# Table of Contents

Editors' Introduction
IRENE LO and INBAL TALGAM-COHEN  
1

SIGecom Winter Meeting 2022 Highlights
EMILY DIANA, MINGZI NIU and GEORGY NOAROV  
3

Algorithmic Fair Allocation of Indivisible Items: A Survey and New Questions
HARIS AZIZ, BO LI, HERVÉ MOULIN and XIAOWEI WU  
24

On the Nisan-Ronen Conjecture
GEORGE CHRISTODOULOU, ELIAS KOUTSOUPIAS and ANNAMARIA KOVACS  
41

On Modeling Human Perceptions of Allocation Policies with Uncertain Outcomes
HODA HEIDARI, SOLON BAROCAS, JON KLEINBERG and KAREN LEVY  
47

Analyzing Data with Systematic Bias
MANOLIS ZAMPETAKIS  
55

Fairness and Equity in Resource Allocation and Decision-Making: An Annotated Reading List
FAIDRA MONACHOU and ANA-ANDREEA STOICA  
64

Economics and Computation Meets Cognitive Biases: A (Biased) Annotated Reading List
SIGAL OREN  
67

Spillover Effects in Online Field Experiments: An Annotated Reading List
YUAN YUAN and TRACY XIAO LIU  
70

Puzzle: Communicating to Plan Noam Nisan’s 60th Birthday Workshop
VINCENT CONITZER  
73

ACM SIGecom Exchanges, Vol. 20, No. 1, July 2022
By submitting an article to the SIGecom Exchanges, and on condition of acceptance by the editor, authors grant to ACM the following non-exclusive, perpetual, worldwide rights:
— to publish the article in print;
— to digitize and post the article in the electronic version of the publication;
— to include the article in the ACM Digital Library and in any Digital Library related services;
— to allow users to make a personal copy of the article for noncommercial, educational or research purposes.

Contributing authors retain copyright to the article, and ACM will refer requests for republication directly to them.
Editors’ Introduction

IRENE LO
Stanford University
and
INBAL TALGAM-COHEN
Technion – Israel Institute of Technology

This summer issue of SIGecom Exchanges begins with coverage of the SIGecom Winter Meeting through a series of interviews with some of the excellent invited speakers. It then includes a survey, three research letters, and three annotated reading lists. It ends with a puzzle in honor of Noam Nisan’s 60th birthday.

We hope many of our readers had a chance to attend the second SIGecom Winter Meeting that took place virtually in January 2022 on the topic of fairness (broadly construed). In this issue, we invite three leading graduate students, Emily Diana, Mingzi Niu and Georgy Noarov, to present the highlights of the event. The students recap the fireside chat between Cynthia Dwork and Sendhil Mullainathan, and proceed with a series of interviews with some of the top researchers who spoke at the meeting: Annie Liang, Ariel Procaccia, Hoda Heidari, Ashesh Rambachan, and Aislinn Bohren. The interviews give a new perspective on the event as well as the topic of fairness in general, and also touch upon how to come up with problems, the grad school experience, and what counts as a good paper.

Haris Aziz, Bo Li, Hervé Moulin and Xiaowei Wu authored a comprehensive survey on algorithmic fair allocation of indivisible items. Their survey highlights common techniques in the design of (approximation) algorithms for allocation, and nicely complements Warut Suksompong’s survey in our last winter issue on fair division under constraints.

A letter from George Christodoulou, Elias Koutsoupias and Annamaria Kovacs presents their recent FOCS’21 breakthrough towards confirming the famous Nisan-Ronen conjecture. Hoda Heidari, Solon Barocas, Jon Kleinberg and Karen Levy describe in their letter their model for comparison among different allocation policies, which focuses on human perceptions of the probability distributions induced by these policies. This work was selected as the Exemplary Paper of the Applied Modeling Track at EC’21. Manolis Zampetakis, winner of the 2020 ACM SIGecom Doctoral Dissertation Award, surveys his exciting recent work on learning from data under systematic bias, where the bias is due to either truncation or self-selection.

For readers wondering which cutting-edge research area to educate themselves on this summer, this issue includes a selection of three annotated reading lists. Faidra Monachou and Ana-Andreea Stoica offer a comprehensive list of resources on fairness and equity in both resource allocation and decision-making. Sigal Oren compiles for the community a list of works on cognitive biases in economics and computation. The final list, by Yuan Yuan and Tracy Xiao Liu, focuses on online

Author’s address: ilo@stanford.edu, italgam@cs.technion.ac.il.
field experiments, following the tutorial on Experimental Design led by Liu at WINE’21.

This issue ends with a puzzle by Vincent Conitzer on the communication complexity of planning a workshop – specifically, one celebrating Noam Nisan’s 60th birthday!

We would like to take this opportunity to thank S. Matthew Weinberg for his outstanding service to our community as co-editor-in-chief of SIGecom Exchanges since 2019. We also extend thanks to Yannai Gonczarowski for his continuing help in putting together the issues of Exchanges. As always, please do not hesitate to reach out to us if you would like to volunteer a letter, survey, annotated reading list or position paper. We hope you find the research showcased in this issue inspiring!
EMILY DIANA  
University of Pennsylvania  
and  
MINGZI NIU  
Rice University  
and  
GEORGY NOAROV  
University of Pennsylvania

Emily Diana is a rising fifth year Ph.D. student in Statistics and Data Science at the Wharton School, University of Pennsylvania, where she is advised by Michael Kearns and Aaron Roth. Her research focuses on the intersection of ethical algorithm design and socially aware machine learning, and she is honored to have been recognized as both a Rising Star in EECS by MIT and a Future Leader in Data Science by the University of Michigan. Before Penn, she received a B.A. in Applied Mathematics from Yale and an M.S. in Statistics from Stanford, and she spent two years as a software developer at Lawrence Livermore National Laboratory.

Mingzi Niu is a rising fifth year Ph.D. student in Economics at Rice University, where she is advised by Mallesh Pai and Hülya Eraslan. Her research interests are primarily in microeconomic theory, with a focus on mechanism design, information theory and behavioral economics. Before Rice, she received a B.A. in Finance and Banking and a B.S. in Mathematics and Statistics at Peking University, and a M.A. in Economics at Duke University.

Georgy Noarov is a rising third year PhD student in Computer and Information Science at the University of Pennsylvania, advised by Michael Kearns and Aaron Roth. Previously, he graduated from Princeton University with a B.A. in Mathematics. His research interests span across the fields of uncertainty quantification, online learning, fairness in machine learning, and algorithmic game theory.

The Second Annual ACM SIGecom Winter Meeting took place virtually on February 23, 2022. Organized by Mallesh Pai and Aaron Roth, it brought together researchers from economics and computation and adjacent communities to focus on the topic of Fairness (broadly construed). The 2022 Winter Meeting featured tutorials and invited speakers spanning many disciplines, as well as a fireside chat and other social activities. We present some highlights from the event, including a recap of the fireside chat with Cynthia Dwork and Sendhil Mullainathan, and interviews we conducted with invited speakers.
Recap of the Fireside Chat with Cynthia Dwork and Sendhil Mullainathan

One of the highlights of the 2022 Winter Meeting was the Fireside Chat, a 30-minute long Q&A session with Cynthia Dwork (Harvard) and Sendhil Mullainathan (UChicago). Both panelists are renowned scholars and authors of seminal research on Fairness in ML. The Fireside Chat was an exciting part of this workshop, giving food for thought to both young researchers and seasoned scientists wishing to enter the field of algorithmic fairness. Here are edited excerpts from several questions that the panelists shared their wisdom on.

**Fair ML theory experts often face the criticism that mathematical fairness research is reductionist and too narrowly focused. As a result, some may argue that it brushes aside real-world structural issues of injustice that do not have straightforward technical solutions. Is this a fair criticism?**

**Mullainathan.** Mathematical fairness researchers think like philosophers. They seek to design a language and framework for discussing and engaging with fairness issues. Mathematics simply extends this philosophical mindset by allowing us to make our statements even more formal and precise. Both theorists and philosophers approach fairness issues broadly, just like humanists would. They search the space of possible definitions and notions of fairness and investigate their interrelationships. In parallel, they constantly perform reality checks on these definitions and look for missing pieces that could be added to the theory. Indeed, any single paper on fairness will typically only look at a specific and narrow aspect of fairness — but together, these papers form a broad and diverse body of research, whose goals and breadth are fundamentally in line with what humanists attempt to do.

**Dwork.** Mathematical fairness research is essential for the field’s future success. This is analogous to how mathematics has revolutionized the field of cryptography: mathematically formal cryptographic protocols and schemes have been instrumental in enabling engineers to build powerful and scalable code for complex cryptosystems. This process of making cryptography rigorous has been taking place since World War II, and has helped us formally reason about crucial questions such as: What exact security guarantees are we trying to achieve? How powerful is the adversary we are defending against? Clearly, modern-day cryptographic software would not have been possible without first attaining this high level of mathematical clarity and precision. Similarly, the field of algorithmic fairness is currently going through the cycle of proposing new mathematical definitions, augmenting them, and proposing new ones. In this manner, we are following a clear path of progress providing an indispensable foundation for concrete, in particular software-based, future fairness solutions.

**Over the last few decades, many predictive models have become central ingredients of automated decision-making tools used by governments and businesses. When deployed in areas such as hiring, lending and law enforcement, the decisions made by these models directly impact people’s lives, potentially in negative ways. For instance, neural network-based facial recognition tools are widely used in law enforcement to identify**
criminals, but they are known to be prone to having baked-in racial biases. Can a researcher developing an ML model predict and forestall any future long-term fairness-related risks that the model may pose once deployed?

**Dwork.** The only, and indispensable, way of identifying and preventing future fairness issues with a model is to take time before deploying it and speak seriously with lots and lots of different groups. This will help you see how what you are doing is received and whether it is perceived as appropriate or inappropriate. This question also directly links to the issue of responsibly scaling AI solutions, which is something that tech companies — including behemoths such as Meta and Google — have been increasingly grappling with.

**Mullainathan.** Our aim should not be to develop extreme degrees of foresight into such future issues. Rather, we should identify fairness flaws in AI models by exercising vigilance. Often, influential AI models can in a matter of 5-10 years become impactful beyond all our initial expectations — and just as they turn out to be unexpectedly powerful, they may become dangerous in various unexpected ways. As a result, we cannot hope to reliably predict fairness-related fallouts — but we can continually monitor the situation to identify any emergent fairness risks.

Can you identify a “Greatest Hits” list of Fair ML papers and books that all researchers entering the field should study?

**Dwork.** The paper *Fairness through Awareness* [Dwork et al. 2012] initiated the study of fairness in machine learning. Among other things, it articulates and elaborates on the difference between individual and group notions of fairness. *Inherent Trade-offs in the Fair Determination of Risk Scores* [Kleinberg et al. 2016] is a seminal paper that demonstrated a fundamental conflict between several very natural definitions of group fairness.

*Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness* [Kearns et al. 2018] and *Multicalibration: Calibration for the Computationally-identifiable Masses* [Hébert-Johnson et al. 2018] contemporaneously introduced multigroup fairness: a setting where fairness guarantees are given for each group in a potentially complex (e.g. large and intersecting) family of population groups. Multigroup fairness can be viewed as providing a bridge between individual and group notions of fairness.

**Mullainathan.** The field of algorithmic fairness is still in its budding stage, so we have ample opportunity to contribute to the literature by coming up with novel models of real-world phenomena we care about. By contrast, in many well-established research areas a lot of modern-day research is a further elaboration of existing models. An excellent general-audience book illustrating how recent fairness research connects with real-world issues and phenomena is *The Ethical Algorithm: The Science of Socially Aware Algorithm Design* [Kearns and Roth 2019].
Interview with Annie Liang

Annie Liang is an Assistant Professor of Economics and Karr Family Assistant Professor of Computer Science at Northwestern University. Her research is in economic theory, and the application of machine learning methods for model building and evaluation.

As an invited speaker at the SIGecom 2022 Winter Meeting, Dr. Liang gave a presentation on her recent work “Algorithmic Design: Fairness vs Accuracy”. This paper is joint work with Jay Lu and Xiaosheng Mu.

Dr. Liang’s talk described an elegant framework addressing the important and delicate issue of balancing accuracy and fairness in automated decision-making. For concreteness, imagine an automated hiring process where a decision-making algorithm receives features of candidates coming from two different population groups, and outputs a binary hiring decision for each candidate. We (the designer) can control which decision-making algorithm is used, as well as regulate what information the algorithm can access about the candidates. The central object of study is the accuracy-fairness Pareto frontier, which characterizes all “optimally fair” ways for us to trade off the algorithm’s performance (i.e. its error rates) on each of the two population groups. Many natural notions of fairness are permitted, including the egalitarian (the group errors must be similar), the Rawlsian (both group errors must be small), and the utilitarian (the overall population error must be small).

Even though it is simple, this framework is surprisingly rich and yields plenty of rigorous qualitative insights on the accuracy-fairness trade-off — in particular, on the implications of the algorithm using or ignoring the candidates’ group identities or group identity correlates. This helps cast new light on some hotly debated real-world topics such as the Ban the Box movement, and the ban on using test scores for admissions purposes implemented by some US colleges.

In our post-meeting interview, Dr. Liang told us more about this exciting research project, and also spoke about her academic trajectory and her perspective on the growth and development of the econ/CS community.

I would like to start by asking you about your academic path so far. What brought you to the econ/CS intersection?

I’ve had an outsider’s respect for CS since undergrad at MIT: I was myself an economics and math major, but computer science was very big there, and I absorbed this idea that computer science was really cool. I didn’t personally get interested in computer science until I was in graduate school in economics. The initial hook for me was different definitions of complexity, but my interests quickly grew from there. And it was a good time to be thinking about CS, since economists were starting to become aware of and get excited about machine learning. Several of the faculty organized reading groups with students to learn about new ideas in CS. And later, I was fortunate enough to do a postdoc at Microsoft Research, where I was exposed still further to computer scientists working at the econ/CS intersection.
What did the econ/CS area look like from your perspective back then? And which directions are you excited about right now?

When I was in grad school, CS/econ was still a niche area — that has changed quite a bit in the last couple of years, which I am really happy about.

I think there are several interesting directions at present. For example, there's the growing area of econometrics and machine learning (related to causal inference), and machine learning is increasingly used as a new tool in empirical economics. I've personally been most involved in the intersection between machine learning and economic theory.

In principle, there's a conflict between the machine-learning, or black-box, way of doing prediction and the way that economic theorists think about model building. Economic models tend to be interpretable theories that offer some narrative or explanation about the underlying behavior, while black box machine learning models are often complicated objects, where it isn’t clear why the black box is predicting what it is. But I’ve always thought that there were potential complementarities between these two methods, and a lot of my recent work has been about how we can use black boxes to better evaluate or improve on economic models.

In the other direction, I think that economics has a lot to add to computer science as well. In the last decade or so, computer science has taken a turn away from developing algorithms with clear, well defined criteria in mind — such as predictive accuracy – to considering these algorithms within a larger social context. Economists can definitely contribute a lot here, because economics has had a long history of developing frameworks and tools for reasoning about social welfare.

What specific benefits do you think economic modeling can contribute to future fairness research?

One of the things I am most excited about regarding the paper that I presented at this workshop is that we were able to import an economics perspective on welfare and preferences in these settings.

Much of the literature in CS on algorithmic fairness literature has focused on a specific optimization criterion—for example, showing how to optimize for efficiency, subject to constraints such as equalized error rates. In the paper that I presented, we defined a broad class of different preferences that the designer might have, varying across many different ways of trading off between fairness and accuracy: from utilitarian preferences to pure egalitarian preferences. Many insights in this paper hold without needing to specify exactly what the objective function is. There are even certain policy recommendations that hold uniformly across this diverse class of fairness-accuracy preferences.

Broadly generalizing, I think there is a tendency in computer science to want to provide a solution to the problem at hand. This is a bit different from the way that economists approach the problems that emerge in social science, where the goal is sometimes not to provide a solution, but rather to figure out how to think about the situation and the inherent trade-offs. There’s clearly value to both approaches.

The paper you presented at the Winter Meeting offers nice geometric insight into the nature of this fairness-accuracy Pareto frontier. At the
same time, this is enabled through the setting being two-dimensional: in particular, the paper mostly deals with two population groups, and the case where you can only undertake binary actions (such as making an accept/reject decision) for each data point. How hard do you think it would be to extend these results to more groups or more actions while keeping the geometric insights intact?

