding. AI is finding new ways to make knowledge that you couldn't codify before codifiable, and it's taking away the constraints on information processing. If we take that seriously, the economic reasoning behind decentralized markets is going away—centralization might be economically efficient. And that might not be great, because part of how people had bargaining power in the economy was that they had knowledge that couldn't be taken from them. So the question becomes: what can we do to give people bargaining power in the new economy? What are the countervailing institutions that we can build now for this completely new kind of economy?

If something goes wrong with AI, who is legally liable? The AI developers, the users or somewhere in between?

Peter Henderson: Even defining what systems are in scope is challenging—for example, autonomous vehicles and systems that act in the real world versus simple systems in video games or apps that might have very different types of downstream harms. To my mind, one of the most interesting short-term areas where we're going to have to figure this out is cybersecurity harms. It's pretty clear that these models can already be repurposed for third-party harms pretty easily, and on the technical safeguards question, it's pretty hard to identify these misuses in real time at scale. Now, if we look to past cybersecurity cases, we actually don't have that much litigation against providers of tools that already provide software to penetration test or to run attacks that can be used both offensively and defensively. I don't think there's going to be this beautiful framework in which we figure out how to allocate exact liability in every scenario. I think in practice it gets really messy really quickly, and we'll just have to wade through many of these cases. In the cybersecurity setting, for example, we do have duty of reasonable care standards and negligence standards that may well cover many of these scenarios and attribute liability in a perfectly reasonable way. I don't know that we need to go in and completely swap in a new framework when we have yet to work through the existing frameworks that may well apply already.

The EC community has had so much interaction with online ad auctions and allocating the resource of attention. Zoë, how might this relate to your remarks on bargaining power?

Zoë Hitzig: I think there is some way of creating the right institution that

gives people power over how they spend their attention. If we think about how we came to be auctioning off people's attention in the first place, it was a property rights issue—when Google said, hey, those search histories are mine. So there is this connection between the data rights issue and who has control over their own attention. As soon as you've given over tons of data, you have opened yourself to a certain kind of manipulation or advertising. There's an amazing book out called *Attensity* that is laying out a way to bring back our attention—it's not in our language, but I think it's a book that we might all be able to learn something from.

Vince Conitzer: Related to Zoë's response, there's an article by Marietje Schaake in the Financial Times about what she calls "botlash." A phrase associated with that is "resist and unsubscribe." The "unsubscribe" part really gets at using your attention for bargaining power.

Doesn't AI make cybersecurity more robust, since we can potentially find exploits before criminals do?

Peter Henderson: The key challenge is that defenders can use AI models to try to patch vulnerabilities, but each patch takes time to deploy. If both parties have equal access to the same model, there's a time delay to patching all the vulnerable systems, whereas the offensive side can start attacking right away. So if there was privileged early access, maybe accounting for this potential time delay, you could have it be a bit more balanced.

Any final thoughts?

Georgios Piliouras: I think change is coming, and that the rate of change is going to keep increasing. I have no predictions for the next five years, because one year already feels intense. We are in a very privileged minority to actually participate in these technical discussions and help steer this technology in a direction that is beneficial for all. I think we should take that opportunity seriously and collaborate.

5. BLAMELESS USERS IN A CLEAN ROOM: DEFINING COPYRIGHT PROTECTION FOR GENERATIVE MODELS (ALONI COHEN)

Aloni Cohen is an assistant professor of computer science and data science at the University of Chicago. Previously, he was a postdoctoral associate at Boston University with a joint appointment in the Institute for Computing and the School of Law. His research explores the interplay between theoretical cryptography, privacy, law, and policy. He completed his PhD at MIT, where he was advised by Shafi Goldwasser.

Cohen's core problem statement highlighted that current generative models risk "regurgitating" copyrighted training data, potentially making unwitting users liable for infringement. His research proposes a formal mathematical framework to define and provably prevent copyright-infringing outputs by focusing on the relationship between model training and "blameless" user behavior. Cohen identified gaps in the definition and conception of existing "Near Access Free" (NAF) models, which rely on a differential privacy-style comparison between models with and without a specific work. Cohen argued that NAF is insufficient because models can still learn a work through its derivatives (parodies, sequels) and because the guarantee fails

to compose across multiple user prompts.

To remedy these gaps, and drawing on the legal concept of "clean room design,"Cohen introduced a new standard for training algorithms. The framework defines a "blameless user" as one who, when provided only with the unprotected *ideas* of a work, has a negligible probability (β) of recreating the protected *expression*. A model is considered "kappa-beta clean" if it guarantees that such a user's probability of copying after interacting with the model remains below a specified threshold (κ). Cohen specified that provable protection under this framework would require two components: strong differential privacy, and golden data.

6. AI SUPPRESSION: E-DISCOVERY SOFTWARE AND BRADY (REBECCA WEXLER)

Rebecca Wexler is a professor of law at Columbia Law School whose teaching and research sit at the intersection of law and technology. She has a specific focus on privacy and secrecy in the context of the criminal legal system. She has served as a senior policy advisor in the White House Office of Science and Technology Policy and has testified before both the House and the Senate Judiciary Committees.

Wexler presented a paper co-authored with Jason Hartline, Liren Shan, and Alex Sun. It explores how machine learning-based e-discovery software, which helps litigators review large amounts of evidence, might either facilitate or hinder the discovery of exculpatory evidence of innocence in criminal cases, depending on how it is configured.

Wexler highlighted the complete lack of regulatory guidance on the use of machine learning (ML) e-discovery tools in criminal proceedings, and specifically regarding compliance with the *Brady* rule that obligates prosecutors to disclose exculpatory evidence to the defense in criminal proceedings. Wexler (and the authors of the paper) share the concern that current ML workflows in the field, specifically Continuous Active Learning (CAL), prioritizes documents it is most confident are relevant to the human reviewer, unlike classic machine learning that might prioritize data near the classifier line to learn. Through simulations on synthetic data, the research identified a technical bias where inculpatory evidence tends to cluster, while exculpatory evidence (such as alibis or impeachment material) is often more dispersed. Because CAL surfaces high-confidence documents first, there is a high probability that a prosecutor will find sufficient evidence of guilt and cease review prematurely before the algorithm surfaces the more "distant" evidence of innocence.

To mitigate the risk of suppressed exculpatory evidence, Wexler proposed three primary interventions for "technology-assisted review" (TAR) in criminal cases:

—**Two-Pass Search Strategy:** Prosecutors should be required to run separate, independent searches for inculpatory and exculpatory materials rather than training a single classifier on both.

—**Multi-calibration:** Implementing multi-calibration ensures the algorithm performs reliably across all data subsets, preventing the neglect of less-obvious exculpatory clusters.

—**Minimization Protocols:** In cases where Fourth Amendment constraints limit searches, protocols must explicitly allow for the "tagging" of Brady material during the document-sorting phase.

In conclusion, the research calls for courts to clarify the strict liability standard for evidence that the government possesses digitally but has not reviewed. The speaker concludes that a failure to resolve these doctrinal ambiguities leaves a gray area that may incentivize poor data hygiene and result in the continued failure to disclose evidence of innocence.

7. ALGORITHMIC COLLUSION WITHOUT THREATS (JUBA ZIANI)

Juba Ziani is an Assistant Professor in the H. Milton Stewart School of Industrial and Systems Engineering and an Adjunct Professor in the School of Computer Science. His research lies at the intersection of Computer Science, Operations Research, and Economics. His paper *Algorithmic Collusion without Threats* is joint work with Eshwar Ram, Natalie Collina, Sampath Kannan, and Aaron Roth.

Juba's talk began with a discussion of the differences between *classic price collusion*, in which collusion arises via human interaction (e.g., explicit agreement), and *algorithmic collusion*, in which collusion is an emergent behavior of algorithms in equilibrium. One mechanism by which algorithmic collusion can arise is if the algorithms involved use *threats* (e.g., grim trigger strategies). The main question their paper asks is: *Are threats the only mechanism by which (individually rational) sellers can engage in autonomous algorithmic collusion?*

They show, contrary to standard economic intuition, that supra-competitive prices can arise in equilibrium, even when both players are using algorithms which do not encode threats. An example such equilbrium involves one seller using a no-swap-regret algorithm, and a second seller setting a fixed price, equal to the Stackelberg leader strategy in the one-shot game. This example is particularly striking because the no-swap-regret property has been proposed as a "definitionally non-threatening" property by which one might audit pricing algorithms for antitrust compliance [Chassang and Ortner 2023; Hartline et al. 2024b]. Juba's talk closed with a discussion highlighting how this work calls for a more nuanced understanding of algorithmic collusion.

8. FIRESIDE CHAT: JASON HARTLINE AND NICOLE IMMORLICA

The 2026 Winter Meeting closed with a fireside chat, in which Nicole Immorlica interviewed Jason Hartline about his perspective on CS&Law.

Nicole Immorlica: What do you see as the main opportunity with CS&Law?

Jason Hartline: I tend to think of computer science and law as having three main touch points: CS technology might advance the law; as CS tools impact society, law needs to keep up; and viewing the legal system itself as a computational system. As computer scientists trying to understand the outcomes of computational systems, we should apply our tools to the legal system to try to understand its outcomes. I see a lot of parallels between this and the early days of EconCS.

Nicole Immorlica: What do you see as the challenges to collaborating with legal scholars, and how can we approach that from the CS or EconCS side?

Jason Hartline: I've been involved in the EC community since its very beginning. At the beginning, economists didn't come. Then there was a switch point around 2010, where we started to get more and more of them trickling over. Now a large fraction of papers in EC are authored by economists. It's been wonderful to

have them on board. One of the things I love about the CS&Law community right now is it feels like EC did back in 2000, where every time you go, there's someone doing a brand new connection between computer science and law that you've never heard of before. There's a lot of creativity happening. Legal scholars are hungry to bring computer science in. They've been doing law and tech stuff for decades, and we haven't been paying attention. So compared to CS and economics, you almost switch roles for CS and law.

Nicole Immorlica: I think one of the main frictions between CS and law is that legal scholarship and CS scholarship look very, very different. Computer scientists tend to have actually done something technical with every new paper we write. That's not the case with legal scholarship—they typically have an interpretation or they're combining ideas in a way that hasn't previously been discussed. Thinking about CS and economics, we might disagree about what approximation means or what efficiency means, but when we write down the math, we then agree. The math says what the math says. In a lot of the legal work, they haven't written down the math yet, so we can't even know if we agree. This insistence on understanding what the definitions are that allow computer scientists to make progress on legal problems is super important. But my impression of law is it's a lot of interpretation, and that's almost by design, because the flexibility in interpretation enables law to modernize. I wonder if the CS framework is going to perhaps box it in too much. Do you have thoughts on that?

Jason Hartline: It's interesting that you would say box it in. We have a law of copyright, and then LLMs come along, and when they invented copyright, they weren't thinking about LLMs. You can ask two kinds of questions: what does the existing law of copyright say about LLMs? And also, if we wanted to have a more LLM-savvy law of copyright, what should it say about LLMs? I think these are both important questions. But I think being future-proof is a strength of computer science. One of my favorite examples is when the designers of the internet came up with TCP/IP, they had no way of anticipating the things that we'd be doing with it today. And somehow those protocols that were invented in the '70s are still working and quite effective. This insistence of designing things that work out of the context that we were designing them for is a strength of computer science and algorithms.

Nicole Immorlica: What are some potential big wins that we could be proud of as an interdisciplinary field?

Jason Hartline: One thing that I've been quite impressed with about talking with legal scholars is that they seem to be much closer to practical impact than we tend to be in computer science. They frequently get put in front of panels of legislatures or judges, and they have to give opinions and write opinions. So the connections between legal scholarship and practice are actually often quite impactful.

I think one possible big win is having algorithmic sanity prevail in legal struggles relating to algorithms. For example, thinking about algorithmic collusion: the standard definition of competitive is Nash equilibrium, and we know by the PPAD completeness result that Nash equilibrium is not something algorithms can find quickly. So if we had a law saying algorithms playing against each other had

to reach Nash equilibrium, otherwise they were colluding, that would be an algorithmically insane law. The big wins are making sure that the regulations are algorithmically sane. A compelling example is a paper by Aloni Cohen and Kobbi Nissim on understanding singling out in the GDPR. The GDPR says that if you can't be singled out, then it's not private data. Their study showed that if you just literally try to interpret what the law is saying, that definition made no sense because obvious things might result in singling out in ways that are statistically uninteresting. They changed the definition to something they call *predicate singling out*, which means: can you single out someone more so than you could by random chance? Then they were able to argue that things like differential privacy protect from being predicate singled out, whereas other commonly recommended practices, like k-anonymity, do not. That kind of work—clarifying what definitions should be—is very impactful.

Nicole Immorlica: Can you pick out another greatest hit?

Jason Hartline: Let's talk about zero knowledge for criminal evidence. This is a paper with Rebecca Wexler and her co-authors. Rebecca Wexler is an expert on privilege—the idea that the law can't make you do certain kinds of things. Some privileges are at odds with getting the truth out in court proceedings. And so the law has to figure out how to get the truth out while respecting privileges. In computer science, we have novel ideas about what a proof is—interactive proofs, zero knowledge proofs. Zero knowledge proofs are proofs where the verifier is able to be certain that the prover can establish something, but doesn't learn anything extra about other stuff that they shouldn't be allowed to learn. So there's a possibility that we could have the best of both worlds: defendants in cases can interrogate the technology that's implicating them, whereas the technology owners and operators don't have to divulge their intellectual property, which they have privilege over. The law is often about getting a proof of something, and theoretical computer science has very sophisticated ideas about what a proof is. I think this is a huge opportunity for work.

Nicole Immorlica: I really appreciate that this paper shifts the conversation from trying to decide what objects must be disclosed, to what must we be able to prove about these objects.

Jason Hartline: And Rebecca Wexler made the point that this paper proves to judges that it should not be enough for someone to say "we have intellectual property, so we don't want to disclose this." Because it is possible to have zero-knowledge proofs, we should demand more from the technologies that people want to use in courts.

This is very reminiscent of the kind of meta-reasoning we like in computer science: this problem is NP-complete, so I don't know that no one can solve it, but all these people who tried to solve other NP-complete problems also can't solve my problem, so we think it's intractable. This way of doing reasoning between computer scientists and legal scholars I think is really quite compelling.

Nicole Immorlica: One thing that might be missing from the CS&Law intersection is the EconCS&Law intersection. Do you think is this CS and law interdisciplinary space needs more injection of other different fields?

Jason Hartline: The SIGecom community has brought computer science and

economics much closer together, and that makes the field of law and economics very accessible to us—we can just readily get involved. We have the same interest in putting formality and arguing about incentives in the law, so I don't see any reason to be separate. Already, economic papers on antitrust and algorithmic collusion are showing up in the law and economics literature.

Nicole Immorlica: What would you recommend we all read if we're wanting to dive into this a little bit more?

Jason Hartline: I would say come next week to the ACM CS&Law Conference in Berkeley.

9. CONCLUSION

This conference concluded with remarks from Inbal Talgam-Cohen, who highlighted the contributions of the numerous participants. Perhaps the most important takeaway from this year's research meeting is that the burgeoning field of CS-law is rich ground for the computer science and economics thinking that the field of AGT offers. And, not only is this field amenable to the AGT toolkit, but, with the rising influence of AI in our society, perhaps, also has a lot to gain from it.

REFERENCES

Bar-Gill, O., Sunstein, C. R., and Talgam-Cohen, I. 2023. Algorithmic harm in consumer markets. *Journal of Legal Analysis 15,* 1, 1–47.

Bitan, D., Canetti, R., Goldwasser, S., and Wexler, R. 2022. Using zero-knowledge to reconcile law enforcement secrecy and fair trial rights in criminal cases. In *Proceedings of the 2022 Symposium on Computer Science and Law*. 9–22.

Brynjolfsson, E. and Hitzig, Z. 2025. *AI's Use of Knowledge in Society*. University of Chicago Press.

Chan, P. K., Carrel, A., Varia, M., and Wang, X. 2025. Murmurs of the silenced: Secure reporting of misconduct settlements. In *Proceedings of the 2025 Symposium on Computer Science and Law*. 121–135.

Chassang, S. and Ortner, J. 2023. Regulating Collusion. *Annual Review of Economics 15,* Volume 15, 2023 (Sept.), 177–204. Publisher: Annual Reviews.

Cohen, A. and Nissim, K. 2020. Towards formalizing the gdpr's notion of singling out. *Proceedings of the National Academy of Sciences 117,* 15, 8344–8352.

Cohen, A., Scheffler, S., and Varia, M. 2022. Can the government compel decryption? don't trust-verify. In *Proceedings of the 2022 Symposium on Computer Science and Law*. 23–36.

Collina, N., Globus-Harris, I., Goel, S., Gupta, V., Roth, A., and Shi, M. 2026. Collaborative prediction: Tractable information aggregation via agreement. In *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 4712–4798.

Dutz, M., Shao, H., Blum, A., and Cohen, A. 2025. A machine learning theory perspective on strategic litigation. *arXiv preprint arXiv:2506.03411*.

Feigenbaum, J. and Weitzner, D. J. 2026. At the intersection of computer science and law.

Fish, S., Gonczarowski, Y. A., and Shorrer, R. I. 2024. Algorithmic collusion by large language models. *arXiv preprint arXiv:2404.00806 7,* 2, 5.

Hadfield-Menell, D. and Hadfield, G. K. 2019. Incomplete contracting and ai alignment. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. 417–422.

Hartline, J. D., Linna Jr, D. W., Shan, L., and Tang, A. 2022. Algorithmic learning foundations for common law. In *Proceedings of the 2022 Symposium on Computer Science and Law*. 109–117.

Hartline, J. D., Long, S., and Zhang, C. 2024a. Regulation of algorithmic collusion. In *Proceedings of the 2024 Symposium on Computer Science and Law*. 98–108.

Hartline, J. D., Long, S., and Zhang, C. 2024b. Regulation of Algorithmic Collusion. In *Proceedings of the 2024 Symposium on Computer Science and Law*. CSLAW '24. Association for Computing Machinery, New York, NY, USA, 98–108.

Kleinberg, J. and Mullainathan, S. 2024. Language generation in the limit. *Advances in Neural Information Processing Systems 37*, 66058–66079.

Lee, K., Cooper, A. F., and Grimmelmann, J. 2024. Talkin"bout ai generation: copyright and the generative-ai supply chain (the short version). In *Proceedings of the 2024 Symposium on Computer Science and Law*. 48–63.

Meding, K. and Sorge, C. 2025. What constitutes a deep fake? the blurry line between legitimate processing and manipulation under the eu ai act. In *Proceedings of the 2025 Symposium on Computer Science and Law*. 152–159.

Saig, E., Einav, O., and Talgam-Cohen, I. 2024. Incentivizing quality text generation via statistical contracts. *Advances in Neural Information Processing Systems 37*, 51196–51222.

Singh, A., Rodemann, J., Verma, R., Chau, S. L., and Muandet, K. 2026. Incentive aware ai regulations: A credal characterisation. *arXiv preprint arXiv:2603.05175*.

How Should Marketplaces Decide Who to Show?

PENG SHI
USC Marshall School of Business

Across three papers, I study how digital marketplaces should decide which providers to recommend when providers set their own prices. The central lesson is that prices let providers manage demand and capacity, so the platform often need not ration attention or track capacities directly. This yields three simple prescriptions. When expected customer surplus is measured well, ranking by that score maximizes social welfare, defined as the sum of customer surplus, provider surplus, and platform revenue. When quality information is unreliable, sell customer attention at a market-clearing fee, or equivalently rank ads by Bid $\cdot$ CTR. When both signals are available, add them: Bid $\cdot$ CTR $+ \mu \cdot \widehat{\text{Quality}}$, interpretable as a quality discount on paid attention. The weight $\mu \in [0, 1]$ reflects how much the platform values customer surplus relative to provider and platform surplus. This letter explains the economic logic and connects it to practice on platforms such as Amazon, Alibaba, Google, Yelp, and HomeAdvisor.

Categories and Subject Descriptors: K.4.4 [**Computers and Society**]: Electronic Commerce

General Terms: Design, Economics, Reliability, Theory

Additional Key Words and Phrases: Market design, matching, ranking, sponsored advertising, two-sided marketplace

1. WHY THE QUESTION MATTERS

Digital platforms now mediate a large and growing share of how customers find goods, services, stays, and freelance professionals: products on Amazon and Alibaba; local services such as plumbers, auto shops, and restaurants on Yelp, Google Maps, and HomeAdvisor; stays on Airbnb; and freelance work on Upwork. The economic footprint is large: Digital Commerce 360 estimates that the top 100 global online retail marketplaces reached roughly $3.8 trillion in gross merchandise value in 2024, having roughly doubled in six years [Digital Commerce 360 2024]. The true scope is even larger, since this aggregate omits the many service, labor, and local-search platforms consumers turn to throughout daily life. Yelp alone reports tens of millions of monthly app users and hundreds of thousands of paying advertisers [Yelp 2026].

I focus on platforms that *recommend* options for the customer to compare and choose from, rather than platforms that assign a single provider (e.g., ride-hail dispatch). Across these platforms, a key value proposition is to elevate providers likely to suit the customer. The interface design varies widely across platforms: Yelp and Google search return ranked lists; Amazon arranges products in a grid of varied prominence, with sponsored cards, brand banners, and video creatives interleaved among conventional listings; Google Maps shows sponsored pins more prominently than organic ones; and lead-generation platforms often forward each customer request to a small assortment of providers. On all of these interfaces, a shared operational decision is: among the providers who could serve a customer, which ones should the platform put in front of that customer?

Author's address: Peng Shi, USC Marshall School of Business, pengshi@usc.edu.

The natural impulse is to show the "best" providers first. But what does "best" mean? Suppose a homeowner searches for a kitchen remodel contractor. Should the platform always promote the contractor with the strongest reviews? Reviews may not capture whether that contractor is a good fit for the homeowner's specific needs. But even if they did, a contractor shown first to everyone may raise prices, schedule far into the future, cherry-pick jobs, or take on more jobs than they can personally supervise. These responses may not be what homeowners want when searching for a hands-on contractor. The platform therefore needs some way to distribute recommendations across providers based on capacity. But providers may have incentives to overstate capacity, since visibility can create pricing power. At the same time, providers also want to avoid poorly matched leads because answering calls and giving quotes are time-consuming.

A natural academic framing is two-sided assortment optimization [Ashlagi et al. 2022; Aouad and Saban 2023; Housni et al. 2026; Rios and Torrico 2026]. This literature was motivated in part by dating markets, where each participant has limited attention and the platform rations how often each profile is shown. The optimization problem is computationally hard, so the literature develops sophisticated submodular optimization algorithms with constant-factor approximation guarantees. These algorithms look quite different from the familiar ranking and ad-allocation systems used in commercial marketplaces. A more fundamental gap is that this literature takes prices, and hence provider attractiveness, as exogenous. The papers discussed here instead study commercial marketplaces where providers set their own prices in response to platform policy, and may also choose how much attention to seek or accept.

The three papers discussed here address this gap. They build a tractable theory of match recommendations in marketplaces where providers choose prices in response to platform policy. The papers are [Shi 2024] (EC'22), [Shi 2026a] (EC'24), and [Shi 2026b] (EC'25, Exemplary Applied Modeling Track Paper Award). The models differ, but share a common idea: when providers set their own prices, they use price to balance demand against their own capacity, so the platform need not artificially ration attention to scarce providers, and simple recommendation and ad-pricing rules suffice. The remainder of the letter develops three prescriptions: rank by expected customer surplus when quality is measured well; use willingness to pay when it is not; and combine the two through quality-adjusted ad pricing. I then explain how the same logic applies across different platform interfaces and close with limitations and open directions.

2. RANKING BY EXPECTED CUSTOMER SURPLUS

What is the ideal metric by which a platform should rank providers? For customer type t and provider j, define

$$\begin{aligned}\text{Quality}_{tj} &:= \mathbb{E}[\text{customer surplus} \mid t, \text{showing } j] \\ &= \text{CTR}_{tj} \cdot \text{CVR}_{tj} \cdot \mathbb{E}[\text{customer surplus} \mid \text{transaction with } j].\end{aligned}$$

In words, Quality_{tj} is the expected customer surplus from directing one unit of attention from a type-t customer to provider j. (Section 5 makes "unit of attention" precise; for now, think of it as one listing impression.) The customer type may

represent a search query, geographic segment, or richer user context. Customer surplus is measured in dollars: it is the value the customer keeps after paying the provider, relative to an outside option such as not purchasing, choosing another provider, or leaving for another platform. On a platform where the customer clicks through to inspect a listing, the score decomposes into click-through rate (CTR), conversion rate conditional on click (CVR), and expected surplus from the resulting transaction. On a platform without a click step, the CTR term can be set to one.

This quality score is broader than ratings or service quality. It depends on price, as surplus is evaluated at the prices customers actually face. A higher provider price may reduce clicks or conversions, and it lowers the surplus captured by customers who still transact. A provider can therefore improve Quality_{tj} by providing a better product or service, or by setting a price that leaves customers more surplus.

When Quality_{tj} is measured exactly, ranking by this score achieves first-best social welfare in all three papers. Here, "social welfare" means the total value generated by the platform, including customer surplus, provider surplus, and platform revenue. "First-best" means the best outcome a perfectly informed central planner could achieve when given the power to choose both recommendations and provider prices. The simple policy of ranking by Quality_{tj} attains the first-best benchmark because the quality-score is evaluated at equilibrium prices and availability. Providers' own decisions internalize their capacity, even when the platform cannot observe it. As providers maximize profit, those with scarce capacity may raise prices, decline jobs, or cap incoming attention — as a busy contractor does by pausing new leads. Raising prices or declining jobs lowers the conversion rate and hence the quality score, while pausing leads reduces attention directly; either way, the provider receives less attention. Therefore, ranking by quality automatically distributes attention based on provider capacity, mimicking the first-best.

In practice, platforms rarely observe Quality_{tj} exactly. CTR and CVR can often be predicted from logs, but dollar-equivalent surplus conditional on transaction is harder to measure. Ratings, reviews, returns, complaints, and post-transaction surveys are noisy, sparse, delayed, and sometimes manipulated. Offline service platforms may not even observe whether a transaction occurred. Shi [2026b] addresses this by using a conservative lower estimate, denoted $\widehat{\text{Quality}}_{tj}$. The true Quality_{tj} is assumed to lie in the range

$$[\widehat{\text{Quality}}_{tj},\ \widehat{\text{Quality}}_{tj}/\rho],$$

where $\rho \in (0, 1]$ summarizes the reliability of the quality estimate. The paper shows that ranking by $\widehat{\text{Quality}}_{tj}$ guarantees at least a ρ fraction of first-best social welfare, and the bound is tight in the worst case. The platform need not know ρ to implement the ranking; ρ governs the welfare guarantee, not the algorithm.

The conservatism matters. If a platform ranks by optimistic estimates, it may reward providers whose reviews are sparse, manipulated, or merely lucky. A conservative score gives credit only to what the platform can support with evidence. This resembles a familiar way to interpret ratings: a provider with 4.1 stars and 1000 reviews is typically ranked above one with 4.8 stars and 5 reviews, because the former signal is more reliable. Better measurement and validation correspond to a higher reliability parameter ρ and a stronger welfare guarantee.

3. RANKING BY WILLINGNESS TO PAY

Some platforms may not have reliable quality estimates. In home services, lead-generation platforms such as Angi, HomeAdvisor, Bark, Google Local Services, Modernize, Porch, and Thumbtack often do not directly observe whether transactions occur, since contractors visit customers' homes, give quotes, and close offline. Reviews may also be compressed near the top of the scale and sparse; Raval [2024] documents that 96% of HomeAdvisor businesses have average ratings above 4 stars. Related measurement problems can arise in other markets with infrequent or hard-to-observe outcomes, or where customers expect to have future face-to-face interactions with the provider and are therefore reluctant to leave bad reviews.

A common response in lead and advertising systems is to charge providers for the opportunity to be shown, using willingness to pay as a signal of who can profitably serve the customer. HomeAdvisor connects each customer with up to a small number of providers and charges category-specific lead fees: a small repair job may cost a few dollars per lead, whereas a large remodel may cost hundreds. In the model, the fee is a market-clearing price f_t for one unit of attention from customer type t. The platform can implement this without detailed information on market parameters, simply by adjusting fees over time — raising the lead fee where leads are oversubscribed, lowering it where they go unsold.

Shi [2026a] analyzes this policy and shows that it achieves a constant fraction of first-best social welfare in the worst case. The policy is quality-agnostic: providers self-select by buying leads only when they expect to convert the customer and earn sufficient margin. Thus, when the reliability ρ of quality estimates is low, willingness to pay may signal provider relevance better than a noisy organic ranker.

The specific constant in the guarantee depends on demand assumptions: under standard regularity conditions on demand, Shi [2026a] obtains $1/(e-1) \approx 58\%$ assuming additively separable customer utilities with a vertical component and a horizontal component whose distribution is symmetric across providers; Shi [2026b] obtains $1/e \approx 37\%$ allowing arbitrarily different demand distributions across providers. Both are tight worst-case guarantees: they hold across arbitrarily many providers, heterogeneous capacities, and multiple customer types.

These partial guarantees reflect a real limitation: willingness to pay is an imperfect welfare signal. A high bid may reflect high margins, strong sales ability, or the ability to extract surplus from customers, not just customer value. Bid ranking is therefore a robust fallback when quality measurement is weak, not a substitute for quality information when it is available.

The same market-clearing logic appears in sponsored advertising. In a reduced-form view of an ad system, ads are ranked by an Ad-Rank score and admitted when the score clears a threshold θ_t, determined by the supply and demand for attention of type t. The threshold θ_t represents the marginal price of customer attention. The cost per click is the minimum bid needed for the Ad-Rank score to clear θ_t. A common baseline formula is

$$\text{AdRank} = \text{Bid} \cdot \text{CTR},$$

which was behind Google's early sponsored-search auctions [Edelman et al. 2007; Varian 2007], and remains a recognizable building block of many advertising sys-

tems. Multiplying the per-click bid by the click-through rate converts it into per-impression willingness to pay. Under this policy, the effective per-impression fee is $f_t = \theta_t$. When the platform sets the Ad-Rank threshold θ_t to be market-clearing, which approximates an auction system with no reserve price, the above Ad-Rank formula achieves the same constant-fraction welfare guarantees as described above.

This view of ads as welfare-enhancing connects to, but is distinct from, the literature on advertising as a signal [Kihlstrom and Riordan 1984; Milgrom and Roberts 1986]. In classic signaling stories, advertising may change what customers infer about product quality. Here, the ad is not assumed to persuade customers; rather, the platform uses providers' willingness to pay to select the most motivated and available providers to show.

The mechanism also depends on customers making informed conversion decisions. In practice, that requires accountability systems such as fraud detection, complaint handling, and removal of bad actors. With those safeguards, having providers pay for customer attention can be a welfare-enhancing discovery mechanism rather than merely a distortion.

4. COMBINING QUALITY AND WILLINGNESS TO PAY

When the platform has both conservative quality estimates and providers' willingness to pay, neither pure organic ranking nor pure bid ranking is best. Shi [2026b] studies a robust-optimization problem: the platform knows only that true quality lies in $[\widehat{\text{Quality}}_{tj},\ \widehat{\text{Quality}}_{tj}/\rho]$, and policies are evaluated by their worst-case fraction of first-best weighted welfare. Here weighted welfare is (supply-side surplus) + $\mu \cdot$ (customer surplus). Supply-side surplus is provider profit before platform fees, equivalently provider surplus plus platform revenue. The parameter $\mu \in [0, 1]$ represents how much the platform values the customer-side relative to the supply-side.[1] Strikingly, across all policies monotone in bid and estimated quality — including nonlinear ones — the best worst-case guarantee is achieved by a simple linear sum whose quality weight is exactly the μ from the weighted objective:

$$\text{AdRank}_{tj} = \text{Bid}_{tj} \cdot \text{CTR}_{tj} + \mu \cdot \widehat{\text{Quality}}_{tj}. \tag{1}$$

Both terms are in dollars per unit of customer attention: the first is per-impression willingness to pay, and the second is a conservative estimate of customer surplus per impression, weighted by μ. Noisier quality evidence lowers the conservative estimate, so bids matter more; more reliable evidence raises the quality term.

The parameter μ is a policy choice. Raising μ moves the platform toward customer surplus; lowering it moves toward supply-side surplus. Setting $\mu = 0$ recovers the policy in the previous section, which in the model attains the first-best supply-side surplus exactly. Extensive simulations further show that the platform can optimize the breakdown of supply-side surplus into provider surplus and platform revenue by utilizing additional transaction fees or commissions. Under optimal choice of such fees, $\mu = 0$ generally achieves the highest platform revenue among the policies considered. Larger μ trades some platform revenue for customer and

[1] For $\mu > 1$, the weighted objective overcounts customer surplus, so a central planner could increase it without bound by lowering prices to transfer surplus from providers to customers, implying that the first-best benchmark is infinite and therefore uninteresting. [Shi 2026b] hence restricts $\mu \leq 1$.

provider surplus, making it more attractive when the platform is willing to forgo short-term revenue for long-term customer value.

The formula also has a price interpretation. A provider's AdRank_{tj} clears the threshold θ_t when their per-impression willingness to pay $\text{Bid}_{tj} \cdot \text{CTR}_{tj}$ exceeds

$$f_{tj} = \theta_t - \mu \cdot \widehat{\text{Quality}}_{tj},$$

the equilibrium per-impression fee. Higher estimated quality therefore acts as a discount on customer attention, which in practice may be delivered through lower advertising costs, reduced commissions, rebates, or credits. Such discounts also give providers a reason to improve the quality estimate in welfare-relevant ways, including by setting prices that leave customers more surplus.

This prescription is concrete: it tells platforms what to estimate (a conservative dollar-valued customer surplus per impression) and how to incorporate it into ad prices (as a linear discount, weighted by μ). Quality-based discounts on ad fees already appear in industry practice: Alibaba's published cost-per-click formula (as of January 2025) discounted advertising costs by a promotion-quality score [Alibaba 2025], and on Amazon, marketing-agency analysis reported that products with stronger organic rank paid much less per click for comparable sponsored positions [Signalytics 2023]. What the theory adds is a welfare rationale for using such discounts and a concrete formula for platforms to calibrate and test.

5. THE SAME LOGIC ACROSS INTERFACES

The framework operates at the level of allocating customer attention. In the formulas above, an "impression" is a unit of customer attention rather than a raw page event, and different placements may correspond to different amounts: a large slot at the top might count as 3.2 impressions, a small one near the bottom as 0.5. This view aligns with a familiar industry decomposition: Varian [2006] writes the raw click-through rate as a multiplicatively separable model with a position-specific factor (how much attention the slot delivers) and an ad-specific factor (how appealing the listing is, given that attention). The CTR_{tj} in the Ad-Rank formula corresponds to the ad-specific factor alone; the position factor is absorbed into the impression count and can be estimated from data.

This abstraction makes the framework implementation-agnostic across both format and billing convention. The Ad-Rank formula in (1) requires only per-impression willingness to pay as its first term. Different formats — ranked lists, map pins, product grids, banners, and video ads — enter through different impression counts, leaving the per-impression economics unchanged. Different billing conventions differ only in the conversion factor that translates a per-unit bid into per-impression willingness to pay: CTR_{tj} does this for per-click billing; under pay-per-transaction, replace it with the rate at which a unit of attention converts to a transaction.

A more speculative extension is LLM-driven search. As AI assistants mediate how people find products and services, "attention" may include the prominence of a mention, the wording used, and whether the assistant offers a direct "book now" action. The same prescription suggests charging per unit of attention, with a discount for conservatively estimated customer surplus. When responses include clickable links, the cost-per-click form in (1) applies directly, with CTR_{tj} capturing

appeal conditional on attention. The existing results do not model conversational persuasion or trust formation; those are open directions.

6. LIMITATIONS AND NEXT STEPS

To derive clean results, the theory fixes the total amount of customer attention and studies how that attention should be allocated. An assortment extension in Shi [2026b] allows providers shown together to affect one another's demand and equilibrium prices, and the above results extend to that setting. This still leaves room for richer models of customer search in which platform policies affect whether customers use the platform in the first place and how much attention they are willing to devote. It would be interesting to analyze the proposed policies in more complex models of endogenous customer search and platform competition.

The theory is also stationary and abstracts from fluctuations in market conditions. Moreover, it characterizes long-run equilibrium outcomes, but not the transient path by which providers learn to bid and platforms adjust estimates. Nevertheless, the underlying economic logic can be incorporated into dynamic implementations, by adding a conservative quality term in the Ad-Rank formula and letting existing auction systems determine allocations. When estimating quality, more weight should be put on recent data so estimates do not become stale. Formally studying learning, cold start, and rate of convergence would further bridge the gap between theory and practice, and is a direction for future work.

The framework also uses a stylized treatment of capacity. Shi [2026b] extends the results to richer settings, including allowing providers to choose their own capacity and accommodating platform-imposed allocation constraints such as diversity requirements. Even so, capacity binds only in expectation: providers are assumed to have enough scheduling flexibility to serve demand as long as the overall rate fits within their capacity. This approximation is less applicable when matches are time-sensitive and commit a large share of a provider's capacity for an extended period — for example, finding a long-term nanny.

The most important implementation challenge is measuring Quality_{tj}. The appendix of Shi [2026b] gives two starting points. A survey-based approach asks customers whether they transacted, with whom, and how much more they would have paid; the responses train a predictive model of expected surplus per impression. A revealed-preference approach uses logged impressions, prices, and transactions to estimate demand and dollar-valued customer surplus, drawing on randomized price variation or valid instruments such as cost shifters. In both cases, the appendix applies conservative shrinkage before the estimate enters the Ad-Rank formula.

Production systems add another layer. Real ranking stacks combine candidate generation, relevance filters, safety rules, pacing, reserves, and reranking. None of this complexity is fatal to the theory's applicability: the formula serves as a benchmark for allocating and pricing attention among eligible providers, after the platform has applied its other operational constraints. Translating that benchmark into a tested production rule remains substantial work. Because the theory describes long-run equilibrium, short A/B tests are unlikely to be informative; field tests should run long enough, and announce policy changes clearly enough, for providers to respond by adjusting prices, ad spend, and capacity. Autobidders and

pricing agents ease this by pushing provider behavior toward the profit optimization the theory assumes; platforms could go further by offering decision-support tools that jointly optimize pricing and ad spend.

These limitations point to a rich agenda spanning empirical estimation, theoretical extensions, and production design. The SIGecom community — with its blend of economic modeling, algorithm design, market design, and empirical platform research — is well placed to advance it, and I welcome collaboration in closing the gap between theory and practice.

REFERENCES

Alibaba. 2025. How will keyword advertising charge me? Available at https://so.alibaba.com/s/ggs/category?categoryId=1000045920&questionId=1000100938. Accessed January 2025.

Aouad, A. and Saban, D. 2023. Online assortment optimization for two-sided matching platforms. *Management Science 69,* 4, 2069–2087.

Ashlagi, I., Krishnaswamy, A. K., Makhijani, R., Saban, D., and Shiragur, K. 2022. Assortment planning for two-sided sequential matching markets. *Operations Research 70,* 5, 2784–2803.

Digital Commerce 360. 2024. Top global online marketplaces: Key data and statistics. Available at https://www.digitalcommerce360.com/top-online-marketplaces-data-stats/. Accessed May 2026.

Edelman, B., Ostrovsky, M., and Schwarz, M. 2007. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review 97,* 1, 242–259.

Housni, O. E., Hennebelle, U., and Torrico, A. 2026. Two-sided assortment optimization: Adaptivity gaps and approximation algorithms.

Kihlstrom, R. E. and Riordan, M. H. 1984. Advertising as a signal. *Journal of Political Economy 92,* 3, 427–450.

Milgrom, P. and Roberts, J. 1986. Price and advertising signals of product quality. *Journal of Political Economy 94,* 4, 796–821.

Raval, D. 2024. Do bad businesses get good reviews? Evidence across several online review platforms. *Working paper*. Available at https://deveshraval.github.io/reviews.pdf.

Rios, I. and Torrico, A. 2026. The dating heuristic: A provably strong matching algorithm for dating platforms. *Manufacturing & Service Operations Management*.

Shi, P. 2024. Optimal match recommendations in two-sided marketplaces with endogenous prices. *Management Science 71,* 9, 7431–7448.

Shi, P. 2026a. The welfare effects of selling leads in a two-sided marketplace. *Management Science, Forthcoming.* Preprint available at https://ssrn.com/abstract=4727198.