Extending this theory to more groups is probably not difficult. One would need to decide on how to generalize the fairness notion: should we be comparing groups individually, or relative to some average, or looking at the worst-off and best-off groups? But ultimately, this choice probably will not have a qualitative impact on the results.

Regarding more than two actions: the full design case in the first half of the paper extends readily. It’s once we bring in the information design problem that we actually start using the binary nature of the actions.

How did this project begin? Did it stem from you and your collaborators thinking about various fairness problems out there, or perhaps from a concrete mathematical problem?

I’ve been aware of the algorithmic fairness literature in computer science for quite a while, and already during my postdoc people were very excited about it. So I have been following and admiring this literature, and I definitely wanted to write a paper on this topic. My collaborators and I were especially intrigued by the trade-off between fairness and accuracy rates: it seemed evident that such a trade-off might occur when groups have different distributions, and we wanted to know if we could say something about it. All three of us have experience working on papers in information economics, and we naturally were also curious about how the information fed into the algorithm affected the nature of this trade-off. So we began by sketching out some formal models to think through these questions, and went from there.

How fast do you anticipate the econ/CS and fairness areas will be developing in the near future? And what does this mean from the perspective of young researchers in these areas who are planning to enter the job market?

There are many signs indicating that this is an emerging and rapidly developing area. In the last couple of years, graduate students in economics have been increasingly going on the job market with econ/CS papers and getting great jobs. I’ve also been noticing more and more job postings explicitly looking for somebody in this intersection.

I know you have participated in organizing several great workshops focused on game theory, and on the social impact of machine learning. Naturally, these events bring together researchers with unique perspectives on fairness and econ/CS. Could you speak about your experience organizing the workshops and bringing together all these different speakers, and any takeaways?
As I mentioned, this intersection is growing very rapidly — but I find especially interesting that it is growing in many different directions simultaneously. In general, I think one shouldn’t picture the econ/CS intersection as economists (as a monolith) interacting with computer scientists (as a monolith); there are many exchanges going on here, and many different subcommunities involved. For example, my initial exposure to computer scientists was through the algorithmic game theory community, and I only realized after a while that there was a separate machine learning community, with a different (but overlapping) group of people and set of conferences. And in the same way, economic theorists are not the same as econometricians who are not the same as empirical economists, although each of these groups has recently been shaped in some way by computer science. So the most interesting takeaway for me so far has been the vast diversity of these synergies between the two fields.

It will be interesting to see how this evolves. Will there ultimately be a CS-Economics field housing all these different people? Or will each area within economics and within CS be influenced by this interaction in a different way?

As a final question, can you say a bit about your hobbies?

I’ve been a learner of Russian since grad school. At some point, I may decide it’s good enough and move on to something else, but right now I am still really enjoying continuing to improve my understanding of the language.
Interview with Ariel Procaccia

Ariel Procaccia joined the Winter Meeting as an invited speaker to discuss his paper “Fair Algorithms for Selecting Citizens’ Assemblies.” Dr. Procaccia is Gordon McKay Professor of Computer Science at Harvard University and works on problems related to artificial intelligence, algorithms, economics, and society, and he is especially excited about projects that involve both interesting theory and direct applications. Most recently, his sortition algorithm and online framework at Panelot.org has been rapidly adopted by government agencies for their selection processes to form citizens’ assemblies.

Dr. Procaccia was gracious enough to agree to speak with us about fair division, Panelot, and several tidbits of his experience as an academic.

Algorithmic fairness has exploded in the past few years. Where do you see the field going? Similarly, what do you think are the most important open problems and areas for future research in the field right now?

Let me mention one direction that I think is important. The study of fairness in machine learning has developed almost independently from fair division, an area that dates back to the 1940s and has very similar goals: to define rigorous notions of fairness and devise methods for achieving them. Not surprisingly, notions developed in fair division can be applied to fair machine learning. For example, the classic notion of envy-freeness can be used to design fair classifiers: the utility of each individual for their own (possibly random) outcome should be at least as high as their utility for any other individual’s outcome (this makes sense when utilities are heterogenous). Going forward, I believe that ideas from fair division, and, more generally, from normative economics, have a much bigger role to play in fair machine learning.

Can you tell us a bit about your experience with Panelot and working with government officials? For example, how did you get interested in algorithms for political fairness purposes? Was it hard to get your algorithm publicized and launched?

I’ve always been excited about the intersection of computer science and democracy. I got interested in sortition – random selection of representatives – specifically, when my Ph.D. student Paul Götz recommended to me an amazing book, “Against Elections” by David Van Reybrouck. In 2019 I wrote an opinion piece about sortition, which led to conversations with practitioners. Eventually Paul and I were invited to a demonstration of an algorithm for selecting citizens’ assemblies, which was developed and presented by Brett Hennig of the UK-based Sortition Foundation. This was the beginning of a wonderful collaboration with the Sortition Foundation, which later facilitated the deployment of our own algorithm and its adoption by other organizations.
How did you handle the code development and professional software production for Panelot?

Our selection algorithm was coded up by Paul, and in the initial deployment we simply plugged his code into the open-source interface created by Brett. The website Panelot.org, which makes the selection algorithm more easily accessible, was mainly created by Gili Rusak, who was a master’s student at Stanford at the time and will start her Ph.D. at Harvard in the fall. Other contributors (who also played key roles in designing the algorithm) include Bailey Flanigan and Anupam Gupta. To summarize, code development and software production were all done on a pro bono basis by our research group, and the code is open source. (That said, there are some expenses, including the design of a professional logo and, more significantly, running Panelot on AWS.)

Do you have any advice for young researchers?

My number one advice for young researchers is “frequently say no.” Academia has an unusual workflow in that one is asked to do many things (reviews, program committees, talks, department service, etc.) by many people who are not aware of each other’s requests. This issue is especially acute for young faculty members, who typically say “yes” to almost everything and end up being inundated with tasks they can’t complete. Be judicious about what you agree to do.

What do you enjoy doing outside of research?

I have three kids (13, 8 and 3) so between family and work I don’t have a lot of free time. But one thing I still greatly enjoy is playing video games. Currently I’m perhaps 60-70 hours into Elden Ring and, disturbingly, the game claims my progress is 20%, so I expect to finish it around 2025.
Interview with Hoda Heidari

For the last talk of the Winter Meeting, we had the pleasure of listening to Hoda Heidari discuss her recent paper with Jon Kleinberg, “Allocating Opportunities in a Dynamic Model of Intergenerational Mobility.”

Dr. Heidari is an Assistant Professor at Carnegie Mellon University with joint appointments in the Machine Learning Department and the Institute for Software Research. She is broadly interested in societal aspects of artificial intelligence and machine learning and, in particular, algorithmic fairness and accountability.

Dr. Heidari was kind enough to give us an inside perspective on her paper and share her broader experiences as an academic in this growing field.

I see that you recently joined Carnegie Mellon as a faculty member. How has your transition been?

It’s been great. My job responsibilities as a faculty member are different compared to when I was a doctoral student or postdoctoral fellow, so definitely the volume and diversity of responsibilities amps up substantially, but I also have the privilege of advising students and teaching classes that I enjoy. So overall, I have more on my plate, but at the same time more autonomy and opportunities to push my research agenda forward and contribute to training the next generation of researchers in my field.

I see you are teaching a class “Machine Learning, Ethics, and Society.” It sounds exciting to be teaching a course on very new material – is it difficult to have it be comprehensive and fit together as one unit?

It definitely is – as you mentioned, the material is very new and the research community as a whole is still trying to figure out its path and purpose. Currently, my approach is to offer a sample of the existing research landscape. I hope that at the end of the semester the students see a common thread, but I don’t offer the topics as sequentially related to one another. That’s something reflective of where the research is, and I think it is, in a sense, liberating. There are not that many standard topics and methods you feel obliged to cover, so you get to shape the syllabus to teach students how to critically evaluate new situations and problems they may face in the future.

And have you been finding that there are specific things that the students get really excited about?

It’s amazing how engaged the students are in the class discussions. I make sure to have multiple open-ended discussions because this is not a topic you can cover through a one-sided lecture in which you tell them what is the correct or right way of looking at the problem. It’s important to stimulate students’ own ways of reasoning about a new scenario. One area that my students often passionately express their thoughts and experiences on is the issue of fairness, and one common theme in
their comments is about the limited and narrow nature of existing definitions of fairness. Their concern is justified as much of the mathematical modeling that has been done around fairness is really about a very narrow definition of parity in predictive outcomes, and they are only valid for very specifically defined decisions. They do not capture a whole lot of other important factors including procedural and social justice considerations. What was the process by which the decision-making system came to be? What is the institution governing system? What are the checks and balances around it? What does fairness even mean in the specific context of decision-making? These are reflected in students’ questions asking why we are focusing on the specific definitions of fairness and the research community hasn’t moved on to much broader notions. I think the students are right. One way in which I think we can address some of these concerns is through effective engagement with stakeholders and impacted communities. So I make sure to have a module in my course that brings in community-oriented approaches and perspectives.

One question that I had about your paper modeling affirmative action policies is whether there was interest in implementing these decision-making procedures and trying to do any behavioral studies. Is there an application coming up?

I am not sure – with a stylized model of the type we proposed, the point is not necessarily to claim that the model is sufficiently realistic to warrant real-world applications. Rather the purpose is to strip away all sorts of nuances and layers of complication from the real world, so you can rigorously analyze a very stylized, hypothetical world instead. There were two key points that we were trying to make with our analysis. One relates to the discussions around the temporary versus perpetual nature of affirmative action. Usually affirmative action-type policies are cast as temporary measures to level the playing field, but it is not clear what that means and what would be a good way of phasing it out. The second point relates to “fairness interventions” – for instance, enforcing statistical parity is not exactly the same as affirmative action, but it clearly is conceptually similar. So we should carefully consider the dynamic consequences of enforcing such fairness constraints. It’s not just that we are employing this intervention today and we are done tomorrow, but rather we should think about what the impact of it is on the underlying population and how we should update the model moving forward.

Finally I should emphasize that for a highly charged topic like affirmative action, implementation is usually something that is impacted by many political considerations, and we definitely don’t think our work on its own is sufficient to inform such decisions.

Yeah, that makes sense. I like that you were able to frame the problem in a way where you weren’t explicitly weighing the moral implications of affirmative action but rather taking a long-term utilitarian perspective. I’m curious what pushback you got in this paper.

If you look at economic models of affirmative action, one aspect that’s usually accounted for is the strategic component of agents’ behaviors. A typical model
would assume that parents have a certain amount of endowment, and they decide what portion of it to invest in their offspring. We basically ignored strategic considerations and instead focused on the temporal aspect, which was something that was absent from prior work. So the absence of strategic modeling was one valid criticism. The other more conceptual criticism was that some readers had a hard time distinguishing between socioeconomic affirmative action and affirmative action based on demographic characteristics such as race or gender, and the moral implications are obviously very different. We always try to be very clear about that distinction when discussing our work.

Your primary location is in computer science, but you also work in the intersection of computer science and economics. Do you find that there are a lot of challenges translating between the two communities? Do you think that they complement each other well?

Well, having thought for a while about the ethical considerations around machine learning and AI (which are topics in humanities and social sciences), I have realized that the synergy between economics and computer science is already great. There is a common language that people in both fields speak. Game theory, for example, is one such tool used by both computer scientists and economists. We may be more interested in algorithmic problems and they may be interested more in the modeling component, but at the end of the day, we all do math. Now that my work has shifted more towards ethical considerations such as fairness and explainability, I have started collaborating with scholars in philosophy, law, sociology and so on, and we are still in the process of forming that common language. There is already a tradition of computer scientists and economists working together, which has been going on for almost two decades now, whereas the collaborations between computer scientists and scholars in social sciences and humanities is very new. So we are currently at a formative stage – definitely uncertain, but at the same time, immensely exciting.

What is your process like for coming up with problems?

I don’t think I have a very systematic way of coming up with problems. Reading and talking to people, that’s my main source of inspiration for choosing research problems. I find that interdisciplinary exchanges – having conversations with colleagues from different fields, people who have different views or experiences on a topic – are fantastic sources of inspiration for research, so I try to maintain those conversations.

Finally, who is one person who has been quite impactful on your career, not counting your doctoral advisors?

Jon Kleinberg has always been a source of inspiration and my academic role model, and I have been lucky enough to collaborate with him closely over the past few years. When I was a doctoral student at Penn I took a microeconomic course taught by George Mailath. He was one of the best teachers I have seen in my life and definitely what I aspire to emulate as a teacher myself ... although I realize it will likely take twenty to thirty years of research and practice to become a teacher of that caliber!
Interview with Ashesh Rambachan

Ashesh Rambachan presented his job market paper, “Identifying Prediction Mistakes in Observational Data” in the Winter Meeting. Dr. Rambachan completed his Ph.D. in economics at Harvard this May, will be a Post-doctoral Researcher at Microsoft Research in 2022–2023, and will join MIT’s economics department as an Assistant Professor in 2023. His job market paper investigates how to identify systematic errors in human decision makers’ prediction-based decisions when their preferences and private information are unknown to the researcher.

After the talk, Dr. Rambachan generously shared with us about his research and graduate study experience.

Your paper “An Economic Perspective on Algorithm Fairness” with Kleinberg, Ludwig and Mullainathan briefly mentions two opposite forces at work in automated decision-making: the algorithm may simply reflect or correct the bias in data. Could you say more about the interaction between bias and machine learning?

People have reasonable intuition that if you train on historical data that is generated by some discriminatory process to then estimate some machine-learning-based model, the model’s predictions and the resulting decisions may be discriminatory as well. So we try to point out the ways in which that intuition is not quite right. After we estimated some model to predict outcomes as a function of observable features, we ultimately have control over the decision rule that translates these predictions into decisions. We can implement a decision rule based on that prediction function to equalize decisions, whatever our notion of fairness is. This is the point we want to make in my paper with Jon, Jens and Sendhil.

In the “bias in, bias out” work, we try to think about under what conditions the discriminatory data generating process leads to a discriminatory prediction function. To give you an example, from the pretrial release system, we ultimately want to predict whether or not a defendant will fail to appear in courts, but we only observe that outcome if a judge decides to release the defendant historically. One model of a discriminatory data generating process is that the judge is taste-discriminating against minority defendants, meaning the judge sets a higher threshold for the minority group when forming predictions about the probability of release. However, a higher threshold for release for the minority group implies that minorities that are released are positively selected on unobservables or the judge’s private information. Then among the released defendants, minority defendants actually show lower risk. In this example, discrimination in the data generating process will actually yield a prediction function that is more favorable to the minority group.

It seems to me that the majority of the literature looks at issues such as potential biases, or under-representation for minority groups, but sidesteps the ethical aspects about whether or not depending important
life decisions solely on measurable data is the right thing to do in the first place. What is your take on this?

I think that is the heart of the problem in the application of data-driven decision-making tools in social policy domains. The data we have to train these models were generated by human decision makers. As a result, economics values and emphasizes two key features of these settings. First, heterogeneity: there may be important differences across individuals in how they make decisions. How does that affect the data, and in turn affect the models we train on that data? The second is unobservables: individuals may observe extra information that is recorded, and so we have to account for them and how they affect the training data. One thing I want to emphasize is that there may be certain conditions under which we can actually formally test whether human decision makers actually have valuable private information available to them. That is what I do in my job market paper, and I think it is an important diagnostic for users of machine learning tools in high-stake policy settings – first ask could there be private information/unobservables that could explain human decision makers’ choices. If so, this may serve as a strong argument for not using automated decisions. Or we could ask what is that private information, could we in principle go and collect such information.

One key result in your job market paper is a sharp partial identification of correct beliefs under expected utility maximization. If there’s no belief in this range that rationalizes observed choices, then the paper interprets it as “incorrect beliefs”. Alternatively, the gap between observations and model implications may arise because the expected utility theory itself is not descriptively accurate. Could it be that people are making the right decisions, and that we researchers try to use some oversimplified model to fit human behaviors?

First, expected utility theory is the natural benchmark that every empirical researcher in economics is going to reach to model decision-making under uncertainty. So understanding what we can and cannot say about beliefs under expected utility framework is a valuable exercise. Second, I also think of it as a normative restriction on behaviors. For example, it is reasonable to say that the policy makers in the pretrial release setting would like judges to act as if they were expected utility maximizers. Now you could question whether we actually want decision makers to be expected utility maximizers in the first place as opposed to some other decision-theoretic criterion. That objection highlights that when we deploy machine learning tools, it becomes even more important for policymakers to be extraordinarily explicit about (a) the objective function and (b) how exactly they want that objective function to be maximized.

Could you share with us how you came up with ideas for your job market paper?

I came up with this idea because I’ve been spending a lot of time thinking about the use of data-driven tools and the related econometric issues. One thing I kept coming back to is a very simple question: why exactly would a policy maker want to replace a decision maker with an algorithm? What are forces that are on the table?
After conversations with a lot of people, I believe there are three key forces at play in these settings. One is that the policy maker may worry that decision makers mispredict based on observed features. Second they may think that decision makers have an objective function that is misaligned with the policy maker’s objective function. Third, decision makers may have private information. So I wanted to think about how I can actually use data to test whether these forces exist, what the magnitude of these forces is and how understanding these forces could in turn inform the design of algorithmic tools. So that’s where this paper came out.

When writing your job market paper, did you come across any major obstacles or changes in direction?

I would say for me the biggest turning point was thinking about an actual application. Only once I had been thinking through the real-world empirical application, which ended up being the pretrial setting that I focus on in that paper, did I really realize that there was a lot of stuff in the theory that I hadn’t fully though through that was important empirically.

Is there any particular result in your paper that you appreciate the most? Is there anything you are not yet satisfied with and may work further on in the future?

The result I got most excited about is that in this pretrial application, you can test whether choices are consistent with expected utility with any accurate belief under some weak assumptions about the decision maker’s utility and the decision maker’s private information. On the empirical side, when I actually applied that identification result to the data and found that a large fraction of judges are actually making decisions that are inconsistent with the model, I felt that was pretty exciting.

In terms of next steps, I really only provide some limited evidence of what we can say about the misspecified beliefs in the current paper – I provide some results that you can bound the extent of beliefs, and define accordingly overreaction and underreaction. But that could be consistent with many behavioral explanations of what’s driving the prediction mistakes. Is it because judges fail to pay attention to all the evidence available to them? Is it because they overweight or underweight some piece of the information? It would be exciting to know whether we can use this sort of data pinning down whether choices are consistent with some type of behavioral mistakes or not. I think understanding the ways in which beliefs are biased will be helpful in understanding how decision makers like judges respond to the introduction of data-driven tools. If we can say something about why their beliefs are misspecified, perhaps it would suggest sensible policy implications.

What do you view as the biggest challenge as a graduate student? Was there anything you found difficult as a graduate student?

I think it is not particularly unique to me, but managing the transition from being a full-time student in the first two years to working on research was definitely challenging. I was very lucky to have great advisors who can help me out along the way, but that was certainly a challenge.
When learning about an interesting topic, it’s easy to gather many relevant papers. Reading is rewarding but time is limited. How do you cope with the problem of having too much to read?

One piece of advice I got early on in graduate school is that you need to know when you have learned enough about existing literature to start working in it, but not enough to overly shape your thinking with how the literature approaches the problem. To answer the question, you may need to do something different. What I found is that, sometimes it is better to start earlier, and start to think about what you would do in this setting, and once you have that written out, and then going back to read more, rather than trying to read them all in one shot.

In retrospect, is there anything in your Ph.D. student life that you feel glad about and think you are doing particularly right? Is there anything you regret and wish you had been better informed about earlier?

One thing I am really glad I did is to start working on research relatively early on. When I started my third year, I had collaborations with faculty members and other graduate students. The way I learned about how to write papers is by working with faculty, with people who had published papers. That was very helpful. Something I still struggle with is when I remember being frustrated with myself during my Ph.D., not realizing that there are limited chunks of truly productive windows in a day or in a week. It’s good to work hard, but you cannot slam your head on your desk to no end; that’s not going to help you do better research.
Interview with Aislinn Bohren

In one of the tutorial sessions, Aislinn Bohren talked about two of her recent papers on the dynamics of discrimination and systemic discrimination. Dr. Bohren is an Associate Professor of Economics at the University of Pennsylvania, and she works actively on topics in microeconomics including discrimination, misspecified learning and information aggregation. Her work has both theoretical and empirical components.

Dr. Bohren’s first paper, “The Dynamics of Discrimination: Theory and Evidence”, introduces a dynamic dimension in the discussion of discrimination. In this paper, the empirical finding of dynamic reversal – the initial discrimination against one particular group ends up working in favor of this group – provides evidence for belief-based discrimination with incorrect belief. More importantly, this paper points out that dynamic reversal does not offset initial discrimination, and thus this sort of discrimination causes systematic under-rating for the initially discriminated group. The second paper, “Systemic Discrimination: Theory and Measurement”, goes beyond the classic economic position of “direct discrimination” — holding all the other observables fixed to isolate the direct effect of group identity — and proposes the concept of “systemic discrimination”, which demonstrates the pronounced indirect effects of earlier or contemporary discrimination.

In an interview after the winter meeting, Dr. Bohren gave further insight on both papers, and also shared some thoughts on doing economic research.

**How did you start to study algorithmic fairness and issues of fairness in automated algorithms?**

My dissertation research started with the project on the dynamics of discrimination. A lot of discrimination research in economics had focused on a very static question: in this period in time, is there discrimination, can we causally identify it, what are the sources, is it caused by some taste preferences or is it caused by beliefs. In a joint project with Alex Imas, we found a really neat online forum where you could test how discrimination evolves across time. Our intuition was discrimination may evolve especially if it’s caused by beliefs. If people are discriminating at entry-level positions, then that could also affect discrimination at the promotion stage of the subsequent hiring stage.

The platform provides a publicly available reputation score, which is a summary of past performance on the site. So we used that to causally test how discrimination varies with positive past performance reviews. In that paper we found a discrimination reversal by gender. Looking at how math questions were evaluated, we generated the posts ourselves and then randomly assigned the quality of posts to each gender. We found discrimination against users with female names at the entry level, but users with female names are actually favored at the higher education level. Paired with the theoretical analysis, we showed that this is consistent with discrimination that stems from inaccurate beliefs.
In the course of that paper, we thought over the standard definition of discrimination used in economics, which identifies discrimination by holding fixed what’s seen at that decision point and comparing populations with similar observables. But actually if a man and a woman who generate the same initial quality post receive different reputation scores after those posts, we should compare people who have similar quality posts, not people with similar reputation, because their reputation has already embedded some discrimination. That’s what motivated our recent systemic paper, which was trying to broaden the definition of discrimination in economics to capture the systemic factors that seed the differences in the reputation aggregated at the decision point.

I think that’s actually quite closely related to a lot of the algorithmic work in computer science. Within an algorithm, what direct discrimination would correspond to is differential treatments for users who have the same observables except for their group identity. That would be an algorithm that’s explicitly using group identity to discriminate but the missing fact is that the algorithm might be trained on data that already has some discrimination baked into it. Even if you’re using a group blind decision rule, say, a male user and a female user with the same observed variables are treated the same in the algorithm, still a male and a female with the same underlying productivity may be treated differently, because they have different observables that stem from discrimination in the past. So it’s very closely connected to this idea that algorithms may lead to differential treatments across groups even if they do not explicitly discriminate against any group.

What would you say is a desirable goal for those conducting economic research on discrimination?

The first best would be to eliminate discrimination, but given that people often can’t, any sort of policies to reduce discrimination, any decision node or any sort of way economists can provide evidence for a hypothesis that can reduce discrimination, will at least be moving things in the right direction. A key component of my paper with Alex Imas as well as Peter Hull is what we call total discrimination — the sum of direct and systemic discrimination. Essentially it needs to be defined against a reference point — that’s a choice variable for the researcher. At the one extreme, your reference point is a constant and you can think of that is measuring all group-based disparities starting from birth or even before birth. On the other extreme, you could set that reference point as all the observables now and that would collapse things to the definition of direct discrimination. But there’s a whole continuum of reference points in between. By choosing a reference point in between, it’s not saying that the discrimination that occurred before doesn’t matter or didn’t occur; it’s measuring discrimination that’s occurred since this point.

Little bits of discrimination are added at each point in the pipeline and so trying to figure out where you can make decisions, then isolating the discrimination that’s occurring along your decision nodes, can help researchers figure out effective policies to target that particular discrimination. Different policies will help eliminate discrimination in elementary school versus in higher university, so breaking it into different pieces will let you more effectively target a bit of it at a time. I really think this research can be useful, and people may not be aware of or have a way of thinking about systemic discrimination. So one of our goals of formalizing
it conceptually is to make people more aware that there may be bias baked into things you’re looking at. Once you have a framework thinking about the systemic discrimination, it can help inform the public about a sense in which ways where discrimination may be baked into decisions.

Apart from economics, other social sciences are studying discrimination seriously as well, such as anthropology, sociology and psychology. How do researchers in these disciplines approach discrimination compared to economists?

Other fields like sociology and psychology have been thinking about this idea of systemic discrimination or algorithm fairness things from a broader perspective. So one of our goals of our systemic paper is to provide a framework to formally define and figure out how to measure and identify, using methods in economics, these more broad definitions of discrimination that other fields and other literature have been considering for a while, to try and bring economics up to speed in terms of thinking beyond just direct discrimination.

I also noticed that you have been doing research on misspecified learning. Do you have any suggestions for someone who wants to learn more about this topic, or any specific techniques or open questions that people should be familiar with?

Yeah, I think it’s a relatively new literature, and I do think there’s a lot of open questions. I’d say if you’re interested in further reading, the literature review section in my paper has a lot of citations about other recent work, so I would start there and read through some of the other papers that have been written in the past. I think there’s a lot of room for interesting applications like taking misspecified models and figuring out conceptually when they’re relevant for particular markets and figuring out what sort of different predictions we can get or what empirical facts can be justified by the correct model in terms of belief updating with errors.

Is there any overarching theme or question that you constantly come back to in your research?

Broadly I’m really interested in how people learn from information. The assumption that people correctly interpret information and have rational expectations using Bayesian rules can be too strong. But once you relax that assumption, you, as a researcher, have a lot of degrees of freedom. It’s really important to try and relax those models grounded on empirical work, observations of what’s actually happening, but in a way that doesn’t give you too much freedom. So I’m interested in providing theoretical foundations for these types of questions, and I also think a really important application is discrimination. In that line I’m also interested in taking these insights to more applied settings, like a discrimination setting, to see how they impact markets. That’s one motivation for systemic discrimination; but, more broadly, I think that the idea of systemic discrimination is opening up a whole new door for really interesting research questions and so I’m hoping to also keep working in that area.
How would you distinguish good projects which truly contribute to the literature from those that are minor extensions from the existing literature?

It’s hard to describe generally; it’s more on a case-by-case basis. Basically, if you’d want to make a change to an existing model, you want to have good motivation for it. So you need to start with either some good psychological motivation or evidence for why this extension is relevant, rather than just saying like “Oh, you know, let me try and extend those elements in a random direction”. You also need to generate new predictions that are testable and plausible from your model if you want to get a lot of mileage out of the new restriction. You want to make some sharp predictions and see it tie back to things that you actually see in markets.

What counts as a good paper for you? Of course, good papers should be written well and might use neat techniques. But other than that, is there any particular factor that you care about when judging a paper?

I like papers that are conceptually creative, papers that start with some evidence from the real world and then sort of conceptually push our boundaries and how we’re thinking about things as economists. I think definitely there’s technical papers that make important contributions too; but in terms of what I enjoy reading, I enjoy the conceptual creativity.

Could you share with us how you handle unproductive periods as a researcher?

I think if you’re stuck on something, it’s always good to work on a couple of projects. You don’t want too many, because then you’ll be too scattered; but also if you have only one and get stuck on it, you might just keep staring at the same thing, whereas if you have something else, you can just take a break and work on a different project for a week. Then you go back with a fresh perspective and can see the bigger picture. So have a couple of things you’re working on, so that you don’t be afraid to put something on for a few days. When you’re just trying to do the same thing over and over, it’s not working. For me, if I’m trying to think through something, I’ll take a walk to figure a way to get around that issue.
REFERENCES


Algorithmic Fair Allocation of Indivisible Items: A Survey and New Questions

HARIS AZIZ
UNSW Sydney
and
BO LI
The Hong Kong Polytechnic University
and
HERVÉ MOULIN
University of Glasgow, Higher School of Economics St. Petersburg
and
XIAOWEI WU
University of Macau

The theory of algorithmic fair allocation is at the center of multi-agent systems and economics in recent decades due to its industrial and social importance. At a high level, the problem is to assign a set of items that are either goods or chores to a set of agents so that every agent is happy with what she obtains. In this survey, we focus on indivisible items, for which exact fairness as measured by envy-freeness and proportionality cannot be guaranteed. One main theme in the recent research agenda is designing algorithms that approximately achieve fairness criteria. We aim at presenting a comprehensive survey of recent progress through the prism of algorithms, highlighting the ways to relax fairness notions and common techniques to design algorithms, as well as the most interesting questions for future research.

Categories and Subject Descriptors: I.2.11 [Distributed Artificial Intelligence]: Multiagent Systems; J.4 [Computer Applications]: Social and Behavioral Sciences—Economics

General Terms: Theory, Algorithms, Economics

Additional Key Words and Phrases: Fair Allocation, Envy-free, Proportional, Maximin Share

1. INTRODUCTION

While fair allocation is an age-old problem and the widely known Divide-and-Choose algorithm can be traced back to the Bible, modern research on fair allocation is regarded to be initiated by Steinhaus at a meeting of the Econometric Society in Washington D.C. in 1947 (Steinhaus, 1948). Since then, a large body of work in economics and mathematics has been directed towards understanding the theory of allocating resources among agents in a fair manner (Moulin, 2003). The recent focus on indivisible items is motivated, in part, by the applications that inherently entail allocation of items that cannot be fractionally allocated, such as assigning computational resources in a cloud computing environment and courses
to teachers in a school. In the last decade, computer science has offered a fresh and practical angle to the research agenda – algorithmic fair allocation. In addition to designing algorithms, computer science has brought many more ideas, such as computational and communication complexity, and informational assumptions, which do not align with the main theme of the current survey. Interested readers can refer to the surveys by Walsh (2020) and Aziz (2020a) for detailed discussion. Fair allocation algorithms have been implemented in the real world; for instance, Course Match is employed for course allocation at the Wharton School in the University of Pennsylvania, and the websites Spliddit (spliddit.org) and Fair Outcomes (fairoutcomes.com) provide online access to fair allocation algorithms.

Although this survey mainly focuses on indivisible items, the study of fair allocation was classically centered around allocating a divisible resource, which is also known as the cake-cutting problem (Brams and Taylor, 1996; Robertson and Webb, 1998). Fairness is mostly captured by envy-freeness and proportionality in the literature. An envy-free allocation (which is also proportional) of a divisible cake always exists and can be found in bounded steps (Aziz and Mackenzie, 2016). Moreover, a competitive equilibrium from equal incomes guarantees envy-freeness and Pareto optimality simultaneously (Varian, 1973). A recent line of research extends the study to chores, such as the computation of envy-free allocations (Dehghani et al., 2018) and competitive equilibria (Boodaghians et al., 2021; Chaudhury et al., 2021a). Unlike divisible items, when items are indivisible, absolutely fair allocations rarely exist. For example, when allocating a single item to two agents, no allocation is envy-free or proportional. Accordingly, an extensively studied subject is to investigate the extent to which these fairness notions or their relaxations can be approximately satisfied.

There are several surveys highlighting different perspectives of fair allocation theory. Moulin (2018) reviewed the theory through the prism of economics. Aleksandrov and Walsh (2020) and Suksompong (2021) respectively focused on the online and constrained settings. Lang and Rothe (2016), Walsh (2020) and Aziz (2020a) reviewed the problem in the perspective of broad computer science. Instead, the angle of the current survey is algorithmic and the focus is particularly on the introduction of common techniques to design (approximation) algorithms. Moreover, we will discuss more sophisticated settings introduced in the last couple of years that uncovered new challenges and open problems in the field of fair allocation.

Roadmap. In the remaining of the survey, we define the model of fair allocation in Section 2 and introduce the widely adopted solution concepts in Sections 3 and 4. In Section 5, we review the commonly used techniques to design fair allocation algorithms. In Section 6, we introduce more sophisticated settings that have been proposed recently. Finally, we discuss two more properties that may be desirable to be satisfied together with fairness, efficiency and truthfulness, in Section 7.