Shi, P. 2026b. Welfare-optimal policies for sponsored advertising in a two-sided marketplace. Working paper, updated May 2026. Available at https://ssrn.com/abstract=5132218.

Signalytics. 2023. What you didn't know about Amazon's "second-price" auction. Available at https://www.linkedin.com/pulse/what-you-didnt-know-amazons-second-price-auction-signalytics/. Accessed January 2025.

Varian, H. R. 2006. The economics of internet search. *Rivista di politica economica 96,* 11/12, 177–191.

Varian, H. R. 2007. Position auctions. *International Journal of Industrial Organization 25,* 6, 1163–1178.

Yelp. 2026. Investor relations overview. Reports 28 million monthly average app unique devices and 496,000 paying advertising locations for 2025. Available at https://www.yelp-ir.com/overview/default.aspx.

An Introduction to Contract Theory for Computer Scientists

PAUL DÜTTING
Google Research, Switzerland
and
MICHAL FELDMAN
Tel Aviv University, Israel
and
INBAL TALGAM-COHEN
Tel Aviv University, Israel

This survey introduces contract design to computer scientists: using outcome-based payments to incentivize effort. We discuss applications that motivate a computational approach to contract design, ranging from online labor markets, to incentive-aware algorithmic classification, to emerging ecosystems of AI agents. We present the classic principal-agent model and introduce basic concepts. We show how optimal contracts are computable via linear programming, and why they can be opaque and brittle. Motivated by this we turn to simpler contract formats, specifically linear contracts. We discuss approximation guarantees, and robust (max–min) optimality of such contracts. We then move beyond the single-agent setting to combinatorial contract settings with multiple agents, mapping the algorithmic landscape of tractability and approximation. Finally, we discuss partial information settings through the lens of learning algorithms, and outline open directions. The goal is a concise, accessible entry point that connects economic foundations with algorithmic tools and highlights opportunities for new theory and applications.

Categories and Subject Descriptors: F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems; J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Economics*

General Terms: Algorithms, Design, Economics, Theory

Additional Key Words and Phrases: Algorithmic Game Theory, Contract Design, Principal-Agent Model, Hidden Action

1. INTRODUCTION

The field of Algorithmic Game Theory (AGT) addresses the challenges and opportunities that arise when algorithms interact with real-life, strategic players. One of the founding papers of the field, by Nisan and Ronen (2001), was motivated by the rising prominence of the internet circa 2000 —"a distributed setting where the participants cannot be assumed to follow the algorithm but rather their own self-interest". Nisan and Ronen suggested a framework for studying algorithms in such settings, based on *mechanism design* from microeconomics. This led to a vast body of research on algorithmic mechanism design, with a focus on payment schemes that incentivize strategic players to *report truthfully* to algorithms.

However, in many instances, strategic reaction to algorithms is unrelated to truth-

Authors' addresses: duetting@google.com, mfeldman@tau.ac.il, inbaltalgam@gmail.com,

fulness and mechanism design. In this survey we introduce *algorithmic contract design*—a new frontier for AGT. Contract design is a renowned subfield of microeconomics. It is similar to mechanism design in its theme of incentivizing strategic agents, but the incentives are not to report truthfully, but rather to exert effort earnestly. A contract is essentially a performance-based payment scheme, which ties an agent's compensation to that agent's performance and created value.

A premier application of contracts is in online markets of services, which, similar to markets of goods in the first two decades of the internet, have recently seen a massive growth both in importance and scale—creating a unique opportunity for computational approaches.

Example 1 (Influencer marketing). *Consider a brand that wants to pay a social media influencer (say a famous TikToker) to post sponsored/branded content. How should the contract be structured? Should it be a flat payment? Or should the payment be tied to how successful the post is? Should it be tied to the number of views, or the number of conversions?*

A second, perhaps less obvious applications of contract theory, is as a framework for understanding how humans react to algorithms such as a classification algorithm (even absent payments). This connection plays a role, for example, in the design of massive open online courses (MOOCs).

Example 2 (Strategic reaction to a Classifier). *Consider a classification algorithm, classifying human subjects according to some features they exhibit. This could be a prediction algorithm that assigns credit scores, or an exam at a university course or MOOC. The subjects of classification may invest strategic effort into developing more favorable features. In other words, an algorithm cannot measure without changing what it measures. Contract design helps design classifiers that anticipate this strategic effect, and can help route the efforts to modify features into productive avenues.*

We are now on the cusp of a new revolution, where AI agents are becoming increasingly capable of autonomously carrying out more and more tasks. The classic, pre-set algorithms of Nisan and Ronen are being replaced with opaque GenAI models, operating autonomously on behalf of their users and interacting with other AI agents. This growing ecosystem of AI agents and LLM providers depends crucially on effective delegation and cooperation, as enabled by contract design.

Example 3 (Paying an AI agent). *Consider a marketing GenAI agent, creating content for a campaign on behalf of a brand. What is the ideal payment scheme? Current LLM monetization by providers like OpenAI or Anthropic is usually per-token or subscription-based. In comparison, a performance-based contract would enable the utilization of key performance indicators like profit or audience engagement, thus leading to aligned incentives and elevated performance.*

The examples demonstrate some of the fundamental computational questions that arise when considering contracts: Perhaps the simplest possible performance-based payment scheme is a fixed fraction of the profit, known as a *linear* contract; how far from optimal is such a scheme, relative to a more complex one that relies on a combination of multiple indicators? What if instead of a single agent,

there is a team of agents; how should they split the profit? What if certain aspects that are relevant to the design of contracts are not fully known, and have to be learned? These and other challenges require a combination of algorithmic and learning techniques with economic theory, and form the new research area of algorithmic contract design.

This survey aims to provide an accessible entry point to this new area for computer scientists. A more comprehensive survey by Dütting et al. (2024) covers additional topics within this new frontier and gives additional details on the topics surveyed here. A separate survey by Feldman (2025) covers in greater depth recent advances in multi-agent and multi-action combinatorial settings.

Organization. We introduce the canonical contract model in Section 2. In Section 3 we explore optimal contracts, and discuss simple, linear contracts in Section 4. Section 5 explores combinatorial, multi-agent contracts. In Section 6, we take a learning perspective. A discussion of what's ahead appears in Section 7.

2. MODEL

In this section, we introduce the hidden-action principal agent model (Holmström, 1979; Grossman and Hart, 1983). Our coverage follows (Dütting et al., 2019).

Setting. There is a principal (she) and an agent (he). The agent can take one of n actions. Each action $i \in \mathcal{A} = [n]$ has a cost $c_i \geq 0$. Actions are sorted by non-increasing cost. The cost of an action is borne by the agent. We model the fact that the agent can opt out of the contract by assuming that one of the actions has zero cost. Each action stochastically leads to one of m outcomes. We write q_{ij} for the probability of outcome j under action i, and use $\mathbf{q}_i$ to denote the probability distribution over outcomes induced by action i. The agent's choice of action is *hidden* from the principal, who can only observe the stochastic outcome of the action. This reflects that the principal is unable or unwilling to keep track of the agent's action. Each outcome j is associated with a reward $r_j \geq 0$, enjoyed by the principal. Rewards are sorted in non-increasing order. We use $R_i := \mathbb{E}_{j\sim\mathbf{q}_i}[r_j] = \sum_{j\in[m]} q_{ij} r_j$ to denote the expected reward of action i.

Example 4 (Running Example). *Consider a principal-agent setting with three actions $i \in [3]$ and three outcomes $j \in [3]$:*

reward:	$r_1 = 0$	$r_2 = 1$	$r_3 = 7$	*cost*
action 1*:*	$q_{11} = 1$	$q_{12} = 0$	$q_{13} = 0$	$c_1 = 0$
action 2*:*	$q_{21} = 0$	$q_{22} = 1/2$	$q_{23} = 1/2$	$c_2 = 1$
action 3*:*	$q_{31} = 0$	$q_{32} = 1/6$	$q_{33} = 5/6$	$c_3 = 2$

If the agent takes action 2*, then he bears a cost of* 1*. The outcome is $j = 2$ or $j = 3$ with equal probability. Note that if the principal observes outcome* 2*, she cannot be certain this is because the agent chose action* 2*, but action* 2 *is more likely than action* 3*.*

It is possible to impose additional structure on the distributions, like (first-order) stochastic dominance, which reflects that more rewarding outcomes become more

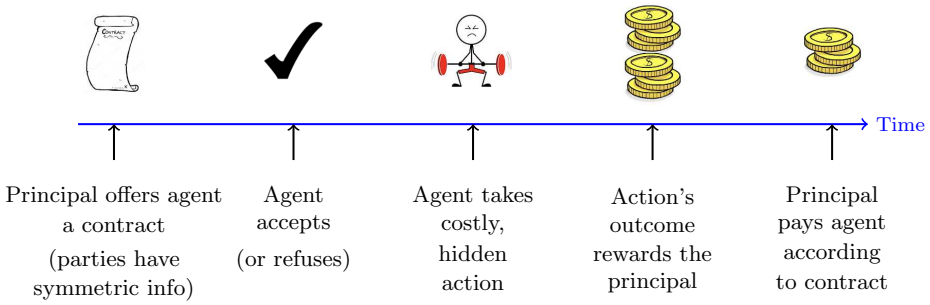


Fig. 1: Timeline.

likely as the agent exerts more costly effort. Other common assumptions include the stronger *Monotone Likelihood-Ratio Property* (MLRP), or the *Concavity of Distribution Function Property* (CDFP).

Contract. The principal enjoys the rewards while the agent bears the cost; these misaligned preferences create an incentive problem called *moral hazard*, which the principal seeks to address by defining a contract. A contract is a payment rule $\mathbf{t}$ that consists of m non-negative payments or *transfers* $(t_1, ..., t_m)$, one for each outcome.[1] The transfers are associated with outcomes rather than actions since the actions are hidden from the principal. For action $i \in [n]$ let

$$T_i := \mathbb{E}_{j \sim \mathbf{q}_i}[t_j] = \sum_{j \in [m]} q_{ij} t_j \tag{1}$$

denote the expected payment from principal to agent for taking action i.

Both the principal and the agent are assumed to be *risk neutral*. For a fixed contract $\mathbf{t}$ and choice of action i, the agent's expected utility and principal's expected utility (a.k.a. *revenue*) are given by

$$U_A(i \mid \mathbf{t}) := T_i - c_i \quad \text{and} \quad U_P(i \mid \mathbf{t}) := R_i - T_i.$$

Notice that the sum of the players' expected utilities is always equal to the expected welfare $W_i := R_i - c_i$ of the action i chosen by the agent. The contract thus influences the agent's choice of the welfare "pie" (through his choice of action), in addition to determining how this pie is divided between the principal and the agent.

Example 4, Continued. *Consider the contract* $\mathbf{t} = (0, 1, 3)$. *The expected payment for action* 1 *under this contract is* $T_1 = 0$, *for action* 2 *it is* $T_2 = 2$, *and for action* 3 *it is* $T_3 = 8/3$. *The agent's expected utility is thus maximized by action* 2, *which yields an expected utility of* $U_A(2 \mid \mathbf{t}) = 1$, *compared to an expected utility of* $U_A(1 \mid \mathbf{t}) = 0$ *for action* 1 *and an expected utility of* $U_A(3 \mid \mathbf{t}) = 2/3$ *for action* 3. *The principal's utility is then* $U_p(2 \mid \mathbf{t}) = 2$.

Incentives. The contract design problem defines a game between the principal and the agent, in which the principal moves first and defines a contract $\mathbf{t}$ and the agent

[1]The requirement that payments should be non-negative is referred to as *limited liability*. It reflects the asymmetric roles of the principal and the agent in contractual relations, and also rules out trivial but unrealistic solutions.

reacts with a utility-maximizing action $i^\star(\mathbf{t})$ (see Figure 1). Formally, for a fixed contract $\mathbf{t}$ let $\mathcal{A}^\star(\mathbf{t}) := \arg\max_{i\in[n]} U_A(i \mid \mathbf{t}) \subseteq [n]$ denote the set of actions that maximize the agent's expected utility. Using this notation, the agent chooses an action

$$i^\star \in \mathcal{A}^\star(\mathbf{t}) = \arg\max_{i\in[n]} U_A(i \mid \mathbf{t}). \tag{2}$$

Any such choice $i^\star$ is *incentive compatible* (IC) for the agent, because it is preferred over any other action. Note that it is also *individually rational* (IR) for the agent, meaning that the agent is guaranteed non-negative utility. This is because payments are non-negative, and we assumed that the agent has a zero-cost action.

Importantly, the principal's utility may be different for different actions $i^\star \in \mathcal{A}^\star(\mathbf{t})$. It is thus important to specify how the agent breaks ties. By default, and as is standard in the contracts literature, we assume that the agent breaks ties in favor of the principal.[2] We refer to this tie-breaking rule as the *canonical* tie-breaking rule.

3. OPTIMAL CONTRACTS

The principal's design problem is to select a contract $\mathbf{t}$ that maximizes her expected utility (i.e., revenue from the contract), given that the agent chooses an action $i^\star \in \mathcal{A}^\star(\mathbf{t})$ that maximizes his own expected utility. This contract is referred to as *optimal*. The next result is an algorithmic reformulation of a characterization of optimal contracts by (Grossman and Hart, 1983), showing that the optimal contract can be computed (in polyomial time) by solving n *linear programs* (LPs), one per action.

Proposition 1 (Grossman and Hart (1983)). *An optimal contract can be found by solving n linear programs, one per action.*

To outline the LP approach, we call an action $i \in [n]$ *implementable* (up to tie-breaking) if there exists a contract $\mathbf{t}$ such that $U_A(i \mid \mathbf{t}) \geq U_A(i' \mid \mathbf{t})$ for every $i' \neq i$. The approach is to construct an LP for each action $i \in [n]$ that determines whether the action is implementable and, if so, to find the minimum expected payment required to implement it. This LP and its dual are given in Figure 2. An optimal contract can then be found by applying this procedure to all actions, and choosing the action that maximizes the principal's utility. The corresponding pair of action and contract can be shown to be compatible with the canonical tie breaking rule.

An important corollary of the LP formulation is a characterization of the actions that the principal can implement.

Proposition 2 (Hermalin and Katz (1991)). *Action i is implementable if and only if there is no convex combination $\{\gamma_{i'}\}_{i'\neq i}$ of the actions other than i that results in the same distribution over outcomes, i.e., $\sum_{i'\neq i} \gamma_{i'} q_{i'j} = q_{i,j}$ for all outcomes j, with lower weighted cost, i.e., $\sum_{i'\neq i} \gamma_{i'} c_{i'} < c_i$.*

The characterization follows from LP duality: MINPAY-LP(i) is infeasible (equiv., action i cannot be implemented) precisely when its dual (which is always feasible)

[2]This tie-breaking rule is justified by the fact that a small perturbation would make the agent strictly prefer that action (see, e.g., Carroll (2015); Dütting et al. (2019) for additional discussion).

$$\begin{aligned} \min \quad & \sum_j q_{ij} t_j & & \\ \text{s.t.} \quad & \sum_j q_{ij} t_j - c_i \geq \sum_j q_{i'j} t_j - c_{i'} & \forall i' \neq i \\ & t_j \geq 0 & \forall j \end{aligned}$$

(a) MINPAY-LP(i)

$$\begin{aligned} \max \quad & \sum_{i' \neq i} \lambda_{i'} (c_i - c_{i'}) & & \\ \text{s.t.} \quad & \sum_{i' \neq i} \lambda_{i'} (q_{ij} - q_{i'j}) \leq q_{ij} & \forall j \\ & \lambda_{i'} \geq 0 & \forall i' \neq i \end{aligned}$$

(b) DUAL-MINPAY-LP(i)

Fig. 2: The MINPAY-LP(i) for action i **(left)** and its dual **(right)**. The primal variables are t_j for $j \in [m]$, and the dual variables are $\lambda_{i'}$ for $i' \in [n] \setminus \{i\}$.

is unbounded. If action i's outcome distribution can be replicated by a lower-cost convex combination of other actions, then a scaled version of the coefficients of this combination both satisfies the dual constraints, and achieves arbitrarily-high dual objective. The opposite direction is also not hard to show, establishing the characterization.

Optimal Contracts in Special Cases. In the special cases of binary actions and binary outcomes, optimal contracts have a particularly simple structure. For binary outcomes, the optimal contract is linear (see Section 4). For binary actions, the optimal contract only pays for a single outcome, the one that maximizes the so-called likelihood ratio. This also holds more generally for instances that, beyond the two actions, have a zero-cost action that leads to a zero-reward outcome with certainty, as in our running example.

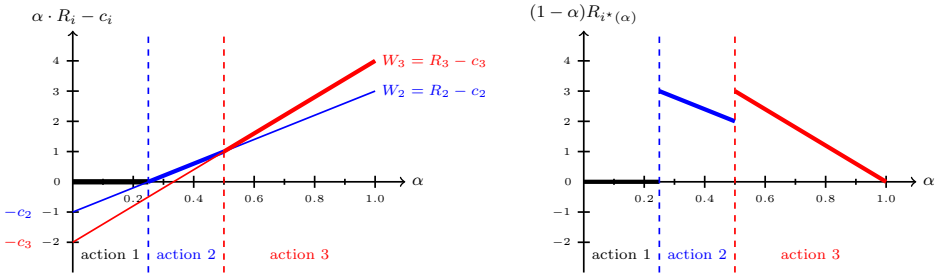
Example 4, Continued. *The optimal contract in our running example incentivizes the agent to take action* 3*, with contract* $t = (0, 0, 3)$ *which has a non-zero payment only for outcome* 3*. Note that outcome* 3 *is the outcome that maximizes the likelihood ratio* q_{3j}/q_{2j} *over all outcomes* j*.*

Shortcomings of Optimal Contracts. Beyond special cases, optimal contracts are often opaque and generally do not lend themselves to a clear, intuitive interpretation. Moreover, they are known to exhibit certain counterintuitive properties from an economic perspective. One particularly notable issue is that optimal contracts are not necessarily *monotone*, meaning that in the optimal contract $\mathbf{t}$, a higher principal reward r_j may entail a lower payment t_j. This is unnatural as the agent would be paid less, even though he achieved a better outcome.

Another drawback of optimal contracts is their reliance on perfect knowledge of inputs, such as distributions $\mathbf{q}$ and costs $\mathbf{c}$. This issue is exacerbated by the fact that they are sensitive to small perturbations. These shortcomings motivate some of the work covered in the following sections.

4. LINEAR CONTRACTS

A contract is *linear* if it specifies a fixed fraction $\alpha \in [0, 1]$ of the rewards to be transferred to the agent by the principal. The parameter α determines all the payments: $t_j = \alpha r_j$ for every outcome $j \in [m]$. The agent's expected utility from the ith action is then $\alpha R_i - c_i$. So as α becomes closer to 1, the agent puts more weight on the expected reward R_i relative to the cost c_i when choosing an action. At the extreme $\alpha = 1$, the agent chooses the welfare-maximizing action. Linear contracts



(a) The agent's perspective

(b) The principal's perspective

Fig. 3: The agent's expected utility as a function of the linear contract's parameter α **(left)**, and the principal's expected utility as a function of α **(right)**, for the principal-agent setting in Example 4.

have several nice properties like simplicity, interpretable payments (e.g., paying per conversion in Example 1), and monotonicity. Due to their widespread use, linear contracts were extensively studied in the economic literature. The computational perspective provides new insights, underscoring the potential of an algorithmic theory of contracts. We describe here two such insights: the approximation guarantees of linear contracts, and their robust optimality to uncertainty about distributions.

A Geometric Perspective. An important tool in the analysis of linear contracts, referred to as the geometric approach (Dütting et al., 2019), utilizes the fact that the agent's utility $\alpha R_i - c_i$ for each action i is a linear function in α. Specifically, the agent's best response to each contract α is given by the action i that results in the highest utility for the agent at that given α. In other words, by tracing the *upper envelope* of the set of functions $\{\alpha R_i - c_i\}_{i\in[n]}$ we can determine the agent's best response for each $\alpha \in [0,1]$. Figure 3a illustrates this for Example 4. From this perspective, it becomes clear that the unit interval of possible linear contracts will be divided into contiguous intervals where the agent's choice of action is constant. Also, if we rename the actions as they appear on the upper envelope (from left to right), then actions are sorted by non-decreasing welfare, reward, and cost.

Switching the roles, when we consider the principal's utility as a function of the contract $\alpha \in [0,1]$, we arrive at a graph like the one in Figure 3b (which is for the same example). We see that the principal's utility is a piece-wise linear function. Each segment corresponds to a different best-response of the agent, and within the segment corresponding to action i the principal's utility decreases at rate $-R_i$. Finally, under the standard tie-breaking rule, at the left-most point of a segment the agent will always choose the action with higher welfare (higher index, after re-indexing). Thus, the principal's utility is continuous when going to the right, while it might "jump" when going to the left. In particular, the best way to incentivize action i is by choosing the α_i that corresponds to the left-most endpoint of action i's interval. By checking the principal's utility at all *critical values* of α, i.e., all endpoints where the agent switches from one action to another, we can determine an optimal linear contract.

Example 4, Continued. *Example 4 demonstrates the suboptimality of linear*

contracts, which can obtain no more than 3 *as expected utility for the principal (see Figure 3b), in comparison to* 3.5 *by the optimal nonlinear contract.*

Approximation Guarantees. A key insight of the algorithmic study of contracts is the following set of tight (multiplicative) approximation guarantees, showing that linear contracts provide a good approximation except in cases where there is simultaneously a large number of actions, a large spread in expected rewards, and a large spread in costs.

Theorem 1 (Dütting, Roughgarden, and Talgam-Cohen (2019)). *Let ρ denote the worst-case ratio between the principal's expected utility under the optimal contract, and the principal's expected utility under the optimal* linear *contract. Then among all principal-agent settings with n actions, $\rho = n$; among all settings with a ratio of H between the highest and lowest expected reward, $\rho = \Theta(\log H)$; and among all settings with a ratio of C between the highest and lowest cost, $\rho = \Theta(\log C)$.*

The upper bounds in Theorem 1 in fact apply against the stronger benchmark of optimal welfare, while the lower bounds apply even under the regularity assumption of MLRP.

The direction that $\rho \leq n$ follows from the characterization of optimal linear contracts above. The tightness direction ($\rho \geq n$) is established for a particular family of contract settings with n actions known as the *equal revenue* setting, parameterized by a small $\epsilon > 0$. This setting has exponentially increasing expected rewards among the actions: $R_i = 1/\epsilon^{i-1}$; as well as exponentially increasing costs: $c_i = R_i - i + \epsilon(i-1)$; but the welfare of the actions increases linearly: $W_i = i - \epsilon(i-1)$. The key observation is that any linear contract extracts utility of 1 for the principal, while the setting can be set up such that a general contract can extract $W_n \approx n$. The other bounds follow by similar constructions.

Robust Optimality. A different approach, spearheaded in the context of contract design by Carroll (2015), is the robust (or max-min) optimality approach. This approach assumes that certain aspects of the contracting problem are *known* to the principal, while other aspects are *unknown*. The principal then aims to design a contract $\mathbf{t}$ that achieves the best-possible worst-case utility over all instances $\mathcal{I}$ compatible with what is known. That is, $\mathbf{t}$ maximizes $U_P(\mathcal{I} \mid \mathbf{t}) = \inf_{I \in \mathcal{I}} U_P(I \mid \mathbf{t})$, where $I \in \mathcal{I}$ is the worst-case compatible instance. It turns out that under a broad range of assumptions, it is a *linear* contract that achieves optimal such worst-case guarantees.

In Carroll's work, the aspect known to the principal is a subset of actions A_0 available to the agent, while the actual set of actions A available to the agent can be any superset of A_0.

Theorem 2 (Carroll (2015)). *Linear contracts are max-min optimal when the principal knows a subset of actions A_0 available to the agent, and $\mathcal{I}$ includes instances in which the full action set is any superset A, i.e., $A_0 \subseteq A$.*

An alternative formulation of the robustness of linear contracts was given by Dütting et al. (2019), in a natural model of moment information that is easy to interpret.

Theorem 3 (Dütting, Roughgarden, and Talgam-Cohen (2019)). *Linear contracts are max-min optimal when the principal knows the costs $c_1, \ldots, c_n$, the rewards $r_1, \ldots, r_m$, and the expected rewards $R_1, \ldots, R_n$ available to the agent, and $\mathcal{I}$ includes all instances in which the distributions over rewards induced by the actions are compatible with the expected rewards.*

In both cases, the proof approach is based on showing that whatever non-linear contract $\mathbf{t}$ is considered as a candidate, an instance $I \in \mathcal{I}$ and a linear contract $\mathbf{t}'$ exist such that $U_P(I \mid \mathbf{t}) \leq U_P(\mathcal{I} \mid \mathbf{t}')$. Combined with $U_P(\mathcal{I} \mid \mathbf{t}) \leq U_P(I \mid \mathbf{t})$, we get that $U_P(\mathcal{I} \mid \mathbf{t}) \leq U_P(\mathcal{I} \mid \mathbf{t}')$, as desired.

5. COMBINATORIAL CONTRACTS

In the basic contracting model (Section 2), a single principal interacts with a single agent, who chooses one of n actions leading to one of m outcomes, and the revenue-optimal contract can be computed efficiently via LPs (see Section 3).

Real-world contract settings, however, are rarely this simple. A brand may engage multiple influencers (agents), each promoting multiple brands (principals) across various platforms (actions), evaluated by diverse metrics (outcomes). Such generalizations introduce new computational challenges, which the growing body of work on algorithmic contract theory aims to address. This body of work provides both efficient algorithms and impossibility results for settings with multiple agents (Babaioff et al., 2012; Dütting et al., 2026), combinatorial actions (Dütting et al., 2025a, 2024; Deo-Campo Vuong et al., 2024; Dütting et al., 2026), combinatorial outcomes (Dütting et al., 2021), multiple principals (Alon et al., 2024), and combinations thereof (Dütting et al., 2025b). As is typical at the interface of economics and computation, the computational perspective reveals key structural insights. Below we present a single paradigmatic example to illustrate the kinds of challenges that arise, and approaches to address them.

Paradigmatic Example: Multiple Agents. A natural extension of the contracting problem arises when a principal seeks to incentivize a team of agents. The classic reference here is Holmström (1982) on moral hazard in teams. The effectiveness of a team depends heavily on its composition, leading to the algorithmic challenge of designing (near-)optimal contracts for subsets of agents. This is already nontrivial in the basic setting where agents can either exert effort or not (the binary action setting), reducing to the question of which agents to contract with.

This line of work was initiated by Babaioff et al. (2012) under the name *combinatorial agency*. We present here the more general model of Dütting et al. (2026). There is a set N of n agents. Each agent $i \in N$ incurs a cost $c_i \in \mathbb{R}_{\geq 0}$ to exert effort, and the project either succeeds (reward 1) or fails (reward 0). A set function $f : 2^N \to [0, 1]$ specifies the success probability as a function of the subset of agents exerting effort. As in the single agent case, in this binary-outcome model one can focus on linear contracts: a vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$, where α_i is the share of the reward agent i receives if the project succeeds. Given a contract $\boldsymbol{\alpha}$ and a set $S \subseteq N$ of agents who exert effort, the principal's utility is $U_P(S \mid \boldsymbol{\alpha}) := (1 - \sum_{i \in N} \alpha_i) f(S)$, and agent i's utility is $U_i(S \mid \boldsymbol{\alpha}) := \alpha_i f(S) - \mathbb{1}[i \in S] \cdot c_i$, where $\mathbb{1}[i \in S] = 1$ if $i \in S$ and $\mathbb{1}[i \in S] = 0$ otherwise.

Each contract thus defines a game among the agents. We focus on pure Nash equilibria, namely action profiles where no agent has an incentive to deviate. A set S is an equilibrium under a contract $\boldsymbol{\alpha}$ if: (i) for every $i \in S$, $\alpha_i f(S) - c_i \geq \alpha_i f(S \setminus \{i\})$ (no incentive to deviate to shirking), and (ii) for every $i \notin S$, $\alpha_i f(S) \geq \alpha_i f(S \cup \{i\}) - c_i$ (no incentive to deviate to exerting effort). It is not too difficult to show that every contract $\boldsymbol{\alpha}$ admits an equilibrium (Deo-Campo Vuong et al., 2024; Dütting et al., 2025b). The problem is to find the contract $\boldsymbol{\alpha}$ that maximizes the principal's utility, in the best equilibrium of $\boldsymbol{\alpha}$.

The optimal choice of $\boldsymbol{\alpha}$ for a given set S is $\alpha_i = \frac{c_i}{f(i|S\setminus\{i\})}$ for $i \in S$, and $\alpha_i = 0$ for $i \notin S$. That is, the payment to an agent in S is proportional to his cost and inverse proportional to his marginal contribution to the success probability, reflecting the fact that an agent whose work matters less requires high compensation to avoid free riding on others' work. Thus, finding the optimal contract reduces to identifying the set $S^\star$ that maximizes the function

$$g : 2^N \to \mathbb{R} \cup \{-\infty\}, \qquad \text{where} \quad g(S) := \left(1 - \sum_{i \in S} \frac{c_i}{f(i \mid S \setminus \{i\})}\right) f(S).$$

Approximation Results. Dütting et al. (2026) study the problem of computing near-optimal contracts, assuming the success function f belongs to the hierarchy of complement-free set functions (e.g., additive, submodular, XOS, subadditive), capturing situations where an agent contribution diminishes as the team grows. They show that even for additive f, the problem is NP-hard, but admits an FPTAS. For more general set functions, they establish approximation results, given access to common primitives: A value query is given a set S and returns $f(S)$. A demand query gets a price vector $p_1, \ldots, p_m$ and returns a set maximizing $f(S) - \sum_{j \in S} p_j$.

Theorem 4 (Dütting, Ezra, Feldman, and Kesselheim (2026)). *For submodular (resp., XOS) success functions f, an $O(1)$-approximation to the optimal contract can be computed using poly-many value queries (resp., value and demand queries).*

The proof of Theorem 4 reveals a (perhaps surprising) connection to pricing. Namely, it can be shown that using a demand query with carefully chosen prices allows one to identify a "good" set of agents—one whose success value $f(S)$ is at least half that of the optimal set. However, to ensure a high $g(S)$ value, the payments must remain low; in particular, agents' marginal contributions to f must be sufficiently large. The crux is therefore a scaling lemma, which shows that it is always possible to remove agents from the set S, while also preserving a sufficiently large fraction of the original marginal contributions. For submodular functions, the latter part is straightforward, since the marginal contributions of the remaining agents can only go up. For XOS functions, however, the marginal contributions of the agents that remain may decrease, so a more careful argument is required to establish such a scaling property.

This proof pattern yields an efficient algorithm with value and demand queries. For submodular functions, the argument can be further refined, by replacing demand queries with an appropriately defined relaxation of demand queries that can be computed with poly-many value queries.

Impossibility Results. Follow-up work by Ezra et al. (2024) shows that for submodular functions, there exists a constant $c > 1$ such that no polynomial-time algorithm with value oracle access can approximate the optimal contract to within a factor better than c, assuming P≠NP. More recently, Dütting et al. (2025b) showed that even with both value and demand oracle access to the submodular function, there exists a constant $c > 1$, such that any algorithm that uses a sub-exponential number of queries returns a c-approximation with probability exponentially-small in n. In addition, for XOS functions, Ezra et al. (2024) show that no algorithm that makes poly-many value queries can approximate the optimal contract (with high probability) to within a factor $\Omega(n^{1/6})$.

6. LEARNING CONTRACTS

Another natural direction for an algorithmic study of contracts pioneered in Ho et al. (2016), is to study the contract design problem from a learning perspective. Here we focus on the online learning problem with bandit feedback studied in Ho et al. (2016), and discuss the state-of-the-art results of Zhu et al. (2023). Other variants of the problem have been studied, including under additional restrictions on the model (small action spaces, distributions satisfying additional conditions, etc.) (Bacchiocchi et al., 2024; Chen et al., 2024), and under different feedback models (Dütting et al., 2023; Chen et al., 2024; Dütting et al., 2025). Finally, Wang et al. (2023) propose a deep neural network architecture for designing near-optimal contracts from samples.

Online Learning with Bandit Feedback. There is a single principal interacting repeatedly with a single agent, whose type is drawn from an unknown type distribution $\mathcal{D}$. The interaction takes place over S rounds. In each round s, the agent's type θ^s is drawn from the underlying type distribution $\mathcal{D}$, independently. Over all rounds, the principal has fixed rewards $\{r_j\}_{j\in[m]}$. The agent's type θ^s determines the cost $c_i^\theta \geq 0$ of each action $i \in [n]$, as well as the probability distribution $\mathbf{q}_i^\theta$ over outcomes $j \in [m]$. It is assumed that both rewards and costs are bounded in $[0, 1]$.[3]

In each round s, the principal posts a contract $\mathbf{t}^s = (t_1^s, \ldots, t_m^s)$ (a non-negative payment for each outcome). The choice of contract may depend on what the principal has observed so far, and may be randomized. We consider two classes of contracts. In a (general) *bounded* contract we have $\mathbf{t}^s \in [0, 1]^m$, while a *linear* contract is defined as before. After the principal has posted contract $\mathbf{t}^s$, a type θ^s is drawn from $\mathcal{D}$, the agent takes a best response action $i^\star(\theta^s, \mathbf{t}^s)$, and an outcome j^s is sampled from $\mathbf{q}_{i^\star(\theta^s,\mathbf{t}^s)}$. The principal learns about the outcome j^s, receives the corresponding reward r_{j^s}, and pays the agent the amount specified by contract $\mathbf{t}^s$ for outcome j^s.

The principal's goal is to minimize *regret* with respect to the best single contract in hindsight. To formally define this, let $\mathcal{T}$ denote a class of contracts. Let $U_P(\theta \mid \mathbf{t})$ denote the expected principal utility for contract $\mathbf{t}$ when the agent's type is θ, and let π be a policy which maps each history $\mathcal{H}^{s-1}$ to a distribution over contracts.

[3]Since regret is an additive metric, we need to specify the range of the key quantities involved. Normalization to $[0, 1]$ can be achieved through appropriate scaling, but also scales the regret with respect to the original unscaled instance.

Then

$$\mathsf{regret}(\pi, \mathcal{T}) := \sup_{\bar{\mathbf{t}} \in \mathcal{T}} \sum_{s=1}^{S} \mathbb{E}_{\mathbf{t}^s \sim \pi(\mathcal{H}^{s-1})} \left(\mathbb{E}_{\theta^s}[U_P(\theta^s \mid \bar{\mathbf{t}})] - \mathbb{E}_{\theta^s}[U_P(\theta^s \mid \mathbf{t}^s)] \right).$$

Separation: General vs. Linear Contracts. The main results of Zhu et al. (2023) are near-tight bounds on the regret, showing that while the regret for (general) bounded contracts has to be essentially linear, linear contracts admit sublinear regret bounds.

Theorem 5 (Zhu, Bates, Yang, Wang, Jiao, and Jordan (2023)).
(1) For (general) bounded contracts there is an online learning algorithm that incurs a regret of at most $\tilde{O}(\sqrt{m} \cdot S^{1-1/(2m+1)})$, and no online learning algorithm can incur a regret better than $\Omega(S^{1-1/(m+2)})$.
(2) For linear contracts there is an online learning algorithm that incurs a regret of at most $\tilde{O}(S^{2/3})$, and no online learning algorithm can incur a regret better than $\Omega(S^{2/3})$.

The intuitive reason for the sharp separation between (general) bounded contracts and linear contracts is the very different geometry of the principal's utility in the two cases. For linear contracts (Figure 3b), the optimal principal utility is a well-behaved, right-continuous function. In particular, there is always a "safe" direction: if $\alpha^\star$ is the optimal linear contract and we over-shoot by ϵ, then we only lose ϵ utility. By contrast, for general (bounded) contracts, the principal's utility admits only a weak continuity property: for each contract $\mathbf{t}$, there exists a direction (a cone) along which the utility does not drop too much. While this enables a cover-by-cones argument, yielding the exponential upper bound, the lower-bound construction shows there is no additional structure that would enable more efficient learning.

Notably, the impossibility for (general) bounded contracts already holds for a fixed agent (i.e, when $\mathcal{D}$ is a point-mass distribution), and requires exponentially many (in m) actions. The case of a fixed agent with a small action space is handled in Bacchiocchi et al. (2024). An appealing open problem is whether the exponential lower bound can be circumvented under additional structure, e.g., MLRP (it can, for instances satisfying FOSD and CDFP (Chen et al., 2024)).

7. LOOKING AHEAD

Attesting to the rapid growth of the new research area of algorithmic contract design, in this brief survey we have not been able to touch upon all lines of research. For example, there has been much interest in *typed* contract design: Bayesian environments where agents have private type information, which call for a unified treatment of mechanism design and contract design (Guruganesh et al., 2021; Alon et al., 2021, 2023; Castiglioni et al., 2025, 2023).

While substantial progress has been made in developing the foundations of algorithmic contract design, several directions remain only partially explored. These additional directions include inspection and monitoring of actions and outcomes (Ezra et al., 2026), the interplay with information revelation (Babichenko et al., 2024; Castiglioni and Chen, 2025), ambiguous contract design (Dütting et al., 2024,

2025), sequential decision making (Ezra et al., 2026), and classification (Kleinberg and Raghavan, 2019; Alon et al., 2020) (as in Example 2). It would also be interesting to further explore the connection to closely related approaches such as scoring rules (Hartline et al., 2023) and delegation (Kleinberg and Kleinberg, 2018).

Finally, circling back to Example 3, the rise of an "agentic era," where autonomous AI agents act as both principals and agents, raises new challenges only beginning to be addressed (Saig et al., 2024). Vast and dynamic action spaces, shifting assumptions about rationality and commitment, and the computational capabilities of AI agents open fresh avenues for contract design. As in algorithmic mechanism design, we expect the computational lens to reveal new structures, tractability frontiers, and insights, shaping the next phase of contract theory.

Acknowledgments

This survey received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 101077862) and the Horizon Europe program for research and innovation (grant agreement No. 101170373), by the Israel Science Foundation (grant No. 3331/24), by the Israel Science Foundation Breakthrough Program (grant No. 2600/24), by the NSF-BSF (grant No. 2020788 and grant No. 2021680), by a Google Research Scholar Award, and by an Amazon Research Award.

REFERENCES

Alon, T., Dobson, M., Procaccia, A. D., Talgam-Cohen, I., and Tucker-Foltz, J. 2020. Multiagent evaluation mechanisms. In *Proc. of AAAI 2020*. AAAI Press, Washington, DC, USA, 1774–1781.

Alon, T., Dütting, P., Li, Y., and Talgam-Cohen, I. 2023. Bayesian analysis of linear contracts. In *Proc. of EC 2023*. ACM, New York, NY, USA. Full version available at: `https://arxiv.org/abs/2211.06850`.

Alon, T., Dütting, P., and Talgam-Cohen, I. 2021. Contracts with private cost per unit-of-effort. In *Proc. of EC 2021*. ACM, New York, NY, USA, 52–69.

Alon, T., Lavi, R., Shamash, E. S., and Talgam-Cohen, I. 2024. Incomplete information VCG contracts for common agency. *Oper. Res. 72,* 1, 288–299. An earlier version appeared in EC 2021.

Babaioff, M., Feldman, M., Nisan, N., and Winter, E. 2012. Combinatorial agency. *J. Econ. Theory 147,* 3, 999–1034. An earlier version appeared in EC 2006.

Babichenko, Y., Talgam-Cohen, I., Xu, H., and Zabarnyi, K. 2024. Information design in the principal-agent problem. In *Proc. EC 2024*. ACM, New York, NY, USA, 669–670.

Bacchiocchi, F., Castiglioni, M., Marchesi, A., and Gatti, N. 2024. Learning optimal contracts: How to exploit small action spaces. In *Proc. of ICLR 2024*. PMLR.

Carroll, G. 2015. Robustness and linear contracts. *Am. Econ. Rev. 105,* 2, 536–63.

Castiglioni, M. and Chen, J. 2025. Hiring for an uncertain task: Joint design of information and contracts. In *Proc. of SODA 2025*. SIAM, Philadelphia, PA, USA, 1758–1794.

Castiglioni, M., Chen, J., Li, M., Xu, H., and Zuo, S. 2025. A reduction from multi-parameter to single-parameter Bayesian contract design. In *Proc. of SODA 2025*. SIAM, Philadelphia, PA, USA, 1795–1836.