2. MODEL

In a fair allocation instance, we allocate a set of $m$ indivisible items $M = \{1, \ldots, m\}$ to a group of $n$ agents $N = \{1, \ldots, n\}$. An allocation is represented by an $n$-partition $X = (X_1, \ldots, X_n)$ of $M$, where $X_i \subseteq M$ is the bundle allocated to agent $i$. It is required that each item is allocated to exactly one agent, i.e., $X_i \cap X_j = \emptyset$ for all
\(i \neq j\) and \(\cup_{i \in N} X_i = M\). If \(\cup_{i \in N} X_i \neq M\), \(\mathbf{X}\) is a partial allocation. We sometimes consider fractional allocations, denoted by \(\mathbf{x} = (x_{ie})_{i \in N, e \in M}\), where \(0 \leq x_{ie} \leq 1\) denotes the fraction of item \(e\) allocated to agent \(i\), and \(\sum_{i \in N} x_{ie} = 1\) for all \(e \in M\). Each agent \(i\) has a valuation function \(v_i : 2^M \rightarrow \mathbb{R}\) that assigns a value to each bundle of items. When \(v_i(S) \geq 0\) for all \(i \in N\) and \(S \subseteq M\), the items are goods; when \(v_i(S) \leq 0\) for all \(i\) and \(S\), the items are chores. For ease of exposition, we mainly discuss the case when the valuations are additive and leave the discussion on more general valuations to Section 6. That is, for any \(i \in N\) and \(S \subseteq M\), we have \(v_i(S) = \sum_{e \in S} v_i(\{e\})\). When there is no confusion, we use \(v_i(e)\) to denote \(v_i(\{e\})\). Further, for any \(S \subseteq M\) and \(e \in M\), we use \(S + e\) and \(S - e\) to denote \(S \cup \{e\}\) and \(S \setminus \{e\}\), respectively. Let \(\mathcal{I} = (N, M, \mathbf{v})\) be a fair allocation instance where \(\mathbf{v} = (v_1, \ldots, v_n)\). When all agents agree on the same ordering of all items in values (i.e., \(v_i(1) \geq \cdots \geq v_i(n)\) for all \(i \in N\)), the instance is called identical ordering (IDO).

Before the extensive study of fairness, efficiency was at the centre of the theory of resource allocation. The utilitarian welfare of an allocation \(\mathbf{X}\) is \(\sum_{i \in N} v_i(X_i)\), by maximizing which the total happiness of the agents is maximized. The egalitarian welfare is \(\min_{i \in N} \{v_i(X_i)\}\), by maximizing which the smallest happiness is maximized. A compromise between utilitarian and egalitarian welfare is Nash welfare, i.e., \(\Pi_{i \in N} v_i(X_i)\). We say an allocation \(\mathbf{X}\) Pareto dominates another allocation \(\mathbf{X}'\) if \(v_i(X_i) \geq v_i(X'_i)\) for all \(i \in N\) and \(v_i(X_i) > v_i(X'_i)\) for some \(i\). An allocation is Pareto optimal (PO) if it is not Pareto dominated by any other allocation.

Naturally, the fairness of an allocation can be evaluated by its egalitarian welfare, as in the Santa Claus problem (Bansal and Sviridenko, 2006) and the load balancing problem (Lenstra et al., 1990). However, in practice, since agents may have heterogeneous valuations, the max-min objective is not enough to satisfy all of them. Thus, various notions were proposed to characterize the fairness of allocations, including envy-freeness (EF) (Foley, 1966), proportionality (PROP) (Steinhaus, 1948) and equitability (EQ) (Dubins and Spanier, 1961). The relationships among these notions are discussed by Amanatidis et al. (2018), Sun et al. (2021) and Chakraborty et al. (2021). Given the vast literature on different fairness notions, this survey only focuses on two of the most widely studied, namely EF and PROP.

3. ENvy-FREEness

We first consider envy-freeness, the study of which dates back to Foley (1966) and Tinbergen (1930), and its relaxations.

**Definition 1** EF. For the allocation of items (goods or chores), an allocation \(\mathbf{X}\) is envy-free (EF) if for any two agents \(i, j \in N\), we have \(v_i(X_i) \geq v_j(X_j)\).

The problem of checking whether a given instance admits an EF allocation is NP-complete even for \(\{0, 1\}\)- or \(\{0, -1\}\)-valued instances (Aziz et al., 2015; Bhaskar et al., 2021). Moreover, the example of allocating a single item between two agents defies any bounded multiplicative approximation of EF, and thus researchers turn their attention to additive approximations. Two of the most popular ones are envy-free up to one item (EF1) and envy-free up to any item (EFX).
The notion of EF1 was first studied for the allocation of goods by Lipton et al. (2004), which allows an agent to envy another agent but requires that the envy can be eliminated by removing an item from the envied agent’s bundle. This notion naturally extends to chores by removing an item from the envious agent’s bundle. For both goods and chores, EF1 allocations always exist and can be efficiently computed by the Round Robin algorithm; see Section 5.

**Definition 2 α-EF1.** For any $\alpha \geq 0$, an allocation $X$ is $\alpha$-approximate envy-free up to one item ($\alpha$-EF1) if for any $i, j \in N$, there exists $e \in X_i \cup X_j$ such that $v_i(X_i - e) \geq \alpha \cdot v_i(X_j - e)$. When $\alpha = 1$, the allocation is EF1.

As with EF1, the EFX relaxation was proposed for the allocation of goods, by Caragiannis et al. (2019b). Informally speaking, the notion of EFX strengthens the fairness by requiring that the envy between two agents can be eliminated by removing any item owned by these two agents.

**Definition 3 α-EFX.** For any $\alpha \geq 0$, an allocation $X$ for goods (resp. chores) is $\alpha$-approximate envy-free up to any item ($\alpha$-EFX) if for any $i, j \in N$ and any $e \in X_j$ (resp. $e \in X_i$), $v_i(X_i) \geq \alpha \cdot v_i(X_j - e)$ (resp. $v_i(X_i - e) \geq \alpha \cdot v_i(X_j)$). When $\alpha = 1$, the allocation is EFX.

Unlike the case of EF1 allocations, the existence of EFX allocations remains unknown. For the case of goods, it was shown by Plaut and Roughgarden (2020) that EFX allocations exist in some special cases: (1) identical (combinatorial) valuations, (2) IDO additive valuations, and (3) $n = 2$. Chaudhury et al. (2020) and Amanatidis et al. (2021a) further extended the existence of EFX allocations to the cases when (4) $n = 3$, and (5) bi-valued valuations. In contrast to the case of goods, the chores counterpart is much less well studied. EFX allocations for chores are known to exist only for a few special cases, e.g., IDO instances (Li et al., 2022) and leveled preference instances (Gafni et al., 2021). The existence of EFX allocations for chores remains unknown even for $n = 3$ agents or bi-valued instance.

**Open Problem 1.** Do EFX allocations always exist (for both goods and chores)?

While the existence of EFX allocations remains unknown for the general cases, there are fruitful results regarding EFX partial allocations (where unallocated items are assumed to be donated to a charity) and approximation of EFX allocations.

Since allocating nothing to the agents is trivially EFX, researchers are interested in finding EFX partial allocations with high efficiency. Caragiannis et al. (2019a) showed that there exists an EFX partial allocation achieving half of the maximum Nash welfare. Chaudhury et al. (2021d) proposed a pseudo-polynomial time algorithm that computes an EFX partial allocation with at most $n - 1$ unallocated items under which no agent envies the charity. This result was improved by Berger et al. (2021), who showed that there is an EFX allocation with at most a single unallocated item for $n = 4$, and $n - 2$ unallocated items for $n \geq 5$.

There are also results that aim at computing approximately EFX allocations. Plaut and Roughgarden (2020) showed that every instance (even with subadditive valuations) admits a $0.5$-EFX allocation. The approximation ratio was improved to $0.618$ under additive valuations by a polynomial time algorithm proposed by Amanatidis et al. (2020). Chaudhury et al. (2021b) proposed a polynomial time
algorithm that computes a \((1 - \epsilon)\)-EFX allocation with \(o(n)\) unallocated items and high Nash welfare. For the allocation of chores, only an \(O(n^2)\) approximation of EFX is known to exist (Zhou and Wu, 2021).

4. PROPORTIONALITY AND MAXIMIN SHARE

Proportionality (PROP) was proposed by Steinhaus (1948), and is the most widely studied threshold-based solution concept. PROP is weaker than EF under additive valuations.

\textbf{Definition 4 PROP.} An allocation \(X\) is proportional (PROP) if for every agent \(i \in N\), we have \(v_i(X_i) \geq \text{PROP}_i\), where \(\text{PROP}_i = \frac{1}{n} \cdot v_i(M)\).

For divisible goods and normalized valuations, the items can be allocated such that every agent has value at least \(1/n\), which is not true for indivisible items. Hill (1987) studied the worst case guarantee that an agent can have as a function of \(n\) and \(\max_{i \in N, e \in M} \{v_i(e)\}\). With two agents, the chores version is equivalent to the goods one; but with three or more agents, the equivalence is far from clear, and may not hold. One drawback of this guarantee is that the value of the function decreases quickly and goes to 0 as \(\max_{i,e} \{v_i(e)\}\) becomes large.

\textbf{Open Problem 2.} For chores, what is the worst case guarantee that an agent has as a function of \(n\) and the values of the agents?

\textit{Maximin Share Fairness.} Besides the worst case guarantee studied by Hill (1987), one popular relaxation of PROP is maximin share fairness, motivated by the following imaginary experiment. If agent \(i\) is the mediator and divides all items into \(n\) bundles, the best way to approximate PROP for \(i\) is to maximize the smallest bundle according to \(v_i\). Formally, define the maximin share (MMS) of \(i\) as

\[\text{MMS}_i(M, n) = \max_{X \in \Pi_n(M)} \min_j \{v_i(X_j)\},\]

where \(\Pi_n(M)\) denotes the set of all \(n\)-partitions of \(M\). When \(M\) and \(n\) are clear from the context, we write \(\text{MMS}_i\) for short. Note that \(\text{MMS}_i \leq \text{PROP}_i\), and the computation of \(\text{MMS}_i\) is NP-complete.

\textbf{Definition 5 \(\alpha\)-MMS.} For any \(\alpha \geq 0\), an allocation \(X\) is \(\alpha\)-approximate maximin share fair (\(\alpha\)-MMS) if for any \(i \in N\), we have \(v_i(X_i) \geq \alpha \cdot \text{MMS}_i\). When \(\alpha = 1\), the allocation is MMS.

Note that the approximation ratio \(\alpha \leq 1\) for goods and \(\alpha \geq 1\) for chores. The definition of MMS fairness was first introduced by Budish (2011), based on the concept of (Moulin, 1990). Unfortunately, it is shown that there exist instances for which no allocation can ensure \(\text{MMS}_i\) value for every agent for the case of goods (Kurokawa et al., 2018) and chores (Aziz et al., 2017b). The best known approximation results are \((3/4 + 1/(12n))\)-MMS for goods (Garg and Taki, 2021) and \(11/9\)-MMS for chores (Huang and Lu, 2021). The best known negative results are that \(\alpha \leq 39/40\) for goods and for \(\alpha \geq 44/43\) for chores by Feige et al. (2021).

\textbf{Open Problem 3.} What are the best possible approximation ratios of MMS allocations (for both goods and chores)?
More Solution Concepts. Motivated by the definition of MMS, Caragiannis et al. (2019b) proposed pairwise MMS (PMMS) and Barman et al. (2018a) proposed groupwise MMS (GMMS). Informally, PMMS is similar to MMS, but instead requires that the allocation is MMS for the instance induced by any two agents. GMMS generalizes both MMS and PMMS and requires that the allocation is MMS for the instance induced by any subset of agents. We refer the readers to, e.g., (Caragiannis et al., 2019b; Barman et al., 2018a; Amanatidis et al., 2020), for more detailed discussions, and we summarize the main open problem as follows.

**Open Problem 4.** Do PMMS allocations always exist? What is the best possible approximation of GMMS?

Finally, similar to EF1 and EFX, we can relax PROP to PROP1 and PROPX. It is known that a PROP1 allocation always exists and can be found in polynomial time when the items are goods (Conitzer et al., 2017; Barman and Krishnamurthy, 2019), chores (Brânzei and Sandomirskiy, 2019) or mixture of goods and chores (Aziz et al., 2020b). Regarding PROPX, when items are goods, PROPX allocations may not exist (Moulin, 2018; Aziz et al., 2020b). However, when items are chores, PROPX allocations exist and can be found efficiently (Moulin, 2018; Li et al., 2022). Recently, Baklanov et al. (2021) further proposed PROPm for goods that sits between PROP1 and PROPX, and is guaranteed to exist.

5. ALGORITHMS AND COMMON TECHNIQUES

In this section, we introduce the techniques to design fair allocation algorithms. Due to the vast literature, we choose some of the most commonly used and powerful ones that are also the basis of more complicated algorithms.

5.1 Divide-and-Choose

Divide-and-Choose is one of the most classic allocation algorithms. The algorithm is very useful and intuitive when there are only two agents. The idea is to let the first agent partition the items into two bundles and the other agent choose her preferred bundle. The remaining bundle is allocated to the first agent, and thus her best strategy is to maximize the value of the smaller bundle, i.e.,

\[
(X_1, X_2) \in \arg \max_{(S_1, S_2) \in \Pi_2(M)} \min\{v_1(S_1), v_1(S_2)\}.
\]

Plaut and Roughgarden (2020) proved that \((X_1, X_2)\) is always MMS and EFX to the first agent if we break ties by maximizing the size of the smaller bundle in \((X_1, X_2)\), which they term the leximin++ allocation. This result holds for any number of agents, which implies the existence of EFX allocations for the case of identical valuations. Since the second agent obtains her preferred bundle in \((X_1, X_2)\), the allocation is EF to her. Therefore, with two agents, Divide-and-Choose algorithm returns an allocation that is MMS and EFX.

5.2 Adjusted-Winner

Adjusted-Winner is another widely used algorithm for the two-agent case (Brams and Taylor, 1996). The idea is to sort the items according to the ratios between...
the utilities that they yield for the two agents, i.e.,

\[
\frac{v_1(1)}{v_2(1)} \geq \frac{v_1(2)}{v_2(2)} \geq \cdots \geq \frac{v_1(m)}{v_2(m)},
\]

and let agent 1 choose a minimal set of consecutive items for which she is EF1 starting from left (the remaining items are given to agent 2). The advantage is that it ensures high social welfare between two agents (Bei et al., 2021c). This allocation is EF1 but not necessarily MMS or EFX.

### 5.3 Sequential Allocation and Round-Robin

A general class of algorithms that are also suitable for a distributed implementation is that of sequential-picking allocation (Brams and Taylor, 2000), which was formally studied in a general and systematic way by Bouveret and Lang (2011). Under these methods, agents have a sequence of turns to pick their most preferred item that is still available. A popular sequence protocol is the Round-Robin, where the picking sequence repeats the pattern 1, ..., n. The Round-Robin algorithm produces allocations that are EF1 (but not necessarily EFX) for both goods and chores\(^1\), but not for mixtures of them. For this, Aziz et al. (2020b) proposed the double Round-Robin method that computes EF1 allocations for mixture of goods and chores. Amanatidis et al. (2016) and Aziz et al. (2020a) designed more involved picking sequences to approximate MMS fairness, for goods and chores respectively.

### 5.4 Envy-cycle Elimination

The Envy-cycle Elimination algorithm is inherently a greedy algorithm, where in each round a new item is assigned to the agent who is at a disadvantage for goods or advantage for chores (Lipton et al., 2004). The main technique of the algorithm is to ensure the existence of an agent that is not envied (for goods) or not envious (for chores) by trading items among agents. We use goods as an illustration. The algorithm is based on an envy graph, where the nodes correspond to agents and there is an edge from agent i to agent j if i is envious of j’s bundle. The algorithm works by assigning, at each step, an unassigned item to an agent who is not envied by any other agents, i.e., a node with in-degree 0 in the envy graph. If no such agent exists, the graph must contain a directed cycle. Then the cycle can be resolved by exchanging the bundles of items along the cycle, i.e., an agent in the cycle gets the bundle of the agent she points to. The algorithm terminates when all items are allocated and outputs an EF1 allocation for arbitrary monotone combinatorial valuation functions (Lipton et al., 2004).

The algorithm and its adaptations are very widely studied, combining with which stronger fairness notions can also be satisfied. For example, the algorithm itself ensures EFX (Plaut and Roughgarden, 2020) and 2/3-MMS (Barman and Krishnamurthy, 2020) for IDO instances. With more involved preprocessing procedures, it can ensure 0.618-EFX, 0.553-GMMS, 0.667-PMMS and EF1 simultaneously (Amanatidis et al., 2020). For chores, it is shown in (Barman and Krishnamurthy, 2020)

---

\(^1\)When items are only goods or only chores, there is a larger class of protocols ensuring EF1. This class of protocols uses a recursively balanced sequence in which at any point, the difference between the number of turns of any of two agents is at most 1.
that the returned allocation is $4/3$-MMS. However, noted by Bhaskar et al. (2021), this allocation may not be EF1 if the cycle is resolved arbitrarily. Instead, they used the top-trading technique (in which each agent only points to the agent she envies the most) to preserve EF1. Later, Li et al. (2022) further showed that with this technique, the returned allocation is PROPX. We can also observe the shadow of the algorithm in more complicated techniques, such as the (group) champion graphs and rainbow cycle number (Chaudhury et al., 2020, 2021b) which enable stronger existence and approximation results of EFX.

5.5 Bag-filling Algorithms

Bag-filling Algorithms are particularly helpful for threshold-based notions of fairness like MMS. The idea is to maintain a bag and keep adding items to it until some agent thinks the bag is good enough (for goods) or about to be too bad to all agents (for chores). Then the bag is taken away by some satisfied agent and the algorithm repeats the procedure with the remaining items. The difficulty is to select a proper threshold for the bag so that the approximation for the agent who takes away a bag is good and there remains sufficiently many (or few) items for the remaining agents. With a more careful design and analysis, the approximation ratio can be improved to $2/3$ (Garg et al., 2019) and further better than $3/4$ (Garg and Taki, 2021) for goods. For chores the approximation ratio can be improved to $11/9$ (Huang and Lu, 2021). There are several nice properties regarding MMS fairness (Amanatidis et al., 2017b; Garg et al., 2019), e.g., scale invariance and a reduction to IDO instances. Interestingly, the second property shows that any algorithm for approximating MMS allocations for IDO instances applies to general instances with the same approximation ratio preserved. Making use of these properties can significantly simplify the design of algorithms.

5.6 Rounding Fractional Solutions

Although competitive equilibria may not exist for indivisible items, we can first compute a market equilibrium by assuming the items are divisible and then carefully round the fractional allocation to an integral one (Barman and Krishnamurthy, 2019; Brânzei and Sandomirskiy, 2019; Garg et al., 2021a). This approach is especially helpful when efficiency is desired along with fairness, e.g., for the computation of EF1+PO or PROP1+PO allocations for goods (Barman et al., 2018b; Barman and Krishnamurthy, 2019), EF1+PO allocations for bi-valued chores (Garg et al., 2021b; Ebadian et al., 2021), PROP1+PO (Aziz et al., 2020b) and approximately MMS+PO allocations for mixture of goods and chores (Kulkarni et al., 2021).

5.7 Eating Algorithms

The Probabilistic-Serial (PS) algorithm of Bogomolnaia and Moulin (2001) is a randomized algorithm for allocating indivisible items in an ex-ante EF manner. Agents eat their most preferred items at a uniform rate and move on to the next item when the previous one is consumed. The probability share of an agent for an item is the fraction of the item eaten by the agent. In recent works (Freeman et al., 2020; Aziz, 2020b), researchers have sought allocation algorithms that simultaneously satisfy ex-ante EF and ex-post EF1 for the allocation of indivisible items that are goods or chores. In particular, the PS-lottery method was proposed that provides an
explicit lottery over a set of EF1 allocations. Aziz and Brandl (2020) presented an eating algorithm that is suitable for any type of feasibility constraint and allocation problem with ordinal preferences.

6. MORE SOPHISTICATED SETTINGS

The past several years have witnessed the emergence of more sophisticated settings that brought new challenges to the design of fair allocation algorithms, including the mixture of goods and chores, weighted agents, partial information and general valuations. In the following, we review their models, as well as the corresponding results and open problems.

6.1 Mixture of Goods and Chores

The general case when items are mixture of goods and chores has recently been studied in (Bogomolnaia et al., 2017, 2019; Aziz et al., 2020b, 2022). This model is particularly interesting because it includes the typical setting when the valuations are not monotone. Aziz et al. (2022) proved that a double Round Robin algorithm is able to compute an EF1 allocation for any number of agents, and a generalized adjusted winner algorithm can find an EF1 and PO allocation for two agents. A natural open question is whether PO and EF1 allocations exist for arbitrary number of agents. Recently, Aziz et al. (2020b) and Kulkarni et al. (2021) designed algorithms that compute PROP1+PO or approximately MMS+PO allocations, respectively. More generally, it is an intriguing future research direction to study the fair allocation problem under other non-monotonic valuations.

6.2 Asymmetric Agents

For most of the aforementioned research works, the agents are assumed to be symmetric in the sense of taking the same share in the system. Motivated by real-world scenarios where people in leadership positions take more responsibilities, some recent works studied the fair treatment of non-equals. The definitions of envy-freeness and maximin share fairness have been adapted to the weighted settings by Farhadi et al. (2019), Aziz et al. (2019a) and Chakraborty et al. (2020). Regarding goods, it is shown by Farhadi et al. (2019) that the best approximation ratio for weighted MMS is $\Theta(n)$ and by Chakraborty et al. (2020) that weighted EF1 allocations exist. Regarding chores, although weighted MMS was studied by Aziz et al. (2019a), the best approximation ratio and the existence of weighted EF1 allocations are still unknown. Novel fairness notions, such as AnyPrice share and $l$-out-of-$d$ maximin share, were proposed and studied by Babaioff et al. (2021a,b) which highlight different perspectives of the weighted setting.

**Open Problem 5.** What are the best possible approximations for these weighted fairness notions? Do weighted EF1 allocations exist for chores?

6.3 With Monetary Transfers

Since fairness notion like envy-freeness cannot be satisfied exactly, there are works studying how to use payments or subsidies to compensate agents and achieve fairness accordingly (Halpern and Shah, 2019; Brustle et al., 2020). The problem has been extensively considered in the economics literature under the context of rent
division problems (Edward Su, 1999). Halpern and Shah (2019) aimed at bounding the amount of external subsidies when the marginal value of each item is at most one for every agent, and Brustle et al. (2020) proved that at most one unit of subsidies per agent is sufficient to guarantee the existence of an envy-free allocation. Caragiannis and Ioannidis (2021) studied the optimization problem of computing allocations that are what they term envy-freeable using the minimum amount of subsidies, and designed a fully polynomial time approximation scheme for instances with a constant number of agents. A more general problem is the fair allocation of mixture of divisible and indivisible items, where the divisible item can be viewed as heterogeneous subsidies (Bei et al., 2021a,b).

6.4 Partial Information

Researchers also care about fair allocation with partial information, and particularly the ordinal preference setting, where the algorithm only knows each agent’s ranking over all items without the cardinal values. For goods, the best possible approximation ratio of MMS allocations using only ordinal preferences is $\Omega(\log n)$ by Amanatidis et al. (2016) and Halpern and Shah (2021); for chores, constant upper and lower bounds are proved by Aziz et al. (2020a). Recently, Hosseini et al. (2021) proposed the ordinal MMS fairness, which is more robust to cardinal values. Another interesting question is to investigate the query complexity of unknown valuations. In this model, the algorithms can access the valuations by making queries to an oracle. Oh et al. (2021) proved that $\Theta(\log m)$ queries suffice to define an algorithm that returns EF1 allocations. In general, it is an important research direction to explore how much knowledge is sufficient to design a fair allocation algorithm.

**Open Problem 6.** Explore the trade-off between the amount of knowledge an algorithm has and the fairness guarantee it ensures.

6.5 General Valuations

Besides additive valuations, we may have more complex and combinatorial preferences that involve substitutabilities and complementarities in the items, including submodular, XOS, and subadditive valuations. Formal definitions and discussions of these valuations can be found in (Nisan, 2000). Some of the results we have discussed in previous sections also apply to general valuations. For example, the envy-cycle elimination algorithm returns an EF1 allocation for monotone combinatorial valuations (Lipton et al., 2004). Plaut and Roughgarden (2020) proved the existence of 0.5-EX allocations for subadditive valuations. Regarding MMS, Barman and Krishnamurthy (2020) and Ghodsi et al. (2018) designed polynomial time algorithms to compute constant-approximate allocations for submodular and XOS valuations, and $O(\log n)$-approximate for subadditive valuations. Chaudhury et al. (2021c) further designed algorithms to compute allocations that are approximately EFX and simultaneously achieve $O(n)$-approximation to the maximum Nash welfare for subadditive valuations.

**Open Problem 7.** Can the approximation ratios regarding EFX and MMS for subadditive valuations be improved?
In addition to the above settings, there are more in the literature, such as constrained resources (Suksompong, 2021), public resources (Conitzer et al., 2017; Fain et al., 2018), group fairness (Suksompong, 2018; Conitzer et al., 2019), and dynamic settings (Aleksandrov and Walsh, 2020). We refer the readers to the mentioned works/surveys and the references therein for more details.

7. BEYOND FAIRNESS: EFFICIENCY AND INCENTIVES

Beyond achieving fairness alone, more and more attention is paid to investigating the extent to which we can design algorithms to compute allocations that are fair and simultaneously satisfy other properties, such as efficiency and truthfulness.

7.1 Computing Fair and Efficient Allocations

Although finding an allocation that maximizes the utilitarian welfare is straightforward (by allocating each item to the agent who has highest value), finding such an allocation within fair allocations is NP-hard (Barman et al., 2019). One interesting research question here is to bound the utilitarian welfare loss by enforcing the allocations to be fair, i.e., the price of fairness (Bei et al., 2021c; Barman et al., 2020). Besides utilitarian welfare, a large body of works studied the compatibility between fairness notions and the weaker efficiency notion of PO. For the case of goods, Caragiannis et al. (2019b) proved that the allocation that maximizes Nash welfare is EF1 and PO. Later, Barman et al. (2018b) designed a pseudopolynomial time algorithm for computing EF1+PO allocations. Truly polynomial time algorithms for the problem remain unknown. Barman and Krishnamurthy (2019) designed a polynomial time algorithm for computing PROP1+PO allocations. Regarding the stronger fairness notion of EFX, Amanatidis et al. (2021a) proved that for bi-valued valuations, the allocation that maximizes Nash welfare is EFX+PO, and Garg and Murhekar (2021) improved this result by giving a polynomial time algorithm. Further, Garg and Murhekar (2021) proved that if the valuations have three different values, EFX+PO allocations may not exist. In contrast, Hosseini et al. (2021) proved that if the valuations are lexicographic, EFX+PO allocations exist and can be found in polynomial time.

Open Problem 8. Can EF1+PO allocations be computed in truly polynomial time?

For chores, most of the problems are still open. The good news is that PROP1+PO allocations can be computed in polynomial time, even if the items are a mixture of goods and chores (Aziz et al., 2020b). However, for EF1 and PROPX, their compatibility with PO are still unknown. Ebadian et al. (2021) and Garg et al. (2021b) proved that for bi-valued instances, EF1+PO allocations always exist and can be found efficiently, which are the only exceptions so far.

Open Problem 9. Do EF1/PROPX + PO allocations always exist for chores?

7.2 Being Fair for Strategic Agents

Fair allocation problems are often faced by strategic agents in real-life scenarios, where an agent may intentionally misreport her values for the items to manipulate the outcome of the algorithm and obtain a bundle of higher value. The goal is
to design truthful algorithms in which agents maximize their utilities by reporting true preferences. For two agents, Amanatidis et al. (2017a) gave a complete characterization of truthful algorithms, using which we have the tight approximation bounds for solution concepts such as EF1 and MMS. For an arbitrary number of agents, Amanatidis et al. (2016) and Aziz et al. (2019b) designed truthful approximation algorithms but the tight bounds are still unknown. The aforementioned works consider the social environment where monetary transfers are not allowed. With monetary transfers, polynomial time truthful mechanisms were designed by Barman et al. (2019) for single-parameter valuations, which maximize the social welfare and approximately satisfy fairness notions such as MMS and EF1.

**Open Problem 10.** What are the best possible approximation ratios of EF1 and MMS for truthful algorithms with an arbitrary number of agents?

Another game-theoretic research agenda is to investigate the agents’ strategic behaviours in algorithms that may not be truthful. For example, Amanatidis et al. (2021b) proved that in the Round-Robin algorithm, the allocations induced by pure Nash equilibria are always EF1 (regarding the true values). Bouveret and Lang (2014) and Aziz et al. (2017a) studied the strategic setting in general sequential allocation algorithms. It is interesting to study agents’ behaviours in other algorithms and with other fairness notions.

Acknowledgments

Bo Li is partially funded by the HKSAR RGC under Grant No. PolyU 25211321, NSFC under Grant No. 62102333, and PolyU Start-up under Grant No. P0034420. Xiaowei Wu is partially funded by FDCT (File no. 0143/2020/A3, SKL-IOTSC-2021-2023), the SRG of University of Macau (File no. SRG2020-00020-IOTSC) and GDST (2020B1212030003).

REFERENCES


Algorithmic Fair Allocation of Indivisible Items: A Survey and New Questions


On the Nisan-Ronen Conjecture

GEORGE CHRISTODOULOU
Aristotle University of Thessaloniki
and
ELIAS KOUTSOUPIAS
University of Oxford
and
ANAMARIA KOVACS
Goethe University, Frankfurt M.

The Nisan-Ronen conjecture states that no truthful mechanism for makespan-minimization when allocating m tasks to n unrelated machines can have approximation ratio less than n. Over more than two decades since its formulation, little progress has been made in resolving it and the best known lower bound was a small constant. This note gives an overview of our recent paper that gives a lower bound of $1 + \sqrt{n} - 1$.

Categories and Subject Descriptors: F.2 [Theory of computation]: Mechanism design
General Terms: Algorithms, Economics, Theory
Additional Key Words and Phrases: Algorithmic mechanism design, Nisan-Ronen conjecture

Mechanism design, a main branch of Game Theory and Microeconomics, studies a special class of algorithms, called mechanisms. Unlike traditional algorithms that get their input from a single user, mechanisms solicit the input from different participants (called agents, players, bidders), in the form of preferences over the possible outputs (outcomes). The difficulty of designing such algorithms stems from the fact that the actual preferences of the participants are private information, unknown to the algorithm. The participants are assumed to be utility maximisers who will provide some input that suits their objective and may differ from their true preferences. A truthful mechanism provides incentives such that a truthful input is the best action for each participant.

The question is what kind of problems can be solved within this framework. In their seminal paper that launched the field of algorithmic mechanism design, Nisan and Ronen [Nisan and Ronen 2001] proposed the scheduling problem on unrelated machines as a central problem to capture the algorithmic and information-theoretic aspects of mechanism design. In the classical form of the scheduling problem, which has been extensively studied from the algorithmic perspective, there are n machines that process a set of m tasks; each machine i takes time $t_{ij}$ to process task j. The objective of the algorithm is to allocate each task to a machine in order to minimize the makespan, i.e., the maximum completion time over all machines. In the mechanism design setting, each machine provides as input its processing time.

Authors’ addresses:
gichristo@csd.auth.gr, elias@cs.ox.ac.uk, panni@cs.uni-frankfurt.de

ACM SIGecom Exchanges, Vol. 20, No. 1, July 2022, Pages 41–46
for each task. The selection by Nisan and Ronen of this version of the scheduling problem to study the limitations that truthfulness imposes on algorithm design was a masterstroke, because it turned out to be an extremely rich and challenging setting.

Nisan and Ronen applied the VCG mechanism [Nisan et al. 2007], the most successful generic machinery in mechanism design, which truthfully implements the outcome that maximizes the social welfare. In the case of scheduling, the allocation of the VCG is the greedy allocation: each task is independently assigned to the machine with minimum processing time. This mechanism is truthful, but has poor approximation ratio, $n$. They boldly conjectured that this is the best guarantee that can be achieved by any deterministic (polynomial-time or not) truthful mechanism. The Nisan-Ronen conjecture has been a central problem in algorithmic mechanism design in the last two decades. This note is about a recent paper [Christodoulou et al. 2021a] that made progress towards this conjecture:

**Theorem 1.** There is no deterministic truthful mechanism with approximation ratio better than $1 + \sqrt{n-1}$ for the problem of scheduling $n$ unrelated machines.

This bound is information-theoretic in the sense that it holds for all deterministic mechanisms, regardless of their running time.