Castiglioni, M., Marchesi, A., and Gatti, N. 2023. Designing menus of contracts efficiently: The power of randomization. *Artif. Intell. 318*, 103881. An earlier version appeared in EC 2022.

Chen, Y., Chen, Z., Deng, X., and Huang, Z. 2024. Are bounded contracts learnable and approximately optimal? In *Proc. of EC 2024*. ACM, New York, NY, USA.

Deo-Campo Vuong, R., Dughmi, S., Patel, N., and Prasad, A. 2024. On supermodular contracts and dense subgraphs. In *Proc. of SODA 2024*. SIAM, Philadelphia, PA, USA, 109–132.

Dütting, P., Ezra, T., Feldman, M., and Kesselheim, T. 2025a. Combinatorial contracts. *SIAM J. Comput. 54,* 6, 1427–1455. An earlier version appeared in FOCS 2021.

Dütting, P., Ezra, T., Feldman, M., and Kesselheim, T. 2025b. Multi-agent combinatorial contracts. In *Proc. of SODA 2025*. SIAM, Philadelphia, PA, USA, 1857–1891.

Dütting, P., Ezra, T., Feldman, M., and Kesselheim, T. 2026. Multi-agent contracts. *J. ACM 73,* 2, 13:1–13:37. An earlier version appered in STOC 2023.

Dütting, P., Feldman, M., and Gal-Tzur, Y. 2024. Combinatorial contracts beyond gross substitutes. In *Proc. of SODA 2024*. SIAM, Philadelphia, PA, USA, 92–108.

Dütting, P., Feldman, M., Gal-Tzur, Y., and Rubinstein, A. 2026. When contracts get complex: Information-theoretic barriers. In *Proc. of SODA 2026*. SIAM, Philadelphia, PA, USA, 6464–6493.

Dütting, P., Feldman, M., Peretz, D., and Samuelson, L. 2024. Ambiguous contracts. *Econometrica 92,* 6, 1967–1992. An earlier version appeared in EC 2023.

Dütting, P., Feldman, M., Ponitka, T., and Soumalias, E. 2025. The pseudo-dimension of contracts. In *Proc. EC 2025*. ACM, New York, NY, USA, 514–539.

Dütting, P., Feldman, M., and Rashti, Y. 2025. Succinct ambiguous contracts. Available at `https://arxiv.org/abs/2503.02592`.

Dütting, P., Feldman, M., and Talgam-Cohen, I. 2024. Algorithmic contract theory: A survey. *Found. Trends Theor. Comput. Sci. 16,* 3-4, 211–412. `https://arxiv.org/abs/2412.16384`.

Dütting, P., Guruganesh, G., Schneider, J., and Wang, J. 2023. Optimal no-regret learning of one-sided Lipschitz functions. In *Proc. of ICML 2023*. PMLR.

Dütting, P., Roughgarden, T., and Talgam-Cohen, I. 2019. Simple versus optimal contracts. In *Proc. of EC 2019*. ACM, New York, NY, USA, 369–387. Full version available at `https://arxiv.org/pdf/1808.03713`.

Dütting, P., Roughgarden, T., and Talgam-Cohen, I. 2021. The complexity of contracts. *SIAM J. Comput. 50,* 1, 211–254. An earlier version appeared in SODA 2020.

Ezra, T., Feldman, M., and Schlesinger, M. 2024. On the (in)approximability of combinatorial contracts. In *Proc. of ITCS 2024*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Wadern, Germany, 44:1–44:22.

Ezra, T., Feldman, M., and Schlesinger, M. 2026. Contract design for sequential actions. In *Proc. of SODA 2026*. SIAM, Philadelphia, PA, USA, 6537–6570.

Ezra, T., Leonardi, S., and Russo, M. 2026. Contract design beyond hidden actions. In *Proc. of SODA 2026*. SIAM, Philadelphia, PA, USA, 2113–2133.

Feldman, M. 2025. Combinatorial contract design: Recent progress and emerging frontiers. *arXiv preprint arXiv:2510.15065*.

Grossman, S. J. and Hart, O. D. 1983. An analysis of the principal-agent problem. *Econometrica 51,* 1, 7–45.

Guruganesh, G., Schneider, J., and Wang, J. 2021. Contracts under moral hazard and adverse selection. In *Proc. of EC 2021*. ACM, New York, NY, USA, 563–582.

Hartline, J. D., Shan, L., Li, Y., and Wu, Y. 2023. Optimal scoring rules for multi-dimensional effort. In *Proc. of COLT 2023*. PMLR, 2624–2650.

Hermalin, B. E. and Katz, M. L. 1991. Moral hazard and verifiability: The effects of renegotiation in agency. *Econometrica 59*, 1735–1753.

Ho, C., Slivkins, A., and Vaughan, J. W. 2016. Adaptive contract design for crowdsourcing markets: Bandit algorithms for repeated principal-agent problems. *J. Artif. Intell. Res. 55*, 317–359. An earlier version appeared in EC 2014.

Holmström, B. 1979. Moral hazard and observability. *Bell J. Econ. 10*, 74–91.

Holmström, B. 1982. Moral hazard in teams. *Bell J. Econ. 13*, 324–340.

Kleinberg, J. and Raghavan, M. 2019. How do classifiers induce agents to invest effort strategically? In *Proc. of EC 2019*. ACM, New York, NY, USA, 825–844.

Kleinberg, J. M. and Kleinberg, R. 2018. Delegated search approximates efficient search. In *Proc. of EC 2018*. ACM, New York, NY, USA, 287–302.

Nisan, N. and Ronen, A. 2001. Algorithmic mechanism design. *Games Econ. Behav. 35,* 1, 166–196. An earlier version appeared in STOC 1999.

Saig, E., Einav, O., and Talgam-Cohen, I. 2024. Incentivizing quality text generation via statistical contracts. In *Proc. of NeurIPS 2024*. PMLR.

Wang, T., Dütting, P., Ivanov, D., Talgam-Cohen, I., and Parkes, D. C. 2023. Deep contract design via discontinuous networks. In *Proc. of NeuRIPS 2023*. PMLR.

Zhu, B., Bates, S., Yang, Z., Wang, Y., Jiao, J., and Jordan, M. I. 2023. The sample complexity of online contract design. In *Proc. of EC 2023*. ACM, New York, NY, USA, 1188.

AI Alignment From Social Choice Perspectives

DANIEL HALPERN
Google Research
EVI MICHA
University of Southern California
ARIEL D. PROCACCIA
Harvard University
BENJAMIN SCHIFFER
Harvard University
ITAI SHAPIRA
Harvard University
and
SHIRLEY ZHANG
Harvard University

Alignment from human feedback uses human judgments about model outputs to steer the behavior of language models after pretraining. When those judgments reflect conflicting views of desirable behavior, the learned objective becomes an aggregate determination of what the model should prefer. We survey recent work that has studied this aggregation problem through the lens of social choice theory. We illustrate how the social choice perspective helps identify failure modes in the feedback aggregation layer and reveals a broader design space for handling disagreement in explicit and principled ways.

1. INTRODUCTION

AI alignment is the problem of ensuring that artificial intelligence systems act in ways consistent with human intentions, preferences, and normative constraints [1–4]. Achieving alignment requires identifying which outputs and behavioral patterns are desirable or unacceptable, and using those judgments to control and steer the system [5–7].

For open-ended language models, operationalizing these goals presents a difficult challenge. The full range of desirable behaviors cannot be explicitly specified [8]; the underlying human judgments rely on tacit knowledge and subtle tradeoffs between objectives, making them virtually impossible to capture in formal terms [1, 9].

Fortunately, humans are far better at recognizing acceptable behavior than they are at formally articulating it [10, 11], an asymmetry that motivates a shift from explicit specification to learned evaluation [12]. In methods for *alignment from human feedback*, most notably *reinforcement learning from human feedback (RLHF)* [12–16], human annotators evaluate concrete model outputs, and a scoring function trained

Authors are listed alphabetically. Correspondence to Itai Shapira (itaishapira@g.harvard.edu).

on this feedback is then used to align the model.[1] Typically, annotators express preferences via pairwise comparisons, selecting the better of two responses to a given prompt. Aggregated across users and contexts, the annotations are distilled into a learned reward model: a parametric scoring function trained to predict the desirability of a given output. This reward model is expected to generalize beyond the labeled training data, capturing implicit norms and tradeoffs among competing objectives [20].

The foregoing methods compress human preferences about model behavior into a single, universally applicable scalar reward signal that ostensibly represents human judgment. This approach implicitly assumes a shared, underlying human intuition that can be statistically recovered by querying human evaluators [21–23]. Under this view, annotators are interchangeable. Conflicting preferences are treated as noisy observations of a common ground truth, rather than evidence of value pluralism worth reflecting in the system [22, 24]. While such methods have proven empirically successful on tasks where evaluators largely agree [13, 15, 25], in many contexts, interpretations of "correct" behavior diverge across backgrounds and cultures [26–31].

When annotators inherently disagree, reward modeling goes beyond statistical estimation; it becomes a form of preference aggregation. It collapses conflicting individual judgments into a single collective preference, implicitly fixing tradeoffs among competing values while obscuring the mechanism used to resolve them [32]. Evaluating this choice requires examining the assumed model structure and the aggregation method itself. These questions cannot be resolved within the training objective alone. Instead, they fall squarely within the domain of *social choice theory* [33–42], the mathematical study of how heterogeneous individual preferences are aggregated into collective decisions.

Viewed through this lens, alignment can be analyzed as a formal aggregation pipeline: from elicited judgments to learned objectives to optimized policies. Social choice provides the precise language to express desired properties explicitly, making the normative assumptions embedded in aggregation procedures more transparent and open to mathematical comparison.

In this paper, we survey three related types of contributions the social choice perspective has brought to recent alignment research. First, this framing identifies the aggregation rules embedded in widely used methods such as RLHF and DPO [16], showing which assumptions and normative priorities these methods build into their objectives. Second, social choice axioms expose structural failure modes of such rules, identifying conditions under which alignment methods provably violate desirable aggregation properties. Third, social choice provides a broad toolbox of well-studied aggregation methods, informing new approaches that encode different normative objectives through various choices of feedback elicitation, aggregation, and policy optimization.

[1] Constitutional AI [17] and its variants inherit the same framing one level up, as the "constitution" is itself the product of preference aggregation over normative principles [18], and discretion must still be exercised in resolving their conflicts [19].

Yet thinking of alignment from human feedback purely in terms of classical social choice misses two distinctive aspects of alignment that shape much of the research surveyed below. The first is *generalization*. Because the candidate-response space is generated and effectively unbounded, the feedback data cover only a sparse set of comparisons over prompt-response pairs, provided by a small subset of evaluators. The learned preference model must then generalize in three directions: from evaluated responses to unseen responses, from curated prompts to unseen prompts, and from sampled annotators to a broader population of users [43].

The second feature is that alignment is ultimately evaluated on the *downstream policy* [44, 45]. The intermediate learned reward model matters only insofar as optimizing against it produces a policy aligned with human judgments. A reward model that satisfies desirable aggregation properties need not induce a policy that does the same.

The goal, then, is to adapt what social choice has long understood about collective decision-making into the practical machinery of model training. AI alignment, in this light, is not merely the problem of steering AI to follow human values, but rather of establishing fair principles for incorporating the diversity of values people actually hold [3, 46, 47]. The sections that follow trace this adaptation.

Survey Roadmap. After Section 2 introduces the notation used throughout this paper, Section 3 identifies the aggregation rule implicit in unconstrained Bradley–Terry reward learning, showing that it behaves like the classic *Borda rule*.

The next two sections examine failure modes that surface when this aggregation rule is placed under the distinctive constraints of alignment. Section 4 explores clone robustness, an inherited social-choice pathology that is especially natural in language-model settings, where generated candidate sets may contain many nearby strings expressing the same substantive answer, making the learned reward sensitive to how that answer is represented in the sample. Section 5 turns to failures that arise from the way the reward function is learned from data, in particular from restricting the reward to a limited reward class.

The following sections move from aggregation to the information and policy-level limits of alignment from feedback. Section 6 studies what sparse feedback can reveal about the underlying population preference distribution and how welfare can be lost when policies are optimized from limited preference information. Section 7 then considers policy-level methods that avoid first compressing pairwise preferences into a scalar reward. It explains how *Nash learning from human feedback* optimizes directly against pairwise preferences and achieves optimal welfare guarantees. We conclude in Section 8 by discussing related directions at the boundary between social choice and alignment.

2. RLHF PRIMER AND NOTATION

This section fixes the notation used throughout the paper and reviews the standard RLHF pipeline at the level needed for our analysis. We first define prompts, responses, annotators, policies, and reward models, then describe how pairwise preference data are aggregated into a scalar reward via *random utility models* and

the Bradley–Terry loss, and finally recall the KL-regularized policy objective used to optimize a language model against the learned reward.

Setup and Notation. Let $\mathcal{X}$ and $\mathcal{Y}$ denote the prompt and response spaces, respectively. A prompt x may encode either a single query or the dialogue history of a multi-turn interaction. Let $\mathcal{A}$ denote the target population of annotators, and let $i \sim \mathcal{A}$ denote a randomly sampled annotator. A stochastic policy π assigns to each prompt $x \in \mathcal{X}$ a distribution $\pi(\cdot \mid x) \in \Delta(\mathcal{Y})$ over responses. We let π_{base} denote a fixed reference policy, typically a pretrained model that has undergone supervised fine-tuning. A reward model is a function $r : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ assigning a real-valued score to each prompt-response pair. When considering a parameterized family of reward models, we write r_θ for the model indexed by $\theta \in \Theta$. When the prompt x is fixed or clear from context, we write $r(y)$ as shorthand for $r(x, y)$, and similarly write $r_\theta(y)$ for $r_\theta(x, y)$. For analyses restricted to a fixed prompt x, let $\mathcal{Y}_x \subseteq \mathcal{Y}$ denote the set of candidate responses considered for that prompt. Finally, let D denote the distribution over prompts used for post-training and policy evaluation.

Preference Data and Reward Learning. Alignment from human feedback can use several forms of supervision, including demonstrations, scalar ratings, critiques, natural-language feedback, and comparisons [48]. In the reward-modeling stage of standard RLHF pipelines, the data are typically comparison labels: for a prompt x and two candidate responses $y, y' \in \mathcal{Y}$, an annotator indicates which response they prefer [13, 15]. Rankings over a slate of responses are often converted into induced pairwise comparison labels [13, 49]. We treat these labels as samples from a population preference relation and write

$$P_x(y \succ y') \coloneqq \Pr_{i \sim \mathcal{A}}[y \succ_i y'] \in [0, 1].$$

That is, $P_x(y \succ y')$ is the population probability that a randomly drawn annotator prefers y to y' on prompt x. Most methods considered in this survey take only this pairwise preference object as input.[2] At times, we view a sampled annotator $i \sim \mathcal{A}$ as having an unobserved implicit reward function $r_i(x, \cdot)$ over responses, which induces the preference relation $\succ_i$.

Reward learning fits a scalar function r_θ whose induced pairwise probabilities approximate these population probabilities. In standard alignment pipelines [12–15], this is done by fitting a parametric reward model via maximum likelihood under a *random utility model (RUM)*, in which observed preferences arise from latent per-response rewards. Specifically,

$$P_x(y \succ y') = F(r(x, y) - r(x, y')),$$

where $F : \mathbb{R} \to (0, 1)$ is an increasing link function satisfying $F(t) = 1 - F(-t)$ and encoding the comparison noise [51–54]. Different noise structures, i.e., different choices of F, lead to different objective functions [55]. The widely used *Bradley–Terry (BT) model* [56] takes F to be the sigmoid $\sigma(t) \coloneqq (1 + e^{-t})^{-1}$, corresponding to Gumbel-distributed noise [57, 58].

[2] In social choice theory, such rules are called *C2 rules* [50].

Concretely, let $\mathcal{D}$ be a dataset of pairwise comparisons, where each triple $(x, y^+, y^-) \in \mathcal{D}$ has $x \in \mathcal{X}$, $y^+, y^- \in \mathcal{Y}$, and records that an annotator preferred y^+ to y^- on prompt x. For a parametric reward model r_θ, BT training minimizes the following logistic loss:

$$\mathcal{L}_{\mathrm{BT}}(\theta) = \sum_{(x,y^+,y^-)\in\mathcal{D}} \log\Big(1 + \exp\big(-(r_\theta(x, y^+) - r_\theta(x, y^-))\big)\Big). \tag{1}$$

This loss function is the negative log-likelihood of the observed preference data under the assumption that preferences satisfy $P_x(y^+ \succ y^-) = \sigma\left(r(y^+) - r(y^-)\right)$. When this holds, we say that the BT model is *correctly specified*.

KL-Regularized RLHF. During post-training, the learned reward function provides scalar feedback to the language model to optimize against. In the classical formulation [14], the policy objective is given by:

$$\max_{\pi}\ \mathbb{E}_{x\sim D}\left[\mathbb{E}_{y\sim\pi(\cdot|x)}\big[r(x,y)\big] - \beta^{-1}\mathrm{KL}\big(\pi(\cdot \mid x)\,\|\,\pi_{\mathrm{base}}(\cdot \mid x)\big)\right]. \tag{2}$$

We abbreviate

$$\mathrm{KL}_D(\pi\,\|\,\pi_{\mathrm{base}}) \coloneqq \mathbb{E}_{x\sim D}\big[\mathrm{KL}\big(\pi(\cdot \mid x)\,\|\,\pi_{\mathrm{base}}(\cdot \mid x)\big)\big].$$

The second term keeps the new policy from drifting too far from the reference policy, and we treat β, the tilt strength (inverse temperature), as a control parameter for this drift. Larger β pushes $\pi(\cdot \mid x)$ more aggressively toward high-reward responses and further from $\pi_{\mathrm{base}}(\cdot \mid x)$. Even a small gap between two responses can then shift a large amount of probability mass when β is large, so aggregation errors that look minor at the reward-learning stage can be magnified by downstream policy optimization.

3. REWARD LEARNING AS IMPLICIT AGGREGATION

When human feedback is heterogeneous, different annotators can give conflicting preference labels for the same prompt. Fitting a single reward then requires the objective to reconcile these judgments into one score per response. BT reward learning is usually read as statistical estimation of latent utilities from noisy pairwise comparisons. However, this framing obscures the reconciliation process, treating it as a technical detail rather than a choice of aggregation rule with inherent normative tradeoffs. This section asks what rule BT reward learning implicitly applies, and shows it to be closely connected to a classical voting rule.

The *Borda rule*, also known as Borda count, is named after Jean-Charles de Borda, who (re)introduced it in the 18th century [59, 60]. In our setting, fix a prompt x and let D_x be the distribution over responses used to generate comparison candidates. Let each annotator induce a preference relation $\succ_i$ over responses in the support of D_x. The *Borda score* of a response y is the population probability that y is

preferred to a randomly drawn comparison candidate:[3]

$$\mathrm{Borda}(y) \coloneqq \mathbb{E}_{y' \sim D_x}\big[P_x(y \succ y')\big]. \tag{3}$$

That is, Borda favors responses that maximize their expected pairwise advantage against a randomly drawn candidate.

In practice, the underlying aggregation rule of BT reward learning is obscured by two structural constraints: the learner observes only a sparse, finite sample of comparisons, and it must fit a parametric reward function to generalize across distinct prompts and responses. To isolate the aggregation step, we remove these constraints and consider an idealized setting. Assume the true population pairwise preference probabilities are known for all pairs in the candidate set $\mathcal{Y}_x$, and allow each response $y \in \mathcal{Y}_x$ to receive an independent, arbitrary scalar reward. The BT objective reduces to the prompt-level population loss

$$\mathcal{L}_x(r) \coloneqq -\mathbb{E}_{y,y' \sim D_x}\big[\, P_x(y \succ y') \, \log \sigma\big(r(y) - r(y')\big)\big], \tag{4}$$

minimized over all $r\colon \mathrm{supp}(D_x) \to \mathbb{R}$. This represents the asymptotic limit of empirical BT learning given infinite comparisons and a fully expressive reward class. Under these conditions, the BT objective has a clean social-choice characterization: its optimizer ranks responses exactly by their Borda scores.

Theorem 1 [61]. *Fix a prompt x, and let D_x be the distribution over responses used to generate comparison candidates. Let $r^\star$ be a finite-valued minimizer of Equation 4. Then, for any $y, y' \in \mathrm{supp}(D_x)$,*

$$\mathrm{Borda}(y) > \mathrm{Borda}(y') \quad \Longleftrightarrow \quad r^\star(y) > r^\star(y')\,.$$

Proof of Theorem 1. For any y with $D_x(y) > 0$, differentiating the BT objective with respect to $r(y)$ gives

$$\frac{\partial \mathcal{L}_x}{\partial r(y)}(r) = D_x(y)\mathbb{E}_{z \sim D_x}\Big[\sigma(r(y) - r(z)) - P_x(y \succ z)\Big].$$

Thus first-order optimality at $r^\star$, together with $D_x(y) > 0$, gives

$$\mathrm{Borda}(y) = \mathbb{E}_{z \sim D_x}\Big[\sigma(r^\star(y) - r^\star(z))\Big]. \tag{5}$$

Let $f(t) = \mathbb{E}_{z \sim D_x}[\sigma(t - r^\star(z))]$. By Equation 5, $\mathrm{Borda}(y) = f(r^\star(y))$ for every y with $D_x(y) > 0$. Since $\sigma(t - r^\star(z))$ is strictly increasing in t for every z, the function f is strictly increasing. Hence, for all y, y' with $D_x(y), D_x(y') > 0$,

$$\mathrm{Borda}(y) > \mathrm{Borda}(y') \iff f(r^\star(y)) > f(r^\star(y')) \iff r^\star(y) > r^\star(y')\,. \qquad \square$$

The lineage of this result is long and somewhat fragmented. The identity in Equation 5 appears already in the work of Zermelo [62], was independently rediscovered by Bradley and Terry [56] and Ford [63], and was restated by Daniels [64] and Jech [65], all assuming that the pairwise preference probabilities are generated

[3] If ties are allowed, one can replace $P_x(y \succ y')$ by $P_x(y \succ y') + \frac{1}{2}P_x(y \sim y')$. If D_x assigns positive probability to $y' = y$, this convention also treats self-comparisons as ties.

by a BT model. Anderson et al. [61] state the result in the form closest to ours. Siththaranjan et al. [66] restate it in the RLHF setting.

Theorem 1 clarifies what kind of disagreement BT reward learning preserves when it approximates Borda-style aggregation. Intuitively, Borda favors breadth of acceptability rather than depth of support or pairwise dominance [60]. A response can receive a high score because it is consistently acceptable against many candidate responses, even when another response has stronger support in a particular direct comparison. In this sense, Borda selects the consensus or compromise response, the one no annotator subgroup strongly objects to, over a response backed by a narrower majority, and it downranks polarizing responses [67, 68]. The classic critique is that broad mid-strength acceptability can systematically advantage bland responses over distinctive ones that elicit both strong support and strong opposition in the annotator population [69].

After the reward model is learned, post-training converts this score into a change in the model's response distribution, shifting probability mass toward responses with higher Borda scores. How strongly any particular response is upweighted depends on both the Borda score learned from annotators and the base policy π_{base}. To make this concrete, we return to the idealized setting and assume that fine-tuning attains the unparameterized optimum, where the decision variable for each prompt x is the conditional distribution $\pi(\cdot \mid x)$ itself. The maximizer then has the closed-form Boltzmann/Gibbs expression [70, 71]

$$\pi^\star(y \mid x) \;\propto\; \pi_{\text{base}}(y \mid x)\, \exp\big(\beta\, r(y)\big),$$

from which it follows [57, 72] that:

$$\text{Borda}(y) > \text{Borda}(y') \quad\Longleftrightarrow\quad \frac{\pi^\star(y \mid x)}{\pi_{\text{base}}(y \mid x)} > \frac{\pi^\star(y' \mid x)}{\pi_{\text{base}}(y' \mid x)}.$$

This implies that the post-training density ratio against the base policy is, up to a monotone transformation, the π_{base}-weighted Borda score of y.

In practice, many prominent alignment deployments intentionally flatten P_x into binary preference labels by taking the majority vote, disregarding the soft-label information as noise [15, 17, 73–78]. Applied before BT fitting, this preprocessing induces the *Copeland rule* [79], which ranks responses by the number of pairwise majority contests they win [80, 81].

4. CLONE ROBUSTNESS

Section 3 identified BT reward learning as a Borda-like aggregation rule on a fixed candidate set. This connection suggests that BT reward learning may inherit some of Borda's pathologies. A particularly relevant one for language-model reward learning is sensitivity to near-duplicate candidates. In voting theory, this means that adding candidates nearly identical to an existing candidate can change the rule's outcome [82]. Since Borda is well known to be sensitive to such near-duplicate candidates, *robustness to approximate clones* becomes a natural first benchmark for BT reward learning.

Approximate clones arise structurally in generated language [35, 83]. For a fixed prompt, candidate responses are sampled from a generative model whose probability mass often concentrates around a small number of semantic and stylistic modes. A single substantive answer can therefore appear in many forms, with differences that are largely surface-level, such as paraphrasing, reordered explanations, or changes in verbosity. The candidate set $\mathcal{Y}_x$ is only a sampled discretization of this broader response space, and the often arbitrary number of variants associated with a given substantive answer reflects the sampling and filtering process as much as human preference. Exact copies can often be extracted and merged before training, but the harder case involves near-duplicates: responses that differ slightly while occupying the same semantic region and expressing the same underlying answer.

Within a near-clone cluster, pairwise comparisons among variants provide little new signal. Annotators may exhibit minor idiosyncratic preferences among them, but these differences are usually not the underlying quality the reward model is meant to capture. A clone-sensitive BT model can nevertheless treat the number of variants in the cluster as meaningful, so adding near-duplicates can change the rewards assigned to individual candidates and the total policy probability assigned to the cluster as a whole. The learned reward, and by extension the policy it induces, can then depend on arbitrary sampling artifacts, such as how densely different answers are represented in $\mathcal{Y}_x$, rather than on human preference [84]. As models generate more fluent paraphrases and subtle stylistic variants of the same substantive answer, this representational multiplicity can become a larger source of reward variation.

The corresponding formal requirement is *robustness to approximate clones*: the learned reward should be stable under the addition of near-duplicates. Adding a response that is nearly identical to an existing one should make the two responses receive nearly the same reward, and should not substantially change the rewards assigned to unrelated responses. Procaccia et al. [85] formalize this requirement using a metric on responses. Let ρ be a metric on $\mathcal{Y}$, where smaller values indicate greater similarity. Fix a prompt x and a finite candidate set $\mathcal{Y}_x \subseteq \mathcal{Y}$. Let $r_{\mathcal{Y}_x}$ denote the reward learned from comparisons over $\mathcal{Y}_x$, for example by minimizing the unrestricted prompt-level population analogue of Equation 1.

Definition 2. A reward-learning procedure is *robust to approximate clones* if for every $\delta > 0$ there exists $\varepsilon > 0$ such that the following holds. For any finite candidate set $\mathcal{Y}_x \subseteq \mathcal{Y}$, any $y \in \mathcal{Y}_x$, and any new response y' satisfying $\rho(y, y') \le \varepsilon$, the rewards learned before and after adding y' satisfy

$$\left| r_{\mathcal{Y}_x \cup \{y'\}}(y) - r_{\mathcal{Y}_x \cup \{y'\}}(y') \right| \le \delta$$

and

$$\left| r_{\mathcal{Y}_x \cup \{y'\}}(z) - r_{\mathcal{Y}_x}(z) \right| \le \delta \qquad \forall z \in \mathcal{Y}_x.$$

The first condition says that the original response and its approximate clone should receive nearly the same reward once both are present. The second says that adding the clone should not substantially change the rewards assigned to the pre-existing responses. Together, the two conditions require the reward-learning rule to treat near-duplicates as redundant representations of the same local region.

Using the connection to Borda, Procaccia et al. [85] show that standard BT reward learning is not robust to approximate clones. They then propose a weighted version of the population BT objective that addresses this problem. The idea is to assign each response a uniqueness weight $w(y)$, measuring the normalized mass of points in the response space for which y is the nearest candidate. Responses in crowded regions receive smaller weight, while responses representing larger regions receive larger weight. Applying these weights to pairwise comparisons yields the weighted BT objective in the following theorem.

THEOREM 3 [85]. *Let $\mathcal{S} \subseteq \mathbb{R}^d$ be a compact response space with finite positive volume. Fix a prompt x, a finite candidate set $\mathcal{Y}_x \subseteq \mathcal{S}$, and $\lambda > 0$. Draw a point s uniformly at random from $\mathcal{S}$, and assign it to one of the nearest candidates in $\mathcal{Y}_x$, breaking ties uniformly at random; let $w_{\mathcal{Y}_x}(y)$ denote the resulting weight of candidate y. Define the weighted BT loss*

$$\mathcal{L}_\lambda^w(r) := -\sum_{\substack{y,y' \in \mathcal{Y}_x \\ y \neq y'}} w_{\mathcal{Y}_x}(y) w_{\mathcal{Y}_x}(y') P_x(y \succ y') \log \sigma(r(y) - r(y')) + \frac{\lambda}{2} \sum_{y \in \mathcal{Y}_x} w_{\mathcal{Y}_x}(y) r(y)^2 .$$

If $P_x(y \succ y')$ is Lipschitz continuous in the responses, then the reward-learning algorithm that minimizes $\mathcal{L}_\lambda^w(r)$ is robust to approximate clones under the ℓ_2 metric.

Theorem 3 should be read as a representation-invariance guarantee for the prompt-level aggregation step. Once an embedding and a metric are fixed, near-duplicate responses are treated as surface representatives of the same region, so adding another does not substantially change the learned rewards. The guarantee reduces the need for exact deduplication. What it does not remove is the dependence on the embedding, metric, and reference space $\mathcal{S}$ that fix the weights. In this sense, the invariance holds relative to a fixed geometry, shifting the design burden from detecting duplicates to specifying when responses count as close [86].

5. PARAMETRIC REWARD LEARNING AND UNANIMITY

Sections 3 and 4 studied BT reward learning under an unconstrained reward class, where each prompt-response pair has its own free scalar score. In that setting, BT reproduces the Borda score and inherits both its guarantees and its pathologies. In practice, the reward must be implemented by a shared parametric function, so the score assigned to one response is tied to the scores assigned to others through the same hypothesis class. This coupling creates a new source of failures beyond the aggregation effects already discussed.

Formally, the learner chooses parameters θ from a parameter space Θ, which induce a reward function $r_\theta : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. The possible rewards are therefore restricted to the hypothesis class $\mathcal{R}_\Theta = \{r_\theta : \theta \in \Theta\}$. In the standard BT pipeline, θ is fitted on the comparison dataset $\mathcal{D}$ by minimizing the logistic loss in Equation 1, or equivalently by maximizing the BT likelihood.

The simplest failure is a violation of *unanimity*, also known as *Pareto efficiency*. If every annotator strictly prefers response y^+ to response y^-, the learned reward should assign y^+ a higher score. This property is worth studying not because perfect

unanimity is likely to occur in practice, but because violating it means ignoring the strongest possible consensus signal, falling below the baseline set by every standard voting rule in the classical setting.

Definition 4. A reward-learning procedure satisfies *unanimity* if, for any pair of responses $y^+, y^- \in \mathcal{Y}_x$, whenever every annotator prefers y^+ to y^-, the learned reward satisfies $r_\theta(y^+) > r_\theta(y^-)$.

Ge et al. [38] established this failure for a linear hypothesis class; Hollender and Kraiczy [87] showed that it extends to richer parametric reward classes. We follow the linear setting of Ge et al., which admits the cleanest analysis.

In the linear model, each response $y \in \mathcal{Y}_x$ is associated with a known feature vector $\phi(y) \in \mathbb{R}^d$, where d is the embedding dimension, and the reward is linear in those features, $r_\theta(y) \coloneqq \langle \theta, \phi(y) \rangle$ with $\theta \in \mathbb{R}^d$. The learner holds ϕ fixed and fits only θ. This corresponds to one way of building a reward model, in which ϕ is the embedding obtained by removing the final layer of a pretrained language model and the reward is a linear head trained on preference data on top of it. In many settings, the embedding ϕ is itself parameterized and trained alongside the last layer rather than held fixed. The linear model described here is best read as an analytically convenient special case of the restricted reward class $\mathcal{R}_\Theta$.

Throughout this section, we fix a prompt $x \in \mathcal{X}$ and a candidate set $\mathcal{Y}_x \subseteq \mathcal{Y}$. We assume that the learner observes the complete pairwise preferences of a set of annotators $\mathcal{A}$. That is, for every annotator $i \in \mathcal{A}$ and every pair of distinct responses $y, y' \in \mathcal{Y}_x$, the learner observes which of the two responses i prefers.

We further assume that the learner fits a loss function to these observed pairwise preferences. Specifically, the analysis in this section concerns the broader family of loss-based reward-learning rules, of which BT is one member. Given any loss function $\ell : \mathbb{R} \to \mathbb{R}$, define

$$\mathcal{L}(\theta; \ell) \coloneqq \sum_{\substack{y^+, y^- \in \mathcal{Y}_x \\ y^+ \neq y^-}} n_{y^+ \succ y^-} \cdot \ell\left(r_\theta(y^-) - r_\theta(y^+)\right),$$

where $n_{y^+ \succ y^-}$ denotes the number of annotators who prefer y^+ to y^-. When $\ell(z) = \log(1 + \exp(z))$, this is the fixed-prompt version of the BT logistic loss in Equation 1; the corresponding population objective replaces the empirical counts $n_{y^+ \succ y^-}$ by the pairwise preference probabilities $P_x(y^+ \succ y^-)$. Hinge loss gives another instance, and more generally the larger the gap $r_\theta(y^-) - r_\theta(y^+)$ on a comparison the rule disagrees with, the larger the penalty. This class of rules is a natural family that contains BT and matches the loss-minimization framing of modern training pipelines.

In the unrestricted setting of Section 3, BT satisfies unanimity as a consequence of its equivalence with Borda. This makes unanimity appear to be a natural property of BT. Yet, the theorem below shows that this guarantee need not survive once the reward is required to generalize through a restricted reward class $\mathcal{R}_\Theta$.

Theorem 5 [38]. *Let $\ell : \mathbb{R} \to \mathbb{R}$ be a loss function satisfying $\inf_z \ell(z) < \ell(0)$, and*

suppose that ℓ is either nondecreasing and weakly convex, or strictly convex. Then there exist a dimension d, a finite candidate set $\mathcal{Y}_x$, feature vectors $\phi(y) \in \mathbb{R}^d$, and a population for which the induced loss-based linear reward rule fails unanimity. In particular, there are responses $y^+, y^- \in \mathcal{Y}_x$ such that every annotator prefers y^+ to y^-, but the learned reward $r_{\theta^\star}$ ranks them in the opposite order: $r_{\theta^\star}(y^-) > r_{\theta^\star}(y^+)$.

At a high level, the loss must find a single set of parameters that applies to all pairwise comparisons at once. Each comparison places two demands on the fit: assigning the preferred response the higher score, and increasing confidence through a larger margin $r(y^+) - r(y^-)$. In the unrestricted setting, these demands can be satisfied independently, since the relevant reward differences can be adjusted without forcing a change in unrelated comparisons. Under a restricted hypothesis class, they interact because all margins are induced by the same trainable parameters.

The loss then chooses a global fit across all observed comparisons, allocating score separation across comparison directions. Some directions matter more for this fit than others, either because many comparisons point along them or because their feature differences have a larger effect on the loss. A unanimous comparison can still be associated with a weak direction in this global problem. The learned model can then fit the dominant comparison directions while assigning the wrong ordering to the unanimous pair.

The same parametric restriction that lets the reward generalize beyond observed comparisons can also break the guarantees of the underlying aggregation rule. In this case, the rule and its learned approximation come apart even on a property as basic as unanimity.

6. WELFARE LOSS AND SPARSE ELICITATION

One can view alignment from human feedback as a limited-information approximation procedure. For each prompt, the pipeline aims to induce the model behavior that is best for a population of annotators, while observing only a partial description of their preferences. The true objective, however, may depend on unobserved information, such as the preference intensities represented by cardinal values. Here, we examine the loss incurred due to this lack of information, starting with whether there is, in fact, a loss.

Identifiability. Fix a prompt x and a candidate set $\mathcal{Y}_x \subseteq \mathcal{Y}$. Suppose an annotator $i \sim \mathcal{A}$ has a reward function $r_i(\cdot)$ over these responses. A standard goal is to maximize *utilitarian welfare*, defined as the expected reward across annotators. Formally, given a reward profile $\mathbf{r} = (r_i)_{i \in \mathcal{A}}$, the welfare of a response y is

$$W_x^{\mathbf{r}}(y) \coloneqq \mathbb{E}_{i \sim \mathcal{A}}\left[r_i(y)\right].$$

For a policy $\pi(\cdot \mid x) \in \Delta(\mathcal{Y}_x)$, we can similarly define its expected welfare as

$$W_x^{\mathbf{r}}(\pi) \coloneqq \mathbb{E}_{y \sim \pi(\cdot|x)}\left[W_x^{\mathbf{r}}(y)\right].$$

Access to these cardinal utilities would make welfare maximization immediate: we could compute the welfare of each response and train the policy accordingly. In

practice, we observe only comparisons, so the question is whether they identify enough cardinal information to choose a high-welfare response.

Assuming annotators respond according to a link function F, we learn only the aggregate win probabilities:

$$P_x(y \succ y') = \mathbb{E}_{i\sim\mathcal{A}}\left[F\left(r_i(y) - r_i(y')\right)\right].$$

This formulation captures BT responses, where choices are made with probabilities proportional to $\exp(r_i(y))$, as well as deterministic responses, where annotators simply select the higher-reward option. Unfortunately, unless F is linear, it is impossible to reliably identify welfare-maximizing candidates from these probabilities alone, even for a single prompt x with just two responses, y and y'. Because a non-linear link function F distorts the magnitude of reward differences, one can construct two distinct reward profiles, $\mathbf{r}$ and $\mathbf{r}'$, that yield identical expected win probabilities but differ in aggregate welfare. Specifically, there exist $\mathbf{r}$ and $\mathbf{r}'$ such that:

$$\mathbb{E}_{i\sim\mathcal{A}}\left[F\left(r_i(y) - r_i(y')\right)\right] = \mathbb{E}_{i\sim\mathcal{A}}\left[F\left(r'_i(y) - r'_i(y')\right)\right]$$

yet

$$\mathbb{E}_{i\sim\mathcal{A}}\left[r_i(y) - r_i(y')\right] \neq \mathbb{E}_{i\sim\mathcal{A}}\left[r'_i(y) - r'_i(y')\right].$$

Consequently, identical pairwise observation data can arise from populations with fundamentally different preferences in terms of welfare.

Distortion. While *exact* identification of welfare-maximizing candidates is impossible, a learned policy may still achieve approximately high welfare using only the observed comparisons. This guarantee is formalized in social choice through the concept of *distortion* [88, 89]. Recent work applies this lens to quantify how much welfare can be lost when a policy is learned from observed comparisons [90–92].

Assume reward functions are nonnegative and normalized with $0 \le r_i(y) \le 1$ for each annotator i and response y. Let $\mathcal{R}_x(P_x)$ denote the set of normalized reward profiles consistent with the pairwise preference object P_x at prompt x. For a policy $\pi(\cdot \mid x) \in \Delta(\mathcal{Y}_x)$, its utilitarian distortion [88, 93] at x is:

$$\mathrm{Dist}_x(\pi; P_x) \coloneqq \sup_{\mathbf{r}\in\mathcal{R}_x(P_x)} \frac{\max_{\pi'\in\Delta(\mathcal{Y}_x)} W_x^{\mathbf{r}}(\pi')}{W_x^{\mathbf{r}}(\pi)}.$$

Conditional on P_x, this quantity measures the worst-case welfare loss that π can incur over normalized reward profiles that could have induced P_x,[4] where the welfare loss is defined by the ratio above. Since compatibility is defined by the elicited feedback, while the chosen policy is determined by the aggregation rule and policy class, the distortion reflects all three components of the decision-making process together.