It is well known [Saks and Yu 2005; Archer and Kleinberg 2008; Bikhchandani et al. 2006] that a mechanism is truthful if its allocation function is monotone in the values of each machine. Monotonicity in one dimension (i.e., a single task) is the usual notion of monotonicity of the allocation function, and for two or more dimensions, it takes a particular very natural form that is tightly related to the theory of convex functions. Thus one can restate the above theorem as “no monotone algorithm, polynomial-time or not, has approximation ratio less than $1 + \sqrt{n-1}$ for the problem of scheduling $n$ unrelated machines.” In contrast, the approximation ratio for the usual (non-monotone) class of algorithms is trivially 1, for exponential-time algorithms, and at most 2 for polynomial-time ones [Lenstra et al. 1990].

This is the first non-constant lower bound of the Nisan-Ronen problem. Previous results include a lower bound of 2 [Nisan and Ronen 2001], which was improved to 2.41 [Christodoulou et al. 2009], and later to 2.61 [Koutsoupias and Vidali 2012], as well as recent improvements to 2.755 [Giannakopoulou et al. 2020] and to 3 [Dobzinski and Shaulker 2020]. Some ideas used in the proof of the above theorem first appeared in a recent publication [Christodoulou et al. 2020], which established a lower bound of $\sqrt{n-1}$ for all deterministic truthful mechanisms, when the cost of processing a subset of tasks is given by a submodular (or supermodular) set function, instead of an additive function of the standard scheduling setting.

A crucial role in the proof of the above theorem is played by use of special inputs in which each task can be reasonably allocated to only two machines. This is achieved by setting the values of the other machines sufficiently high so that every algorithm with relatively small approximation ratio will avoid them. With this, we focus on a special case of the scheduling problem, the multi-graph scheduling problem [Christodoulou et al. 2021b]: the input is a multi-graph of $n$ nodes (machines) and each edge (task) has two values, one value for each of its nodes; the mechanism must allocate each edge to one of its nodes. Actually the proof of the above theorem employs multi-stars. Using graphs instead of general instances allows us to take
advantage of a useful characterization of mechanisms for 2 machines; no such good characterization is known for 3 or more machines.

1. OUTLINE OF THE PROOF

We provide an outline for a slightly worse\(^1\) lower bound of \(\sqrt{n - 1}\).

1.1 The construction

We consider instances, with \(n\) players (machines) and \(m\) tasks that are partitioned into \(n - 1\) clusters \(C_1, \ldots, C_{n-1}\). Each cluster \(C_i\) contains \(\ell\) tasks and is associated with player \(i \in [n - 1]\); the number \(\ell\) of tasks per cluster needs to be at least exponential in \(n\) for the proof to work. The processing time for a task \(j \in C_i, i \in [n - 1]\), is described by two values: \(t_j\) of player 0 and \(s_j\) of player \(i\); these values are usually in \((0, 1]\). The processing time of every other player \(k \not\in \{0, i\}\) for a task \(j \in C_i\) is sufficiently high, so that no mechanism with bounded approximation ratio would ever allocate \(j\) to them. Hence, we describe an instance \(T\) by only two values \([t_j, s_j]\) per task \(j\).

1.2 Definitions

The proof relies on a characterization of \(2 \times 2\) mechanisms [Christodoulou et al. 2020] that concerns two players and two tasks. Although we consider a multi-player setting, we are able to use it, by fixing all other values except for the values of two tasks \(p\) and \(p'\) of the same cluster (which we call siblings). We refer to this set of instances as a \((p, p')\)-slice; the resulting allocation for \(p, p'\) corresponds to an allocation of a \(2 \times 2\) truthful mechanism.

The central part of the argument that shows the lower bound, uses an induction on the number \(k\) of clusters. The values of the tasks in the remaining \(n - k - 1\) clusters, which we call trivial clusters, play a limited role, but it is important that they do not affect substantially the approximation ratio. Intuitively, a cluster is called trivial if the optimal allocation for all tasks of the cluster has cost sufficiently small (say at most \(n^{-2}\)).

We usually select a single task from each non-trivial cluster, and we call such a selection of tasks regular. A set of instances is called standard for a set of clusters \(C\) if the value of every task \(j \in \cup C\) is \([t_j = 0, s_j = 1]\), and the remaining clusters are trivial. The following definition of a good set of tasks is at the heart of the proof.

**Definition 2 Good set of tasks – bad task.** Consider a truthful mechanism and let \(\alpha = 1/\sqrt{n - 1}\). Fix a standard instance \(T\) for a set of \(k\) clusters \(C\), and a set of regular tasks \(P = \{p_1, \ldots, p_k\}\) from \(C\). The set of tasks \(P\) is called good,\(^2\) if

\(^1\)We expose the main ideas of the proof, and try to provide intuition, hiding intricate details in our definitions and statements which inevitably sacrifices the rigour of many of the statements. We refer the reader to the full version of our paper for complete arguments and for the slightly improved bound of \(1 + \sqrt{n - 1}\).

\(^2\)Our goal is to communicate the high level idea, as a result this definition is oversimplified and far from the more intricate definition that we have in the full version. For example, what we actually need is that there exists a vector \(V\) of open intervals (around the values of the tasks in \(P\)) such that the mechanism allocates all tasks in \(P\) to player 0 for every instance in the set. We call such sets of instances \(V\)-perturbations – and witness of goodness in particular –, which are crucial for correctness and which we consider an important conceptual contribution of our work.
when we replace the value of every task \( p_j \in P \) with \([t_j = \alpha, s_j = 1]\), the mechanism allocates all tasks in \( P \) to player 0. If the latter property is not satisfied, we call \( P \) a bad set. A singleton bad set is simply called bad task. (See Fig 1 for an illustration.)

\[
T = \begin{bmatrix}
0 & 0 & \alpha^* & 0 & 0 & \alpha^* & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}, \quad T' = \begin{bmatrix}
0 & \alpha & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1^* & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
\]

Fig. 1. \( T \) gives an example of a good set \( P = \{3, 4, 7\} \) of size 3. Indeed, tasks 3 \( \in C_1 \), 4 \( \in C_2 \), 7 \( \in C_3 \) are regular, as they belong to distinct clusters, and have values \([t_j = \alpha, s_j = 1]\). The other values are trivial, and the mechanism allocates the tasks in \( P \) to the 0 player (indicated with ‘*’). In \( T' \), task 2 is a bad task, as it has values \([t_2 = \alpha, s_2 = 1]\), all other tasks are trivial, and the task is allocated to player 1.

**Lemma 3 (Main Lemma).** At least one of the following two properties hold for every truthful mechanism:

(i) there exists a bad task

(ii) there exists a good set of \( n - 1 \) tasks.

The Main Lemma immediately implies the desired lower bound on the approximation ratio. The existence of a bad task (Property (i)) means that there is only one non-trivial task \( j \) with values \([t_j = \alpha, s_j = 1]\) and the mechanism does not give it to player 0, which would be the optimal decision. In this case, the approximation ratio is approximately \(1/\alpha = \sqrt{n-1}\). Finally, a good set of \( n - 1 \) tasks (Property (ii)) has approximation ratio \((n-1)/\alpha = \sqrt{n-1}\), giving the desired result.

To obtain a proof of the Main Lemma, we show that there exists a good set of \( k \) tasks for every \( k \in [n-1] \), by induction on \( k \). We start with some regular set of \( k \) tasks, which we call potentially-good set, such that all its subsets of \( k - 1 \) tasks are good. To satisfy all the requirements in the proof, the precise structure of a potentially-good set of tasks is complicated and it is detailed in the full version of our paper.

### 1.3 Outline of the proof of the Main Lemma

We now give a rough outline of the argument that establishes the Main Lemma (Lemma 3). We consider regular instances \( T \) for sets of \( k \) clusters \( C \). If there is a bad task, then we are done. Hence, we show that otherwise, for \( k = n - 1 \) there is a good set of tasks. By induction on \( k \), we show the stronger claim that there are many sets of tasks that are good, the base case \((k = 1)\) being true due to the fact that there are no bad tasks.

We use a probabilistic argument to show that the probability \( b_k \) of a random regular set of tasks \( P = \{p_1, \ldots, p_k\} \) being a bad set is small. In particular we conclude that \( b_{n-1} < 1 \), which establishes the existence of a good set of \( n - 1 \) tasks.
Showing that $b_k$ is small. We show that $b_k$ is small by establishing the following two facts:

**Fact a.** With high probability a randomly selected regular set $P = (p_1, \ldots, p_k)$ is potentially-good.

**Fact b.** If $P = (p_1, \ldots, p_k)$ is a potentially-good set of tasks, either $P$ is good itself or $(P_{-k}, p'_k)$ is good with sufficiently high probability, where $p'_k$ is a random sibling of $p_k$. Roughly speaking, either $P$ is good or almost all other sets are good (with exponentially small probability of the negative event).

Taking into account that $b_1 = 0$, since there are no bad tasks, we can combine these two facts to get that $b_k$ is bounded above by a decreasing function on the number $\ell$ of tasks per cluster. By selecting $\ell$ to be sufficiently large, this establishes that $b_{n-1} < 1$.

Showing Fact b. The difficult part is to establish the second of the above two facts (Fact b). Let’s assume that $P$ is potentially-good but not good. We show that $(P_{-k}, p'_k)$ is good for many $p'_k$’s, as follows:

— First, we observe that none of the tasks in $P = (p_1, \ldots, p_k)$ is given to player 0. This essentially follows from the definition of potentially-good and weak monotonicity.

— Let $p'_k$ be a sibling of $p_k$ and consider the $(p_k, p'_k)$-slice mechanism. This is exactly the point where we exploit the $2 \times 2$ characterization. The proof proceeds by treating carefully all possible cases, as they appear in the characterization [Christodoulou et al. 2020]:

1. **Affine minimizers:** we show that the mechanism is not an affine minimizer almost surely

2. **Relaxed affine minimizers:** we show that the probability that the mechanism is a relaxed affine minimizer is at most $2n^2/\ell$

3. **1-dimensional and constant mechanisms:** we show that 1-dimensional and constant mechanisms do not occur, or the approximation ratio is high

4. **Task independent or relaxed task independent mechanisms:** we show that if the mechanism is task independent or relaxed task independent for each of $k$ appropriately selected random instances from the witness, then $(P_{-k}, p'_k)$ is good

We conclude that for a random sibling $p'_k$, the mechanism must be either task independent or relaxed task independent for all these $k$ instances with probability at least $1 - k 2n^2/\ell$; therefore $(P_{-k}, p'_k)$ is good with probability at least $1 - 2n^3/\ell$.

The first item, i.e., to show that affine minimizers are sparse, exploits an interesting use of (potential-)goodness and linearity, the latter being an important implication of affine maximization. The proof of relaxed affine minimizers uses the same machinery, but it has an extra layer of difficulty, as such mechanisms may have non-linear parts which are hard to handle. In fact, we might have a positive probability (at most $2n^3/\ell$) to pick a wrong sibling $p'_k$ due to this deficiency. The proof of the last item about task independent and relaxed task independent mechanisms has very similar flavor. It is essentially this part that takes away the complications that arise from having to deal with an additive domain.
2. CONCLUSION

The major problem left open is to settle the Nisan-Ronen conjecture. We expect the techniques of this work to be helpful in this direction. The case of randomized or fractional mechanisms appears also very challenging; the best known lower bound of the approximation ratio is $2$ [Mu’alem and Schapira 2018; Christodoulou et al. 2010], embarrassingly lower than the best known upper bound $(n + 1)/2$. The bottleneck of applying the techniques of the current work to these variants appears to be the lack of a good characterization of $2 \times 2$ fractional mechanisms. Finally, although the result of this work indicates that mechanisms constitute a limited subclass of allocation algorithms, a more direct demonstration would be to find a useful characterization of mechanisms for the domain of scheduling and its generalizations.

REFERENCES


On Modeling Human Perceptions of Allocation Policies with Uncertain Outcomes

HODA HEIDARI
Carnegie Mellon University
and
SOLON BAROCAS
Microsoft Research & Cornell University
and
JON KLEINBERG
Cornell University
and
KAREN LEVY
Cornell University

Many policies allocate harms or benefits that are *uncertain* in nature: they produce distributions over the population in which individuals have different probabilities of incurring harm or benefit. Comparing different policies thus involves a comparison of their corresponding probability distributions, and we observe that in many instances the policies selected in practice are hard to explain by preferences based only on the expected value of the total harm or benefit they produce. In cases where the expected value analysis is not a sufficient explanatory framework, what would be a reasonable model for societal preferences over these distributions? Here we investigate explanations based on the framework of *probability weighting* from the behavioral sciences, which over several decades has identified systematic biases in how people perceive probabilities. We show that probability weighting can be used to make predictions about preferences over probabilistic distributions of harm and benefit that function quite differently from expected-value analysis, and in a number of cases provide potential explanations for policy preferences that appear hard to motivate by other means. In particular, we identify optimal policies for minimizing perceived total harm and maximizing perceived total benefit that take the distorting effects of probability weighting into account, and we discuss a number of real-world policies that resemble such allocational strategies. Our analysis does not provide specific recommendations for policy choices, but is instead interpretive in nature, seeking to describe observed phenomena in policy choices.

Categories and Subject Descriptors: J.4 [Social and Behavioral Sciences]: Economics

General Terms: Economics, Human factors, Theory

Additional Key Words and Phrases: probability weighting, uncertain allocations, human perceptions, harm minimization, benefit maximization

1. INTRODUCTION

Societies frequently wrestle with tough decisions regarding the allocation of benefits or burdens among their populations (see, e.g., [Calabresi and Bobbitt 1978; Viscusi 2018]). These decisions—particularly those that involve harm—are immensely dif-
ficult yet often unavoidable. As Sunstein points out, governments regularly pursue policies that lead to harms, including death, among the public: “If government allows new highways to be built, it will know that people will die on those highways; if government allows new power plants to be built, it will know that some people will die from the resulting pollution. [...] Of course it would make sense, in most or all of these domains, to take extra steps to reduce risks. But that proposition does not support the implausible claim that we should disapprove, from the moral point of view, of any action taken when deaths are foreseeable.” [Sunstein 2003] These considerations remain true even when the prospective harms are reduced as much as possible; to the extent that harms remain, we must reason about the impact of policies that produce foreseeable harms.

To make matters more complicated, many of these allocations deal in probabilities of some outcome occurring: when we raise the speed limit by a certain amount, for example, we can estimate to some approximate level the number of additional traffic fatalities that will result [Farmer 2019], but we can say much less about who in particular will die. Thus, for matters involving harm, the policy process necessarily involves a set of choices (even if these choices arise only implicitly) between different distributions of harm over the population. For example, policy $P$ might produce a probability $p_i$ that individual $i$ is harmed, while policy $Q$ might produce a probability $q_i$ that individual $i$ is harmed, for each individual in the population. (To keep the discussion simple, we will think about a single kind of “harm” that can befall people as a result of the policy, rather than adding the complexity of different types or degrees of harm.)

How should we compare the two distributions of harm that arise from policies $P$ and $Q$? Much of the work that underpins mathematical models in these domains, including many of the loss functions that go into algorithmic decisions, tend to be based on expected cost—the idea that we should favor the policy that produces the lower expected harm. In our case, policy $P$ produces a sequence of probabilities $(p_1, p_2, \ldots, p_n)$ over the $n$ members of the population, and its expected harm is the sum $p_1 + p_2 + \cdots + p_n$; we can write a similar expression for the probabilities of harm $(q_1, q_2, \ldots, q_n)$ produced by policy $Q$.

Of course, real-life policymaking is complex, and it is not clear that minimization of expected harm is typically the chief criterion in selecting among policy options. But there is a more basic problem with using expected harm as the criterion: many policy questions about competing distributions of harm begin after we’ve already reduced the total amount of harm to a roughly fixed, low target level, and so the debate is among distributions that all have the same expected level of harm. How, then, should we think about preferences among these competing policy proposals?

1.1 A Real-life Example

We can see the outlines of such debates in a number of settings where a risk of harm is being allocated across a population. In the policies for drafting people into the military in the United States, for example, the government has considered a number of different implementations for randomizing the selection of inductees. (Here, required service in the military is the cost, or harm, that is being allocated according to a probability distribution.) Under a given policy $P$, individual $i$ would learn that they had a probability $p_i$ of being drafted. Crucially, difficult questions
about the implementations of draft systems persist regardless of the desired size of the military; that is, for a given size of the military, the sum of the draft probabilities \( p \) over the population is pinned to this number, but some distributions of these probabilities have been nonetheless viewed as preferable to others.