Theorem 6 [90]. *Fix a prompt x and a finite response set $\mathcal{Y}_x$ with $|\mathcal{Y}_x| \ge 3$. Suppose annotator comparisons are well-specified by a BT model with inverse-*

[4]When $W_x^{\mathbf{r}}(\pi) = 0$, the corresponding ratio is interpreted as ∞.

temperature parameter η, so that for each annotator $i \in \mathcal{A}$,

$$\Pr[y \succ_i y' \mid x] = \sigma\left(\eta \cdot (r_i(y) - r_i(y'))\right).$$

In other words, the link function is $F(t) \coloneqq \sigma(\eta \cdot t)$. For each pairwise preference object P_x, let $\pi_0(P_x)$ be the policy that BT reward learning returns once the KL penalty is removed; by Theorem 1, it places all mass on the Borda winner under P_x. Then

$$(1 - o(1))\,\eta \;\le\; \sup_{P_x} \mathrm{Dist}_x\left(\pi_0(P_x); P_x\right) \;\le\; O(\eta^2),$$

where P_x ranges over pairwise preference objects generated by a normalized reward profile $\mathbf{r}$ and a comparison-pair sampling distribution. The upper bound holds for every such P_x; the lower bound is the rate as $\eta \to \infty$.

The lower bound should be read against the minimax lower bound for the same information model. Gölz et al. [90] show that, when each annotator contributes a single comparison, every rule mapping the comparisons to a policy incurs distortion at least $\left(\frac{1}{2} + o(1)\right)\eta$ on some reward profile. Strikingly, this lower bound is achieved by a framework known as Nash learning from human feedback, which we define and analyze in Section 7 (see Theorem 7).

Richer Elicitation. Theorem 6 demonstrates that approximating welfare is possible, though imperfect, at least under BT responses. However, utilitarian welfare is not the only valid objective in social choice. Consider two hypothetical responses y and y': y provides a utility of 2 to all annotators, while y' provides a utility of 3 to half the annotators and 1 to the other half. Utilitarian welfare treats these outcomes as equivalent. By contrast, more egalitarian choices, such as *Nash welfare* — the product of utilities — would strictly prefer y, favoring broad approval over polarizing outcomes that benefit one group at the expense of another. How, then, can we optimize for these alternative choices?

Chidambaram et al. [94] and Ge et al. [95] consider linear social choice models (similar to the one described in Section 5) where responses follow BT and deterministic link functions, respectively. Both show that even in these restricted settings, a single pairwise comparison per annotator cannot reveal which of the two responses a more egalitarian objective should prefer. Without preference intensities, a single comparison cannot separate a universally indifferent population from one that is evenly split between strong opposing preferences; in both, each candidate is preferred equally often across the population. Thus, there is no hope of optimizing more egalitarian welfare functions that would strictly prefer broad-appeal candidates. On the other hand, under mild structural conditions, eliciting two comparisons per annotator yields enough information to completely identify the voter type distribution. This identification enables direct optimization for *any* desired social welfare function.

Similarly, Cherapanamjeri et al. [96] demonstrate in a latent-utility model that moving from pairwise choices to best-of-three queries provides essentially complete identifiability under appropriate structural conditions.

Richer information can also come from passively recorded signals. For example, the

time taken to provide a comparison label can carry information about preference intensities that binary labels discard [97, 98].

While these models rely on different structural assumptions, each points to the same fact that slightly richer elicitation can, in principle, make far more welfare objectives optimizable than pairwise comparison allows.

7. DIRECT ALIGNMENT FROM PAIRWISE PREFERENCES

Earlier sections studied alignment rules that infer a scalar reward from pairwise preference data and then optimize a policy against that reward. This scalar reward projects pairwise comparisons among responses onto one ordered axis, which forces the aggregate relation into a transitive ranking. This section asks what is lost when the aggregate relation is instead intransitive, and how alignment objectives can keep the pairwise structure directly.

Such intransitivity can arise even when every annotator is internally consistent. Averaging their preferences may produce a *Condorcet cycle*, where y_a is majority-preferred to y_b, y_b to y_c, and yet y_c to y_a [60]. In such cases there is no *Condorcet winner*, meaning no response defeats every other response by majority comparison. Any deterministic target must therefore break the cycle somewhere, motivating objectives that reason over the pairwise relation itself.

Maximal Lotteries. One way to avoid arbitrarily breaking cycles is to select a distribution over responses rather than a single response. This shifts the question from which response should win to which randomized target should be chosen. *Maximal lotteries* answer this by comparing each candidate distribution against alternatives using the population majority margin: draw one response from each distribution and measure which the population prefers on average. This solution concept was introduced by Kreweras [99], developed systematically by Fishburn [100],[5] and later applied in many settings, including AI evaluation [103, 104].

Fix a prompt x and a finite candidate set $\mathcal{Y}_x \subseteq \mathcal{Y}$. Let

$$M_x(y, y') := P_x(y \succ y') - P_x(y' \succ y)$$

denote the *majority margin* — the net pairwise preference for y over y' in the population. A maximal lottery is a distribution $p^\star \in \Delta(\mathcal{Y}_x)$ whose expected margin against every response distribution $q \in \Delta(\mathcal{Y}_x)$ is nonnegative:

$$\mathbb{E}_{y\sim p^\star, y'\sim q}[M_x(y, y')] \geq 0.$$

Since M_x is skew-symmetric, the maximal lotteries are exactly the maximin strategies of the zero-sum game with payoff matrix M_x:[6]

$$p^\star \in \arg \max_{p\in\Delta(\mathcal{Y}_x)} \min_{q\in\Delta(\mathcal{Y}_x)} p^\top M_x q = \arg \max_{p\in\Delta(\mathcal{Y}_x)} \min_{q\in\Delta(\mathcal{Y}_x)} \mathbb{E}_{y\sim p,\, y'\sim q}\left[M_x(y, y')\right].$$

[5]The same solution has been rediscovered independently under other names, including the "game theory method" in voting [101] and the von Neumann winner in contextual dueling bandits [102].
[6]Existence follows from von Neumann's minimax theorem [105]. For a generic majority-margin matrix, the maximal lottery is unique, although degenerate ties can yield multiple maximal lotteries.

When the majority relation is cyclic, the maximal lottery spreads mass across several responses that compete in the cycle, so the target itself records the unresolved disagreement. When a strict Condorcet winner exists, no mixing is needed and the unique maximal lottery places all mass on that response. Maximal lotteries also satisfy other desirable axiomatic properties [106]. In particular, they are invariant to *exact* clones (duplicate responses). However, the introduction of *approximate* clones — as defined in Section 4 — can affect the lottery, even if the approximation is arbitrarily precise.

Nash Learning from Human Feedback. *Nash Learning from Human Feedback (NLHF)* [40, 107] applies the same maximin idea directly at the level of policies. Instead of compressing pairwise comparisons into a scalar reward, NLHF uses the pairwise preference model as the payoff of a two-player game between policies and trains toward the equilibrium of that game.

The population preference P_x induces a preference between two policies by comparing the responses they generate,

$$P_x(\pi \succ \pi') \coloneqq \mathbb{E}_{y\sim\pi(\cdot|x),\, y'\sim\pi'(\cdot|x)}[P_x(y \succ y')]\,.$$

Thus, $P_x(\pi \succ \pi')$ is the probability that π produces a response preferred to one produced by π', at prompt x. We write

$$P(\pi \succ \pi') \coloneqq \mathbb{E}_{x\sim D}[P_x(\pi \succ \pi')]$$

for the corresponding comparison after averaging over prompts drawn from D. This induces a preference relation over policies

$$\pi \succ \pi' \quad \Longleftrightarrow \quad P(\pi \succ \pi') > \frac{1}{2}.$$

NLHF chooses a policy with the best worst-case comparison probability against alternative policies:

$$\pi^\star \in \arg\max_{\pi} \min_{\pi'} P(\pi \succ \pi').$$

For a fixed prompt with finite candidate support and no regularization, $\pi^\star$ coincides with the maximal lottery of that prompt's margin game. In this setting, the maximal lottery can be computed by the linear program for the row player's maximin strategy. In the open-ended parametric setting, NLHF instead approaches the corresponding policy-level equilibrium through self-play, optimizing a policy against opponents generated from its own iterates. This equilibrium-seeking template underlies several recent preference-optimization methods [108–113].

Welfare Interpretation. The welfare-loss perspective of Section 6 gives a precise sense in which maximal lotteries are optimal. The following result bounds their worst-case utilitarian distortion under anonymous pairwise feedback.

Theorem 7 [90]. *Suppose annotator comparisons are generated from normalized rewards by the BT model in Theorem 6, with inverse-temperature η. Let $\pi_x^{\mathrm{NLHF}}(P_x)$*

be any maximal lottery of the margin game M_x. Then

$$\sup_{P_x} \mathrm{Dist}_x\left(\pi_x^{\mathrm{NLHF}}(P_x); P_x\right) \le \left(\frac{1}{2} + o(1)\right)\eta,$$

where P_x ranges over pairwise preference objects generated by normalized reward profiles and comparison-pair sampling distributions, and the $o(1)$ term is as $\eta \to \infty$.

Together with the minimax lower bound of Section 6, Theorem 7 shows that the maximal lottery attains the smallest possible worst-case distortion, up to lower-order terms, among rules observing anonymous BT pairwise feedback.

Regularization. As with RLHF, deployed NLHF regularizes the learned policy toward a reference policy π_{base} with a KL penalty. The regularized preference between policies is

$$P_\tau(\pi \succ \pi') \;:=\; P(\pi \succ \pi') - \tau\,\mathrm{KL}_D(\pi \,\|\, \pi_{\mathrm{base}}) + \tau\,\mathrm{KL}_D(\pi' \,\|\, \pi_{\mathrm{base}}),$$

where the KL terms penalize each policy for diverging from π_{base}, symmetrically across the two players.[7] The training target remains the Nash equilibrium of this modified game, which exists and is unique [107].

Theorem 7's optimality extends to this KL-regularized objective. Since the penalty keeps the learned policy near π_{base}, the relevant comparison is with policies that are allowed to move the same distance from the reference. That is, the regularized equilibrium is compared with the highest welfare policy whose KL divergence from π_{base} is no larger than its own. Against this benchmark, NLHF retains the same optimal distortion guarantee, for any reference policy and any regularization strength [90].

8. DISCUSSION

Taken together, the sections above give a social-choice account of alignment from human feedback as a sequence of design choices. Human judgments must be elicited, aggregated into a model, generalized beyond the observed data, and translated into policy behavior. We conclude by discussing several related research threads that extend this pipeline view beyond the formal results surveyed earlier.

Preference Collapse. Fine-tuning a base policy toward a learned reward signal can narrow the policy's behavior even when the reward itself supports a broader range of responses. With stronger optimization, probability mass can concentrate around a small set of reward-favored outputs, producing *preference collapse* [114], where majority views are further amplified [115, 116] and response diversity is reduced [117–119]. This is one way in which the downstream policy can sharpen the aggregation choices made by the learned reward: biases in the preference data, once encoded in the reward signal, may be amplified by subsequent optimization [120, 121]. As a result,

[7] Viewing each player's payoff separately, a policy receives its preference payoff minus a KL penalty for its own divergence from π_{base}: $R(\pi;\pi') = P(\pi \succ \pi') - \tau\,\mathrm{KL}_D(\pi \,\|\, \pi_{\mathrm{base}})$, with the symmetric expression for π'.

the learned policy may fail to represent the full range of normative considerations expressed across populations, domains, and interaction contexts [122, 123].

Multiple Policies. A growing number of proposals advocate for *pluralistic alignment*: modeling multiple perspectives in parallel so as to better capture the breadth of human judgments [21, 30, 124, 125]. Work in this direction differs in where it relaxes the standard single-reward pipeline. Some approaches retain a single learned reward while limiting how aggressively the policy optimizes against it, reducing the policy-level amplification of the aggregate objective [121, 126]. Others enrich the reward model so that it can encode preference heterogeneity more directly [127–129].

A more direct approach represents divergent preferences by training multiple reward models, each matched to a distinct segment of the population [124, 130–132]. The typical approach clusters annotators, by demographics or inferred preference patterns, and fits a group-specific reward to each cluster. Halpern et al. [22] avoid this intermediate clustering step by fitting a compact reward ensemble whose aggregate pairwise choices match the observed preference P_x, so each component induces a coherent policy while the mixture preserves disagreement.

Once such an ensemble is learned, its components can be deployed in several ways to serve pluralistic goals [21]. For example, the system can present multiple outputs as an Overton-style slate [133], distill them into a consensus statement [134–136], select the policy best matched to a user's stated preference, or sample from the policy mixture. Over repeated use, mixture sampling preserves population-level diversity and counters the tendency of aligned models to recycle a small set of high-reward responses [116, 118, 137], which can be particularly valuable in creative workflows.

Personalization. At a finer level of granularity, the alignment objective can be personalized to individual annotators, treating each user as carrying a unique reward function [138]. Since fitting a separate model per user is infeasible under sparse per-user data, these methods place each user's reward in a shared low-dimensional space, recovering a per-user latent variable [139] or a weighting over common reward components [140, 141] from a handful of preferences. Lightweight user embeddings or low-rank adapters then specialize a shared model to the individual. Even so, personalizing to each user risks narrowing their information diet and reinforcing prior views, prompting calls to bound how far personalization should extend [123, 142].

AI for Social Choice and Democracy. The work surveyed in this paper is primarily *social choice for AI*: it uses social choice theory to diagnose how alignment pipelines aggregate heterogeneous feedback, and to design learned rewards or policies that handle disagreement more explicitly. A complementary agenda runs in the reverse direction, asking how AI systems can support collective decision-making, deliberation, and democratic representation. One direction uses AI to help groups deliberate and converge on shared positions, through *generative social choice* [135, 136], deliberation mediators such as the *Habermas machine* [143], formal accounts of common ground [134, 144], and platforms for large-scale opinion aggregation [145]. More broadly, progress in social choice for AI and AI for social choice may prove mutually reinforcing, as advances in one direction generate both the concepts and the tools needed to advance the other.

REFERENCES

[1] Jan Leike, David Krueger, Tom Everitt, Miljan Martic, Vishal Maini, and Shane Legg. Scalable Agent Alignment via Reward Modeling: A Research Direction, 2018. arXiv:1811.07871. 47

[2] Jiaming Ji, Tianyi Qiu, Boyuan Chen, Jiayi Zhou, Borong Zhang, Donghai Hong, Hantao Lou, Kaile Wang, Yawen Duan, Zhonghao He, Lukas Vierling, Zhaowei Zhang, Fanzhi Zeng, Juntao Dai, Xuehai Pan, Hua Xu, Aidan O'Gara, Kwan Ng, Brian Tse, Jie Fu, Stephen McAleer, Yanfeng Wang, Mingchuan Yang, Yunhuai Liu, Yizhou Wang, Song-Chun Zhu, Yike Guo, Yaodong Yang, and Wen Gao. AI Alignment: A Contemporary Survey. *ACM Computing Surveys*, 58(5):132:1–132:38, 2025. 47

[3] Iason Gabriel. Artificial Intelligence, Values, and Alignment. *Minds and Machines*, 30(3):411–437, 2020. 47, 49

[4] Stuart J. Russell. *Human Compatible: Artificial Intelligence and the Problem of Control*. Allen Lane, 2019. 47

[5] Jessica Taylor, Eliezer Yudkowsky, Patrick LaVictoire, and Andrew Critch. Alignment for Advanced Machine Learning Systems. In S. Matthew Liao, editor, *Ethics of Artificial Intelligence*, pages 342–382. Oxford University Press, 2020. 47

[6] Zachary Kenton, Tom Everitt, Laura Weidinger, Iason Gabriel, Vladimir Mikulik, and Geoffrey Irving. Alignment of Language Agents, 2021. arXiv:2103.14659. 47

[7] Iason Gabriel and Geoff Keeling. A Matter of Principle? AI Alignment as the Fair Treatment of Claims. *Philosophical Studies*, 182(7):1951–1973, 2025. 47

[8] Dan Hendrycks, Nicholas Carlini, John Schulman, and Jacob Steinhardt. Unsolved Problems in ML Safety, 2021. arXiv:2109.13916. 47

[9] Stephen Casper, Xander Davies, Claudia Shi, Thomas Krendl Gilbert, Jérémy Scheurer, Javier Rando, Rachel Freedman, Tomek Korbak, David Lindner, Pedro Freire, Tony Tong Wang, Samuel Marks, Charbel-Raphael Segerie, Micah Carroll, Andi Peng, Phillip J. K. Christoffersen, Mehul Damani, Stewart Slocum, Usman Anwar, Anand Siththaranjan, Max Nadeau, Eric J. Michaud, Jacob Pfau, Dmitrii Krasheninnikov, Xin Chen, Lauro Langosco, Peter Hase, Erdem Biyik, Anca Dragan, David Krueger, Dorsa Sadigh, and Dylan Hadfield-Menell. Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback. *Transactions on Machine Learning Research*, 2023. 47

[10] Andrew P. Clark, Kate L. Howard, Andy T. Woods, Ian S. Penton-Voak, and Christof Neumann. Why Rate When You Could Compare? Using the "EloChoice" Package to Assess Pairwise Comparisons of Perceived Physical Strength. *PLOS ONE*, 13(1):e0190393, 2018. 47

[11] Yisong Yue, Josef Broder, Robert Kleinberg, and Thorsten Joachims. The k-Armed Dueling Bandits Problem. *Journal of Computer and System Sciences*, 78(5):1538–1556, 2012. 47

[12] Paul F. Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep Reinforcement Learning from Human Preferences. In

The Thirty-first Annual Conference on Neural Information Processing Systems, 2017. 47, 50

[13] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training Language Models to Follow Instructions with Human Feedback. In *The Thirty-sixth Annual Conference on Neural Information Processing Systems*, 2022. 47, 48, 50

[14] Daniel M. Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B. Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. Fine-Tuning Language Models from Human Preferences, 2020. arXiv:1909.08593. 47, 50, 51

[15] Nisan Stiennon, Long Ouyang, Jeff Wu, Daniel M. Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul Christiano. Learning to Summarize with Human Feedback. In *The Thirty-fourth Annual Conference on Neural Information Processing Systems*, 2020. 47, 48, 50, 53

[16] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. Direct Preference Optimization: Your Language Model Is Secretly a Reward Model. In *The Thirty-seventh Annual Conference on Neural Information Processing Systems*, 2023. 47, 48

[17] Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, Carol Chen, Catherine Olsson, Christopher Olah, Danny Hernandez, Dawn Drain, Deep Ganguli, Dustin Li, Eli Tran-Johnson, Ethan Perez, Jamie Kerr, Jared Mueller, Jeffrey Ladish, Joshua Landau, Kamal Ndousse, Kamile Lukosuite, Liane Lovitt, Michael Sellitto, Nelson Elhage, Nicholas Schiefer, Noemi Mercado, Nova DasSarma, Robert Lasenby, Robin Larson, Sam Ringer, Scott Johnston, Shauna Kravec, Sheer El Showk, Stanislav Fort, Tamera Lanham, Timothy Telleen-Lawton, Tom Conerly, Tom Henighan, Tristan Hume, Samuel R. Bowman, Zac Hatfield-Dodds, Ben Mann, Dario Amodei, Nicholas Joseph, Sam McCandlish, Tom Brown, and Jared Kaplan. Constitutional AI: Harmlessness from AI Feedback, 2022. arXiv:2212.08073. 48, 53

[18] Saffron Huang, Divya Siddarth, Liane Lovitt, Thomas I. Liao, Esin Durmus, Alex Tamkin, and Deep Ganguli. Collective Constitutional AI: Aligning a Language Model with Public Input. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 2024. 48

[19] Maarten Buyl, Hadi Khalaf, Claudio Mayrink Verdun, Lucas Monteiro Paes, Caio Cesar Vieira Machado, and Flavio du Pin Calmon. AI Alignment at Your Discretion. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 2025. 48

[20] Abigail Z. Jacobs and Hanna Wallach. Measurement and Fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021. 48

[21] Taylor Sorensen, Jared Moore, Jillian Fisher, Mitchell Gordon, Niloofar Mireshghallah, Christopher Michael Rytting, Andre Ye, Liwei Jiang, Ximing Lu, Nouha Dziri, Tim Althoff, and Yejin Choi. Position: A Roadmap to Pluralistic Alignment. In *Proceedings of the 41st International Conference on Machine Learning*, 2024. 48, 63

[22] Daniel Halpern, Evi Micha, Ariel D. Procaccia, and Itai Shapira. Pairwise Calibrated Rewards for Pluralistic Alignment. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. 48, 63

[23] Tan Zhi-Xuan, Micah Carroll, Matija Franklin, and Hal Ashton. Beyond Preferences in AI Alignment. *Philosophical Studies*, 182(7):1813–1863, 2025. 48

[24] Taylor Sorensen, Liwei Jiang, Jena D. Hwang, Sydney Levine, Valentina Pyatkin, Peter West, Nouha Dziri, Ximing Lu, Kavel Rao, Chandra Bhagavatula, Maarten Sap, John Tasioulas, and Yejin Choi. Value Kaleidoscope: Engaging AI with Pluralistic Human Values, Rights, and Duties. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, 2024. 48

[25] Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, Nicholas Joseph, Saurav Kadavath, Jackson Kernion, Tom Conerly, Sheer El-Showk, Nelson Elhage, Zac Hatfield-Dodds, Danny Hernandez, Tristan Hume, Scott Johnston, Shauna Kravec, Liane Lovitt, Neel Nanda, Catherine Olsson, Dario Amodei, Tom Brown, Jack Clark, Sam McCandlish, Chris Olah, Ben Mann, and Jared Kaplan. Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback, 2022. arXiv:2204.05862. 48

[26] Jiahao Yuan, Zixiang Di, Shangzixin Zhao, Zhiqing Cui, Hanqing Wang, Guisong Yang, and Usman Naseem. Cultural Palette: Pluralising Culture Alignment via Multi-Agent Palette, 2024. arXiv:2412.11167. 48

[27] Vinodkumar Prabhakaran, Rida Qadri, and Ben Hutchinson. Cultural Incongruencies in Artificial Intelligence. In *First Workshop on Cultures in AI/AI in Culture, NeurIPS 2022*, 2022. 48

[28] Usman Anwar, Abulhair Saparov, Javier Rando, Daniel Paleka, Miles Turpin, Peter Hase, Ekdeep Singh Lubana, Erik Jenner, Stephen Casper, Oliver Sourbut, Benjamin L. Edelman, Zhaowei Zhang, Mario Günther, Anton Korinek, Jose Hernandez-Orallo, Lewis Hammond, Eric J. Bigelow, Alexander Pan, Lauro Langosco, Tomasz Korbak, Heidi Chenyu Zhang, Ruiqi Zhong, Seán Ó hÉigeartaigh, Gabriel Recchia, Giulio Corsi, Alan Chan, Markus Anderljung, Lilian Edwards, Aleksandar Petrov, Christian Schroeder de Witt, Sumeet Ramesh Motwani, Yoshua Bengio, Danqi Chen, Philip Torr, Samuel Albanie, Tegan Maharaj, Jakob Nicolaus Foerster, Florian Tramèr, He He, Atoosa Kasirzadeh, Yejin Choi, and David Krueger. Foundational Challenges in Assuring Alignment and Safety of Large Language Models. *Transactions on Machine Learning Research*, 2024. 48

[29] Hannah Rose Kirk, Alexander Whitefield, Paul Röttger, Andrew Michael Bean, Katerina Margatina, Rafael Mosquera, Juan Manuel Ciro, Max Bartolo, Adina Williams, He He, Bertie Vidgen, and Scott A. Hale. The PRISM Alignment Dataset: What Participatory, Representative and Individualised Human Feedback Reveals About the Subjective and Multicultural Alignment of Large Language Models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2024. 48

[30] Shangbin Feng, Taylor Sorensen, Yuhan Liu, Jillian Fisher, Chan Young Park, Yejin Choi, and Yulia Tsvetkov. Modular Pluralism: Pluralistic Alignment via Multi-LLM Collaboration. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 2024. 48, 63

[31] Michael J.Q. Zhang, Zhilin Wang, Jena D. Hwang, Yi Dong, Olivier Delalleau, Yejin Choi, Eunsol Choi, Xiang Ren, and Valentina Pyatkin. Diverging Preferences: When Do Annotators Disagree and Do Models Know? In *Proceedings of the 42nd International Conference on Machine Learning*, 2025. 48

[32] Kanad Shrikar Pardeshi, Itai Shapira, Ariel D. Procaccia, and Aarti Singh. Learning Social Welfare Functions. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 48

[33] Amartya Sen. The Possibility of Social Choice. *American Economic Review*, 89(3):349–378, 1999. 48

[34] Amartya Sen. Social Choice Theory. In Kenneth J. Arrow and Michael D. Intriligator, editors, *Handbook of Mathematical Economics*, volume 3, pages 1073–1181. Elsevier, 1986. 48

[35] Vincent Conitzer, Rachel Freedman, Jobst Heitzig, Wesley H. Holliday, Bob M. Jacobs, Nathan Lambert, Milan Mossé, Eric Pacuit, Stuart Russell, Hailey Schoelkopf, Emanuel Tewolde, and William S. Zwicker. Position: Social Choice Should Guide AI Alignment in Dealing with Diverse Human Feedback. In *Proceedings of the 41st International Conference on Machine Learning*, 2024. 48, 54

[36] Jessica Dai and Eve Fleisig. Mapping Social Choice Theory to RLHF. In *ICLR 2024 Workshop on Reliable and Responsible Foundation Models*, 2024. 48

[37] Abhilash Mishra. AI Alignment and Social Choice: Fundamental Limitations and Policy Implications, 2023. arXiv:2310.16048. 48

[38] Luise Ge, Daniel Halpern, Evi Micha, Ariel D. Procaccia, Itai Shapira, Yevgeniy Vorobeychik, and Junlin Wu. Axioms for AI Alignment from Human Feedback. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 48, 56

[39] Parand A. Alamdari, Soroush Ebadian, and Ariel D. Procaccia. Policy Aggregation. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 48

[40] Roberto-Rafael Maura-Rivero, Marc Lanctot, Francesco Visin, and Kate Larson. Jackpot! Alignment as a Maximal Lottery, 2025. arXiv:2501.19266. 48, 61

[41] Vincent Conitzer, Walter Sinnott-Armstrong, Jana Schaich Borg, Yuan Deng, and Max Kramer. Moral Decision Making Frameworks for Artificial Intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1), 2017. 48

[42] Mahendra Prasad. Social Choice and the Value Alignment Problem. In *Artificial Intelligence Safety and Security*. Chapman and Hall/CRC, 2018. 48

[43] Leo Gao, John Schulman, and Jacob Hilton. Scaling Laws for Reward Model Overoptimization. In *Proceedings of the 40th International Conference on Machine Learning*, 2023. 49

[44] Xueru Wen, Jie Lou, Yaojie Lu, Hongyu Lin, Xing Yu, Xinyu Lu, Ben He, Xianpei Han, Debing Zhang, and Le Sun. Rethinking Reward Model Evaluation: Are We Barking Up the Wrong Tree? In *The Thirteenth International Conference on Learning Representations*, 2025. 49

[45] Evan Frick, Tianle Li, Connor Chen, Wei-Lin Chiang, Anastasios Nikolas Angelopoulos, Jiantao Jiao, Banghua Zhu, Joseph E. Gonzalez, and Ion Stoica. How to Evaluate Reward Models for RLHF. In *The Thirteenth International Conference on Learning Representations*, 2025. 49

[46] Iason Gabriel and Vafa Ghazavi. The Challenge of Value Alignment: From Fairer Algorithms to AI Safety. In Carissa Véliz, editor, *Oxford Handbook of Digital Ethics*, pages 336–355. Oxford University Press, 2023. 49

[47] Seth D. Baum. Social Choice Ethics in Artificial Intelligence. *AI & Society*, 35(1):165–176, 2020. 49

[48] Timo Kaufmann, Paul Weng, Viktor Bengs, and Eyke Hüllermeier. A Survey of Reinforcement Learning from Human Feedback. *Transactions on Machine Learning Research*, 2025. 50

[49] Banghua Zhu, Michael Jordan, and Jiantao Jiao. Principled Reinforcement Learning with Human Feedback from Pairwise or K-Wise Comparisons. In *Proceedings of the 40th International Conference on Machine Learning*, 2023. 50

[50] Peter C. Fishburn. Condorcet Social Choice Functions. *SIAM Journal on Applied Mathematics*, 33(3):469–489, 1977. 50

[51] Louis L. Thurstone. A Law of Comparative Judgment. *Psychological Review*, 34(4):273–286, 1927. 50

[52] R. Duncan Luce. *Individual Choice Behavior*. John Wiley, Oxford, England, 1959. 50

[53] Daniel McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*, pages 105–142. Academic Press, New York, 1974. 50

[54] Ritesh Noothigattu, Dominik Peters, and Ariel D. Procaccia. Axioms for Learning from Pairwise Comparisons. In *The Thirty-fourth Annual Conference on Neural Information Processing Systems*, 2020. 50

[55] W. Bradley Knox, Stephane Hatgis-Kessell, Serena Booth, Scott Niekum, Peter Stone, and Alessandro G. Allievi. Models of Human Preference for Learning Reward Functions. *Transactions on Machine Learning Research*, 2023. 50

[56] Ralph Allan Bradley and Milton E. Terry. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, 39(3/4):324–345, 1952. 50, 52

[57] Sang T. Truong, Andreas Haupt, and Sanmi Koyejo. *Machine Learning from Human Preferences*. Stanford University, 2025. 50, 53

[58] John I. Yellott. The Relationship Between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the Double Exponential Distribution. *Journal of Mathematical Psychology*, 15(2):109–144, 1977. 50

[59] Jean-Charles de Borda. Mémoire sur les élections au scrutin. In *Mémoires de l'Académie Royale des Sciences année 1781*, pages 657–665. l'Imprimerie Royale, Paris, 1784. 51

[60] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia. *Handbook of Computational Social Choice*. Cambridge University Press, 2016. 51, 53, 60

[61] Lowell Bruce Anderson, Helena Dandurova, James E. Falk, and Lana Yeganova. Relationships Between Borda Voting and Zermelo Ranking. *Social Choice and Welfare*, 32(3):355–365, 2009. 52, 53

[62] Ernst Zermelo. Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 29:436–460, 1929. 52

[63] Lester R. Ford. Solution of a Ranking Problem from Binary Comparisons. *The American Mathematical Monthly*, 64(8):28–33, 1957. 52

[64] Henry E. Daniels. Round-Robin Tournament Scores. *Biometrika*, 56(2): 295–299, 1969. 52

[65] Thomas Jech. The Ranking of Incomplete Tournaments: A Mathematician's Guide to Popular Sports. *The American Mathematical Monthly*, 90(4):246–266, 1983. 52

[66] Anand Siththaranjan, Cassidy Laidlaw, and Dylan Hadfield-Menell. Distributional Preference Learning: Understanding and Accounting for Hidden Context in RLHF. In *The Twelfth International Conference on Learning Representations*, 2024. 53

[67] Eric Maskin. Borda's Rule and Arrow's Independence Condition. *Journal of Political Economy*, 133(2):385–420, 2025. 53

[68] Benjamin Reilly. Social Choice in the South Seas: Electoral Innovation and the Borda Count in the Pacific Island Countries. *International Political Science Review*, 23(4):355–372, 2002. 53

[69] Michel L. Balinski and Rida Laraki. *Majority Judgment: Measuring, Ranking, and Electing*. MIT Press, 2010. 53

[70] Emanuel Todorov. Linearly-Solvable Markov Decision Problems. In *The Twentieth Annual Conference on Neural Information Processing Systems*, 2006. 53

[71] Jan Peters, Katharina Mülling, and Yasemin Altun. Relative Entropy Policy Search. *Proceedings of the AAAI Conference on Artificial Intelligence*, 24(1): 1607–1612, 2010. 53

[72] Ali Shirali, Arash Nasr-Esfahany, Abdullah Omar Alomar, Parsa Mirtaheri, Rediet Abebe, and Ariel D. Procaccia. Direct Alignment with Heterogeneous Preferences. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. 53

[73] Hannah Rose Kirk, Andrew M. Bean, Bertie Vidgen, Paul Röttger, and Scott A. Hale. The Past, Present and Better Future of Feedback Learning in Large Language Models for Subjective Human Preferences and Values. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023. 53

[74] Angelica Chen, Sadhika Malladi, Lily H. Zhang, Xinyi Chen, Qiuyi Zhang, Rajesh Ranganath, and Kyunghyun Cho. Preference Learning Algorithms Do Not Learn Preference Rankings. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 53

[75] Kyuyoung Kim, Ah Jeong Seo, Hao Liu, Jinwoo Shin, and Kimin Lee. Margin Matching Preference Optimization: Enhanced Model Alignment with Granular Feedback. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, 2024. 53

[76] Nathan Lambert, Thomas Krendl Gilbert, and Tom Zick. The History and Risks of Reinforcement Learning and Human Feedback, 2023. arXiv:2310.13595. 53

[77] Vinodkumar Prabhakaran, Aida Mostafazadeh Davani, and Mark Diaz. On Releasing Annotator-Level Labels and Information in Datasets. In *Proceedings of the Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop*, 2021. 53

[78] Deep Ganguli, Liane Lovitt, Jackson Kernion, Amanda Askell, Yuntao Bai, Saurav Kadavath, Ben Mann, Ethan Perez, Nicholas Schiefer, Kamal Ndousse, Andy Jones, Sam Bowman, Anna Chen, Tom Conerly, Nova DasSarma, Dawn Drain, Nelson Elhage, Sheer El-Showk, Stanislav Fort, Zac Hatfield-Dodds, Tom Henighan, Danny Hernandez, Tristan Hume, Josh Jacobson, Scott Johnston, Shauna Kravec, Catherine Olsson, Sam Ringer, Eli Tran-Johnson, Dario Amodei, Tom Brown, Nicholas Joseph, Sam McCandlish, Chris Olah, Jared Kaplan, and Jack Clark. Red Teaming Language Models to Reduce Harms: Methods, Scaling Behaviors, and Lessons Learned, 2022. arXiv:2209.07858. 53

[79] Arthur H. Copeland. A Reasonable Social Welfare Function. Mimeographed notes from the Seminar on Applications of Mathematics to the Social Sciences, University of Michigan, Ann Arbor, 1951. 53

[80] Jiancong Xiao, Zhekun Shi, Kaizhao Liu, Qi Long, and Weijie J. Su. Theoretical Tensions in RLHF: Reconciling Empirical Success with Inconsistencies in Social Choice Theory, 2025. arXiv:2506.12350. 53

[81] Zhiyu An, Duaa Nakshbandi, and Wan Du. Differential Voting: Loss Functions for Axiomatically Diverse Aggregation of Heterogeneous Preferences, 2026. arXiv:2601.18824. 53

[82] T. Nicolaus Tideman. Independence of Clones as a Criterion for Voting Rules. *Social Choice and Welfare*, 4(3):185–206, 1987. 53

[83] Ratip Emin Berker, Sílvia Casacuberta Puig, Isaac Robinson, and Christopher Ong. Obvious Independence of Clones. In *Artificial Intelligence for Research and Democracy*, 2024. 54

[84] Damien Berriaud and Roger Wattenhofer. Clone-Robust Weights in Metric Spaces: Handling Redundancy Bias for Benchmark Aggregation. In *Proceedings of the 25th International Conference on Autonomous Agents and Multiagent Systems*, 2026. 54

[85] Ariel D. Procaccia, Benjamin Schiffer, and Shirley Zhang. Clone-Robust AI Alignment. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025. 54, 55

[86] Ameet Deshpande, Carlos Jimenez, Howard Chen, Vishvak Murahari, Victoria Graf, Tanmay Rajpurohit, Ashwin Kalyan, Danqi Chen, and Karthik Narasimhan. C-STS: Conditional Semantic Textual Similarity. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023. 55

[87] Alexandros Hollender and Sonja Kraiczy. Enforcing Axioms for AI Alignment Under Loss-Based Rules. In *The Fourteenth International Conference on Learning Representations*, 2026. 56

[88] Ariel D. Procaccia and Jeffrey S. Rosenschein. The Distortion of Cardinal Preferences in Voting. In *Cooperative Information Agents X*, 2006. 58

[89] Craig Boutilier, Ioannis Caragiannis, Simi Haber, Tyler Lu, Ariel D. Procaccia, and Or Sheffet. Optimal Social Choice Functions: A Utilitarian View. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, 2012. 58

[90] Paul Gölz, Nika Haghtalab, and Kunhe Yang. Distortion of AI Alignment: Does Preference Optimization Optimize for Preferences? In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. 58, 59, 61, 62

[91] Kazusato Oko, Annie S. Ulichney, Nika Haghtalab, and Han Bao. Distortion of AI Alignment Revisited: RLHF Is a Decent Utilitarian Aligner. In *Algorithmic Fairness Across Alignment Procedures and Agentic Systems*, 2026. 58

[92] Luise Ge, Gregory Kehne, and Yevgeniy Vorobeychik. Optimized Distortion in Linear Social Choice. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(20):16930–16937, 2026. 58

[93] Elliot Anshelevich, Aris Filos-Ratsikas, Nisarg Shah, and Alexandros A. Voudouris. Distortion in Social Choice Problems: The First 15 Years and Beyond. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, 2021. 58

[94] Keertana Chidambaram, Karthik Vinay Seetharaman, and Vasilis Syrgkanis. Direct Preference Optimization with Unobserved Preference Heterogeneity: The Necessity of Ternary Preferences. In *The 29th International Conference on Artificial Intelligence and Statistics*, 2026. 59

[95] Luise Ge, Daniel Halpern, Gregory Kehne, and Yevgeniy Vorobeychik. Linear Social Choice with Few Queries: A Moment-Based Approach, 2026. arXiv:2603.19510. 59

[96] Yeshwanth Cherapanamjeri, Constantinos Daskalakis, Gabriele Farina, and Sobhan Mohammadpour. Learning Correlated Reward Models: Statistical Barriers and Opportunities. In *The Fourteenth International Conference on Learning Representations*, 2026. 59

[97] Federico Echenique, Alireza Fallah, and Michael I. Jordan. A General Framework for Estimating Preferences Using Response Time Data. In *Proceedings of the 27th ACM Conference on Economics and Computation*, 2026. 60

[98] Federico Echenique, Alireza Fallah, Baihe Huang, and Michael I. Jordan. Response Time Enhances Alignment with Heterogeneous Preferences, 2026. arXiv:2605.06987. 60

[99] Germain Kreweras. Aggregation of Preference Orderings. In *Mathematics and Social Sciences I: Proceedings of the Seminars of Menthon-Saint-Bernard*, 1965. 60

[100] Peter C. Fishburn. Probabilistic Social Choice Based on Simple Voting Comparisons. *The Review of Economic Studies*, 51(4):683–692, 1984. 60

[101] Ronald L. Rivest and Emily Shen. An Optimal Single-Winner Preferential Voting System Based on Game Theory. In *Proceedings of the Third International Workshop on Computational Social Choice (COMSOC-2010)*, 2010. 60

[102] Miroslav Dudík, Katja Hofmann, Robert E. Schapire, Aleksandrs Slivkins, and Masrour Zoghi. Contextual Dueling Bandits. In *Proceedings of The 28th Conference on Learning Theory*, 2015. 60

[103] Marc Lanctot, Kate Larson, Yoram Bachrach, Luke Marris, Zun Li, Avishkar Bhoopchand, Thomas Anthony, Brian Tanner, and Anna Koop. Evaluating Agents Using Social Choice Theory, 2025. arXiv:2312.03121. 60

[104] Hadi Khalaf, Flavio Calmon, Daniel Halpern, Ariel D. Procaccia, Itai Shapira, and Serena Lutong Wang. Robust AI Evaluation Through Maximal Lotteries. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. 60

[105] John von Neumann. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320, 1928. 60

[106] Florian Brandl, Felix Brandt, and Hans Georg Seedig. Consistent Probabilistic Social Choice. *Econometrica*, 84(5):1839–1880, 2016. 61

[107] Remi Munos, Michal Valko, Daniele Calandriello, Mohammad Gheshlaghi Azar, Mark Rowland, Zhaohan Daniel Guo, Yunhao Tang, Matthieu Geist, Thomas Mesnard, Côme Fiegel, Andrea Michi, Marco Selvi, Sertan Girgin, Nikola Momchev, Olivier Bachem, Daniel J. Mankowitz, Doina Precup, and Bilal Piot. Nash Learning from Human Feedback. In *Proceedings of the 41st International Conference on Machine Learning*, 2024. 61, 62

[108] Gokul Swamy, Christoph Dann, Rahul Kidambi, Zhiwei Steven Wu, and Alekh Agarwal. A Minimaximalist Approach to Reinforcement Learning from Human Feedback. In *Proceedings of the 41st International Conference on Machine Learning*, 2024. 61

[109] Yue Wu, Zhiqing Sun, Huizhuo Yuan, Kaixuan Ji, Yiming Yang, and Quanquan Gu. Self-Play Preference Optimization for Language Model Alignment. In

The Thirteenth International Conference on Learning Representations, 2025. 61

[110] Daniele Calandriello, Zhaohan Daniel Guo, Remi Munos, Mark Rowland, Yunhao Tang, Bernardo Avila Pires, Pierre Harvey Richemond, Charline Le Lan, Michal Valko, Tianqi Liu, Rishabh Joshi, Zeyu Zheng, and Bilal Piot. Human Alignment of Large Language Models Through Online Preference Optimisation. In *Proceedings of the 41st International Conference on Machine Learning*, 2024. 61

[111] Corby Rosset, Ching-An Cheng, Arindam Mitra, Michael Santacroce, Ahmed Awadallah, and Tengyang Xie. Direct Nash Optimization: Teaching Language Models to Self-Improve with General Preferences, 2024. arXiv:2404.03715. 61

[112] Daniil Tiapkin, Daniele Calandriello, Denis Belomestny, Eric Moulines, Alexey Naumov, Kashif Rasul, Michal Valko, and Pierre Menard. Proximal Point Nash Learning from Human Feedback, 2025. arXiv:2505.19731. 61

[113] Benjamin Heymann. Adaptive Preference Aggregation, 2025. arXiv:2503.10215. 61

[114] Jiancong Xiao, Ziniu Li, Xingyu Xie, Emily Getzen, Cong Fang, Qi Long, and Weijie J. Su. On the Algorithmic Bias of Aligning Large Language Models with RLHF: Preference Collapse and Matching Regularization. *Journal of the American Statistical Association*, 120(552):2154–2164, 2025. 62

[115] Alexander Shypula, Shuo Li, Botong Zhang, Vishakh Padmakumar, Kayo Yin, and Osbert Bastani. Evaluating the Diversity and Quality of LLM Generated Content. In *Second Conference on Language Modeling*, 2025. 62

[116] Robert Kirk, Ishita Mediratta, Christoforos Nalmpantis, Jelena Luketina, Eric Hambro, Edward Grefenstette, and Roberta Raileanu. Understanding the Effects of RLHF on LLM Generalisation and Diversity. In *The Twelfth International Conference on Learning Representations*, 2024. 62, 63

[117] Thom Lake, Eunsol Choi, and Greg Durrett. From Distributional to Overton Pluralism: Investigating Large Language Model Alignment. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, 2025. 62

[118] Muhammad Khalifa, Hady Elsahar, and Marc Dymetman. A Distributional Approach to Controlled Text Generation. In *The Ninth International Conference on Learning Representations*, 2021. 62, 63

[119] Ganqu Cui, Yuchen Zhang, Jiacheng Chen, Lifan Yuan, Zhi Wang, Yuxin Zuo, Haozhan Li, Yuchen Fan, Huayu Chen, Weize Chen, Zhiyuan Liu, Hao Peng, Lei Bai, Wanli Ouyang, Yu Cheng, Bowen Zhou, and Ning Ding. The Entropy Mechanism of Reinforcement Learning for Reasoning Language Models, 2025. arXiv:2505.22617. 62

[120] Itai Shapira, Gerdus Benade, and Ariel D. Procaccia. How RLHF Amplifies Sycophancy. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. 62

[121] Stewart Slocum, Asher Parker-Sartori, and Dylan Hadfield-Menell. Diverse Preference Learning for Capabilities and Alignment. In *The Thirteenth International Conference on Learning Representations*, 2025. 62, 63

[122] Shibani Santurkar, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. Whose Opinions Do Language Models Reflect? In *Proceedings of the 40th International Conference on Machine Learning*, 2023. 63

[123] Hannah Rose Kirk, Bertie Vidgen, Paul Röttger, and Scott A. Hale. Personalisation within Bounds: A Risk Taxonomy and Policy Framework for the Alignment of Large Language Models with Personalised Feedback, 2023. arXiv:2303.05453. 63

[124] Souradip Chakraborty, Jiahao Qiu, Hui Yuan, Alec Koppel, Dinesh Manocha, Furong Huang, Amrit Bedi, and Mengdi Wang. MaxMin-RLHF: Alignment with Diverse Human Preferences. In *Proceedings of the 41st International Conference on Machine Learning*, 2024. 63

[125] Kihyun Kim, Jiawei Zhang, Asuman E. Ozdaglar, and Pablo A. Parrilo. Beyond RLHF and NLHF: Population-Proportional Alignment Under an Axiomatic Framework. In *The Fourteenth International Conference on Learning Representations*, 2026. 63

[126] Anthony GX-Chen, Jatin Prakash, Jeff Guo, Rob Fergus, and Rajesh Ranganath. KL-Regularized Reinforcement Learning for Generative Modelling Is Designed to Mode Collapse. In *The Fourteenth International Conference on Learning Representations*, 2026. 63

[127] Ilgee Hong, Zichong Li, Alexander Bukharin, Yixiao Li, Haoming Jiang, Tianbao Yang, and Tuo Zhao. Adaptive Preference Scaling for Reinforcement Learning with Human Feedback. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 63

[128] Jiashuo Wang, Haozhao Wang, Shichao Sun, and Wenjie Li. Aligning Language Models with Human Preferences via a Bayesian Approach. In *The Thirty-seventh Annual Conference on Neural Information Processing Systems*, 2023. 63

[129] Nuoya Xiong and Aarti Singh. Projection Optimization: A General Framework for Multi-Objective and Multi-Group RLHF. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025. 63

[130] Chanwoo Park, Mingyang Liu, Dingwen Kong, Kaiqing Zhang, and Asuman E. Ozdaglar. RLHF from Heterogeneous Feedback via Personalization and Preference Aggregation. In *ICML 2024 Workshop on Theoretical Foundations of Foundation Models*, 2024. 63

[131] Daiwei Chen, Yi Chen, Aniket Rege, Zhi Wang, and Ramya Korlakai Vinayak. PAL: Sample-Efficient Personalized Reward Modeling for Pluralistic Alignment. In *The Thirteenth International Conference on Learning Representations*, 2025. 63

[132] Taylor Sorensen, Pushkar Mishra, Roma Patel, Michael Henry Tessler, Michiel A. Bakker, Georgina Evans, Iason Gabriel, Noah Goodman, and Verena Rieser. Value Profiles for Encoding Human Variation. In *Proceedings*

of the 2025 Conference on Empirical Methods in Natural Language Processing, 2025. 63

[133] Elinor Poole-Dayan, Jiayi Wu, Taylor Sorensen, Jiaxin Pei, and Michiel A. Bakker. Benchmarking Overton Pluralism in LLMs. In *The Fourteenth International Conference on Learning Representations*, 2026. 63

[134] Michiel A. Bakker, Martin J. Chadwick, Hannah R. Sheahan, Michael Henry Tessler, Lucy Campbell-Gillingham, Jan Balaguer, Nat McAleese, Amelia Glaese, John Aslanides, Matthew M. Botvinick, and Christopher Summerfield. Fine-Tuning Language Models to Find Agreement Among Humans with Diverse Preferences. In *The Thirty-sixth Annual Conference on Neural Information Processing Systems*, 2022. 63

[135] Sara Fish, Paul Gölz, David Parkes, Ariel Procaccia, Gili Rusak, Itai Shapira, and Manuel Wuthrich. Generative Social Choice. *Journal of the ACM*, 73(2): 11:1–11:52, 2026. 63

[136] Niclas Boehmer, Sara Fish, and Ariel D. Procaccia. Generative Social Choice: The Next Generation. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025. 63

[137] Ethan Perez, Saffron Huang, Francis Song, Trevor Cai, Roman Ring, John Aslanides, Amelia Glaese, Nat McAleese, and Geoffrey Irving. Red Teaming Language Models with Language Models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 2022. 63

[138] Sriyash Poddar, Yanming Wan, Hamish Ivison, Abhishek Gupta, and Natasha Jaques. Personalizing Reinforcement Learning from Human Feedback with Variational Preference Learning. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 63

[139] Gihoon Kim and Euntai Kim. Swap-Guided Preference Learning for Personalized Reinforcement Learning from Human Feedback. In *The Fourteenth International Conference on Learning Representations*, 2026. 63

[140] Andre Barreto, Vincent Dumoulin, Yiran Mao, Mark Rowland, Nicolas Perez-Nieves, Bobak Shahriari, Yann Dauphin, Doina Precup, and Hugo Larochelle. Capturing Individual Human Preferences with Reward Features. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. 63

[141] Avinandan Bose, Zhihan Xiong, Yuejie Chi, Simon Shaolei Du, Lin Xiao, and Maryam Fazel. LoRe: Personalizing LLMs via Low-Rank Reward Modeling. In *Second Conference on Language Modeling*, 2025. 63

[142] Hannah Rose Kirk, Liu Leqi, Fanzhi Zeng, Henry Davidson, Bertie Vidgen, Christopher Summerfield, and Scott A. Hale. PRISM-X: Experiments on Personalised Fine-Tuning with Human and Simulated Users, 2026. arXiv:2605.13307. 63

[143] Michael Henry Tessler, Michiel A. Bakker, Daniel Jarrett, Hannah Sheahan, Martin J. Chadwick, Raphael Koster, Georgina Evans, Lucy Campbell-Gillingham, Tantum Collins, David C. Parkes, Matthew Botvinick, and Christopher Summerfield. AI Can Help Humans Find Common Ground in Democratic Deliberation. *Science*, 386(6719):eadq2852, 2024. 63

[144] Jay Chooi, Paul Gölz, Ariel D. Procaccia, Benjamin Schiffer, and Shirley Zhang. Finding Common Ground in a Sea of Alternatives, 2026. arXiv:2603.16751. 63

[145] Aviv Ovadya. Generative CI Through Collective Response Systems, 2023. arXiv:2302.00672. 63

A Game Theory Toolkit for Voting Rules

MOSES CHARIKAR
Stanford University
and
PRASANNA RAMAKRISHNAN
Stanford University
and
KANGNING WANG
Rutgers University

Classical voting theory is often viewed as a field of impossibilities, where even the most basic criteria can be unsatisfiable. A growing line of work seeks to sidestep these hurdles by considering voting rules that output randomized distributions over candidates. In this survey, we explore a particular family of randomized voting rules derived from equilibria of simple two-player games, most notably maximal lotteries and stable lotteries. We survey a flurry of recent work using these ideas to prove several positive results in voting theory. These include improved bounds for randomized voting rules in the *distortion* problem, and progress on *committee selection* questions that appear deterministic in nature, such as Condorcet winning sets, approximately stable committees, and approximately dominating sets.

Categories and Subject Descriptors: [**Theory of Computation**]: Algorithmic Game Theory and Mechanism Design

General Terms: Algorithms, Economics, Theory

Additional Key Words and Phrases: Computational Social Choice, Game Theory, Distortion, Committee Selection

1. INTRODUCTION

Voting is a fundamental tool for aggregating individual preferences into a collective decision. While elections are most associated with representative democracy, the same abstraction is at the heart of many other settings: an organization making new hires, judges choosing prize winners, academic conferences choosing papers, a group of friends choosing a restaurant, or human labelers fine-tuning a language model towards better responses. In each case, voters express preferences over a set of candidates, and the goal is to select an outcome that is in some sense fair, representative, or socially efficient.

The persistent challenge in voting theory is that even very modest desiderata are not always satisfiable. This issue is prominently exemplified by the impossibility theorems of Arrow and Gibbard–Satterthwaite, but the earliest example is Condorcet's paradox from 1785, which says that a candidate that beats each other candidate in a head-to-head majority vote (called a *Condorcet winner*) does not always exist. That is, there are elections where no matter which candidate wins, a majority of voters will prefer an alternative.

Authors' addresses: moses@cs.stanford.edu, pras1712@stanford.edu, kn.w@rutgers.edu.

One approach to circumventing these negative results is to use *randomized* voting rules that choose a distribution over candidates rather than a single winning candidate. For example, we can use game theory to prove the existence of a randomized analogue of a Condorcet winner called a *maximal lottery*, which beats each candidate for half the voters in expectation. Although the richer design space of randomized voting rules opens up new possibilities, the idea of deciding elections by chance can still be instinctively unsettling, especially given the close association between voting and high-stakes, politically charged decisions. But randomization can feel much more natural in lower-stakes decisions among friends, in processes that are already inherently noisy, or even in settings involving language models, which are stochastic by design. Even in political contexts, randomized or fractional solutions are not without precedent. For example, ancient Athens used sortition to fill many public offices by lot, and the Roman Republic divided executive authority between two annually elected consuls.

In this survey, we will explore a particular family of randomized voting rules that can be derived from the mixed-strategy Nash equilibria of simple two-player games, like maximal lotteries. We will see how these voting rules have played pivotal roles in new results from seemingly different parts of computational social choice: *committee selection* and *distortion*. A recurring theme is that these randomized solutions are useful not only as an end in themselves, but also as a tool to show the existence of *deterministic* structures with desirable properties.

2. MAXIMAL LOTTERIES AND STABLE LOTTERIES

Imagine that an election is taking place, and two gamblers decide to play the following game. Each gambler independently chooses a candidate to bet on, and then they ask a random voter to compare their chosen candidates. Whoever chose the candidate that the voter prefers is the winner of the game. (If the gamblers happen to choose the same candidate, they flip a coin to decide the winner.)

Intuitively, the gamblers want to choose the "best" candidate — betting on an unpopular candidate is clearly a losing proposition. Conversely, it might be reasonable to think that effective strategies for choosing candidates in the game (given knowledge of how voters' preferences are distributed) correspond to voting rules that choose more popular candidates. Arguably, the optimal strategy would be to *randomize* over the candidates according to the *mixed-strategy Nash equilibrium* of the game. The equilibrium distribution is identical for each gambler since the game is symmetric, and playing according to the equilibrium guarantees a win with probability at least $\frac{1}{2}$ against any fixed choice of the opponent.

This guarantee underpins the concept of a *maximal lottery*. To give a more formal definition, we first introduce some basic notation. An *election* (or *preference profile*) $(V, C, \succ_V)$ consists of a set V of n voters, a set C of m candidates, and a strict linear order $\succ_v$ over the candidates C for each voter $v \in V$. We use $\frac{1}{n}|a \succ b|$ to denote the fraction of voters who prefer candidate a over candidate b (i.e., $\frac{1}{n}|a \succ b| = \Pr_{v \sim V}[a \succ_v b]$). For simplicity, we adopt the convention that $\frac{1}{n}|a \succ a| = 0$, but note that it can also be natural to set $\frac{1}{n}|a \succ a| = \frac{1}{2}$ (akin to the coin-flip tiebreaker between the gamblers). Results will be stated so that they are true with either convention (and are often slightly stronger with the latter).

Theorem 2.1 ([Kreweras 1965; Fishburn 1984]). *In any election, there exists a distribution D_{ML} over candidates such that for all candidates $a \in C$,*

$$\mathop{\mathbb{E}}_{b \sim D_{\mathrm{ML}}} [\tfrac{1}{n}|a \succ b|] \leq \frac{1}{2}.$$

A distribution satisfying this condition is called a maximal lottery.

The proof is a fairly straightforward application of von Neumann's minimax theorem applied to our game between the gamblers. We will defer a proof until Theorem 2.2, which is a more general result.

For some illustrative examples, consider elections with three candidates a, b, c, and voters of three types A, B, C with preferences $a \succ b \succ c$, $b \succ c \succ a$, and $c \succ a \succ b$ respectively. With one voter of each type, this is the classic example of Condorcet's paradox, and the unique maximal lottery is the uniform distribution over the candidates. If we have two voters of types A and B each and one voter of type C, then the maximal lottery would choose (a, b, c) with probabilities $(\frac{3}{5}, \frac{1}{5}, \frac{1}{5})$. If we have three voters of type A, two voters of type B, and one voter of type C, any distribution which chooses (a, b, c) with probabilities $(p, 0, 1-p)$ for $p \in [\frac{2}{3}, 1]$ is a maximal lottery. In elections that have a Condorcet winner (a candidate a such that $\frac{1}{n}|a \succ b| > \frac{1}{2}$ for all candidates $b \neq a$), the maximal lottery deterministically chooses that candidate. If one takes the view that Condorcet's paradox is unlikely[1], then a maximal lottery can be thought of as like any other deterministic Condorcet-consistent voting rule, with a natural randomized tie-breaking scheme if needed.

Maximal lotteries were originally introduced by Kreweras [1965] in a short paper primarily about Condorcet's paradox and Arrow's theorem. In discussing ways of escaping these challenges at the very end, he mentions that using game theory, one can show that the set of convex combinations of candidates no longer suffers from the curse of intransitivity. Fishburn [1984] gave a more in-depth treatment of maximal lotteries, formally proving their existence, analyzing their properties, and also defining the broader class of "maximal lottery methods" which can be derived from variants of the earlier zero-sum game over candidates. For example, if the gamblers decide the winner by using a majority vote instead of a random voter's preference, the resulting equilibrium is a *C1-maximal lottery* (named much later by Brandl et al. [2018]). In surveying randomized voting rules, Brandt [2017] vividly recounts that maximal lotteries have been rediscovered numerous times by researchers across academic disciplines.

Recently, Cheng et al. [2020] generalized maximal lotteries to *stable lotteries*, which are distributions over *sets* of candidates (called *committees*). If we modify our game between the gamblers so that one is allowed to pick k candidates and wins if a random voter prefers *any* of their candidates over the opponent's candidate, then their equilibrium distribution is a stable lottery.

To give a formal description, we introduce some notation. We say that voter v prefers candidate a over committee S, or $a \succ_v S$, if $a \succ_v b$ for each $b \in S$. We define $\frac{1}{n}|a \succ S|$ to be the fraction of voters that prefer candidate a over S (i.e.,

[1]For example, with single-peaked preferences, or in an impartial culture (see Gehrlein [2002] for more).

$\frac{1}{n}|a \succ S| = \Pr_{v\sim V}[a \succ_v S]$). In Section 4.1, we briefly use $|a \succ S|$ to denote the *number* of voters that prefer a over S.

If $a \in S$, we use the convention $\frac{1}{n}|a \succ S| = 0$, but there is also a natural alternate convention[2] which breaks ties between duplicate candidates uniformly,

analogous to setting $\frac{1}{n}|a \succ a| = \frac{1}{2}$. Like before, results will be stated to be true with either convention.

Theorem 2.2 (Stable Lottery [Cheng et al. 2020]). *In any election, there exists a distribution $D_{k\text{-SL}}$ over committees of size k such that for all candidates $a \in C$,*

$$\mathop{\mathbb{E}}_{S\sim D_{k\text{-SL}}}[\tfrac{1}{n}|a \succ S|] \leq \frac{1}{k+1}.$$

A distribution satisfying this condition is called a stable k-lottery.

Proof. Consider the following zero-sum game. The *attacker* chooses a distribution D_a over candidates, the *defender* chooses a distribution D over committees of k candidates, then we sample $a \sim D_a, S \sim D$, and the defender pays the attacker $\frac{1}{n}|a \succ S|$ (one can imagine, a penny for each voter that prefers the attacker's candidate over the defender's committee). Since the players can randomize, the minimax theorem [von Neumann 1928] shows that under optimal play, the order of the players does not change the expected value of the game. That is,

$$\max_{D_a}\min_{D}\mathop{\mathbb{E}}_{\substack{a\sim D_a\\S\sim D}}[\tfrac{1}{n}|a \succ S|] = \min_{D}\max_{D_a}\mathop{\mathbb{E}}_{\substack{a\sim D_a\\S\sim D}}[\tfrac{1}{n}|a \succ S|].$$

If the defender plays second, then they can simply sample from the attacker's distribution k times (setting $D = D_a^k$). The defender's samples are just as likely to be a voter's favorite as the attacker's sample, so the defender pays at most $\frac{1}{k+1}$ in expectation.

□

Before diving into the various ways of applying maximal lotteries and stable lotteries, we briefly mention some useful structural facts.

The first concerns efficient computation of these distributions. The equilibria of zero-sum games can be computed using linear programming in time polynomial in the size of the payoff matrix. Thus, maximal lotteries can be computed in $\text{poly}(m)$ time, and more generally, stable lotteries can be computed in $\text{poly}(m^k)$ time given the payoff matrix as input (which can be computed in time linear in n given the preferences). Cheng et al. [2020] show that the exponential dependence on k can be avoided at the cost of approximation; using tools from no-regret learning, they show that ε-approximate stable k-lotteries can be computed in $\text{poly}(m, \frac{1}{\varepsilon})$ time.

Second, while stable lotteries are defined as distributions over committees, one can show that there always exists a stable lottery that is a *product distribution*. That is, in any election there is a distribution D over candidates such that D^k is a stable k-lottery.

[2]This convention also makes it easier to generalize S to be a *multiset*, which is cleaner when S is sampled from a distribution over candidates with replacement, as in the proof of Theorem 2.2.

This fact is proven implicitly by Charikar et al. [2025], and explicitly by Charikar et al. [2025]. It turns out that if we take $k \to \infty$, then D tends to the distribution which chooses a uniformly random voter's favorite candidate (called a *random dictatorship*).

With this in mind, one can alternatively view stable k-lotteries as interpolating between a maximal lottery ($k = 1$) and a random dictatorship ($k \to \infty$). In some settings where a random dictatorship is useful (like in the metric distortion setting considered by Charikar et al. [2025] and Cai et al. [2026]), we can use a stable k-lottery instead and achieve comparable quantitative guarantees while satisfying additional normative properties.

Finally, it is worth mentioning that some papers do not define stable lotteries in exactly the same way we do here. Cheng et al. [2020] originally used a weaker definition with $< \frac{1}{k}$ in place of $\leq \frac{1}{k+1}$, to align with the existing notion of a "stable committee" when the distribution is deterministic.[3] Cai et al. [2026] instead define a stable lottery directly as the Nash equilibrium of the game described in the proof of Theorem 2.2, following Charikar et al. [2024], which does the same with maximal lotteries.[4] Ultimately, all these choices are reasonable, and their differences have essentially no impact on results.

3. APPLICATION 1: COMMITTEE SELECTION

The committee selection problem addresses social choice settings where the goal is to choose a *committee* of k winners rather than a single winner. For two familiar motivating examples, consider an academic department choosing candidates to hire (or interview), or a conference choosing papers to accept.[5] One commonly studied goal is to choose committees such that no large group of voters would have preferred a candidate that is excluded from the committee. What it means for a group to be "large" and to "prefer an excluded candidate" is open to interpretation, and different choices lead to a variety of definitions which are more or less attainable.

In this section, we will explore how maximal lotteries, stable lotteries, and related randomized voting rules have been used to establish positive results on the existence of three closely related solution concepts in committee selection: Condorcet winning sets, approximately stable committees, and approximately dominating sets.

The fact that these solution concepts are all fundamentally deterministic makes it notable that probabilistic tools have been critical in understanding them. The techniques are applied in similar ways across the three problems, so we will showcase them in one context (Condorcet winning sets), and briefly discuss

the others.

Lastly, we mention that committee selection is a rich field of study, and the setting of ranked preferences and the desiderata that we focus on here are only a thin slice. For more in-depth surveys on committee selection, we refer the reader to Faliszewski et al. [2017] and Lackner and Skowron [2023].

[3] See Section 3.2 for more details.

[4] Though, a maximal lottery as defined in Theorem 2.1 is always an equilibrium of the game, but the same is not necessarily true of a stable lottery as defined in Theorem 2.2.

[5] Of course, it would often be unreasonable to elicit full ranked preferences in these settings, but it is still a helpful abstraction for understanding what objectives are attainable with enough effort.

3.1 Condorcet winning sets

Condorcet's paradox points to an unsettling possibility — no matter which candidate wins, a majority of voters may be upset. For example, if a department can only hire one candidate, most of the faculty may think another candidate would have been better. Even if the department can make *multiple* hires, there could be a risk of an even more embarrassing situation: most faculty think some other candidate would have been better than *all* the hires. Of course, the department could avoid this kind of disaster by hiring every applicant, but can it do so without having to make too many hires?

In this example, the department is hoping to find what Elkind et al. [2015] called a *Condorcet winning set*: a committee of candidates S such that no candidate a is preferred over all members of S by a majority of voters. More formally, we have the following definition.

Definition 3.1. A set S of candidates is a *Condorcet winning set* if for all candidates $a \notin S$, $\frac{1}{n}|a \succ S| < \frac{1}{2}$. The *Condorcet dimension* of an election is the size of its smallest Condorcet winning set.

Elkind et al. [2015] asked: how large can the Condorcet dimension of an election be? Condorcet's paradox shows that there are elections with Condorcet dimension at least 2, and Elkind et al. [2015] constructed a simple example of an election with Condorcet dimension at least 3. On the positive side, they showed that every election with m candidates has Condorcet dimension at most $\lceil \log_2 m \rceil + 1$, and conjectured that the Condorcet dimension must grow with the number of candidates in the election.

This conjecture turned out to be false. Charikar et al. [2025] observed that earlier work by Jiang et al. [2020] on approximately stable committees (see Section 3.2) implies that the Condorcet dimension of any election is at most 32, and introduced some new ideas to improve the result to 6. This bound was further sharpened to 5 by Song et al. [2026], whose substantially simpler proof remains the state of the art.

Theorem 3.2 (Song et al. [2026]). *Every election has Condorcet dimension at most* 5.

Charikar et al. [2025] conjectured that these results can be pushed even farther, and that the lower bound of Elkind et al. [2015] is tight.

Conjecture 3.3 (Charikar et al. [2025]). *In every election, there exists a committee of at most k candidates such that for all candidates $a \notin S$,*

$$\frac{1}{n}|a \succ S| < \frac{2}{k+1}.$$

In particular, all elections have Condorcet dimension at most 3.

In this section, we will start by motivating the common ideas in the proofs from Jiang et al. [2020], Charikar et al. [2025], and Song et al. [2026]. To demonstrate the proof strategy, we will give a simple proof that the Condorcet dimension of any election is at most 15 (following Jiang et al. [2020] with minor modifications). For

the interested reader, we have written a short companion note with a full proof of Theorem 3.2 [Charikar et al. 2026b].

At a very high level, the three papers all proceed via the probabilistic method: construct a certain distribution D over committees or candidates, and argue that with positive probability, we can get a Condorcet winning set by sampling from D. One natural starting point would be to use a stable lottery as D, since Theorem 2.2 says that voters tend to prefer samples from $D_{k\text{-SL}}$ over any candidate a (in expectation, $\frac{1}{n}|a \succ S|$ is small when $S \sim D_{k\text{-SL}}$). To get a single committee that has this kind of property against all candidates, a naive approach would be to use a concentration bound to get a high-probability guarantee for each candidate a, and then union bound over all candidates. However, one quickly finds that it is impossible to shake a dependence on the number of candidates, even if the sharpest concentration bounds could apply.

The key idea is this: compare the sampled committee S against D *itself*, instead of each candidate a.

Building on this idea, all three papers use the following general proof strategy.

(1) Construct a distribution D over committees (or candidates) such that voters tend to prefer samples from D over any candidate a.
(2) Show that if we sample the committee S from D, with positive probability, voters tend to prefer S over samples from D.
(3) Argue that if voters tend to prefer S over samples from D, and samples from D over a, then a majority of voters cannot prefer a over S.

The distribution used by Jiang et al. [2020] is precisely a stable lottery. Charikar et al. [2025] use an equilibrium distribution derived from a similar game, but with some modifications to the payoffs that complement the way the distribution is used in the proof and allow for sharper bounds. Song et al. [2026] use a distribution derived from the equilibrium of a carefully designed market rather than a game (in particular, their distribution adapts the well-studied *Lindahl equilibrium*).

To distill the way that voters compare candidates and committees against a distribution D, we use the notion of *rank*, introduced by Charikar et al. [2025].

Given a distribution D over committees, voter v's rank of candidate a with respect to D is the probability that v prefers a over a committee sampled from D.[6] Formally,

$$\text{rank}_v(a; D) \coloneqq \Pr_{S' \sim D}[a \succ_v S'].$$

To define rank for committees, we extend each voter's preference $\succ_v$ over candidates to a preference over committees as follows. For committees $S \neq T$, we say that $S \succ_v T$ if and only if v prefers their favorite candidate in $S \setminus T$ over their favorite candidate in $T \setminus S$. In another view, each voter v can encode a committee S as a binary string $s_v \in \{0,1\}^m$ such that $s_v(i) = 1$ if and only if v's ith favorite candidate is in S. Voter v's preference over committees is precisely the lexicographic order of the corresponding binary strings.

[6]The intention with this terminology is that "higher ranked" candidates are more preferred, while "lower ranked" candidates are less preferred, though we acknowledge the potential for confusion since colloquially, high ranks can correspond to low numerical ranks.

We then define the rank of a committee S as the probability that v *weakly* prefers S over a committee drawn from D. Formally,

$$\mathrm{rank}_v(S; D) \coloneqq \Pr_{S' \sim D}[S \succeq_v S'].$$

The ranks are a convenient primitive to work with because they encapsulate the voter's preferences over candidates and committees in a way that neatly fits the demands of the proof template. In particular, one can check that the ranks satisfy the following two properties.

Proposition 3.4. *The ranks are consistent with preferences: if* $a \succ_v S$ *then* $\mathrm{rank}_v(a; D) \geq \mathrm{rank}_v(S; D)$.

Proposition 3.5. *The random variable* $\mathrm{rank}_v(S; D)$ *with* $S \sim D$ *stochastically dominates* $r \sim \mathrm{Unif}(0, 1)$.

Proposition 3.4 allows us to use the ranks to interface with voters' preferences over candidates and committees. For example, if we can show that $\mathrm{rank}_v(a; D)$ is small for most voters and $\mathrm{rank}_v(S; D)$ is large for most voters, then we can argue that most voters cannot prefer a over S. Proposition 3.5 is critical in the second step of the proof: it gives us a simple way of understanding the distribution of $\mathrm{rank}_v(S; D)$ when S is sampled from D.

The easiest way to understand why the ranks satisfy the two properties above is with a picture.

Imagine creating a block for each committee S, whose width is the probability mass of S in D, and having each voter arrange these blocks along the interval $[0, 1]$ in increasing order of preference (see Figure 1 for an example). Then $\mathrm{rank}_v(S; D)$ is precisely the top (rightmost point) of the block corresponding to S, and $\mathrm{rank}_v(a; D)$ sits just above the blocks of committees that only consist of candidates that v does not prefer to a. This view explains why we define ranks for candidates and committees as we did, making the ranks for candidates as small as possible, and the ranks for committees as large as possible, while ensuring that if $a \succ_v S$ then $\mathrm{rank}_v(a; D) \geq \mathrm{rank}_v(S; D)$ (precisely Proposition 3.4).

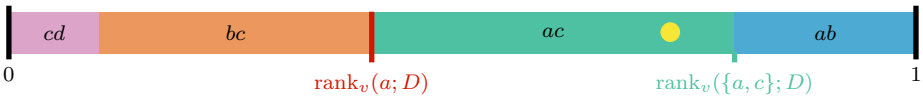


Fig. 1. Depiction of the ranks for a voter v with preference $a \succ b \succ c \succ d$ with respect to a distribution D which chooses the committees $\{c, d\}, \{a, b\}, \{b, c\}, \{a, c\}$ with probabilities $0.1, 0.2, 0.3, 0.4$ respectively. The yellow dot represents $r \sim \mathrm{Unif}(0, 1)$, which can be used to sample from D.

The pictorial representation also gives us a simple way of interpreting samples from D. From each voter's perspective, $S \sim D$ is equivalent to sampling a real number $r \sim \mathrm{Unif}(0, 1)$ and choosing the committee whose block the number r lands in. (Formally, choose the committee S whose $\mathrm{rank}_v(S; D)$ is minimal but at least r.) Proposition 3.5 is immediate from this perspective.

We are now ready to give a constant upper bound on the Condorcet dimension of any election.

THEOREM 3.6. *Every election has Condorcet dimension at most* 15.

PROOF. First, we claim that the average rank of any candidate with respect to a stable k-lottery is at most $\frac{1}{k+1}$. That is, if $D_{k\text{-SL}}$ is the distribution of a stable k-lottery, then for all candidates a,

$$\frac{1}{n}\sum_{v\in V}\text{rank}_v(a; D_{k\text{-SL}}) \le \frac{1}{k+1}. \tag{1}$$

This claim follows by a straightforward translation of Theorem 2.2. In particular for any distribution D over committees, we have

$$\mathop{\mathbb{E}}_{S\sim D}[\tfrac{1}{n}|a \succ S|] = \mathop{\mathbb{E}}_{S\sim D}\left[\frac{1}{n}\sum_{v\in V}\mathbf{1}[a \succ_v S]\right] = \frac{1}{n}\sum_{v\in V}\Pr_{S\sim D}[a \succ_v S] = \frac{1}{n}\sum_{v\in V}\text{rank}_v(a; D).$$

With these equalities, (1) is equivalent to the statement of Theorem 2.2.

Next, we claim that some committee S in the support of the stable lottery has high ranks for a large fraction of voters. In fact, for any distribution over committees D, a random committee $S \sim D$ has $\text{rank}_v(S; D) > \frac{1}{4}$ for at least a $\frac{3}{4}$ fraction of voters v in expectation. This claim follows immediately from the fact that $\text{rank}_v(S; D)$ with $S \sim D$ stochastically dominates $r \sim \text{Unif}(0, 1)$, since r is greater than $\frac{1}{4}$ with probability $\frac{3}{4}$. Figure 2 may be helpful for visual intuition.

Finally, we claim that if $k \ge 15$, then this S (with $D_{k\text{-SL}}$ in place of D) is a Condorcet winning set. Suppose towards a contradiction that for some candidate a, we have that $\frac{1}{n}|a \succ S| \ge \frac{1}{2}$. Since $\text{rank}_v(S; D_{k\text{-SL}}) > \frac{1}{4}$ for at least a $\frac{3}{4}$ fraction of voters v, it means that for at least a $\frac{1}{4}$ fraction of voters v, we have that $\text{rank}_v(S; D_{k\text{-SL}}) > \frac{1}{4}$ *and* $a \succ_v S$. But then, $\text{rank}_v(a; D_{k\text{-SL}}) > \frac{1}{4}$ for this $\frac{1}{4}$ fraction of voters v. It follows that

$$\frac{1}{n}\sum_{v\in V}\text{rank}_v(a; D_{k\text{-SL}}) > \frac{1}{4}\cdot\frac{1}{4} = \frac{1}{16},$$

which contradicts (1) for $k \ge 15$. □

3.2 Approximately stable committees

While Condorcet winning sets focus on finding the best possible *small* committees, one can hope to get strong asymptotic results as a function of the committee size as well. The gold standard is for a committee to be *stable* (following the classic notion of *core stability* from cooperative game theory [Moulin 1988]). The normative basis for stable committees is that if we are to choose a committee S of size k, then any coalition of $\frac{1}{k}$ fraction of voters deserves to pick one candidate on the committee. As such, if there is a candidate a that a $1/k$ fraction of voters prefers over each candidate in S, then the committee S is considered *unstable*. (Perhaps these voters would want to break away from the larger group, and make their decisions separately, like an underrepresented academic community forming a new conference.) If there is no such candidate a, then S is stable.

Definition 3.7. A committee S of size k is *stable* if for all candidates $a \notin S$, $\frac{1}{n}|a \succ S| < \frac{1}{k}$.

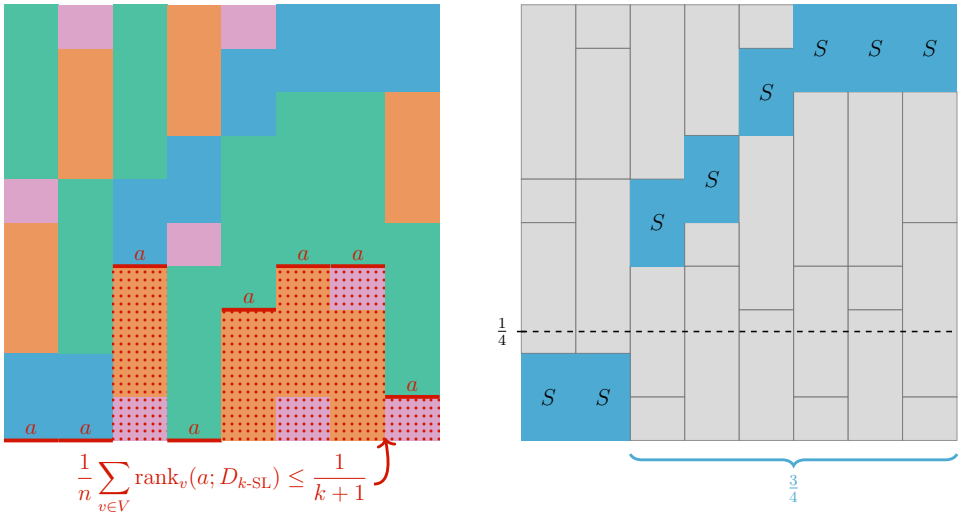


Fig. 2. A visual depiction of the first two steps of the proof. Within the square, each column represents a voter. Each committee is a block in the column, with height corresponding to their probability mass in $D_{k\text{-SL}}$, ordered by the preference of the voter (higher is more preferred). Each candidate a sits at height $\text{rank}_v(a; D_{k\text{-SL}})$ in the column for voter v. The left diagram shows how each candidate is ranked low: the area under each a is at most $\frac{1}{k+1}$. The right shows how some committee S must be high: the top of block S reaches height above a quarter for three quarters of the voters.

Note that the definition is identical to that of a Condorcet winning set (Definition 3.1), but with the majority $\frac{1}{2}$ threshold replaced with $\frac{1}{k}$. The aforementioned counterexample of Elkind et al. [2015] against Condorcet winning sets of size 2 also shows that for each $k > 1$, stable committees of size k need not exist.

Instead, Jiang et al. [2020] proposed the following notion of *approximately* stable committees, and showed that they always exist.

Theorem 3.8 (Jiang et al. [2020]). *A committee S of k candidates is a c-*stable committee *if for all candidates a,*

$$\tfrac{1}{n}|a \succ S| < \frac{c}{k}.$$

There exists a universal constant c such that for all positive integers k, every election has a c-stable committee of size at most k.

Jiang et al. [2020] proved this theorem with the particular constant $c = 16$. It is not hard to see that a c-stable committee of size $2c$ is a Condorcet winning set, and so this result is already enough to establish that all elections have Condorcet dimension at most 32. With some additional ideas[7] the correspondence can be made to work both ways. Charikar et al. [2025] improve the constant c to 9.82 and the techniques of Song et al. [2026] can be used to improve it to 4.91.[8]

[7]The key idea is to sample part of the committee, and then recurse on the subset of voters that are unhappy with the sampled committee (have low ranks).
[8]Song et al. [2026] cite a bound of 5.03 from prior work of Song and Nguyen [2026], but this bound can be improved marginally; see Charikar et al. [2026b] for more details.

Finally, we briefly mention that the interest in (approximately) stable committees extends far beyond the setting of ranked preferences that we focus on in this survey. With any monotone preference structure over committees (meaning $S \succeq_v S'$ if $S' \subseteq S$), the notion of approximately stable committees naturally extends as follows.

Definition 3.9. A committee S of k candidates is a *c-stable committee* if for all committees S' of k' candidates, $\frac{1}{n}|S' \succ S| < c \cdot \frac{k'}{k}$.

Intuitively, a coalition of $\frac{k'}{k}$ fraction of voters "deserves" k' candidates on the committee, so if they prefer some committee S' of size k' over S, then S is unstable. Jiang et al. [2020] proved a general result that for *any* monotone preferences, 32-stable committees of size at most k always exist.

Along with ranked preferences, *approval* preferences are a special case of significant interest. In this setting, each voter v has a subset A_v of candidates that they approve. Voters prefer committees that have more candidates that they approve (i.e., $S \succeq_v S'$ if $|S \cap A_v| \geq |S' \cap A_v|$.) In contrast to the ranking setting, it remains a major open question whether an exactly stable committee always exists (see, e.g., Peters [2025] and Becker et al. [2026]). It is known that 3.65-stable committees exist, due to Gao et al. [2026].

3.3 Approximately dominating sets

A common feature of Condorcet winning sets and approximately stable committees is that they compare a committee *collectively* against each candidate, but in some applications the resulting conditions are far too weak. For example, it would not be reasonable to justify the set of accepted papers for a conference by arguing that less than half of the program committee would think any rejected paper would be a best paper candidate if it were accepted. Instead we could aim for a stronger condition which asks voters to compare excluded candidates *head-to-head* against a single candidate on the committee. One natural proposal would be to choose a committee that is a *dominating set*.