What accounts for these preferences? We note that discussions of revisions to the draft framed uncertainty itself as a cost being borne by members of the population. As the U.S. Selective Service System notes, prior to the introduction of a structured process for randomization, men knew only that they were eligible to be drafted from the time they turned 18 until they reached age 26: “[this] lack of a system resulted in uncertainty for the potential draftees during the entire time they were within the draft-eligible age group. All throughout a young man’s early 20’s he did not know if he would be drafted” [Selective Service System 2020]. The systems that were subsequently introduced specified priority groups according to age, which had the effect of deliberately producing non-uniform probabilities of being drafted in any given year; under these systems, some people were selected with higher-than-average probability and others with lower-than-average probability.\(^1\) Viewed in terms of distributions, these policy changes had the effect of concentrating the probabilities more heavily on a subset of the eligible population each year, rather than diffusing the probabilities more evenly across everyone.

The quote from the Selective Service System points out that a process that diffuses probabilities too widely seems to create unnecessary (and harmful) levels of uncertainty; but there are, of course, corresponding objections that could be raised to processes that concentrate probabilities too heavily on too small a group.

An abstraction of these questions would therefore consider multiple probability distributions of harm—for example, policy \( P \) producing \((p_1, p_2, \ldots, p_n)\), policy \( Q \) producing \((q_1, q_2, \ldots, q_n)\), and perhaps others—and ask which of these should be preferred as a choice for society. In posing such questions, we are guided by the belief that studying reactions to distributions of harm should draw closely on those parts of the behavioral sciences that have considered how people subjectively evaluate probabilities. We therefore develop a framework based on the concept of probability weighting from behavioral economics.

Our model will allow us to evaluate the Selective Service System’s argument, and similar arguments in other domains, at a broad level—the contention that completely uniform randomization over the draft-eligible population is a sub-optimal policy because the cumulative level of uncertainty felt by the population is unnecessarily high. At first glance, this argument is counter-intuitive: since the size of the military is the same under all the draft policies being considered, isn’t the cumulative level of uncertainty felt by the population also the same under all policies? On closer inspection, though, we find that this decision—to shift the probabilities in a non-uniform direction, and to interpret this as reducing cumulative uncertainty—is

---

\(^1\)Specifically, men were drafted according to “priority year,” with the youngest men being drafted first. During the year a man was 20 years old, he was in the top priority group, with reduced likelihood of being called up each subsequent year. Within each group, call-up order was randomized by lottery according to birthday [Selective Service System 2020]. This prioritization based on a known random ordering of birthdays served as an additional way of concentrating the probabilities on a subset of the population.
very much consistent with the predictions of probability weighting.

2. MOTIVATING THE MODEL

We can adapt our discussions about harm allocations—and complex scenarios such as the military draft—into a stylized example in which a fixed amount of harm must be allocated across a given population. We will argue that different allocations of harm have very different subjective resonances, and it is these differences that behavioral theories of probability weighting aim to illuminate.

Thus, as a thought experiment, consider the following hypothetical example. Suppose we need to allocate 1 unit of harm among 100 individuals. For simplicity, let’s assume all 100 individuals are equally deserving and willing to bear the harm. We might allocate the harm to one specific person (say, Bob), while giving the other 99 people certainty that they are not at risk—hence the probability distribution \((1, 0, \cdots, 0)\). Feeling sorry for Bob, we might instead divide the risk between him and another member of the population, Chloe—and ultimately flip a coin to decide which of them is to bear the harm, while the other 98 people are free and clear; i.e. the distribution \((1/2, 1/2, 0, \cdots, 0)\). Or we could have a third person, David, join Bob and Chloe in the risk pool, lowering the risk for each of them to one-third \((1/3, 1/3, 1/3, 0, \cdots, 0)\). Finally, we might allocate the risk evenly among all 100 individuals, and select the recipient of the harm by random lottery: \((0.01, \cdots, 0.01)\).

How might a policymaker select among these policies? Each of them, ultimately, results in the same amount of harm (1 unit) befalling the population, yet they strike us as intuitively quite different. We may consider it blatantly unfair to single Bob out as a certain victim by concentrating the risk completely on him; and indeed, a long line of work in psychology on the so-called identifiable victim effect suggests that we tend to find such outcomes particularly troubling [Jenni and Loewenstein 1997]. On the other hand, a random lottery distributes the risk equally among all 100 individuals—but in the interim, it forces everybody to worry about their chances of being harmed. (This is the form of uncertainty, and corresponding psychological cost, that the Selective Service System was concerned with in our example of the draft lottery.) The second and third options provide intermediate alternatives. In the second alternative, no one person is harmed with certainty, while, at the same time, the smallest possible number of individuals need bear the risk.

The fact that we may prefer some of the above alternatives to others immediately suggests that a cost-benefit analysis based on expected harm is not sufficient to capture our intuitions—since all the options involve the same expected amount of harm. Likewise, our intuitive reactions to these different proposals do not neatly map onto common concerns with distributive justice, where we tend to worry about the relative impact of allocations on different social groups or subgroups within

\[\text{Philosophy has also grappled with the observation that we tend to recoil at the idea of, for example, harvesting one person's organs to save the lives of five other people. Such cases reveal an intuitive distaste for distributions that aim to reduce the overall amount of harm experienced by a population by focusing those harms on a small subset of people [Thomson 1976]. Note that our framework does not apply to these cases because concentrating costs in these instances actually reduces the total cost (e.g., reducing the total number of deaths from five to one); in our settings, the way a policy allocates harms does not affect the amount of harm imposed on the overall population.} \]
the population, given existing social inequalities. In this case, our reactions have nothing to do with any details about who Bob, Chloe, and Dave happen to be or the social groups to which they belong. What we perceive to be the more desirable allocation instead seems to rest on how we perceive the benefits or harms of being subject to uncertain outcomes.  

An interpretive analysis: Our intention in exploring people’s subjective perceptions of risk probabilities is, emphatically, not to prescribe a “best” mode of allocating probabilities of risk, nor to endorse the underlying policy decisions that give rise to a need to allocate such risk in the first place, nor to treat superficially the variety of other procedural and moral concerns that attend the allocation of harms and benefits to people. Ours is a purely interpretive undertaking; we find that preferences for certain allocation policies involving probabilities are difficult to explain unless we take probability weighting into account.

Policy experts disagree about the extent to which cognitive errors ought to be explicitly incorporated into account in public decision-making. While some consider it inappropriate to base policies on what are essentially misunderstandings, others suggest that we might reasonably consider the “psychic benefits” to the public of protecting against “imaginary” risks [Schneier 2008; Viscusi 2018; Portney 1992; Pollak 1998]. We stake no claim in this debate; our goal is to explore descriptively how people’s subjective perceptions of probabilities might impact preferences regarding such allocations—and how these impacts potentially explain peculiar real-life allocation policies. In this way, our work follows a style of research that seeks to shed light on observed policy outcomes by linking them to our behavioral understanding of latent human preferences for certain types of outcomes over others (see, e.g., [Srivastava et al. 2019; Zhu et al. 2018; Lee et al. 2019] for earlier work in this genre).

All of this still leaves us with a basic question. We have seen examples so far (with others to come) of policy-making favoring some level of randomization, while also steering away from completely uniform randomization that would spread risk of harm diffusely across a population. Is there a model that predicts this type of “intermediate” position that avoids both a concentration of risk on identifiable victims as well as too diffuse a distribution over the whole population? And can such a model be derived from known psychological models of human behavior? In this work we will argue that a preference for these types of intermediate distributions of risk can be derived naturally from the concept of probability weighting, one of the most empirically well-grounded human biases studied in behavioral economics [Kahneman and Tversky 2013], to which we now turn.

---

3To put it differently, the purpose of our work is not to argue that probability weighting tends to result in distributions that disproportionately harm members of specific social groups. Rather, we study human perceptions toward distributions that allocate the same type of harm unevenly across otherwise-equal individuals (without specifying their group memberships). Given the centrality of probability weighting in the empirical study of behavioral biases around uncertainty, we believe that showing how a range of distributional considerations arises purely from probability weighting is of interest independent of the possibility of additional biases.
2.1 A Model Based on Probability Weighting

Motivated by the premise that understanding people’s perceptions of harm/benefit allocations are crucial in designing acceptable policies, we posit that models that solely rely on expected value comparisons may miss crucial aspects of human perceptions toward uncertain allocations—which are in part shaped by probability weighting. As a result, expected value-optimizing algorithms may produce allocations that are behaviorally repugnant to people. Our model can partially explain these reactions using one of the fundamental principles in behavioral sciences. To our knowledge, our work is the first to explore the attractiveness of different uncertain allocation policies by exploring optimal allocations under probability weighting. We make several connections between the optimal allocation patterns suggested by our theory and real-world policy choices that would be otherwise difficult to explain.

Probability weighting begins from the qualitative observation that people tend to overweight small probabilities—behaving as though they are larger than they actually are—and tend to underweight large probabilities—behaving as though they are smaller than they actually are. More generally, probability weighting is the premise that when faced with an uncertain event of probability \( p \), people will tend to behave with respect to this event—for example, when determining risks or evaluating gambles involving the event—as though its probability were not \( p \) but a value \( w(p) \), the weighted version of the probability. This weighting function \( w(p) \) has the two properties noted above: that \( w(p) \) is larger than \( p \) when \( p \) is small, and \( w(p) \) is smaller than \( p \) when \( p \) is large. If we think in terms of the graph of \( w(p) \) as a function of \( p \), people refer to these properties as the “inverse S-shaped” nature of the probability weighting curve. There are a number of different models that derive inverse S-shaped probability weighting curves from simple observations; one influential functional form was provided by Prelec [Prelec 1998], who derived it from a set of underlying axioms about preferences for different types of gambles. The concept of probability weighting has been invoked to explain a number of peculiar behavioral patterns; one of the canonical examples is people’s participation in gambling and lotto games [Quiggin 1991; Kahneman 2011].

We use probability weighting here to ask the following basic question. Suppose there are \( r \) units of harm to be allocated across a population of \( n \) people, and we are evaluating policies that assign individual \( i \) a probability \( p_i \) of receiving harm, subject to the constraint that the sum of \( p_i \) over all individuals \( i \) is \( r \). In the motivating settings discussed so far, it is natural to think of the cost borne by individual \( i \) as the perceived probability \( w(p_i) \)—either because individual \( i \) perceives it this way (via the psychological cost of their own uncertainty) or because the rest of society views it this way (via our discomfort at the idea that \( i \) is an identifiable victim with a perceived probability \( w(p_i) \) of being harmed). We can therefore ask: which probability distribution minimizes this total cost, the sum of \( w(p_i) \) over all

---

4To elaborate further on this connection, note that the cost of buying a lotto ticket is always set to be higher than the expected benefit (i.e., the likelihood of winning times the prize); otherwise, lottery operators would lose money. Nonetheless, people participate in these games in large numbers. Work in behavioral economics has advanced probability weighting as one explanation for this irrational behavior, via the tendency to over-weigh small probabilities—here, the chance of winning the lottery [Quiggin 1991; Kahneman 2011].
individuals $i$? Notice that this question allows for distinctions among probability distributions that all produce the same total expected harm for the population: in particular, all the distributions under consideration have a total expected harm of $r$, but they can nevertheless differ substantially in the sum of $w(p_i)$ over all individuals $i$.

3. OVERVIEW OF FINDINGS

We find that the distributions minimizing the weighted sum of harm probabilities $w(p_i)$ in fact correspond to intermediate distributions of the type we have been discussing qualitatively: distributions that concentrate the risk on a subset of the population, such that each member of the at-risk subset has a probability of harm that is strictly less than 1, while most of the population has a probability of harm equal to 0. The analysis leading to this conclusion involves some subtlety: sums of $S$-shaped functions do not exhibit the nice properties that simpler function classes do, and so minimizing them requires additional complexity in the analysis.

With this model in place, we can also explore the natural complement to this dynamic. Our discussion thus far has focused on probabilities of harm, but there is an analogous class of questions about distributing probabilities of benefit across a population—for example, in the availability of opportunities like higher education or financial assistance programs. Suppose there are $r$ units of benefit available to the population as a whole, and we are considering policies that assign a probability $p_i$ that individual $i$ receives the benefit. Which distributions maximize the sum of $w(p_i)$ over all individuals $i$—that is, maximizing the total perceived benefit? As with risks of harm, we do not argue that such a policy is necessarily desirable, only that it may have added or diminished attractiveness in its perceived impact; to the extent that such policies are favored in practice, the theory of probability weighting might therefore offer a suggestive description.

We find that the distributions maximizing this sum of perceived probabilities of benefit are quite different from the distributions minimizing the sum of perceived probabilities of harm. In particular, when the total available benefit $r$ is small relative to the size of the population under consideration, the maximizing distribution is a uniform lottery which assigns all $n$ people a probability of $r/n$; but as $r$ increases, the maximizing distribution changes abruptly to one in which a subset of the population receives a portion of the benefit with certainty, and the rest of the population is given a uniform lottery for the remainder.

4. IMPLICATIONS

Given that a society developing policy seems to favor some probability distributions of harm or benefit over others, even when they have the same expected value, it is natural to ask whether a model based on probability weighting can shed light on the nature of these preferences. Our modeling activity thus works out what the favored policies would look like if society were seeking to maximize or minimize the total weighted probability. As we discuss in our work, properties of these minimizing and maximizing distributions can be observed in a variety of real-world settings. We consider a number of allocation policies that have been adopted in practice that involve distributions of uncertain harms and benefits that closely resemble...
what our model suggests are optimal under probability weighting. Because the attractiveness of these policies is difficult to explain otherwise, we present them as inductive evidence that probability weighting may be playing a meaningful role in guiding societal preferences for certain allocations and in determining the actual distributions of harms and benefits in society.

REFERENCES


In many data analysis problems, we only have access to biased data due to some systematic bias of the data collection procedure. In this letter, we present a general formulation of systematic bias in data as well as our recent results on how to handle two very fundamental types of systematic bias that arise frequently in econometric studies: truncation bias and self-selection bias.

Categories and Subject Descriptors: J.4 [Social and Behavioral Sciences]: Economics; G.3 [Probability and Statistics]: Multivariate Statistics
General Terms: Algorithms, Economics, Theory
Additional Key Words and Phrases: bias, truncation, censoring, self-selection

1. INTRODUCTION

Many problems in data analysis involve the estimation of a property of an unknown probability distribution $P$, given a set of finite samples from $P$. In some settings, the goal is to find a full description of the cumulative distribution function or the probability density function of $P$ (density estimation). In other settings, each sample has the form $(x, y)$ and the goal is to estimate the distribution of $y$ given $x$ (regression or classification).

A key assumption underlying many widely used methods is that we have access to samples that are independently and identically distributed (i.i.d.)—throughout the sampling process, each sample is drawn under the same conditions, it does not affect the rest of the samples, and it is guaranteed to be drawn from the distribution of interest, $P$. However, this assumption ignores many challenges in the data collection procedure that lead to biased or corrupted datasets. The presence of bias or corruption in the data can lead to statistical conclusions that are fallacious or unfair [Mehrabi et al. 2021]. As a result, identifying the sources of bias or corruption, and most importantly developing ways to perform statistical analysis even in the presence of bias is a fundamental problem with many applications in a wide range of scientific areas, including econometrics [Maddala 1986].

Our goal in this letter is to formulate and understand the effects of systematic bias in data analysis. In particular, our main focus is on two fundamental types of systematic bias: truncation bias, and self-selection bias which we introduce in Section 1.1 and Section 1.2, respectively. In Section 2 we provide a general framework that captures systematic bias and we show how truncation and self-selection bias can be realized in this framework. In Section 3, we present some of our recent results on addressing these types of bias for density estimation problems. Finally, we provide open directions in Section 4.
1.1 Truncation Bias

Truncation bias occurs when samples falling outside a subset of the support of the population are not observed. The problem of analyzing data with truncation bias has myriad manifestations in economics, social sciences, and all areas of physical sciences, and dates back to famous statisticians like Pearson, Lee, and Fisher [Pearson 1902; Lee 1914; Fisher 1931]. Since then, it has been a central focus of many studies in econometrics [Maddala 1986], epidemiology see [Klein and Moeschberger 2003], and many other scientific fields. Some real world instantiations of truncation bias are the following:

**Negative Income Tax Experiment [Hausman and Wise 1977].** In this example, we have a dataset that consists of information about households with low wage rates. The goal of this study is to understand the effect of education level in the annual income of the households. In the dataset of [Hausman and Wise 1977], there was an artificial truncation in the data collection process. In particular, no data points were collected if the dependent variable (annual income) was below 1.5 times the poverty rate. The traditional method for this data analysis tasks is to use the **ordinary least squares (OLS)** regression. It is not hard to see though that in the presence of truncation the OLS outputs solutions that are biased and can lead to fallacious conclusions about the relationship between annual income and education level as was shown in [Hausman and Wise 1977].