Definition 3.10. A committee S is a *dominating set* if for each candidate $a \notin S$, there exists $b \in S$ such that $\frac{1}{n}|a \succ b| \leq \frac{1}{2}$.

The terminology comes from graph theory, where a dominating set is a subset of vertices S such that every vertex has a neighbor in S. If one considers the *majority graph* of an election, which is the directed graph whose vertices are candidates and where we have a directed edge $a \to b$ if a majority of voters prefers a over b, then the two notions of dominating sets are aligned.

Unfortunately, unlike Condorcet winning sets and approximately stable committees, dominating sets may need to be arbitrarily large. This fact is a consequence of two classic results. The first is McGarvey's theorem [McGarvey 1953], which shows that *any* tournament graph[9]

is the majority graph of some election. The second is a result due to Erdős [1963] that in random tournament graphs on m vertices, the smallest dominating sets have size $\Theta(\log m)$.

[9]A directed graph obtained by assigning directions to the edges of a complete graph.

Instead of relaxing the condition by allowing collective domination, perhaps we can relax the $\frac{1}{2}$ threshold in the definition of a dominating set.

Definition 3.11 *(Approximately dominating sets).* A committee S is a $(\frac{1}{2}+\varepsilon)$-*dominating set* if for each candidate $a \notin S$, there exists $b \in S$ such that $\frac{1}{n}|a \succ b| \leq \frac{1}{2}+\varepsilon$.

It turns out that for any constant ε, $(\frac{1}{2}+\varepsilon)$-dominating sets of constant size exist. In particular, the following theorem was proven independently by Bourneuf et al. [2025] and Charikar et al. [2026a].

Theorem 3.12. *Every election has a $(\frac{1}{2}+\varepsilon)$-dominating set of size $O(1/\varepsilon^2)$.*

Intriguingly, the techniques in the two papers are substantially different. Bourneuf et al. [2025] prove the result as one of many applications of a new tool they call the *dense neighborhood lemma*, which bounds a variant of the VC-dimension in several combinatorial structures. We will focus on the proof from Charikar et al. [2026a], which constructs *approximate* maximal lotteries with small support size.

Theorem 3.13 (Approximate maximal lotteries). *In any election, there exists a distribution D over candidates with support size at most $(1+o(1))\frac{\pi}{8\varepsilon^2}$ such that for all candidates $a \in C$,*

$$\mathop{\mathbb{E}}_{b\sim D}[\tfrac{1}{n}|a \succ b|] \leq \frac{1}{2}+\varepsilon.$$

A distribution satisfying this condition is called an ε-approximate maximal lottery.

It is not hard to see that this theorem generalizes Theorem 3.12, since the support of D must be a $(\frac{1}{2}+\varepsilon)$-dominating set. On their own, approximate maximal lotteries can be an attractive solution concept even in comparison to dominating sets. In choosing papers for a conference, for example, it may still be unreasonable to justify rejecting a paper by arguing that some accepted paper is preferred by most program committee members. But if many members would think the rejected paper is worse than the *average* accepted paper (or even the average accepted *borderline* paper, since the other decisions are straightforward), rejection seems much more defensible.

Approximate maximal lotteries are also interesting for their connection to approximate Nash equilibria. In fact, an approximate maximal lottery exactly corresponds to an approximate Nash equilibrium in the gambler game at the start of Section 2. However, while ε-approximate Nash equilibria have support size $\Theta(\frac{\log m}{\varepsilon^2})$ in the worst case (where m is the number of pure strategies), the particular structure of elections allows for approximate maximal lotteries to avoid the $\log m$ factor.

At a high level, the proof works by arguing that if we draw k samples from a maximal lottery, then with positive probability, the empirical distribution over the samples is an $O(1/\sqrt{k})$-approximate maximal lottery. Like with Condorcet winning sets in Section 3.1, a naive argument would use concentration inequalities and a union bound over candidates, but a logarithmic dependence on m appears unavoidable. Once again, one can get a better handle on the structure of voters' preferences over candidates using the ranks, and get a bound independent of the number of candidates with a simple proof.

Finally, we mention that the right dependence on ε for the support size of approximate maximal lotteries (and $(\frac{1}{2}+\varepsilon)$-dominating sets) remains open. Charikar et al. [2026a] explain that a construction of Alon et al. [2006] implies a lower bound of $\Omega(\frac{1}{\varepsilon})$, and recently Lin et al. [2026] sharpened this lower bound to $\frac{1}{2\varepsilon}$. Charikar et al. [2026a] conjectured that the upper bound is tight.

Conjecture 3.14 (Charikar et al. [2026a]). *For each $\varepsilon > 0$, there is an election in which all $(\frac{1}{2}+\varepsilon)$-dominating sets have size $\Omega(1/\varepsilon^2)$.*

4. APPLICATION 2: DISTORTION

One takeaway from the history of social choice theory is that no voting rule is perfect, and the best we can hope is to find voting rules that are reasonably good. In other words, if we cannot always make optimal choices, can we at least make *approximately* optimal choices?

The *distortion* framework, introduced by Procaccia and Rosenschein [2006], gives a way of understanding what it means for a voting rule to be approximately optimal. In this setting, we imagine that the voters have some cardinal utilities (or costs) for the candidates. If we knew these cardinal values, the optimal choice would be the candidate that maximizes total social welfare (or minimizes the total social cost). However, the voting rule only has access to the *ordinal* preferences of the voters. With only this limited information, can a voting rule always choose an approximately optimal candidate? What is the best approximation factor (called the *distortion*) that can be guaranteed? (See Figures 3 and 4 for illustrative examples.)

Under completely arbitrary utilities or costs, the distortion can be unbounded, but with mild structural assumptions, it is possible to prove meaningful positive results. The literature has focused on two structural models that have each received substantial attention over the last two decades: *utilitarian distortion* and *metric distortion*. At a high level, utilitarian distortion imposes very mild normalization assumptions that voters' utilities are on a similar scale, while metric distortion imagines that the preferences are derived from the distances in a metric. The tradeoff is that in utilitarian distortion, the structural assumption is weaker, but the distortion guarantees still blow up as the number of candidates grows, while metric distortion imposes stronger structural assumptions and gets distortion guarantees that stay constant with any number of candidates.

Instead of covering the vast literature (for which we defer to a survey of Anshelevich et al. [2021]), we will focus on the slice of results that crucially rely on maximal lotteries and stable lotteries. As it happens, these results are also the state of the art for some of the most central questions in both utilitarian distortion and metric distortion.

4.1 Utilitarian distortion

In *utilitarian distortion* [Procaccia and Rosenschein 2006], each voter v has an underlying utility function $u_v : C \to \mathbb{R}_{\geq 0}$ satisfying the condition that if $a \succ_v b$ then $u_v(a) \geq u_v(b)$. The utility functions are also assumed to be normalized,[10] typically so that each voter's utilities sum to one ($\sum_{a \in C} u_v(a) = 1$ for each voter

[10] We refer the reader to Aziz [2019] for compelling normative arguments in favor of normalization.

v). The *social welfare* of a candidate a is $\mathsf{SW}(a) \coloneqq \sum_{v \in V} u_v(a)$, and a voting rule has distortion α if it chooses a candidate b such that $\mathsf{SW}(b) \geq \frac{1}{\alpha} \max_{a \in C} \mathsf{SW}(a)$ in any election, and with any underlying utility functions that are consistent with the preferences. For a randomized voting rule that chooses a distribution D over candidates, we replace $\mathsf{SW}(b)$ with $\mathbb{E}_{b \sim D}[\mathsf{SW}(b)]$.

Preferences	**Possible Utilities**				**Social Welfare**
$v_1: a \succ b \succ c$		a	b	c	$a:$ 1/3
$v_2: b \succ c \succ a$	v_1	1/3	1/3	1/3	$b:$ 5/6
$v_3: c \succ a \succ b$	v_2	0	1/2	1/2	$c:$ 11/6
	v_3	0	0	1	

Fig. 3. An example of the utilitarian distortion in an election with three candidates and three voters with cyclical preferences. The choice of utilities in the middle results in candidate a having distortion 11/2.

Procaccia and Rosenschein [2006] showed that for deterministic voting rules, the plurality rule (which chooses the most common first choice of the voters) achieves the best possible utilitarian distortion of $\Theta(m^2)$.

As such, the problem is most interesting for *randomized* voting rules. Trivially, choosing a uniformly random candidate has distortion $O(m)$, but this is far from optimal. Boutilier et al. [2015] designed an intricate rule with distortion $O(\sqrt{m} \log^* m)$ and showed that any randomized voting rule has distortion at least $\Omega(\sqrt{m})$. This $\log^* m$ gap persisted for several years, until Ebadian et al. [2024] showed that a simple rule which mixes between a random candidate and a stable lottery gets distortion $O(\sqrt{m})$.

Theorem 4.1 (Ebadian et al. [2024]). *Consider the following voting rule: with probability $\frac{1}{2}$, choose a uniformly random candidate, and with probability $\frac{1}{2}$, sample a committee S from a stable $\sqrt{m}$-lottery, and choose a uniformly random candidate from S. This rule has distortion $O(\sqrt{m})$ in any election with unit-sum utilities.*

Note that Ebadian et al. [2024] prove the same result for a much broader class of utility functions, but we focus on the unit-sum case for simplicity.

Here is some rough intuition for the argument. Imagine that we just sample a committee S of size k from a stable k-lottery, and then choose a uniformly random candidate from S. Since we expect that nearly all voters ($\frac{k}{k+1}$ fraction) will prefer some candidate in S over the optimal candidate $a^\star$, we can argue that a uniformly random candidate from S captures nearly a $1/k$ fraction of the welfare of $a^\star$. There is an additive loss of at most $\frac{n}{k(k+1)}$, coming from the voters who prefer $a^\star$ over S. (The factor $1/k$ in both these terms comes from the fact that we choose one of the k candidates of S at random.) If we set $k \approx \sqrt{m}$ then the roughly n/m additive loss can be made up for by mixing with a uniform lottery, which alone always captures utility at least n/m (since the total utility of all candidates is n). For the curious reader, the full proof is below.

Proof. Suppose that $a^\star$ is the candidate with the maximum social welfare. Let D_S be the distribution that chooses a uniformly random candidate from the committee S of size k. The key claim is that

$$\mathsf{SW}(D_S) \geq \frac{\mathsf{SW}(a^\star) - |a^\star \succ S|}{k}.$$

If a voter v prefers some candidate $b \in S$ over $a^\star$ (i.e., $a^\star \not\succ_v S$), then we have $u_v(a^\star) \leq u_v(b)$. Even more loosely, these voters satisfy

$$u_v(a^\star) \leq \sum_{b \in S} u_v(b).$$

For the remaining voters v satisfying $a^\star \succ_v S$, we can default to the weak bound $u_v(a^\star) \leq 1$. Putting the two together, we have

$$\begin{aligned}\mathsf{SW}(a^\star) &= \sum_{v : a^\star \succ_v S} u_v(a^\star) + \sum_{v : a^\star \not\succ_v S} u_v(a^\star) \\ &\leq |a^\star \succ S| + \sum_{v \in V} \sum_{b \in S} u_v(b) \\ &= |a^\star \succ S| + \sum_{b \in S} \mathsf{SW}(b).\end{aligned}$$

Since $\mathsf{SW}(D_S) = \frac{1}{k}\sum_{b \in S} \mathsf{SW}(b)$, the claim follows after rearranging.

Let $D_{k\text{-SL}}$ be the distribution of a stable k-lottery, and let D_k be the distribution which samples $S \sim D_{k\text{-SL}}$ and then chooses a uniformly random candidate from S. Using the guarantee of Theorem 2.2, we get

$$\mathsf{SW}(D_k) = \mathop{\mathbb{E}}_{S \sim D_{k\text{-SL}}}[\mathsf{SW}(D_S)] \geq \frac{\mathsf{SW}(a^\star) - \mathbb{E}_{S \sim D_{k\text{-SL}}}[|a^\star \succ S|]}{k} \geq \frac{\mathsf{SW}(a^\star)}{k} - \frac{n}{k(k+1)}.$$

In particular, $\mathsf{SW}(D_{\sqrt{m}}) \geq \frac{1}{\sqrt{m}}\mathsf{SW}(a^\star) - \frac{n}{m}$. Finally, the social welfare of the uniform distribution over candidates D_{Unif} is exactly n/m, since

$$\mathsf{SW}(D_{\text{Unif}}) = \frac{1}{m}\sum_{a \in C}\sum_{v \in V} u_v(a) = \frac{1}{m}\sum_{v \in V}\sum_{a \in C} u_v(a) = \frac{n}{m}$$

by the unit-sum assumption. Averaging over the two distributions, we get

$$\mathsf{SW}(\tfrac{1}{2}D_{\sqrt{m}} + \tfrac{1}{2}D_{\text{Unif}}) \geq \frac{1}{2}\left(\frac{1}{\sqrt{m}} \cdot \mathsf{SW}(a^\star) - \frac{n}{m}\right) + \frac{1}{2} \cdot \frac{n}{m} = \frac{1}{2\sqrt{m}} \cdot \mathsf{SW}(a^\star)$$

which gives a distortion bound of $O(\sqrt{m})$ as claimed. □

4.2 Metric distortion

In *metric distortion* [Anshelevich et al. 2015], we imagine that the voters and candidates lie in a metric space with a distance metric $d(\cdot, \cdot)$. Voters prefer candidates that are close to them, meaning that if $a \succ_v b$ then $d(v, a) \leq d(v, b)$. This idea is inspired by a vast literature on proximity spatial models of voting [Enelow and Hinich 1984; Enelow and Hinich 1990; Merrill and Grofman 1999; Armstrong et al. 2020], which includes concepts like the "political spectrum." The

social cost of a candidate a is $\mathsf{SC}(a) := \frac{1}{n}\sum_{v\in V} d(a,v)$, and a voting rule has distortion α if it chooses a candidate b such that $\mathsf{SC}(b) \le \alpha \min_{a\in C} \mathsf{SC}(a)$ in any election, and with any underlying metric space that is consistent with the preferences. For a randomized voting rule that chooses a distribution D over candidates, we replace $\mathsf{SC}(b)$ with $\mathbb{E}_{b\sim D}[\mathsf{SC}(b)]$.

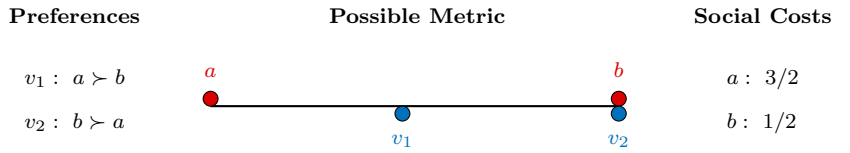


Fig. 4. An example of the metric distortion in an election with two candidates and two disagreeing voters. Candidates and voters lie on a line with a and b at -1 and 1, and one voter is impartial (at 0) while the other is collocated with their preferred candidate. In this case, one candidate has 3 times the cost of the other.

In contrast to the utilitarian distortion model, two differences stand out in the literature on metric distortion. First, despite the broad array of possible metric spaces, their structure is enough for several voting rules to have *constant* distortion, independent of the number of candidates. On the other hand, previously studied voting rules all fall short of the *optimal* distortion, and successive improvements in the best known distortion have come from designing several natural new voting rules.

Anshelevich et al. [2015] initially showed that all deterministic voting rules have distortion at least 3 (due to the simple example in Figure 4), and a long line of work culminated in proving a matching upper bound with simple voting rules [Gkatzelis et al. 2020; Kizilkaya and Kempe 2022; Kizilkaya and Kempe 2023]. Beyond these results, 3 turns out to be a critical threshold for the metric distortion problem as a whole. In particular, a distortion 3 candidate is in a way analogous to a Condorcet winner,[11] and other important classes of voting rules also have optimal distortion 3.[12]

With the distortion of deterministic voting rules resolved, attention turned to *randomized* voting rules. It turns out that getting distortion 3 in this setting is easy: Anshelevich and Postl [2017] and Feldman et al. [2016] independently showed that Random Dictator (which chooses a uniformly random voter's favorite candidate) achieves this. Both papers also noted that no randomized voting rule can have distortion less than 2 (by the example in Figure 4), which was conjectured to be optimal [Goel et al. 2017]. This conjecture was refuted independently by Charikar and Ramakrishnan [2022] and Pulyassary and Swamy [2021], with the former giving the best known lower bound of 2.112. Finally, Charikar et al. [2024] designed a new randomized voting rule with distortion 2.753, separating the best possible distortion of deterministic and randomized voting rules and breaking the barrier of 3.

[11] If a majority of voters prefers b over a then $\mathsf{SC}(b) \le 3\mathsf{SC}(a)$.
[12] Rules that exclusively choose from top-ranked candidates, and those that only need the aggregate preferences between pairs of candidates (called *weighted tournament rules*).

The voting rule used by Charikar et al. [2024] works by randomizing between a maximal lottery and a new voting rule called RaDiUS, which chooses a random voter's favorite candidate in the *Weighted Uncovered Set* (introduced by Munagala and Wang [2019]). Rather than detail the technical proof, we will highlight the way in which maximal lotteries are uniquely well-suited to this application.

Charikar et al. [2024] start by pinning down the distortion of a maximal lottery in isolation.

THEOREM 4.2 (CHARIKAR ET AL. [2024]). *In any election, a maximal lottery has distortion 3.*

Theorem 4.2 is not surprising if one thinks of a maximal lottery as a randomized analogue of a Condorcet winner, but the proof is not as straightforward as this correspondence would suggest. Instead of the full proof, we give an explanation of the additional structure the proof implies that is not present in proofs that other voting rules have distortion 3.

Suppose that our goal is to show that $\mathsf{SC}(b) \leq 3\mathsf{SC}(a^\star)$, or equivalently that $\mathsf{SC}(b) - \mathsf{SC}(a^\star) \leq 2\mathsf{SC}(a^\star)$ (for technical reasons, this rearrangement is cleaner to work with).

We can use the trick of expressing a mean as an integral over tails to write

$$\mathsf{SC}(b) - \mathsf{SC}(a^\star) = \int_0^\infty \Pr_{v\sim V}[d(v,b) - d(v,a^\star) > t]\,\mathrm{d}t$$

and

$$2\mathsf{SC}(a^\star) = \int_0^\infty \Pr_{v\sim V}[2d(v,a^\star) > t]\,\mathrm{d}t.$$

For the sake of clarity, we point out that in the first equation, it would be more correct to have the integral range from $-\infty$ to ∞, or for the $=$ to be replaced by $\leq$. It turns out that in the worst-case metrics, equality actually holds ($d(v,b) \geq d(v,a^\star)$ for all v). Also, for a distribution D in place of b, we would instead write

$$\mathop{\mathbb{E}}_{b\sim D}[\mathsf{SC}(b)] - \mathsf{SC}(a^\star) = \int_0^\infty \mathop{\mathbb{E}}_{b\sim D}\left[\Pr_{v\sim V}[d(v,b) - d(v,a^\star) > t]\right]\mathrm{d}t.$$

The proofs that earlier voting rules[13] have distortion 3 can all be interpreted as proving something stronger than $\mathsf{SC}(b) - \mathsf{SC}(a^\star) \leq 2\mathsf{SC}(a^\star)$. They actually show that for all $t \geq 0$,

$$\Pr_{v\sim V}[d(v,b) - d(v,a^\star) > t] \leq \Pr_{v\sim V}[2d(v,a^\star) > t]. \tag{2}$$

For readers familiar with this line of work, we take a beat to note that these proofs are not typically written in terms of (2). The usual objective is to construct a perfect matching M between two copies of the set of voters V such that for each pair $(v,u) \in M$, v weakly prefers b over u's favorite candidate. It is not hard to use the triangle inequality to show that such a pair must satisfy $d(v,b) - d(v,a^\star) \leq 2d(u,a^\star)$,

[13]Random Dictator [Anshelevich and Postl 2017; Feldman et al. 2016], Plurality Matching [Gkatzelis et al. 2020], Plurality Veto [Kizilkaya and Kempe 2022], Simultaneous Plurality Veto [Kizilkaya and Kempe 2023].

and with a little more effort, one can see that constructing the matching M is equivalent to showing (2).

The *pièce de résistance* is that the proof that maximal lotteries have distortion 3 shows that in the *worst-case* metrics (termed *biased* metrics by Charikar and Ramakrishnan [2022]), something even stronger than (2) holds. For all $t \geq 0$

$$\mathop{\mathbb{E}}_{b \sim D_{\mathrm{ML}}} \left[\Pr_{v \sim V} [d(v,b) - d(v,a^\star) > t] \right] \leq \min \left(\frac{1}{2}, \Pr_{v \sim V} [2d(v,a^\star) > t] \right). \tag{3}$$

To give a window into the proof, the $\frac{1}{2}$ in this expression is no coincidence. It comes directly from Theorem 2.1.

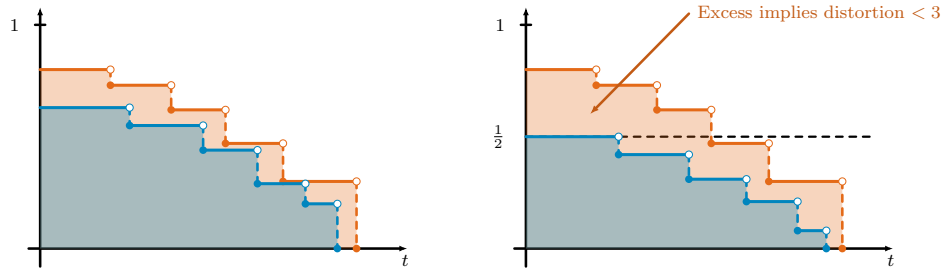


Fig. 5. The blue regions show $\mathsf{SC}(b) - \mathsf{SC}(a^\star) = \int_0^\infty \Pr_{v \sim V}[d(v,b) - d(v,a^\star) > t]\,\mathrm{d}t$, and the orange regions show $2\mathsf{SC}(a^\star) = \int_0^\infty \Pr_{v \sim V}[2d(v,a^\star) > t]\,\mathrm{d}t$. The diagram on the left illustrates the condition (2) usually satisfied when b is chosen by a voting rule with distortion at most 3, while the right illustrates the condition (3) satisfied when b is sampled from a maximal lottery.

The power of the stronger condition (3) is that it implies a precise structure in the metrics where maximal lotteries have high distortion (close to 3). Figure 5 helps show why: if the distortion of maximal lotteries is close to 3, then the function $\Pr_{v \sim V}[2d(v,a^\star) > t]$ (in orange) must be close to $\frac{1}{2}$ for small values of t, meaning that nearly half the voters are tightly clustered around the optimal candidate $a^\star$. The RaDiUS rule complements maximal lotteries by having particularly small distortion in these structured metrics, and so mixing the two voting rules together covers all the bases to beat 3.

Finally, we conclude by mentioning a couple of follow-up results that use stable lotteries in addition to maximal lotteries.

Charikar et al. [2025] study the metric distortion of *k-tournament* rules, whose input is the aggregate preferences of the voters for each set of k candidates (the quantities $\frac{1}{n}|a_1 \succ \cdots \succ a_k|$ for each $(a_1, \ldots, a_k) \in C^k$). They show that a k-tournament rule can have distortion $3 + \tilde{O}(\frac{1}{k^{1/4}})$ without randomness, and strictly less than 3 with randomness even for $k = 3$. Both results crucially make use of stable lotteries, leveraging the intuition that they are similar to random dictatorships for large k.

Cai et al. [2026] study the minimal amount of randomness needed to beat distortion 3, and show that it is possible while randomizing over $O(1)$ candidates. Their voting rule is again built on top of maximal lotteries and stable lotteries, using

the fact that they can be approximated with small support (generalizing Theorem 3.13). An interesting open question raised by this work is whether a voting rule which randomizes between at most two candidates can have distortion less than 3.

Conjecture 4.3 (Cai et al. [2026]). *There exists a voting rule which randomizes between at most two candidates and has distortion* $3 - \varepsilon$ *for some fixed* $\varepsilon > 0$.

An exciting aspect of this conjecture is that resolving it plausibly requires completely new techniques.

5. CONCLUSION

Throughout this survey, we have seen myriad ways of using maximal lotteries and stable lotteries to prove positive results. To illustrate the point once again, even though Condorcet's paradox is one of the oldest and most fundamental challenges in voting theory, we discussed at least *four* different relaxations of a Condorcet winner that are universally achievable (maximal lotteries, Condorcet winning sets, approximately dominating sets, distortion 3 and below). The broad applicability of these tools is also incredibly exciting from the perspective of a theoretician. It suggests that they are not overly sensitive to the idiosyncrasies of a particular setting, and instead may capture some part of the more fundamental principles that the field broadly aims for, such as representation, efficiency, and fairness. For this reason, maximal lotteries, stable lotteries, and their relatives deserve a prominent place in any voting theorist's toolkit.

What is especially promising is that so many of these results are recent, and there is no indication that we have found all or even most of the domains of social choice where game-theoretic rules could be impactful. One possible domain on the horizon is pluralistic AI alignment. Recently, researchers have proposed alignment methods based on maximal lotteries, and several papers have made a compelling case for these approaches, both theoretically and experimentally [Munos et al. 2024; Maura-Rivero et al. 2025; Khalaf et al. 2026; Gölz et al. 2026].[14]

Finally, we hope that these new results and directions also motivate the community to step back and develop a deep and comprehensive theory of probabilistic tools in social choice. In doing so, we can find ideas that are flexible enough to apply across the diverse range of scenarios that voting can model, and benefit from the compounding effects of rich theory and positive results.

Acknowledgments. We thank editors-in-chief Nick Arnosti and Sam Taggart for helpful feedback that greatly improved the exposition of the survey.

REFERENCES

Alon, N., Brightwell, G., Kierstead, H. A., Kostochka, A. V., and Winkler, P. 2006. Dominating sets in k-majority tournaments. *Journal of Combinatorial Theory, Series B 96,* 3, 374–387.

[14]See also [Conitzer et al. 2024; Ge et al. 2024; Fish et al. 2026] for some broader agenda-setting work at the intersection of social choice and AI.

Anshelevich, E., Bhardwaj, O., and Postl, J. 2015. Approximating optimal social choice under metric preferences. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI)*. 777–783. 91

Anshelevich, E., Filos-Ratsikas, A., Shah, N., and Voudouris, A. A. 2021. Distortion in social choice problems: The first 15 years and beyond. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*. 4294–4301.

Anshelevich, E. and Postl, J. 2017. Randomized social choice functions under metric preferences. *J. Artif. Intell. Res. 58*, 797–827. 93

Armstrong, D. A., Bakker, R., Carroll, R., Hare, C., Poole, K. T., and Rosenthal, H. 2020. *Analyzing spatial models of choice and judgment*. Chapman and Hall/CRC. 91

Aziz, H. 2019. Justifications of welfare guarantees under normalized utilities. *SIGecom Exch. 17,* 2, 71–75.

Becker, P., Greger, M., and Peters, D. 2026. Core existence in approval-based committee elections with up to five voter types.

Bourneuf, R., Charbit, P., and Thomassé, S. 2025. A dense neighborhood lemma: Applications of partial concept classes to domination and chromatic number. In *Proceedings of the 66th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 1–37.

Boutilier, C., Caragiannis, I., Haber, S., Lu, T., Procaccia, A. D., and Sheffet, O. 2015. Optimal social choice functions: A utilitarian view. *Artif. Intell. 227*, 190–213.

Brandl, F., Brandt, F., and Stricker, C. 2018. An analytical and experimental comparison of maximal lottery schemes. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI)*. ijcai.org, 114–120.

Brandt, F. 2017. Rolling the dice: Recent results in probabilistic social choice. In *Trends in computational social choice*. AI Access, Lulu, 3–26.

Cai, Z., Gao, D. D., Ramakrishnan, P., and Wang, K. 2026. Distortion of metric voting with bounded randomness. In *Proceedings of the 27th ACM Conference on Economics and Computation (EC)*.

Charikar, M., Lassota, A., Ramakrishnan, P., Vetta, A., and Wang, K. 2025. Six candidates suffice to win a voter majority. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing (STOC)*. ACM, 1590–1601.

Charikar, M. and Ramakrishnan, P. 2022. Metric distortion bounds for randomized social choice. In *Proceedings of the 2022 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2986–3004.

Charikar, M., Ramakrishnan, P., Tan, Z., and Wang, K. 2025. Metric distortion for tournament voting and beyond. In *Proceedings of the 26th ACM Conference on Economics and Computation (EC)*. ACM, 790–818.

Charikar, M., Ramakrishnan, P., and Wang, K. 2026a. Approximately dominating sets in elections. In *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 1747–1760.

Charikar, M., Ramakrishnan, P., and Wang, K. 2026b. An exposition of five candidates suffice for a majority. 83

Charikar, M., Ramakrishnan, P., Wang, K., and Wu, H. 2024. Breaking the metric voting distortion barrier. *Journal of the ACM 71,* 6, 1–33.

Cheng, Y., Jiang, Z., Munagala, K., and Wang, K. 2020. Group fairness in committee selection. *ACM Transactions on Economics and Computation (TEAC) 8,* 4, 23:1–23:18. 80

Conitzer, V., Freedman, R., Heitzig, J., Holliday, W. H., Jacobs, B. M., Lambert, N., Mossé, M., Pacuit, E., Russell, S., Schoelkopf, H., Tewolde, E., and Zwicker, W. S. 2024. Position: social choice should guide AI alignment in dealing with diverse human feedback. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*. 9346–9360. 95

Ebadian, S., Kahng, A., Peters, D., and Shah, N. 2024. Optimized distortion and proportional fairness in voting. *ACM Transactions on Economics and Computation (TEAC) 12,* 1, 3:1–3:39.

Elkind, E., Lang, J., and Saffidine, A. 2015. Condorcet winning sets. *Social Choice and Welfare 44,* 3, 493–517.

Enelow, J. M. and Hinich, M. J. 1984. *The spatial theory of voting: An introduction.* CUP Archive. 91

Enelow, J. M. and Hinich, M. J. 1990. *Advances in the spatial theory of voting.* Cambridge University Press. 91

Erdős, P. 1963. On a problem in graph theory. *The Mathematical Gazette 47,* 361, 220–223.

Faliszewski, P., Skowron, P., Slinko, A., and Talmon, N. 2017. Multiwinner voting: A new challenge for social choice theory. *Trends in computational social choice 74,* 2017, 27–47.

Feldman, M., Fiat, A., and Golomb, I. 2016. On voting and facility location. In *Proceedings of the 2016 ACM Conference on Economics and Computation (EC)*. 269–286. 93

Fish, S., Gölz, P., Parkes, D., Procaccia, A., Rusak, G., Shapira, I., and Wuthrich, M. 2026. Generative social choice. *Journal of the ACM 73,* 2, 1–52. 95

Fishburn, P. 1984. Probabilistic social choice based on simple voting comparisons. *The Review of Economic Studies 51,* 4, 683–692. 79

Gao, D. D., Sun, Y., and Vondrák, J. 2026. Computation of approximately stable committees in approval-based elections. In *Proceedings of the 27th ACM Conference on Economics and Computation (EC)*.

Ge, L., Halpern, D., Micha, E., Procaccia, A. D., Shapira, I., Vorobeychik, Y., and Wu, J. 2024. Axioms for AI alignment from human feedback. In *Conference on Neural Information Processing Systems (NeurIPS)*. 95

Gehrlein, W. V. 2002. Condorcet's paradox and the likelihood of its occurrence: Different perspectives on balanced preferences. *Theory and Decision 52,* 2, 171–199.

Gkatzelis, V., Halpern, D., and Shah, N. 2020. Resolving the optimal metric distortion conjecture. In *Proceedings of the 61st IEEE Annual Symposium on Foundations of Computer Science (FOCS)*. 1427–1438. 92, 93

Goel, A., Krishnaswamy, A. K., and Munagala, K. 2017. Metric distortion of social choice rules: Lower bounds and fairness properties. In *Proceedings of the 2017 ACM Conference on Economics and Computation (EC)*. 287–304. 92

Gölz, P., Haghtalab, N., and Yang, K. 2026. Distortion of AI alignment: Does preference optimization optimize for preferences? *Advances in Neural Information Processing Systems (NeurIPS) 38*, 25969–26007. 95

Jiang, Z., Munagala, K., and Wang, K. 2020. Approximately stable committee selection. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*. ACM, New York, 463–472.

Khalaf, H., Wang, S. L., Halpern, D., Shapira, I., du Pin Calmon, F., and Procaccia, A. D. 2026. Robust AI evaluation through maximal lotteries. In *Proceedings of the 43rd International Conference on Machine Learning (ICML)*. 95

Kizilkaya, F. E. and Kempe, D. 2022. Plurality veto: A simple voting rule achieving optimal metric distortion. In *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI)*. 349–355. 92, 93

Kizilkaya, F. E. and Kempe, D. 2023. Generalized veto core and a practical voting rule with optimal metric distortion. In *Proceedings of the 24th ACM Conference on Economics and Computation (EC)*. ACM, 913–936. 92, 93

Kreweras, G. 1965. Aggregation of preference orderings. In *Mathematics and Social Sciences I: Proceedings of the seminars of Menthon-Saint-Bernard, France (1–27 July 1960) and of Gösing, Austria (3–27 July 1962)*. 73–79. 79

Lackner, M. and Skowron, P. 2023. *Multi-winner voting with approval preferences.* Springer Nature.

Lin, Y., Qin, S., Wang, K., and Xia, L. 2026. Winning in the limit: Average-case committee selection with many candidates. In *Proceedings of the 27th ACM Conference on Economics and Computation (EC)*.

Maura-Rivero, R.-R., Lanctot, M., Visin, F., and Larson, K. 2025. Jackpot! Alignment as a maximal lottery. *arXiv preprint arXiv:2501.19266*. 95

McGarvey, D. C. 1953. A theorem on the construction of voting paradoxes. *Econometrica: Journal of the Econometric Society*, 608–610. 87

Merrill, S. and Grofman, B. 1999. *A unified theory of voting: Directional and proximity spatial models*. Cambridge University Press. 91

Moulin, H. 1988. *Axioms of Cooperative Decision Making*. Econometric Society Monographs, vol. 15. Cambridge University Press. 85

Munagala, K. and Wang, K. 2019. Improved metric distortion for deterministic social choice rules. In *Proceedings of the 2019 ACM Conference on Economics and Computation (EC)*. 245–262.

Munos, R., Valko, M., Calandriello, D., Azar, M. G., Rowland, M., Guo, Z. D., Tang, Y., Geist, M., Mesnard, T., Fiegel, C., Michi, A., Selvi, M., Girgin, S., Momchev, N., Bachem, O., Mankowitz, D. J., Precup, D., and Piot, B. 2024. Nash learning from human feedback. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*. 95

Peters, D. 2025. The core of approval-based committee elections with few seats. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*. ijcai.org, 4014–4022.

Procaccia, A. D. and Rosenschein, J. S. 2006. The distortion of cardinal preferences in voting. In *Proceedings of the 10th International Workshop on Cooperative Information Agents (CIA)*. 317–331. 89

Pulyassary, H. and Swamy, C. 2021. On the randomized metric distortion conjecture. *arXiv preprint arXiv:2111.08698*.

Song, H. and Nguyen, T. 2026. Ordinal Lindahl equilibrium for voting. In *Proceedings of the 27th ACM Conference on Economics and Computation (EC)*.

Song, H., Nguyen, T., and Lin, Y. 2026. A few good choices. In *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 4861–4874.

von Neumann, J. 1928. Zur theorie der gesellschaftsspiele. *Mathematische annalen 100,* 1, 295–320. 80

An Online Learning Perspective on Bilateral Trade

MARTINO BERNASCONI
Bocconi University
and
FEDERICO FUSCO
Sapienza University of Rome

Bilateral trade models the mediation between two strategic agents—a buyer and a seller—wishing to exchange a good. While both economists and computer scientists have extensively studied this problem in the Bayesian setting, recent efforts have increasingly focused on mechanisms that operate *without* prior knowledge of the agents' valuations. This article surveys these recent advancements with a focus on *online learning*. We present the literature along three primary axes: gain-from-trade maximization, profit maximization, and related models. Beyond summarizing state-of-the-art results, we aim to build intuition for the rich technical toolkit emerging from this research—techniques we believe will prove valuable well beyond the specific domain of bilateral trade.

1. INTRODUCTION

In the bilateral trade problem, two strategic agents—a seller and a buyer—wish to exchange a good. Both agents hold private valuations for the item and seek to maximize their respective quasi-linear utilities, while the task of designing a mechanism to reach agreement is usually delegated to a third party. This scenario arises naturally in various digital platforms, such as ride-sharing services like Uber or Lyft, where trades between sellers (drivers) and buyers (riders) are managed by a centralized mechanism.

An extensive line of research, building on the seminal work of Myerson and Satterthwaite [1983], has investigated this problem from a Bayesian perspective: the agents' valuations are drawn from known and independent distributions, and the mechanism designer aims to maximize either efficiency or profit while enforcing incentive compatibility and maintaining budget balance (i.e., without subsidizing the agents). The Bayesian approach relies heavily on the assumption that agents draw their valuations from independent distributions whose exact laws are known to the mechanism designer. In this paper, we survey a recent body of work that relaxes this assumption by adopting a machine learning perspective, addressing the following central question:

Under which conditions and how fast is it possible to learn a good mechanism for bilateral trade?

In particular, we focus on the online learning framework introduced by Cesa-Bianchi et al. [2021], where a good mechanism must be learned *on the fly* while minimizing the regret incurred from mistakes made along the way. Beyond summarizing the core results of this literature, this survey also aims to convey the fundamental technical

Authors' addresses: martino.bernasconi@unibocconi.it, fuscof@diag.uniroma1.it

ACM SIGecom Exchanges, Vol. 24, No. 1, June 2026, Pages 99–122

tools developed in this space, which we believe are of general interest to the broader community.

The remainder of this survey is organized as follows. First, Section 2 formally introduces the bilateral trade problem and its online learning formulation. Next, Sections 3 and 4 present the primary results and techniques for gain-from-trade and profit maximization, respectively. Section 5 then reviews alternative models that have stemmed from the foundational online learning protocol. Throughout the text, we highlight prominent open problems and promising directions for future research.

2. BILATERAL TRADE

We start by introducing the "one-shot" version of bilateral trade, where the agents' valuations are fixed but private. Note that the main focus of this survey is the online learning version of this model, in which a learner is presented with a sequence of T instances of this problem.

In the (static) bilateral trade problem, a seller and a buyer wish to trade a good via an intermediary. The seller holds a single item and is willing to trade it for at least $v_s \in [0,1]$, while the buyer is willing to pay at most $v_b \in [0,1]$ in exchange. The valuations of the two agents are private, so they may misreport them strategically; indeed, they submit bids $(b_s, b_b) \in [0,1]^2$ (not necessarily truthfully) to an intermediary who decides whether the trade happens and under what conditions. More precisely, the intermediary runs a mechanism M, characterized by an allocation region $A \subseteq [0,1]^2$, and pricing rules $p, q : [0,1]^2 \to [0,1]$. The trade happens if and only if the bids (b_s, b_b) belong to the allocation region A, while the payments are made according to p and q. For simplicity, we require $p(b_s, b_b) = q(b_s, b_b) = 0$ whenever $(b_s, b_b) \notin A$, i.e., there is no trade.

The agents' utilities have the standard quasi-linear structure: the buyer's and seller's utilities with valuations v_b and v_s under bids $(b_s, b_b) \in [0,1]^2$ are:

$$U_s(b_s, b_b) = v_s - \mathbb{1}_{\{(b_s,b_b)\in A\}} \cdot v_s + p(b_s, b_b) \qquad \text{(Seller's utility)}$$
$$U_b(b_s, b_b) = \mathbb{1}_{\{(b_s,b_b)\in A\}} \cdot v_b - q(b_s, b_b). \qquad \text{(Buyer's utility)}$$

We focus on dominant-strategy incentive compatible (DSIC) and individually rational (IR) mechanisms: each agent maximizes its utility by truthfully reporting its actual private valuation—regardless of the other player's bid—and the utility from participating in the mechanism is at least as high as that from not participating in the mechanism. In formulae,

$$\begin{aligned}
\text{DSIC:} \quad & U_s(v_s, b_b) \geq U_s(b_s, b_b) && \forall\, v_s \in [0,1], (b_s, b_b) \in [0,1]^2 \\
& U_b(b_s, v_b) \geq U_b(b_s, b_b) && \forall\, v_b \in [0,1], (b_s, b_b) \in [0,1]^2 \\
\text{IR:} \quad & U_s(v_s, b_b) \geq v_s,\ U_b(b_s, v_b) \geq 0 && \forall\, (v_s, v_b) \in [0,1]^2, (b_s, b_b) \in [0,1]^2
\end{aligned}$$

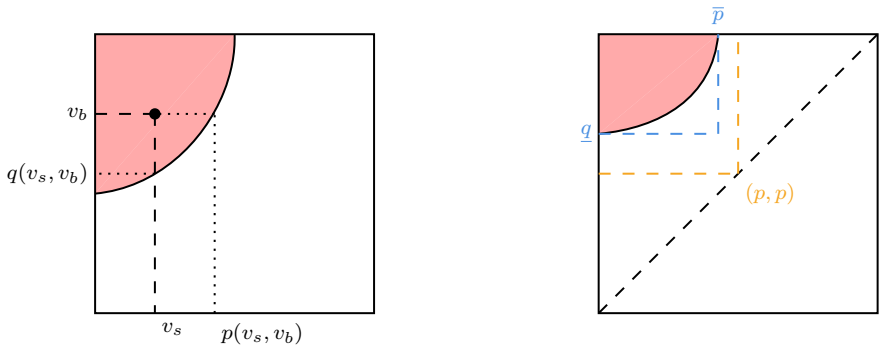


Fig. 1: Visualization of Proposition 2.2 (left) and Proposition 2.5 (right).

2.1 Characterization of DSIC and IR Mechanisms.

Standard mechanism design arguments prescribe that every (deterministic)[1] DSIC and IR mechanisms have a distinctive structure: the payments are uniquely induced by the allocation region, which, in turn, has to respect a monotonicity property. We denote the class of all DSIC and IR mechanisms by $\mathcal{M}$ and restrict to it throughout the rest of the paper, so that we consistently assume that the players' bids are equal to their respective valuations.

Definition 2.1 Monotone Regions & Myerson Payments. *An allocation region $A \subseteq [0,1]^2$ is monotone if for any $x = (x_1, x_2) \in A$ and $y = (y_1, y_2) \in [0,1]^2$ with $x_1 \geq y_1$ and $x_2 \leq y_2$, it holds that $y \in A$. A mechanism is monotone if its allocation region is monotone. Given a monotone allocation region, the associated Myerson payments are defined as follows:*

$$p(v_s, v_b) = \mathbb{1}_{\{(v_s,v_b)\in A\}} \cdot \max\{x \in [0,1] : (x, v_b) \in A\} \qquad \forall\, (v_s, v_b) \in [0,1]^2$$
$$q(v_s, v_b) = \mathbb{1}_{\{(v_s,v_b)\in A\}} \cdot \min\{y \in [0,1] : (v_s, y) \in A\} \qquad \forall\, (v_s, v_b) \in [0,1]^2$$

In other words, for an allocation region to be monotone, it should be "closed" in a north-west direction. That is, if a point (v_s, v_b) is in A then any point (v_s', v_b) with $v_s' \leq v_s$ and any point (v_s, v_b') with $v_b' \geq v_b$ should also be in A. The payments of a point $(v_s, v_b) \in A$, in turn, correspond to the "east" projection minus the "south" one onto the allocation boundary. We refer to Figure 1 (left) for visualization. Note that for the above definition to be well-posed, we require (without loss of generality) all allocation regions to be topologically closed. By standard mechanism-design arguments [Myerson 1981; Myerson and Satterthwaite 1983], monotone allocation regions and Myerson payments do characterize $\mathcal{M}$ (explicit proof is provided in the Appendices of Di Gregorio et al. [2025]).

Proposition 2.2. *A mechanism for bilateral trade is dominant-strategy incentive compatible and individually rational if and only if its allocation region is monotone and implements Myerson payments.*

[1]Given the *deterministic* nature of the constraints we consider in this survey, we restrict without loss of generality to deterministic mechanisms. Indeed, once the agents' valuations are fixed, any randomized DSIC and IR mechanism can be written as a distribution over deterministic DSIC and IR mechanisms.

2.2 Gain From Trade, Profit, and Budget Balance

Proposition 2.2 describes how a mechanism for bilateral trade should look for the agents to participate truthfully in it. Switching perspectives, we present the most common metrics for assessing a mechanism's quality from the intermediary's perspective: social welfare and profit.

The social welfare induced by a mechanism is equal to the valuation of the agent holding the good *in the end*. More precisely, the social welfare obtained by a generic mechanism $M \in \mathcal{M}$ with allocation region A, when the underlying valuations are (v_s, v_b) is

$$\mathsf{SW}(M, v_s, v_b) = v_b \cdot \mathbb{1}_{\{(v_s,v_b)\in A\}} + v_s \cdot \mathbb{1}_{\{(v_s,v_b)\notin A\}}.$$

An important metric related to social welfare is the gain from trade, which measures the *increase* in social welfare relative to the initial state in which the seller holds the good. In formula, the gain from trade is

$$\mathsf{GFT}(M, v_s, v_b) = (v_b - v_s) \cdot \mathbb{1}_{\{(v_s,v_b)\in A\}}.$$

Note these two metrics are just an additive factor away from each other, as $\mathsf{SW}(M, v_s, v_b) = \mathsf{GFT}(M, v_s, v_b) + v_s$. Social welfare and gain from trade both measure *economic efficiency*, and they are maximized at the same mechanism (while from the "multiplicative" approximation perspective, the latter is more challenging than the former).

The profit extracted by the intermediary from a mechanism $M \in \mathcal{M}$ when the underlying valuations are (v_s, v_b) is simply the difference between the price paid *by* the buyer and the amount paid *to* the seller. In formula:

$$\mathsf{profit}(M, v_s, v_b) = (q(v_s, v_b)) - p(v_s, v_b)) \cdot \mathbb{1}_{\{(v_s,v_b)\in A\}}.$$

So far, it seems there is nothing special about bilateral trade within the realm of mechanism design. However, there is a crucial catch, which was already highlighted in the seminal paper by Vickrey [1961]. Let's say you want to design a mechanism to maximize social welfare, while still enforcing incentive compatibility and individual rationality. Then it makes sense to give the good to the buyer as soon as $v_b \geq v_s$.

Example 2.3. *Consider running the mechanism that allocates in the triangle* $\{(v_s, v_s) \in [0,1]^2 : v_b \geq v_s\}$. *If the agents' valuations are* $(v_s, v_b) = (1/3, 2/3)$, *then the trade happens but induces a* negative profit *for the intermediary, as it pays* $2/3$ *to the seller and receives only* $1/3$ *from the buyer.*

Stated differently, *perfect efficiency may require the mechanism designer to lose money*. To avoid situations like this, where the intermediary *subsidizes* the market, the economic literature has introduced the budget balance constraint, which requires that a social-welfare-oriented mechanism should not incur a negative profit [Myerson and Satterthwaite 1983].

Definition 2.4. *A mechanism* $M \in \mathcal{M}$ *respects the budget balance condition if* $\mathsf{profit}(M, v_s, v_b) \geq 0$, *for all* $(v_s, v_b) \in [0,1]^2$.

In some cases, one may want to require an even stronger condition, so that the profit extracted by the mechanism is exactly 0. This latter condition is called *strong* budget balance (while the one in the above definition is sometimes referred to as

weak budget balance). Surprisingly, the subfamily of $\mathcal{M}$ that enforces strong budget balance has a very simple characterization. Consider a generic mechanism $M \in \mathcal{M}$, and let $\overline{p}$, respectively $\underline{q}$, be the price paid to the seller, respectively by the buyer, under valuations $(0, 1)$; then M enforces budget balance if and only if $\underline{p} \leq \overline{q}$.[1]

A crucial role is played by the family of *fixed-price-mechanisms*. A mechanism $M_{p,q}$ belongs to this family if its allocation region is a rectangle of the form $[0, p] \times [q, 1]$, i.e., if it posts price p to the seller and q to the buyer and the trade happens if and only if both agents accept their price. If the two prices coincide, i.e., $p = 1$, then we omit one of the two prices and use M_p to denote it.

Fixed-price mechanisms enjoy the desirable properties of being simple and requiring minimal communication between the intermediary and the agents. Moreover, they somewhat characterize budget-balance mechanisms.[2] More precisely, fixed-price mechanisms are welfare-optimal among all DSIC mechanisms enforcing budget balance and individual rationality.

PROPOSITION 2.5. *Let $M \in \mathcal{M}$ be any budget-balanced mechanism, and denote with $\overline{p}$ and $\underline{q}$ the prices corresponding to valuation $(0, 1)$. Then the following inequalities hold for any $(v_s, v_b) \in [0, 1]^2$ and any $p \in [\overline{p}, \underline{q}]$:*

$$\mathsf{SW}(M, v_s, v_b) \leq \mathsf{SW}(M_{\overline{p},\underline{q}}, v_s, v_b) \leq \mathsf{SW}(M_p, v_s, v_b).$$

There is a simple way of picturing the above proposition, given that none of the mechanisms forces a trade with negative gain from trade (i.e., when $v_s > v_b$): the allocation region of the initial mechanism A is contained into the $[0, \overline{p}] \times [\underline{q}, 1]$ square which, in turns, is contained in the $[0, p] \times [p, 1]$ rectangle, for any $p \in [\overline{p}, \underline{q}]$. We refer to Figure 1 (right) for a "proof-by-picture". We conclude by noting that, even though fixed-price mechanisms are welfare-optimal, they are not generally profit-optimal among budget-balanced mechanisms. To see this, consider the following example, where the private valuations are not fixed but drawn from some distribution.

EXAMPLE 2.6. *Consider the uniform distribution over valuations in the $(0, 1/2)$–$(\delta, 1)$ segment, where δ is an arbitrarily small parameter. The mechanism allocating in the $(0, 1/2)$–$(\delta, 1)$–$(0, 0)$ triangle would extract an expected profit of $\approx 3/4$, while any fixed price mechanism would not extract more than $\approx 1/2$.*

2.3 A Quick Digression: The Bayesian Perspective

Although this is a survey on bilateral trade and online learning, we devote some quick words to describe the Bayesian perspective and the corresponding results. In that line of work, the agents' private valuations are drawn *once* from some independent and *known* distributions. First, we recall that the seminal paper by Myerson and Satterthwaite [1983] shows that incentive compatibility, individual rationality, and budget balance are incompatible with full efficiency. Actually, this

[1] One implication is easy, since a budget balanced M should not lose money for valuations $(0, 1)$. The other implication follows by monotonicity. The price $\underline{p}$ upper bounds the payments to the seller under *any* valuation in the allocation region, and similarly $\overline{q}$ lower bounds the payments made by the buyer in the same situation.

[2] Already Hagerty and Rogerson [1987] showed that fixed-price mechanisms that post a single price to both agents are the only possible DSIC, IR, and strong budget-balanced mechanisms.

result holds for the weaker notions of (Bayesian) incentive compatibility, (interim) individual rationality, and budget balance which only need to hold *in expectation*.

Given this impossibility result, the most natural question that arises is whether a constant *multiplicative* factor of perfect efficiency is attainable while enforcing the other constraints. For social welfare, Blumrosen and Dobzinski [2021] provide a $1-1/e$ approximation, later improved to $1 - 1/e + 10^{-4}$ in Kang et al. [2022]. The current state of the art is that the optimal approximation factor achievable is contained in the $[0.72, 0.7381]$ interval [Cai and Wu 2023; Liu et al. 2023]. Interestingly, it turns out that fixed-price mechanisms are insufficient to extract a constant factor of the optimal gain from trade [Blumrosen and Mizrahi 2016]. Instead, the random-offerer mechanism, which randomly delegates the pricing power to the seller or the buyer, can provide such a guarantee [Deng et al. 2025] (note, the random-offerer mechanism is no longer DSIC, but only ensures truthfulness *in expectation*). The state of the art places the optimal gain-from-trade approximation achievable in the $[1/3.15, 2/e]$ interval [Fei 2022; Blumrosen and Mizrahi 2016]. We also mention a recent paper [Hajiaghayi et al. 2025] studying the interplay between gain from trade and profit maximization, investigating how much gain from trade a profit-maximizing intermediary can ensure.

We finally note that the online learning approach is not the only approach adopted in the literature to overcome the Bayesian assumption (i.e., *independent* agents' distributions that are *perfectly known* to the mechanism designer). For instance, a line of work has investigated the number of samples (either one or more) needed to get a meaningful multiplicative approximation to the optimal mechanisms [Kang et al. 2022; Hajiaghayi et al. 2025; Dütting et al. 2026], while Dobzinski et al. [2025] investigate the one-shot problem when the agents' valuations may exhibit correlation.

2.4 The Learning Protocol

We are ready to introduce the main character of the survey: the learning protocol for bilateral trade, originally introduced in the conference version of Cesa-Bianchi et al. [2024a]. At each time step t, a new pair of agents arrives, with private valuations (v_s^t, v_b^t), while the learner/intermediary proposes a mechanism $M^t \in \mathcal{M}$. The trade then happens according to the valuations and the mechanism, with the learner observing some feedback z_t. We refer to the pseudocode for further details. For simplicity, we use $\mathsf{GFT}_t(M)$, respectively $\mathsf{profit}_t(M)$, to denote the gain from

The Learning Protocol

for time $t = 1, 2, \dots$ **do**
- a new pair of agents arrives with private valuations $(v_s^t, v_b^t) \in [0,1]^2$
- the learner declares a mechanism $M^t \in \mathcal{M}$
- the trade takes place according to M^t and (v_s^t, v_b^t)
- the learner observes some feedback z_t
- the learner is awarded either $\mathsf{GFT}(M^t, v_s^t, v_b^t)$ or $\mathsf{profit}(M^t, v_s^t, v_b^t)$

trade, respectively profit, of the generic mechanism M on the valuations (v_s^t, v_b^t)

at time t. The learning protocol offers many degrees of freedom beyond the profit vs. gain-from-trade dichotomy. We list here some of them:

- **Feedback Models.** There are many meaningful feedback models. For instance, for a general mechanism, it makes sense to directly observe both valuations, as they are needed to compute Myerson payments. If we restrict ourselves to fixed-price mechanisms, however, then a simple yes/no answer to the proposed prices may be enough.
- **Data-Generation Models.** It is possible to study many adversaries generating the sequence of valuations. Beyond the standard stochastic i.i.d. and the adversarial ones, an interesting model is provided by the σ-smooth adversary.
- **Budget Balance Constraints.** In gain-from-trade maximization scenarios, it makes sense to investigate the impact of budget balance constraints on the learnability of the problem. Beyond strong and weak *per-round* budget balance, the literature has also investigated more global versions, such as constraining the overall (i.e., cumulative) negative profit.

As is standard in online learning, a learning mechanism is measured by its *regret* relative to a fixed benchmark. In formula, the regret of a gain-from-trade maximizing algorithm $\mathcal{A}$ against an adversary $\mathcal{S}$ is

$$R_T(\mathcal{A}, \mathcal{S}) = \sup_{M \in \mathcal{M}_{\text{BB}}} \mathbb{E}\left[\sum_{t=1}^{T} \mathsf{GFT}_t(M^t) - \mathsf{GFT}_t(M)\right],$$

where the algorithm chooses budget-balanced mechanisms $M^t \in \mathcal{M}$, while the adversary generates the valuation pairs (v_s^t, v_b^t) (and thus $\mathsf{GFT}_t(\cdot)$). Here we use $\mathcal{M}_{\text{BB}}$ to stress that the benchmark is in $\mathcal{M}$ and is also restricted to respect budget balance. In practice, Proposition 2.5 tells us that the benchmark is the best fixed price mechanism for $\mathcal{S}$. The regret of algorithm $\mathcal{A}$ against a certain family of adversaries (e.g., stochastic i.i.d) is denoted as the worst case regret within that family: $R_T(\mathcal{A}) = \sup_{\mathcal{S}} R_T(\mathcal{A}, \mathcal{S})$. Note that gain from trade and social welfare are identical up to an additive v_s; therefore, studying either of the two is perfectly equivalent in the regret minimization world.

For profit-maximizing algorithms, the definition is analogous, with the sole difference that there is no need to explicitly restrict the algorithm and the benchmark to budget-balanced mechanisms:

$$R_T(\mathcal{A}, \mathcal{S}) = \sup_{M \in \mathcal{M}} \mathbb{E}\left[\sum_{t=1}^{T} \mathsf{profit}_t(M^t) - \mathsf{profit}_t(M)\right].$$

3. GAIN FROM TRADE

In this section, we present the line of research relative to gain-from-trade-maximizing learning algorithms. Motivated by Proposition 2.5, these works focus on fixed-price mechanisms, which is particularly interesting from the online learning perspective. Indeed, fixed-price mechanisms allow for peculiar feedback models: when the learner/intermediary posts two prices, one for the seller and one for the buyer, it is natural to only observe whether the agents accept or not, without eliciting their actual valuations. The literature mainly studies two feedback models: the one-bit

model, where the learner only observes whether the trade happens or not, and the two-bit one, where the learner separately observes the answers of both agents.[1]

In the presentation, we put more emphasis on the most challenging (and arguably natural) setting: one-bit feedback and non-stationary valuations under different budget-balance conditions. Beyond these two *partial* feedback models, the literature has also investigated the paradigmatic full feedback setting as a reference.

We introduce some simplified notation for the fixed-price setting: At each time t, the learner proposes a price p_t to the seller (with valuation v_s^t) and a price q_t to the buyer (with valuation v_b^t). Then the gain from trade simplifies to

$$\mathsf{GFT}_t(p_t, q_t) = (v_b^t - v_s^t)\mathbb{1}_{\{v_s^t \le p_t \wedge q_t \le v_b^t\}},$$

Similarly, the profit becomes $\mathsf{profit}_t(p_t, q_t) = (q_t - p_t)\mathbb{1}_{\{v_s^t \le p_t \wedge q_t \le v_b^t\}}$. Given Proposition 2.5, it seems redundant to study posted price mechanisms with two different prices. However, one surprising result of this line of work is that posting two prices actually helps in learning settings.

3.1 Technical Challenges of GFT Maximization

We start by presenting two features that differentiate GFT maximization from other online learning settings.

The first one is a "lack of observability" under partial feedback (i.e., one- or two-bit feedback). While the performance of a learner is evaluated in terms of the total gain from trade accumulated, such an objective is never observed *directly* by the learner, as the valuations v_s, v_b remain hidden even if the trade is accepted by both parties. For instance, if the learner posts price $p = 1/2$ to both agents and observes that the trade is accepted, it cannot discriminate between the case in which $v_s = 0, v_b = 1$ (and thus the GFT is 1) and the case in which $v_s = v_b = 1/2$ (and thus the GFT is 0).

Second, the action space $[0, 1]^2$ is continuous, and the reward is highly irregular, i.e., it is neither continuous, one-sided Lipschitz nor concave. This might lead to the so-called "needle in a haystack" problem. For instance, consider an adversary selecting an hidden "needle" x arbitrarily close to $1/2$ and sampling valuations uniformly and independently at random from $\{(0, x), (x, 1)\}$. Then, posting any price $p \neq x$ leads to a GFT of $\approx 1/4$. However, in this instance, the optimal price is exactly the needle x, which leads to a GFT of $\approx 1/2$.

3.2 Strong Budget Balance

The strong budget balance condition requires that the learner posts the same price to both agents, at each time step, i.e., $p_t = q_t$ for all $t = 1, \dots, T$. This means that the platform does not extract any profit from the market, nor does it subsidize it, which is reasonable in many contexts that require a platform that only guarantees facilitating the trade between participants.

With one bit of feedback at each time step, posting one single price is not enough

[1] In principle, we could also learn a fixed-price mechanism by employing more general mechanisms and committing only to a fixed-price one once valuations have been learned; however, this would require a stronger feedback model, which contrasts with the advantage of using fixed-price mechanisms, and, to the best of our knowledge, this hybrid model has not been studied.

	Stationary i.i.d. Settings				Non-Stationary Settings	
	i.i.d.	+ bd	+ ind.	+ bd + ind.	σ-smooth	Adversarial
Full	$\tilde{\Theta}(\sqrt{T})$					$\Omega(T)$
2-bit	$\Omega(T)$			$\tilde{\Theta}(T^{2/3})$	$\Omega(T)$	
1-bit	$\Omega(T)$					

Table I: Minimax rates achievable for **gain-from-trade** maximization in bilateral trade with **strong budget balance**. The rows correspond to feedback models (full feedback, two- and one-bit feedback), while the columns correspond to data generation models (in particular, bd denotes bounded density, while ind. denotes that the seller and buyer distributions are independent).

for the learner to achieve sublinear regret under one-bit feedback *even against the stochastic i.i.d. adversary*, see Theorem 5 of Cesa-Bianchi et al. [2024a]. The lower bound construction is simple and is based on the "lack of observability" property: it is possible to construct two distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ over valuations which induce the same feedback distribution, i.e., for any $p \in [0,1]$, the feedback $z_t = \mathbb{1}_{\{v_s \le p\}} \cdot \mathbb{1}_{\{v_b \ge p\}}$ follows the same law under both $\mathcal{D}_1$ and $\mathcal{D}_2$. At the same time, the optimal prices under the two distributions are far apart, so that any algorithm suffers linear regret because it is forced to take the same actions in both instances, and thus it will incur linear regret in at least one of the two.

We mention some results beyond one-bit feedback due to Cesa-Bianchi et al. [2024a]. In the adversarial setting *even full feedback* is not enough to achieve sublinear regret, while in the stochastic i.i.d. setting no-regret is attainable. In particular, under two-bit feedback, sublinear regret is achievable if and only if the agents' distributions are independent and admit bounded density. Similarly, if the learner has access to full feedback, then no-regret is achievable in the stochastic i.i.d. setting without assuming anything about the distributions.

The bounded density assumption can be generalized in a non-stationary setting using the notion of σ-smooth adversary [Haghtalab et al. 2020], which is a popular beyond-worst-case input assumption [Roughgarden 2020]. In particular, a σ-smooth adversary for bilateral trade chooses at each time step a distribution $\mathcal{D}_t$ over $[0,1]^2$ which admits a density bounded by σ^{-1}.[1]

In a follow-up work, Cesa-Bianchi et al. [2024b] show that no-regret is attainable under this assumption in the full feedback setting (thus yielding a positive result in a non-stationary setting), while partial feedback still remains useless. The exact minimax rates of the known results for strong budget balance learning algorithms are summarized in Section 3.2. For simplicity, here and in the following, we only report the time-horizon dependency in the regret rates, without specifying the impact of the smoothness term $1/\sigma$. Indeed, while the algorithms exhibit regret rates that are polynomial in $1/\sigma$, the known lower-bound rates are independent of this term.

[1]Or, alternatively, if μ denotes the Lebesque measure and $\mathcal{B}([0,1]^2)$ the Borel sigma-algebra of the unit square, $\mathcal{D}_t$ respects that $\mathbb{P}_{v\sim\mathcal{D}}(v \in A) \le \frac{\mu(A)}{\sigma}$, for all $A \in \mathcal{B}([0,1]^2)$.

3.3 Weak Budget Balance

Proposition 2.5 states that, in the (one-shot) Bayesian setting, there is no point in posting two different prices. Surprisingly, from the online learning perspective, this is not the case.

For fixed-price mechanisms, the (weak) budget balance condition requires that the seller price p_t and the buyer price q_t respect $p_t \leq q_t$, thus guaranteeing that $\mathsf{profit}_t \geq 0$ for all t. Although suboptimal from an optimization point of view (posting any price $p \in [p_t, q_t]$ yields at least the same gain-from-trade), it turns out that posting two prices allow to estimate the gain-from-trade with one-bit feedback. More formally, Azar et al. [2024] show that for every target price $p \in [0,1]$, there exists a randomized procedure that produces two random variables $\hat{p}, \hat{q} \in [0,1]$ such that the probability of observing the trade is exactly $\mathsf{GFT}(p, v_s, v_b)$ for the underlying hidden valuations (v_s, v_b). In formulas,

$$\mathbb{P}(v_s \leq \hat{p}, \hat{q} \leq v_b) = \mathsf{GFT}(p, v_s, v_b).$$

Algorithm 1 GFT estimator with weakly budget balanced prices

Require: $p \in [0,1]$

1. With probability p: $\hat{p} = U$, $\hat{q} = p$, where $U \sim \text{Uniform}[0, p]$
2. With probability $1 - p$: $\hat{p} = p$, $\hat{q} = V$, where $V \sim \text{Uniform}[p, 1]$

return $(\hat{p}, \hat{q})$

The simple estimator is shown in Algorithm 1. Having an unbiased estimator of the reward solves the first challenge of GFT maximization with partial feedback, i.e., the lack of observability of the reward. However, this alone does not solve the "needle in a haystack" problem, so that no-regret is still unattainable in the adversarial setting or the stochastic i.i.d. one with either one or two bit feedback (see Cesa-Bianchi et al. [2024b] for the adversarial setting, while the independent lower bound from Cesa-Bianchi et al. [2024a] still holds for weak budget balance). However, it provides a first step towards achieving positive results in a non-stationary environment and partial feedback.

Notably, the estimator only needs one bit of feedback, and has been used in Cesa-Bianchi et al. [2024b] to derive a positive result for the σ-smooth adversary, that we briefly present here. The main property given by σ-smoothness is that it turns the GFT into a Lipschitz function. More formally, Lemma 1 of Cesa-Bianchi et al. [2024b] ensures that $p \mapsto \mathbb{E}[\mathsf{GFT}_t(p)]$ is $1/\sigma$-Lipschitz. This assumption addresses the "needle in a haystack" problem: playing only the discretized prices $\{0, \varepsilon, 2\varepsilon, \ldots, 1\}$ leads to a discretization error (the difference between the maximum cumulative GFT on $[0,1]$ and the maximum on the discretized prices) of at most $\sigma^{-1}\varepsilon T$. Then, one can combine the estimator in Algorithm 1 with the block-decomposition algorithm of Awerbuch and Mansour [2003]. The resulting algorithm divides the time horizon T into N blocks, and in each of the N blocks, it assigns randomly $O(1/\varepsilon)$ steps (one for each discrete price) to perform the estimation procedure, and at the end of each block, it updates some full-feedback regret minimization algorithm (e.g., Hedge [Cesa-Bianchi and Lugosi 2006]) and probes from it the next price to play in

	Stationary i.i.d. Settings		Non-Stationary Settings	
	i.i.d.	+ bd	σ-smooth	Adversarial
Partial Feedback	$\Omega(T)$	$\tilde{\Theta}(T^{3/4})$	$\tilde{\Theta}(T^{3/4})$	$\Omega(T)$

Table II: Minimax rates achievable for **gain-from-trade** maximization in bilateral trade with **budget balance**. Partial feedback refers indifferently to either one-bit or two-bit feedback, while the columns correspond to data generation models (in particular, bd denotes bounded density).

the next block (modulo the estimation steps). Informally, by fixing $\varepsilon \approx T^{-1/4}$ and $N \approx \sqrt{T}$, one can derive the following regret bound:

$$R_T \lesssim \underbrace{\frac{T}{N}}_{\text{scale}} R_N^{\text{HEDGE}} + \underbrace{\frac{N}{\varepsilon}}_{\substack{\text{estimation}\\ \text{rounds}}} + \underbrace{\frac{1}{\sigma}\varepsilon T}_{\substack{\text{discretization}\\ \text{error}}} \lesssim O\left(\frac{1}{\sigma} T^{\frac{3}{4}}\right),$$

where we use that the regret of Hedge over $O(1/\varepsilon)$ experts and N time steps is $R_N^{\text{HEDGE}} \in \tilde{O}(\sqrt{\log(1/\varepsilon) N})$, which gets multiplied for the scale of the rewards at each time block (T/N). It turns out that this rate is essentially tight, even with two-bit feedback. This is due to a lower bound of Cesa-Bianchi et al. [2024b] which holds for constant σ. The known results for weak budget balance algorithms are summarized in Table II; note that in full feedback there is no point in posting different prices to seller and buyer (so that the minimax rates are the same as in Section 3.2).

3.4 Global Budget Balance

The σ-smooth assumption is useful because it solves the "needle in a haystack" problem. Bernasconi et al. [2024b] use another approach to solve this problem, without weakening the adversary, but by relaxing the budget balance constraint. Inspired by the bandit with knapsack (BwK) framework with non-monotone resource consumption [Kumar and Kleinberg 2022; Slivkins et al. 2024; Bernasconi et al. 2024a], they define a *long-term* constraint on the budget, i.e., they require a *global* budget balance condition:

$$B_T \geq 0, \quad \text{where} \quad B_0 = 0, B_{t+1} = B_t + \mathsf{profit}_t(p_t, q_t).$$

Let's see how allowing for a controlled amount of negative profit helps finding the needle in the haystack. For instance, recall the hard distribution which outputs uniformly at random either $(v_s, v_b) = (0, 1/2)$ or $(1/2, 1)$. By only playing budget balanced prices (p, q) with $p \leq q$, only $p = q = 1/2$ has GFT of 1, while any other price has GFT of $1/2$. However, if we allow p to be slightly greater than q (let's say $p = q + \varepsilon$) then playing $(q + \varepsilon, q)$ for any $q \in [1/2 - \varepsilon, 1/2]$ induces a GFT of 1, by only paying ε negative profit. This simple observation leads to a global budget-balanced algorithm with one-bit feedback, adversarial valuations, and $\Theta(T^{3/4})$ regret. The upper bound comes from Bernasconi et al. [2024b] while the tight lower bound is due to two independent and concurrent works: Lunghi et al. [2026a] and [Chen et al. 2025]. The idea of the algorithm is to divide the time horizon into two phases: the first phase allows accumulating a profit budget β, and the second phase allows spending that budget by playing slightly non-budget-balanced prices.

The second phase almost exactly models the weak budget balanced constraint, i.e.,

	Stationary i.i.d. Settings		
	i.i.d.	+ ind.	Adversarial
Full Feedback	$\tilde{\Theta}(\sqrt{T})$		
Partial Feedback	$\tilde{\Theta}(T^{3/4})$	$\tilde{\Theta}(T^{2/3})$	$\tilde{\Theta}(T^{3/4})$

Table III: Minimax rates achievable for **gain-from-trade** maximization in bilateral trade with **global budget balance**. Partial feedback refers indifferently to either one-bit or two-bit feedback, while the columns correspond to data generation models (in particular, ind. denotes that the seller and buyer distributions are independent).

to use an estimator for non-budget-balanced prices $H_K = \{(i/K, (i-1)/K)\}_{i\in[K]}$ and the block decomposition algorithm. The only difference is that the estimator for the prices in H_K is slightly biased (by an additive $O(1/K)$ term). The key technical challenge is to show that, while maximizing the profit in the first phase, the algorithm does not lose too much GFT. Concretely, Bernasconi et al. [2024b, Lemma 4.1] provides a multiplicative discrete grid F_K (of size $\tilde{O}(K)$) of budget-balanced prices and show that

$$\max_{p\in[0,1]} \sum_{t\in[T]} \mathsf{GFT}_t(p) \le O(\log T) \max_{(p,q)\in F_K} \sum_{t\in[T]} \mathsf{profit}_t(p,q) + O\left(\frac{T}{K}\right).$$

This result is crucial in relating the global budget constraint with the (gain-from-trade) regret, ensuring that the profit-maximizing phase does not induce "too much regret". More precisely, this shows that the optimum price on the grid F_K in terms of profit approximates the optimum price in terms of GFT(up to log factors). The first phase is conducted until a budget of β is accrued (call the stopping time τ). In this phase, the algorithm instantiates any regret minimizer over F_K. This let us conclude that the maximum GFT until τ is small, i.e.

$$\max_{p\in[0,1]} \sum_{t\in[\tau]} \mathsf{GFT}_t(p) \le O(\log T)(\beta + \sqrt{TK}) + O\left(\frac{T}{K}\right).$$

Choosing $\beta \in \Theta(T^{3/4})$, $K \in \Theta(T^{1/4})$ and $N \in \Theta(T^{1/2})$, we can show that $B_T \ge 0$ (since the algorithm loses at most $T/K \le \beta$ profit in the second phase), and thus global budget balance is satisfied. In the first phase, with our choice of parameters, $\max_{p\in[0,1]} \sum_{t\in[T]} \mathsf{GFT}_t(p) \in \tilde{O}(T^{3/4})$ and in the second phase we can mimic the calculations done in the weak budget balance section which also give a regret of $\tilde{O}(T^{3/4})$, for a total regret is $O(T^{3/4})$.

Beyond the matching adversarial lower bound, Chen et al. [2025] provides tight results for other data generation models. All known minimax rates are summarized in Table III.

3.5 Other Works on GFT Maximization

Beyond the models presented so far, other works provide a broader view on gain-from-trade maximization in bilateral trade, across several dimensions: budget violations, stronger benchmarks, and intermediate feedback models.

3.5.1 *Other Budget Balance Notions.* The two-phase algorithm of Bernasconi et al. [2024b] first accumulates the profit budget and then spends it by posting mildly subsidizing prices. A natural question is what happens when the learner is allowed to violate the global budget constraint. Lunghi et al. [2026a] study the impact of budget violation allowed and the achievable regret. In particular, they show that for a violation of the order of $O(T^\beta)$ (i.e., if the learner is allowed to lose up to T^β profit throughout the time horizon), with $\beta \in [3/4, 6/7]$, a tight regret of order $\tilde{O}(T^{1-\beta/3})$ can be obtained. On the other hand, for $\beta \leq 3/4$, violations are not useful, since the regret of $\tilde{O}(T^{3/4})$ cannot be improved upon, while also violating more than $\beta \geq 6/7$ does not lead to an improvement since $O(T^{5/7})$ is tight in that regime.

3.5.2 *Other Benchmarks.* The global budget balance constraint motivates stronger benchmarks than the best fixed price. If the agent is allowed to balance profit and losses over the whole horizon, it is natural to compare against the best *distribution* over fixed prices whose expected profit is non-negative. This benchmark can be strictly stronger than the best deterministic price: randomizing across prices may use profitable trades to finance subsidized trades and thereby increase total gain from trade. However, Bernasconi et al. [2024b] shows that they cannot be further than a multiplicative factor of 2, namely that, for any sequence of valuations:

$$\sum_{t\in[T]} \mathbb{E}_{(p,q)\sim\gamma^*}[\mathsf{GFT}_t(p,q)] \leq 2 \sum_{t\in[T]} \mathsf{GFT}_t(p^*),$$

where γ^* is the best feasible distribution in expectation and p^* is the best fixed price. This benchmark is too strong in the adversarial model, indeed, Bernasconi et al. [2024b] show a linear $(1 + \Omega(1))$-regret lower bound.[1] The bounded constant multiplicative ratio mentioned above shows that any no-regret algorithm for the single fixed price is a 2-regret algorithm against the best feasible distribution, so there is a gap between the upper bound and lower bound for this particular benchmark and adversarial valuations. This impossibility motivates studying stochastic models.

Lunghi et al. [2026a] shows that stochastic valuations are not enough, and that there is a linear regret lower bound. However, they also show that for a valuation distribution with bounded density, the minimax regret rate is $\tilde{\Theta}(T^{3/4})$. In a subsequent work, Lunghi et al. [2026b] construct an algorithm that scales with the amount of perturbation C, which interpolates between the stochastic (with bounded density), which is the case $C = 0$, to the adversarial one, which is the case $C = T$, and obtains a regret of $\tilde{O}(T^{3/4}) + O(C \log T)$, against the distributional benchmark. Moreover, the algorithm does not need to know the perturbation parameter C, which is defined as the sum over $t \in [T]$ of total variation distance between the distribution choose at time t by the adversary and the unperturbed distribution.

Finally, Azar et al. [2024] studies 2-regret (with respect to the best fixed-price) under different feedback schemes and under the strong and weak budget-balanced conditions. In particular, they prove that sublinear 2-regret is not achievable with one- or two-bit feedback under the strong budget-balanced condition. On the other

[1]For $\alpha > 1$, we define α-regret as the normal regret in which the baseline is divided by α, in the spirit of competitive ratio.

hand, under the weak budget-balanced condition, sublinear 2-regret is achievable even with one-bit feedback.

3.5.2.1 *Other feedback models.* There are other intermediate feedback models for GFT maximization. Bacchiocchi et al. [2025] studies an *asynchronous* mechanism that queries the seller only if the buyer has already accepted the offer, which is a feedback in between the one-bit and two-bit one, as it receives a clean bit from the buyer and a censored one from the seller. Under this feedback model and a strong budget balance, the only known algorithm that achieved this was that of Cesa-Bianchi et al. [2024a], which achieved regret of $\Omega(T^{2/3})$ under the bounded-density assumption. Bacchiocchi et al. [2025] matches this under this more restricted feedback.

Finally, Chen et al. [2025], Jin [2026] and Lunghi et al. [2025] study upper and lower bounds for an intermediate feedback (called semi-feedback in Chen et al. [2025] and Jin [2026], and asymmetric-feedback in Lunghi et al. [2025]) in which one of the two valuations is observed directly, for instance, the buyer's valuation v_b, while for the seller's valuation v_s, we either observe the outcome of the seller's decision or the overall success of the trade.

In the global budget balance regime, they prove a tight $T^{2/3}$ regret bound, improving from the $T^{3/4}$ regret obtainable under just one-bit feedback.

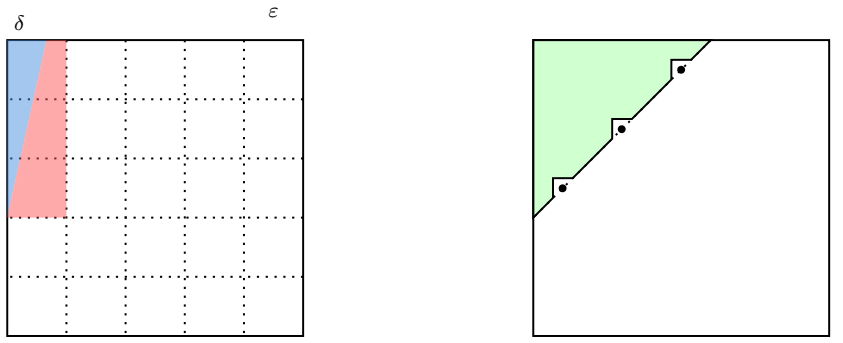
4. PROFIT

In this section, we present the results for profit maximization. Although the profit function resembles GFT, we do not have an equivalent characterization as in Proposition 2.5. This means that we cannot restrict our attention only to fixed-price mechanisms (see Example 2.6) and, to compute Myerson payments, we need the agents to reveal their valuations. In online learning jargon, this means that profit-maximizing algorithms receive *full feedback*. Compared to gain-from-trade maximization, we then have an incredibly larger action space (all monotone allocation regions vs. all fixed prices) but a more informative feedback (full vs. one- or two-bit). These two structural differences call for new technical tools.

When faced with a large action space, the natural approach in online learning is to discretize it (e.g., in Lipschitz bandits [Kleinberg et al. 2019] or pricing [Kleinberg and Leighton 2003]). This method splits the problem into two parts: first, finding *offline* a finite subset of actions that well approximates the optimal one *uniformly over all possible inputs*, and then running *online* a discrete learning algorithm on the discretized action space. The most natural discretization of the mechanism space $\mathcal{M}$ is to cast a uniform ε-grid on the $[0,1]^2$ square and consider all the $\binom{2/\varepsilon}{1/\varepsilon}$ mechanisms $\mathcal{M}_\varepsilon^\perp$ with monotone allocation regions that can be described as the union of the cells of the grid[1]. Unfortunately, as we show in the following example, this rich class of mechanisms does not provide a uniform approximation for our problem. This is due to the non-continuity (and thus non-Lipschitzness) of our objective function.

Example 4.1 (Fixed Discretizations Fail). *Consider the uniform distribution $\mathcal{D}$ over the segment $(0, 1/2)$-$(\delta, 1)$, where δ is an arbitrarily small parameter*

[1]The family $\mathcal{M}_\varepsilon^\perp$ is a uniform ε-coverage of the mechanism space, for a suitable definition of distance between mechanisms.



(a) Visualization of Example 4.1.

(b) Hard Instance of Example 4.2.

Fig. 2: Supporting figures for the hard examples.

(as in Example 2.6). The corresponding optimal mechanism M allocates in the $(0, 1/2)$-$(\delta, 1)$-$(0, 1)$ triangle, for an expected profit of approximately $3/4$. Consider now any fixed discretization parameter $\varepsilon > \delta$, the support of $\mathcal{D}$ is contained in the first column of the grid, so the profit extracted by any mechanism $M \in \mathcal{M}_\varepsilon^\perp$ only depends (up to $O(\delta)$ payment to the seller) on the horizontal segment that crosses such column, for a maximum profit of $1/2$ (see also Figure 2a for a visualization).

This simple construction is at the core of the adversarial lower-bound construction in Di Gregorio et al. [2025], showing that no learning algorithm can achieve sublinear regret when the task is maximizing profit in the adversarial setting. In the same paper, the authors study the problem in the stochastic i.i.d. setting and show that learning is indeed possible there, with a regret of $\tilde{O}(\sqrt{T})$, which is tight up to poly-logarithmic terms. Finally, Di Gregorio et al. [2026] complete the picture of natural data-generation models, proving that nearly tight $\tilde{O}(\sqrt{T})$ is achievable also against a non-stationary adversary, at the cost of assuming σ-smoothness. The rest of the section is devoted to presenting briefly the challenges and the techniques used to achieve the $\tilde{O}(\sqrt{T})$ regret in the stochastic i.i.d. setting. We refer to Table IV for a summary of these results.

4.1 Profit Maximization against the i.i.d. Stochastic Adversary

The well-known online-to-offline reduction (e.g., Cesa-Bianchi et al. [2004]) shows that finding the best mechanism online in the stochastic i.i.d. setting *under full feedback* is equivalent to studying the standard offline PAC learning framework. In particular, deriving a $\sqrt{T}$ regret bound is equivalent to showing that, given access to approximately $1/\varepsilon^2$ i.i.d. samples from the underlying distribution $\mathcal{D}$, one is capable of finding an $M \in \mathcal{M}$ such that

$$\sup_{\mathcal{M}^\star \in \mathcal{M}} \mathbb{E}\left[\mathsf{profit}(\mathcal{M}^\star, v)\right] \lesssim \mathbb{E}\left[\mathsf{profit}(M, v)\right] + O(\varepsilon), \tag{1}$$

where v denotes a fresh valuation drawn from $\mathcal{D}$. It turns out that even with this simplification, the task poses significant challenges: the family of "profit functions" $\{\mathsf{profit}(M, \cdot)\}_{M \in \mathcal{M}}$ is statistically extremely complicated! In particular, it exhibits unbounded fat-shattering dimension for constant values of the margins (and thus it also has unbounded pseudo-dimension) [Di Gregorio et al. 2026]. Instead of proving

Stochastic i.i.d. Setting	Smooth Adversary	Adversarial Setting
$\tilde{\Theta}(\sqrt{T})$	$\tilde{\Theta}(\sqrt{T})$	$\Omega(T)$

Table IV: Minimax regret rates achievable for profit-maximizing algorithms in bilateral trade.

these impossibility results, we propose here a convincing and direct exhibit of the complexity of learning this class of mechanisms: it does not enjoy the uniform convergence property. Stated differently, regardless of the number of samples a learner has access to, there may be some mechanisms whose empirical profit is far from the expected one.

Example 4.2 (No Uniform Convergence). *Let $\mathcal{D}$ be the uniform distribution on the $(0, 3/4)$-$(1/4, 1)$ segment, and consider any realization of n i.i.d. samples S from $\mathcal{D}$. Denote with $\delta > 0$ a small parameter we set later, and let M be the mechanism allocating in the $(0, 3/4)$, $(1/4, 1)$, $(0, 1)$ triangle, minus the δ-radius ℓ_∞ balls $B_\delta(v)$ for v in S. By construction, the empirical profit of M on the samples is zero, while the expected profit on $\mathcal{D}$ is at least $1/4$. In fact, S is finite, so it is possible to take δ small enough so that at least half of the support of $\mathcal{D}$ falls in the allocation region of M. For such δ, a trade happens with probability at least $1/2$, with an expected profit of at least $1/4$. We refer to Figure 2b for a visualization.*

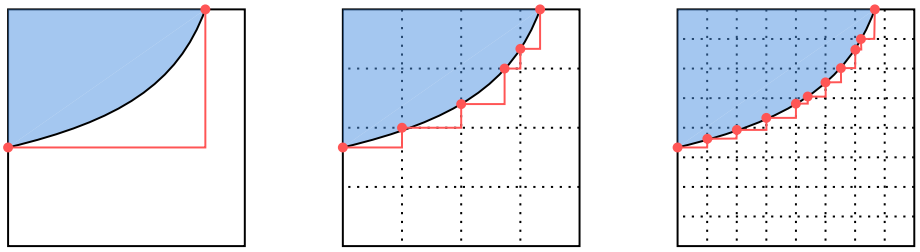
From Example 4.2, we know that past samples are not enough to estimate *at the same time* all mechanisms' expected performance. Note, the absence of uniform convergence is peculiar to our problem, and separates it to other mechanism design learning tasks such as, e.g., one-sided pricing [Kleinberg and Leighton 2003], revenue maximization in second price auctions with reserve [Cesa-Bianchi et al. 2015; Morgenstern and Roughgarden 2016], and gain-from-trade maximization with a single fixed price [Cesa-Bianchi et al. 2024a]. Moreover, from Example 4.1, we know we cannot focus on learning *only* mechanisms on a *finite and fixed* discretization! Di Gregorio et al. [2025] overcome this challenge by constructing a data and distribution dependent family of mechanisms which can be learned very fast (with $1/\varepsilon^2$ samples) and which well approximates the optimal mechanism.

A crucial role in the construction is played by the families $\mathcal{M}_\eta$ of η-simple mechanisms for $\eta \in (0, 1]$, i.e., such that the boundary of their allocation regions are determined by at most $O(1/\eta)$-axis parallel segments each (see Figure 3). In particular, for any mechanism $M \in \mathcal{M}$ there exists an η-simple mechanism in $\mathcal{M}_\eta$ which "mimics" its allocation region with $O(1/\eta)$ corners such that

$$\mathbb{E}\left[\mathsf{profit}(M, v)\right] \leq \mathbb{E}\left[\mathsf{profit}(M_\eta, v)\right] + O(\eta). \tag{2}$$

Focusing on the ε-simple mechanisms already yields some positive results, as a delicate but direct union-bound argument shows that approximately $1/\varepsilon^3$ samples are enough to estimate up to precision $O(\varepsilon)$ all the mechanisms in $\mathcal{M}_\varepsilon$.

To do better, Di Gregorio et al. [2025] resort to a delicate application of probabilistic chaining [Talagrand 2014], which allows for a more fine-grained control of the concentration argument. Denote with $H = \log 1/\varepsilon$, where ε is the target precision. Using an involved "good event" analysis, for any sample S of n i.i.d. valuations $v^1, v^2, \ldots, v^n$, Di Gregorio et al. [2025] shows the existence of a target sub-family $\mathcal{M}^\star$ of the ε-simple mechanisms which can be used to perform the chaining analysis. In



(a) Boundary of $\phi_0(M)$. (b) Boundary of $\phi_2(M)$. (c) Boundary of $\phi_3(M)$.

Fig. 3: Visualization of the chain construction.

particular, $\mathcal{M}^\star$ can be partitioned at geometrically decreasing "levels" of precisions $(1, 1/2, 1/4, \dots, 2^{-H})$ ensuring that (i) there exists a target mechanism $M^\star \in \mathcal{M}^\star$ which is $O(\varepsilon)$ far from optimum, (ii) each level of the partition has cardinality $\binom{2^{2h}}{2^h}$, and (iii) the following L^1-net property is respected: for any $M \in \mathcal{M}^\star$ and any precision 2^{-h} there exists a mechanism $\phi_h(M) \in \mathcal{M}^\star \cap \mathcal{M}_{2^{-h}}$ such that

$$\left| \frac{1}{n} \sum_{v^i \in S} \left(\mathsf{profit}(M, v^i) - \mathsf{profit}(\phi_h(M), v^i) \right) \right| \leq O(2^{-h}). \tag{3}$$

These ingredients can then be combined in the standard probabilistic chaining analysis to get uniform convergence in the $\mathcal{M}^\star$ family with the desired sample complexity of $\tilde{O}(1/\varepsilon^2)$. In particular, the algorithm, which first examines the realized samples, then computes the empirical profit maximizers, and finally transforms them so that they lie in the $\mathcal{M}^\star$ family, achieves the desired learning outcome.

We conclude by mentioning that, interestingly, a similar *chaining-based* approach is adopted in Di Gregorio et al. [2026] against the smooth adversary. There, given the non-stationary nature of the adversary, the authors need to resort to the "algorithmic" version of chaining, inspired by Cesa-Bianchi et al. [2017].

5. GENERALIZATIONS AND RELATED MODELS

In this section, we present some models for bilateral trade—or related problems—that extend or vary the basic learning protocol.

5.1 Contextual Bilateral Trade

Gaucher et al. [2025] introduce a natural tweak: *contexts*. Namely, at each time step t, the agents are characterized by d-dimensional vectors (x_s^t, x_b^t), and their actual valuations at that time are given by the inner product with unknown but fixed weight vectors v_s and v_b.

In the contextual setting, bilateral trade takes the form of searching the space of possible weight vectors, using a specific type of queries: the learner posts two prices, and observes whether the agents accept or not. Taking inspiration from the Ellipsoid Pricing method for (one-sided) pricing [Cohen et al. 2020], Gaucher et al. [2025] shows that, for gain-from-trade, it is possible to achieve *logarithmic* regret with respect to a very powerful benchmark—the best omniscient pricing strategy—as long as two-bit feedback is available.

Following up on this work, Cosson et al. [2026] derive nearly minimax regret rates for both profit and gain-from-trade, leveraging the notion of Steiner potentials [Liu et al. 2021]. Surprisingly, it turns out that gain-from-trade maximization is "equivalent" to contextual search ($\tilde{\Theta}(d)$ regret)[1] while profit-maximization in bilateral trade and one-sided pricing yields the same rate ($\tilde{\Theta}(d \log \log T)$). When the learner has only access to one-bit feedback, the regret landscape becomes less clear.

Gaucher et al. [2025] also introduce the *noisy* version of the problem, where the agents' valuations are given by the context-weight inner product plus some i.i.d. noise. In this case, they prove a tight (up to polylog(T) terms and dimension dependence) $\Theta(T^{2/3})$ regret bound for gain-from-trade maximization. Note, given the stochastic nature of the data-generation model, the benchmark is less powerful: the standard fixed-price-in-hindsight adversary. [Coccia et al. 2026] generalizes the model further, by considering agents whose valuations are given by a fixed *Lipschitz function*. This problem is non-parametric in nature, but the authors are able to show a tight $\tilde{O}(T^{d/d-1})$ regret rate for gain-from-trade maximization, using only one-bit feedback, under the strong budget balance condition. This positive result is achieved via a geometric decomposition of the action space, in a way that is somehow reminiscent of *adaptive zooming* [Slivkins 2014] and *algorithmic chaining* [Cesa-Bianchi et al. 2017].

These works provide positive results for various notions of budget balance or suboptimal regret rates, but a clean answer remains elusive. For instance, we find the following question especially interesting: can we achieve sublinear regret when the valuations are linear and noisy, under the strongly budget-balanced condition, with only one-bit feedback?

5.2 Brokerage

Bolić et al. [2024] introduce the brokerage problem: a version of bilateral trade in which the identity of the seller and the buyer are not fixed in advance, but depend on the relative ordering of their valuations. Namely, there exists an underlying fixed distribution over $[0, 1]$ and, at each iteration, two independent samples are drawn from it; the larger of the two samples becomes the buyer, while the smaller becomes the seller. Thus, the two sides of the market are coupled by the same underlying distribution, and the agents' roles are endogenous.

In the full-feedback model, Bolić et al. [2024] prove a tight $\sqrt{T}$ regret rate for a general distribution. Under the σ-smoothness assumption, they further derive a surprising $1/\sigma \cdot \log T$ regret rate, which comes from a clever rewriting of the gain-from-trade objective in terms of the variance of the empirical mean of the samples observed. This rewriting allows the authors to exploit the exponential decay of such variance via standard concentration arguments. Bolić et al. [2024] also consider the weaker one-bit-feedback model, where they prove that no learning is achievable for general distributions, while smoothness allows for a $\sqrt{T/\sigma}$.

Following up on this initial work, Bachoc et al. [2025a] and Bachoc et al. [2025b] investigate the contextual version of the brokerage problem, from both the parametric and non-parametric perspectives. In another modification, Cesari and Colomboni [2025] consider the same brokerage model, with the difference that the learner is

[1]Here $\tilde{\Theta}$ hides terms polynomial in the dimension d

now interested in maximizing the number of trades (and no longer the resulting gain from trade).

A closely related mediated model is the repeated-mediated newsvendor problem in Bolić et al. [2025]. Here, the learner is again a mediator between two sides of a market, but the interaction is richer than single-unit bilateral trade: a supplier has a private production cost, a retailer has a private utility function, the mediator posts a unit trading price, and the retailer's response determines the quantity purchased. This model can be viewed as a newsvendor-type generalization of the same mediated-trade viewpoint underlying brokerage and bilateral trade.

Finally, Cesa-Bianchi et al. [2025] studies market making, where agents arrive sequentially, and the mechanism posts prices that determine whether an arriving agent trades as a seller or as a buyer. This is not a direct bilateral-trade generalization, but it is a natural market-microstructure counterpart of brokerage: the side of the trade is again determined endogenously by the relationship between the agent's valuation and the posted prices.

5.3 Fair Bilateral Trade

One drawback of the gain-from-trade metric is that it measures only the *increase* in social welfare and provides no information about each agent's utility. For instance, consider a seller with valuation $v_s = 0$, and a buyer with $v_b = 1$; whichever price the mechanism may post in the $[0, 1]$ interval results in a trade with an increase in social welfare of 1. However, some prices may disproportionately benefit one agent at the expense of the other. Posting $p = 1$ increases the seller's utility from 0 to 1, while the buyer's utility remains constant at 0. To address this shortcoming, Bachoc et al. [2024] introduce the notion of *fair* gain from trade. Formally, the fair-gain-from-trade FGFT under valuations v_s and v_b for posting price p is:

$$\mathsf{FGFT}(p, v_s, v_b) = \min\{(v_b - p)_+, (p - v_s)_+\},$$

where $(a)_+ = \max\{0, a\}$. With this notion, the best price with valuations $(0, 1)$ becomes $1/2$, which divides the increase in utility between the two agents equally.

Bachoc et al. [2024] investigates various feedback and data generation models for the online version of the fair bilateral trade problem. In particular, for full feedback and stochastic i. adversary they show that the minimax regret is $\tilde{\Theta}(\sqrt{T})$ in the stochastic i.i.d. setting, while for one-bit-feedback, they show that no regret is only achievable assuming independence of the seller and buyer valuations, in which case the minimax rate is $\tilde{\Theta}(T^{2/3})$.

Building on Bachoc et al. [2024], Bachoc et al. [2026] develop an axiomatic fairness framework for bilateral trade, leading to a Rawls-to-Nash family of fair-gain objectives obtained by aggregating the seller's and buyer's gains through nonpositive Hölder means. This class leads to a rather involved statistical problem: estimating these objectives from threshold feedback requires handling two-dimensional singular-kernel identities. They obtain tight (up to logarithmic factors) regret guarantees, as well as PAC guarantees that hold uniformly and simultaneously over the fairness parameter.

5.4 Query Complexity

In Castiglioni et al. [2026], the authors investigate a fundamental learning problem that turns out to be intimately related to bilateral trade. Their goal is to get a *uniform* estimate of the multi-dimensional CDF of a distribution on the d-dimensional cube, using only "pricing queries" [Leme et al. 2023]. A pricing query takes in input d prices $p_1, \dots, p_d$, draws a fresh sample X from the underlying distribution, and return Yes if and only if $X \in [0, p_1] \times \cdots \times [0, p_d]$ (or, alternatively, it answer yes, if each coordinate of X *accepts* its price). [Castiglioni et al. 2026] shows how to solve the problem using $1/\varepsilon^3 (\log 1/\varepsilon)^{O(d)}$ pricing queries, which nearly-matches the 1-dimension lower bound of $1/\varepsilon^3$ implicit in Kleinberg and Leighton [2003].

Let's get back to bilateral trade and consider the task of finding the *fixed-price* mechanism that maximizes profit given a certain valuation distribution. In the i.i.d. setting, the task becomes finding the pair of prices $p \leq q$ maximizing $\mathbb{E}\left[\mathsf{profit}(M_{(p,q)}, v)\right] = (q-p) \cdot \mathbb{P}\left(v_s \leq p, v_b \geq q\right)$, while receiving one-bit feedback if the type $\mathbb{1}_{\{v_s \leq p, v_b \geq q\}}$. This problem is very similar to the one studied in Castiglioni et al. [2026], so that $\tilde{\Theta}(1/\varepsilon^3)$ samples are necessary and sufficient to get a uniform ε-estimate of the profit function! This implies that a suitably tuned explore-then-commit algorithm yields a $\tilde{O}(T^{3/4})$ regret: use the algorithm in Castiglioni et al. [2026] for $\approx T^{3/4}$ time steps to construct an $T^{-1/4}$-uniform estimate of the CDF of the distribution underlying (v_s, v_b), and then commit to the best pair of prices on the estimate.

An interesting direction for future work is pinpointing the right minimax regret achievable. Indeed, the best known lower bound is the $\Omega(T^{2/3})$ impossibility result for single-dimensional pricing in Kleinberg and Leighton [2003]. Notably, solving the stochastic setting is only the first step, as no tight result is known for the adversarial setting (where, with a bit of work, one could recover the same $T^{3/4}$ rate via a Lipschitzness argument).

5.5 Private Bilateral Trade

Some of the learning algorithms presented so far have the undesirable property of being too dependent on a few data points. For instance, the $\sqrt{T}$ algorithm for profit maximization in Di Gregorio et al. [2025] heavily relies on the realized points to construct the simple mechanism to play. This dependency leads to algorithms that are fragile to perturbed samples and, crucially, may publicly reveal information about the observed valuations. To address this weakness, Di Gregorio et al. [2026] investigate the bilateral trade problem through the lens of differential privacy (DP) within the PAC-learning framework. The authors establish a foundational impossibility result, proving that achieving both DP and near-optimality is inherently unattainable under general valuation distributions for both profit and gain-from-trade maximization. To bypass this barrier, they consider σ-smooth distributions and deliver nearly tight sample complexity bounds, showing that efficiency can be privately maximized using $\tilde{O}(1/\alpha^2 + 1/\epsilon\alpha)$ samples, where α is the approximation parameter, while ε is the privacy one. For profit maximization they need $\tilde{O}(1/\epsilon\sigma\alpha^2)$ samples. Beyond these statistical guarantees, a technical highlight of the work is algorithmic: they overcome the challenge of sampling from an exponentially large mechanism space by reducing to online shortest paths and exploiting a sampling

trick from Takimoto and Warmuth [2003].

5.6 Beyond Bilateral Trade: Two-Sided Markets

Bilateral trade focuses on the interactions of *two* agents: a single seller and a single buyer. It is then natural to ask what happens in more complex *two-sided* markets. While the notions of incentive compatibility, individual rationality, and budget balance are easily generalizable, the induced class of mechanisms gets richer and loses the simple structure described in Propositions 2.2 and 2.5. In particular, gain-from-trade maximization now becomes richer than simply finding the best fixed price. In the online learning setting, two-sided markets have been investigated in Babaioff et al. [2024], Lunghi et al. [2025], and Feng et al. [2026]. In particular, Babaioff et al. [2024] show that even in small markets of three agents (one seller and two buyers, where at most one buyer can purchase the good from the seller), the problem is hopeless in the stochastic i.i.d. setting. Conversely, making the assumption that all the agents' valuations are independent is enough to guarantee a sample complexity of $\approx 1/\varepsilon^2$ (which corresponds, in the online setting, to $\approx \sqrt{T}$). Lunghi et al. [2025] instead focuses on larger markets, featuring one seller and an arbitrary number of buyers, but only restricts their attention to a simple family of mechanisms: second-price auction with reserve on the buyers' side and fixed price on the seller's side. For this family, they prove tight minimax results under various feedback and data generation models. Finally, Feng et al. [2026] extends the positive results for contextual bilateral trade to a more complex setting in which a single seller interacts with multiple buyers. Given these (partial) results, the study of two-sided markets in the online learning setting remains an exciting open direction to pursue.

5.7 Beyond DSIC mechanisms: Approximating the First Best

So far, the machine learning literature has mainly focused on studying the DSIC version of the problem, where incentive compatibility is ensured *regardless* of the other agent's behavior. Motivated by the fact that no DSIC and IR mechanism ensures a constant factor of gain-from-trade with budget balance [Blumrosen and Mizrahi 2016], Deng et al. [2025] investigates the learnability of the seller- and buyer-pricing mechanisms [Deng et al. 2022]. Indeed, it is known that uniformly randomizing between these two mechanisms (which enforce incentive compatibility only in expectation) yields a constant factor approximation of the so-called first best (i.e., $(v_b - v_s)_+$, see also discussion in Section 2.3).

Deng et al. [2025] considers the situation in which agents' valuations are drawn independently from unknown distributions. One of the two agents then observes its realized private valuation and is tasked with proposing a price to the other, using *only* sample access to the other agent's distribution. They investigate various sample-based pricing strategies and observe that, under some conditions, the resulting mechanisms yield, in expectation, a constant-factor approximation to the first best.

While Deng et al. [2025] adopts a different perspective than the one studied in the rest of the survey (there is "offline" sample access and the learner is one of the agents, and not the intermediary), it points to an exciting new research direction, namely investigating bilateral trade in the online learning framework *beyond* the DSIC + IR world, which is now fairly well understood. As a final (and related)

research opportunity, we note that all previous work has made the simplifying assumption that an intermediary plays with a fresh pair of agents at each time step. This "myopic" version of incentive compatibility—incentives are measured only at each time step, without considering the possibility that agents might strategize about future time steps—is natural in some applications (e.g., a large population of "hasty" agents) and has allowed for a comprehensive (and satisfying) theoretical study. It is, however, natural to investigate what happens in a non-myopic world, where the agents are, for instance, no-regret learners (à la Braverman et al. [2018]); this problem is still wide open and is a natural next step in the study of bilateral trade.

Acknowledgment

The authors would like to thank Matteo Castiglioni, Tom Cesari, and Roberto Colomboni for their valuable feedback on preliminary versions of the survey.

REFERENCES

Awerbuch, B. and Mansour, Y. 2003. Adapting to a reliable network path. In *PODC*. ACM, 360–367.

Azar, Y., Fiat, A., and Fusco, F. 2024. An α-regret analysis of adversarial bilateral trade. *Artif. Intell. 337*, 104231.

Babaioff, M., Frey, A., and Nisan, N. 2024. Learning to maximize gains from trade in small markets. In *EC*. ACM, 195.

Bacchiocchi, F., Castiglioni, M., Colomboni, R., and Marchesi, A. 2025. Online bilateral trade with minimal feedback: Don't waste seller's time. In *NeurIPS*. Curran Associates, Inc.

Bachoc, F., Cesa-Bianchi, N., Cesari, T., and Colomboni, R. 2024. Fair online bilateral trade. In *NeurIPS*.

Bachoc, F., Cesari, T., and Colomboni, R. 2025a. A parametric contextual online learning theory of brokerage. In *ICML*. Proceedings of Machine Learning Research. PMLR / OpenReview.net.

Bachoc, F., Cesari, T., and Colomboni, R. 2025b. A tight regret analysis of non-parametric repeated contextual brokerage. In *AISTATS*. Proceedings of Machine Learning Research. PMLR, 2836–2844.

Bachoc, F., Colomboni, R., and Kaufmann, E. 2026. Repeated bilateral trade: The quest for fairness. *CoRR abs/2606.15369*.

Bernasconi, M., Castiglioni, M., Celli, A., and Fusco, F. 2024a. Bandits with replenishable knapsacks: the best of both worlds. In *ICLR*. OpenReview.net.

Bernasconi, M., Castiglioni, M., Celli, A., and Fusco, F. 2024b. No-regret learning in bilateral trade via global budget balance. In *STOC*. ACM, 247–258.

Blumrosen, L. and Dobzinski, S. 2021. (almost) efficient mechanisms for bilateral trading. *Games Econ. Behav. 130*, 369–383.

Blumrosen, L. and Mizrahi, Y. 2016. Approximating gains-from-trade in bilateral trading. In *WINE*. Lecture Notes in Computer Science. Springer, 400–413.

Bolić, N., Cesari, T., and Colomboni, R. 2024. An online learning theory of brokerage. In *AAMAS*. International Foundation for Autonomous Agents and Multiagent Systems / ACM, 216–224.

Bolić, N., Cesari, T., Colomboni, R., and Paravalos, C. 2025. Online Learning in the Repeated Mediated Newsvendor Problem. In *NeurIPS*.

Braverman, M., Mao, J., Schneider, J., and Weinberg, S. M. 2018. Selling to a no-regret buyer. In *EC*. ACM, 523–538.

Cai, Y. and Wu, J. 2023. On the optimal fixed-price mechanism in bilateral trade. In *STOC*. ACM, 737–750.

Castiglioni, M., Lunghi, A., and Marchesi, A. 2026. The sample complexity of uniform approximation for multi-dimensional cdfs and fixed-price mechanisms. *To appear at STOC 2026, preprint on the arXiv, CoRR abs/2602.10868*.

Cesa-Bianchi, N., Cesari, T., Colomboni, R., Foscari, L., and Pathak, V. 2025. Market making without regret. In *COLT*. Proceedings of Machine Learning Research. PMLR, 799–837.

Cesa-Bianchi, N., Cesari, T., Colomboni, R., Fusco, F., and Leonardi, S. 2024a. Bilateral trade: A regret minimization perspective. *Math. Oper. Res. 49,* 1, 171–203.

Cesa-Bianchi, N., Cesari, T., Colomboni, R., Fusco, F., and Leonardi, S. 2024b. Regret analysis of bilateral trade with a smoothed adversary. *J. Mach. Learn. Res. 25*, 234:1–234:36.

Cesa-Bianchi, N., Cesari, T. R., Colomboni, R., Fusco, F., and Leonardi, S. 2021. A regret analysis of bilateral trade. In *EC*. ACM, 289–309.

Cesa-Bianchi, N., Conconi, A., and Gentile, C. 2004. On the generalization ability of on-line learning algorithms. *IEEE Trans. Inf. Theory 50,* 9, 2050–2057.

Cesa-Bianchi, N., Gaillard, P., Gentile, C., and Gerchinovitz, S. 2017. Algorithmic chaining and the role of partial feedback in online nonparametric learning. In *Conference on Learning Theory*. PMLR, 465–481.

Cesa-Bianchi, N., Gentile, C., and Mansour, Y. 2015. Regret minimization for reserve prices in second-price auctions. *IEEE Trans. Inf. Theory 61,* 1, 549–564.

Cesa-Bianchi, N. and Lugosi, G. 2006. *Prediction, learning, and games*. Cambridge university press.

Cesari, T. and Colomboni, R. 2025. An online learning theory of trading-volume maximization. In *ICLR*. OpenReview.net.

Chen, H., Jin, Y., Lu, P., and Zhang, C. 2025. Tight regret bounds for fixed-price bilateral trade. *CoRR abs/2504.04349*.

Coccia, E., Bernasconi, M., and Celli, A. 2026. Nonparametric contextual online bilateral trade. In *ICLR*. OpenReview.net.

Cohen, M. C., Lobel, I., and Leme, R. P. 2020. Feature-based dynamic pricing. *Manag. Sci. 66,* 11, 4921–4943.

Cosson, R., Fusco, F., Gupta, A., Leonardi, S., Leme, R. P., and Russo, M. 2026. Contextual online bilateral trade. *To appear at EC'26, preprint on the arXiv CoRR abs/2602.12903*.

Deng, Y., Mao, J., Sivan, B., and Wang, K. 2022. Approximately efficient bilateral trade. In *STOC*. ACM, 718–721.

Deng, Y., Mao, J., Sivan, B., Wang, K., and Wu, J. 2025. Approximately efficient bilateral trade with samples. In *EC*. ACM, 206–223.

Di Gregorio, S., Dütting, P., Fusco, F., and Schwiegelshohn, C. 2025. Nearly tight regret bounds for profit maximization in bilateral trade. In *FOCS*. IEEE, 1570–1594.

Di Gregorio, S., Dütting, P., Fusco, F., and Schwiegelshohn, C. 2026. Profit maximization in bilateral trade against a smooth adversary. *CoRR abs/2605.12664*.

Di Gregorio, S., Fusco, F., Leonardi, S., and Schwiegelshohn, C. 2026. Private learning in bilateral trade. *CoRR abs/2606.02050*.

Dobzinski, S., Eden, A., Goldner, K., Shaulker, A., and Tsilivis, T. 2025. Bilateral trade with interdependent values: Information vs. approximation. In *EC*. ACM, 641–665.

Dütting, P., Fusco, F., Lazos, P., Leonardi, S., and Reiffenhäuser, R. 2026. Efficient two-sided markets with limited information. *SIAM J. Comput. 55,* 1, 65–92.

Fei, Y. 2022. Improved approximation to first-best gains-from-trade. In *WINE*. Lecture Notes in Computer Science. Springer, 204–218.

Feng, Y., Ma, M., Peng, B., and Wan, Z. 2026. Searching for optimal prices in two-sided markets. *To appear at EC'26, preprint on the arXiv CoRR abs/2602.11691*.

Gaucher, S., Bernasconi, M., Castiglioni, M., Celli, A., and Perchet, V. 2025. Feature-based online bilateral trade. In *ICLR*. OpenReview.net.

Hagerty, K. M. and Rogerson, W. P. 1987. Robust trading mechanisms. *Journal of Economic Theory 42,* 1, 94–107.

Haghtalab, N., Roughgarden, T., and Shetty, A. 2020. Smoothed analysis of online and differentially private learning. In *NeurIPS*.

Hajiaghayi, I., Hajiaghayi, M., Peng, G., and Shin, S. 2025. Gains-from-trade in bilateral trade with a broker. In *SODA*. SIAM, 4827–4860.

Jin, Y. 2026. Tight regret bounds for bilateral trade under semi feedback. *CoRR abs/2601.16412*.

Kang, Z. Y., Pernice, F., and Vondrák, J. 2022. Fixed-price approximations in bilateral trade. In *SODA*. SIAM, 2964–2985.

Kleinberg, R., Slivkins, A., and Upfal, E. 2019. Bandits and experts in metric spaces. *J. ACM 66,* 4, 30:1–30:77.

Kleinberg, R. D. and Leighton, F. T. 2003. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *FOCS*. IEEE Computer Society, 594–605.

Kumar, R. and Kleinberg, R. 2022. Non-monotonic resource utilization in the bandits with knapsacks problem. In *NeurIPS*.

Leme, R. P., Sivan, B., Teng, Y., and Worah, P. 2023. Pricing query complexity of revenue maximization. In *SODA*. SIAM, 399–415.

Liu, A., Leme, R. P., and Schneider, J. 2021. Optimal contextual pricing and extensions. In *SODA*. SIAM, 1059–1078.

Liu, Z., Ren, Z., and Wang, Z. 2023. Improved approximation ratios of fixed-price mechanisms in bilateral trades. In *STOC*. ACM, 751–760.

Lunghi, A., Castiglioni, M., and Marchesi, A. 2025. Online two-sided markets: Many buyers enhance learning. *CoRR abs/2503.01529*.

Lunghi, A., Castiglioni, M., and Marchesi, A. 2026a. Better regret rates in bilateral trade via sublinear budget violation. In *SODA*. SIAM, 6494–6536.

Lunghi, A., Castiglioni, M., and Marchesi, A. 2026b. Regret minimization in bilateral trade with perturbed markets. *CoRR abs/2605.10475*.

Morgenstern, J. and Roughgarden, T. 2016. Learning simple auctions. In *COLT*. JMLR Workshop and Conference Proceedings, vol. 49. JMLR.org, 1298–1318.

Myerson, R. B. 1981. Optimal auction design. *Mathematics of operations research 6,* 1, 58–73.

Myerson, R. B. and Satterthwaite, M. A. 1983. Efficient mechanisms for bilateral trading. *Journal of economic theory 29,* 2, 265–281.

Roughgarden, T. 2020. *Beyond the Worst-Case Analysis of Algorithms*. Cambridge University Press, Cambridge, UK.

Slivkins, A. 2014. Contextual bandits with similarity information. *Journal of Machine Learning Research 15*, 2533–2568.

Slivkins, A., Zhou, X., Sankararaman, K. A., and Foster, D. J. 2024. Contextual bandits with packing and covering constraints: A modular lagrangian approach via regression. *J. Mach. Learn. Res. 25*, 394:1–394:37.

Takimoto, E. and Warmuth, M. K. 2003. Path kernels and multiplicative updates. *J. Mach. Learn. Res. 4*, 773–818.

Talagrand, M. 2014. *Upper and lower bounds for stochastic processes*. Vol. 60. Springer, Berlin, Heidelberg, Germany.

Vickrey, W. 1961. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance 16,* 1, 8–37.

Differential Privacy for Strategic Information Sharing and Learning: An Annotated Reading List

M. AMIN RAHIMIAN
University of Pittsburgh
and
YUXIN LIU
University of Pittsburgh

Differential privacy has evolved from a technical framework for protecting individual-level data into a useful language for studying privacy in economic and strategic environments. Beyond limiting what can be inferred about any single data contributor, differential privacy can shape incentives, participation, information disclosure, and data-driven decisions. This annotated reading list highlights a small set of papers that connect the foundations of differential privacy to recent work on data acquisition, platform design, and operational decision-making under privacy constraints.

General Terms: Algorithms, Economics, Theory
Additional Key Words and Phrases: Differential Privacy, Mechanism Design, Data Acquisition, Platform Decisions, Operational Decisions

1. INTRODUCTION

Differential privacy (DP) was originally developed as a rigorous framework for protecting individual-level data while enabling statistical analysis. The standard formulation of DP requires that a randomized mechanism produce similar output distributions on neighboring datasets [Dwork et al. 2006]. We call two datasets neighboring if they differ in the data of a single individual. Formally, a randomized mechanism M is ε-differentially private if, for any neighboring datasets D and D' and for any possible output set S,

$$\Pr[M(D) \in S] \leq e^{\varepsilon} \Pr[M(D') \in S].$$

This definition limits how much the participation of any single individual can change the distribution of observable outcomes.

For the economics and computation community, this stability property is useful not only as a technical privacy guarantee, but also as a way to reason about incentives and strategic behavior. When the influence of any single report is limited, an agent's ability to manipulate outcomes is also limited. This observation connects differential privacy to mechanism design, large games, and other settings in which agents strategically decide what information to reveal.

Building on this perspective, recent work has used differential privacy to study a broader range of economic and operational questions. Privacy constraints can affect who is willing to contribute data, how much data must be purchased, how platforms use information for personalization, and how firms make data-driven pric-

Authors' addresses: rahimian@pitt.edu, yul435@pitt.edu

ing, recommendations, and inventory decisions. Thus, DP is not only a constraint on statistical release; it can also be a design feature that shapes participation, information use, and operational performance.

This annotated reading list highlights a small set of papers that connect the foundations of differential privacy to recent work on data acquisition, platform design, and operational decision-making under privacy constraints. The selected papers are not intended to be comprehensive. Rather, they offer a concise path from foundational ideas to applications in economic, strategic, and operational settings.

2. ANNOTATED READING LIST

I. Foundations and Game-Theoretic Perspectives

1. *Calibrating Noise to Sensitivity in Private Data Analysis* [Dwork et al. 2006]

This paper introduces differential privacy and the Laplace mechanism, establishing the sensitivity-based approach to rigorous privacy guarantees. It provides the formal starting point for understanding privacy as a stability constraint: the participation of any single individual should have only a limited effect on the distribution of released outcomes.

2. *Mechanism Design via Differential Privacy* [McSherry and Talwar 2007]

This paper made one of the earliest connections between differential privacy and incentives. By introducing the exponential mechanism, it showed that DP can be used not only to protect data but also to stabilize agents' incentives by limiting the influence of any single report on the outcome. Because the potential gain from manipulation is bounded to be near zero, truthful reporting becomes an approximate dominant strategy.

3. *Selling Privacy at Auction* [Ghosh and Roth 2015]

This paper studies agents who have explicit concerns about the privacy loss associated with the use of their personal data. It formulates private-data collection as a mechanism design problem, where agents may need to be compensated for privacy loss before their data can be used to estimate a population statistic. This perspective connects privacy concerns with participation, payments, and the cost of obtaining useful data.

II. Privacy, Data Acquisition, and Participation

4. *Optimal Data Acquisition with Privacy-Aware Agents* [Cummings et al. 2023]

This paper studies data acquisition when agents have heterogeneous privacy costs and derive value from the quality of the learned model. It highlights a central trade-off: stronger privacy can reduce statistical accuracy by adding noise, but it can also increase participation by making agents more willing to share data.

5. *Optimal and Differentially Private Data Acquisition: Central and Local Mechanisms* [Fallah et al. 2024]

This paper studies data acquisition from privacy-sensitive agents who have heterogeneous and privately known privacy costs. The platform must choose privacy

losses, payments, and an estimator to induce truthful participation while estimating an underlying parameter. The paper compares central and local privacy architectures. In the central model, agents share data with a trusted curator, who forms a weighted average and adds aggregate Laplace noise. In the local model, each agent adds Laplace noise before sharing data, so the platform averages already privatized reports. This distinction changes both estimation and mechanism design: the central model supports a more direct score-based mechanism, whereas the local model leads to a harder optimization problem over independently privatized reports.

6. *The Privacy Paradox and Optimal Bias–Variance Trade-offs in Data Acquisition* [Liao et al. 2024]

This paper studies data acquisition when privacy concerns may be correlated with the data being collected. It shows how participation decisions can create bias and develops mechanisms that balance the bias introduced by selective participation against the variance introduced by privacy noise.

III. Differential Privacy in Operational and Platform Decisions

7. *Privacy-Preserving Personalized Revenue Management* [Lei et al. 2024]

This paper studies differential privacy in personalized revenue management. It shows how firms can use historical customer data to make personalized revenue decisions while limiting the disclosure of individual information. A central insight is that, when sufficient historical data are available, privacy protection can sometimes be achieved with only a small additional loss relative to the statistical cost of learning demand.

8. *An Algorithmic Approach to Managing Supply Chain Data Security: The Differentially Private Newsvendor* [Chen and Chua 2026]

This paper brings differential privacy into supply-chain and inventory decisions. In a data-driven newsvendor problem, the ordering decision itself may reveal sensitive demand information. The paper studies how to design private decision rules that protect data while maintaining near-optimal operational performance.

9. *Privacy-Preserving Dynamic Personalized Pricing with Demand Learning* [Chen et al. 2022]

This paper studies dynamic personalized pricing when a firm learns demand from consumer data over time. It shows how privacy constraints interact with demand learning and revenue maximization in terms of regret and develops pricing policies that protect consumer information while preserving useful personalization.

10. *Privacy-Preserving Personalized Recommender Systems* [Fu et al. 2026]

This paper studies personalized recommendation under differential privacy. It analyzes how a platform can use customer data to improve recommendations while limiting the disclosure of individual information. It also illustrates the broader role of differential privacy in platform decisions that depend on personalization.

REFERENCES

Chen, D. and Chua, G. A. 2026. An algorithmic approach to managing supply chain data security: The differentially private newsvendor. *Operations Research 74,* 2, 958–983.

Chen, X., Simchi-Levi, D., and Wang, Y. 2022. Privacy-preserving dynamic personalized pricing with demand learning. *Management Science 68,* 7, 4878–4898.

Cummings, R., Elzayn, H., Pountourakis, E., Gkatzelis, V., and Ziani, J. 2023. Optimal data acquisition with privacy-aware agents. In *2023 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*. IEEE, 210–224.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*. Springer, 265–284.

Fallah, A., Makhdoumi, A., Malekian, A., and Ozdaglar, A. 2024. Optimal and differentially private data acquisition: Central and local mechanisms. *Operations Research 72,* 3, 1105–1123.

Fu, X., Chen, N., Gao, P., and Li, Y. 2026. Privacy-preserving personalized recommender systems. *Manufacturing & Service Operations Management 28,* 1, 271–289.

Ghosh, A. and Roth, A. 2015. Selling privacy at auction. *Games and Economic Behavior 91*, 334–346.

Lei, Y., Miao, S., and Momot, R. 2024. Privacy-preserving personalized revenue management. *Management Science 70,* 7, 4875–4892.

Liao, G., Su, Y., Ziani, J., Wierman, A., and Huang, J. 2024. The privacy paradox and optimal bias–variance trade-offs in data acquisition. *Mathematics of Operations Research 49,* 4, 2749–2767.

McSherry, F. and Talwar, K. 2007. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*. IEEE, 94–103.