**Schooling and Earnings of Low Achievers [Hansen et al. 1970].** In this example, we have samples of people that have been rejected from military, i.e. they scored low to the Armed Forces Qualification Test (AFQT), and we wish to estimate the following equation

\[ y = f(\text{education}, \text{age}, \ldots) \]

where \( y \) corresponds to the AFQT score. Again, the data available are truncated due to the fact that the collected data correspond only to people that received a low AFQT score. This truncation makes again the OLS estimator biased, and leads to fallacious conclusions.

**Hubble’s Law in Astronomy [Woodroffe 1985].** In this example, we have access to astronomical data and our goal is to estimate the relationship between the absolute luminosity \((M)\) of a star and its observed luminosity \((m)\) which also depends on another parameter called the redshift. One main issue with estimating this relationship between \(m\) and \(M\) is that we only have access to truncated data and in particular we may observe one star only if \(m \geq t\) where \(t\) is some threshold that depends only on the measurement device.

**Efficacy of COVID-19 Vaccine [Dagan et al. 2021].** In many biomedical and epidemiological studies truncation or censoring is a classical type of bias. In this particular study of the efficacy of a COVID-19 vaccine, truncation occurs in the control group of non-vaccinated people because after a while they receive the vaccine.

1.2 Self-selection Bias

Following the example of [Roy 1951], consider a village with two possible occupations: hunting and fishing. Everyone in the village chooses the occupation that maximizes their earning based solely on their own capabilities. Consider now a
statistician collecting observations of the earnings and occupations from this village with the goal of estimating a model that predicts the earnings that a particular person would make as a fisherman or as a hunter. It is not hard to see that applying naive statistical analysis techniques, which ignore the self-selection bias, would produce wrong results. Self-selection bias arises in different forms in many fundamental settings:

**Imitation Learning.** Consider the problem of learning an optimal policy in some contextual bandit setting wherein we observe the arms (e.g., treatments) pulled by an expert (e.g., doctor) in different contexts (e.g., patients). Modeling the reward (e.g., efficacy) from each arm \( j \) as an unknown function \( f_{w_j}(x, \epsilon_j) \) of the context \( x \) and additional randomness \( \epsilon_j \) that the expert might observe but we do not, we assume that the expert selects the arm \( j \) with the highest reward \( \max_j \{ f_{w_j}(x, \epsilon_j) \} \).

Our goal is to learn the underlying models \( w_1, \ldots, w_k \) of the arms by observing the expert make decisions in different contexts. This scenario is an instantiation of a statistical analysis task with self-selection bias since we do not get to observe the reward of all the arms \( f_{w_1}(x, \epsilon_1), \ldots, f_{w_k}(x, \epsilon_k) \), but only the reward of the arm that was selected from the expert, i.e., \( \max_j \{ f_{w_j}(x, \epsilon_j) \} \).

**Learning from Strategically Reported Data.** A widely studied setting featuring self-selected data is one wherein agents strategically chose which data to report. This is a standard challenge in econometrics, which has recently received increased attention in machine learning literature due to the impact of learning-mediated decisions in various contexts; see, e.g., [Hardt et al. 2016; Krishnaswamy et al. 2020; Liu and Garg 2021] and their references. A common example is the reporting of standardized test scores in college admissions, where applicants have a variety of standardized tests available to them, and are only required to report a chosen subset of them.

**Learning from Market Data.** Following [Fair and Jaffee 1972], consider a linear model of the a market, wherein there is a supply function \( S(x, \epsilon_S) = w^S_S x + \epsilon_S \) and a demand function \( D(x, \epsilon_D) = w^D_S x + \epsilon_D \), where \( x \) corresponds to a feature vector of the market, \( w_S \) and \( w_D \) are the coefficients that determine linear functions \( S \) and \( D \), and \( \epsilon_S, \epsilon_D \) correspond some random noise. If the market is in disequilibrium then supply does not equal demand, i.e., \( S(x, \epsilon_S) \neq D(x, \epsilon_D) \). So the quantity transacted is \( Q(x) = \min\{ S(x, \epsilon_S), D(x, \epsilon_D) \} \). If we want to estimate \( w_S, w_D \) from data of the form \( \{(x^{(t)}, Q(x^{(t)}))\} \), then this problem can be expressed as a problem of linear regression with self-selection bias [Cherapanamjeri et al. 2022b].

**Learning from Auction Data.** [Athey and Haile 2002] and a large body of literature in Econometrics consider the problem of learning bid (and valuation) distributions from auction data with partial observability, wherein only the winner of each auction and the price they paid are observed. Consider such observations in repeated first-price auctions. We can cast this problem as an instance of self-selection problem since we only get to see the maximum of the bids at ever iteration. A body of work in the literature has provided estimation and identification results in this setting [Athey and Haile 2007], including recent work of [Cherapanamjeri et al. 2022a] which demonstrates algorithms for estimating the bid distributions non-parametrically (see also Informal Theorem 3.2).

---

1These applications of self-selection bias are from the work of [Cherapanamjeri et al. 2022b]
2. SYSTEMATIC BIAS IN DATA

For simplicity of exposition we present our general framework for the density estimation problem only, but similar formulation can be obtained for regression and classification problems as well, as we discuss in Section 3.1.

Assume that there is an unknown distribution $P$ with support $S \subseteq \mathbb{R}^d$. The traditional density estimation problem can be formulated as follows.

**Definition 2.1 (Density Estimation).** Let $\varepsilon > 0$, $P$ a probability distribution that belongs to a family of probability distributions $D$ and has support $S \subseteq \mathbb{R}^d$. Given as input $n$ i.i.d. samples $x_1, \ldots, x_n$ drawn from $P$, our goal is to compute a distribution $Q$ such that $\text{dist}(P, Q) \leq \varepsilon$ with probability of failure at most 1%, where $\text{dist}(\cdot, \cdot)$ is a distance metric or a divergence between probability distributions, e.g., the total variation distance, or the Kolmogorov distance, or the KL-divergence.

If we specify the family of distributions $D$ and the distance metric $\text{dist}$ and there exist a smallest number $f(d, \varepsilon)$ such that for every $n > f(d, \varepsilon)$ the above problem is solvable, then we call $f(d, \varepsilon)$ the sample complexity of this density estimation problem. The running time of the fastest algorithm that takes as input $x_1, \ldots, x_n$ and outputs $Q$ is the time complexity of this problem.

**Remark.** A simplification that we make in the formulation above is that we ignore the dependence on the probability of failure, which for the purposes of this letter we assume is a constant, e.g., 1%. In virtually all the settings that we discuss, the probability of failure can be decreased to $\delta$ if we pay an additional $\log(1/\delta)$ factor in sample and time complexity.

**Example (DKW Inequality).** Assume that $d = 1$, $\text{dist}(A, B)$ is the Kolmogorov distance, i.e., the maximum difference of the cumulative distribution functions of $A$ and $B$, and $P$ is the set of all probability distributions over $\mathbb{R}$. In this setting the celebrated DKW inequality [Dvoretzky et al. 1956] provides a simple algorithm for solving this density estimation problem, with sample and time complexity $O(1/\varepsilon^2)$ which is known to be tight [Massart 1990].

Next, we define the density estimation problem with adversarial corruptions which is a very well studied problem in statistics and machine learning [Huber 2011; Diakonikolas and Kane 2019]. This problem provides some intuition for our formulation of systematic bias.

**Definition 2.2 (Density Estimation with Corruptions).** Let $\alpha > 0$, $P$ a probability distribution that belongs to a family of probability distributions $D$ and has support $S \subseteq \mathbb{R}^d$. We are given as input $n$ corrupted samples $z_1, \ldots, z_n$ such that $z_i = h_i(x)$ where $x_1, \ldots, x_n$ are i.i.d. samples drawn from $P$ and $h_i : S \rightarrow S$ are arbitrary unknown functions.\(^2\) For any meaningful estimation to be possible we require that at least $(1 - \alpha) \cdot n$ of the $h_i$’s equal to the identity, i.e., $h_i(x) = x$ for all $x \in S$. Our goal is to compute a distribution $Q$ such that $\text{dist}(P, Q) \leq g(\alpha)$ with probability of failure at most 1%. If we specify the family of distributions $D$, the distance metric $\text{dist}$, and $g$ and there exist a smallest number $f(d, \alpha)$ such that

\(^2\)Observe that since $h_i$ is arbitrary and unknown, the choice of $h_i$ may depend on the rest of the samples $x_1, \ldots, x_{i-1}, x_{i+1}, x_n$ as well.

ACM SIGecom Exchanges, Vol. 20, No. 1, July 2022, Pages 55–63
for every \( n > f(d, \alpha) \) the above problem is solvable then we call \( f(d, \alpha) \) the sample complexity of this density estimation problem. The running time of the fastest algorithm that takes as input \( x_1, \ldots, x_n \) and outputs \( Q \) is the time complexity of this problem.

**Example (DKW Inequality Continued).** It is easy to see that the the empirical cumulative distribution function that is used to prove the DKW inequality is robust to \( \alpha \) fraction of adversarial corruptions. This means that we can solve the density estimation problem with corruptions with \( g(\alpha) = \alpha \) and sample and time complexity \( O(1/\alpha^2) \).

There are a few things to observe about Definition 2.2.

1. The problem formulation allows for a very general class of corruptions since we have no restrictions or knowledge about a small fraction of the \( h_i \)'s.
2. Most importantly the fraction of corrupted data, i.e., data \( z_i \) for which \( h_i \) is not equal to the identity, determines the accuracy that we can achieve, i.e., we cannot hope to estimate \( P \) in distance \( \varepsilon \) unless \( \varepsilon \geq g(\alpha) \).

The second point above is what makes the adversarial corruptions framework not applicable in many settings where the data collection procedure introduces bias to all the data and not just a small fraction of them, as the examples from Sections 1.1 and 1.2 illustrate. This leads us to the formulation of density estimation with systematically biased data.

**Definition 2.3 (Density Estimation with Systematic Bias).** Let \( \varepsilon > 0 \), \( P \) a probability distribution that belongs to a family of probability distributions \( \mathcal{D} \) and has support \( S \subseteq \mathbb{R}^d \). We are given as input \( n \) systematically biased samples \( z_1, \ldots, z_n \) such that \( z_i = h(x_i) \) where \( x_1, \ldots, x_n \) are i.i.d. samples drawn from \( P \) and \( h : S \to T \cup \{ \bot \} \) is a function that is known to belong to a known family of functions \( \mathcal{H} \). Depending on the model we might observe or not observe any \( z_i \) with \( z_i = \bot \). In that case \( n \) is the total number of \( z_i \)'s observed. Our goal is to compute a distribution \( Q \) such that \( \text{dist}(P, Q) \leq \varepsilon \) with probability of failure at most 1%. If we specify the family of distributions \( \mathcal{D} \), a family of functions \( \mathcal{H} \), the distance metric \( \text{dist} \) and there exist a smallest number \( f(d, \varepsilon) \) such that for every \( n > f(d, \varepsilon) \) the above problem is solvable then we call \( f(d, \varepsilon) \) the sample complexity of this density estimation problem with systematic bias \( \mathcal{H} \). The running time of the fastest algorithm that takes as input \( x_1, \ldots, x_n \) and outputs \( Q \) is the time complexity of this problem.

The main differences of Definition 2.2 and Definition 2.3 are the following.

1. In the corruption framework we do not have knowledge about \( h_i \) other than that the fraction of corruption is limited, whereas in the systematic bias framework we know that the bias is the same for all the samples, this is why we call it systematic, and also we know the set \( \mathcal{H} \) that \( h \) belongs to.
2. On the other hand the important feature of the systematic bias is that in many settings, it allows for estimation up to arbitrarily small error \( \varepsilon \), assuming that we have enough samples, even though the bias function \( h \) applies to all the data.
We are now ready to show how truncation bias, censoring, and self-selection bias can be expressed in the framework of systematic bias.

**Truncation Bias – Censoring.** If we know the truncation/censoring then the set of functions \( \mathcal{H} \) contains only one function \( h \) where \( h(x) \) is equal to \( x \) if \( x \) is inside the survival set \( K \subseteq S \) and is equal to \( \perp \) otherwise. The difference between truncation and censoring is that for the former we do not observe the \( z_i = \perp \) points whereas for the latter we observe them as well.

**Self-selection Bias.** The density estimation instantiation of self-selection bias appears when \( x_i \) is a \( d \)-dimensional vector \( x_i = (x_{i1}, \ldots, x_{id}) \) and

\[
    h(x_i) = (\max_{j \in [d]} x_{ij}, \arg \max_{j \in [d]} x_{ij}).
\]

This corresponds to observing the highest bid and the identity of the highest bidder in a repeated first-price auction and trying to estimate the distribution of each individual agent.

Both of the above problems are impossible if we allow the family of distributions \( D \) unrestricted. As we will see in the next section, to get algorithms with small sample and time complexities we need to assume that \( D \) contains smooth distributions for the case of truncation bias or product measures for the case of self-selection bias.

### 3. RESULTS

Both truncation bias and self-selection bias introduce difficult estimation questions even for the fundamental case where \( D \) is the set of Gaussian distributions, but for simplicity and for consistency with the examples that we presented for classical density estimation and density estimation with corruption we will present our result in the case where \( D \) cannot be parameterized from a small set of parameters. For our results for the Gaussian case we refer to [Daskalakis et al. 2018].

First we need the definition of smooth probability distributions.

**Definition 3.1.** Let \( d = 1, S = [0, 1] \), then we say that the family of distributions \( D \) with support \( S \) is a smooth family of distributions if for every \( P \in D \) it holds that

1. \( P \) has a density,
2. the logarithm of the density of \( P \) is an infinitely differentiable function,
3. the \( i \)-th derivative of the log-density of \( P \) is upper bounded by \( M^i \), for some constant \( M \).

In order to get some intuition of the family of smooth distributions we observe that it contains all the distributions with log-density of the form \( f_1(x) + \cdots + f_k(x) \) for a finite number \( k \), where \( f_i \) can be any of the following:

- a polynomial \( \text{poly}(x) \) of constant degree,
- \( f_i(x) = \exp(\text{poly}(x)) \),
- \( f_i(x) = \sin(\text{poly}(x)) \).
An example of a class of distributions that are not smooth according to the above definition is the class of distributions with log-density equal to \( a \cdot \log(x) \).

We are now ready to present our results for density estimation from truncated data.

**Informal Theorem 3.1.** Assume that we observe \( n \) truncated samples \( z_1, \ldots, z_n \) from \( P \in \mathcal{D} \), where \( \mathcal{D} \) is a smooth family of distributions with support \([0, 1]\) and the truncation is with respect to a known survival set \( K \subseteq [0, 1] \). If the measure of \( K \) with respect to \( P \) is at least 1%, then there exists an algorithm to estimate \( P \) from truncated samples with error \( \varepsilon \) in total variation distance and with sample and time complexity \( \text{poly}(1/\varepsilon) \).

The surprising conclusion of this theorem is that if we know that the unknown distribution \( P \) is smooth, as per Definition 3.1, then we can identify \( P \) in its whole support \([0, 1]\), even though we observe samples only from a subset \( K \) of \([0, 1]\). In other words we can extrapolate and estimate \( P \) even in a region that is completely hidden to us due to the truncation. An intuition for why this is possible comes from Taylor’s theorem in calculus. Taylor’s theorem suggests that if a function \( f \) is sufficiently smooth then the knowledge of the values of all the derivatives of \( f \) at a point \( x_0 \) is sufficient to determine the values of \( f \) in the whole interval \([0, 1]\).

The technical contribution of our work is to provide a statistical version of Taylor’s theorem where instead of having access to the values of the derivatives, we have access only to truncated samples of the unknown distribution. For more details about Informal Theorem 3.1 and its multi-dimensional generalizations, we refer to [Daskalakis et al. 2021].

Next, we present our results for density estimation from data with self-selection bias.

**Informal Theorem 3.2.** Assume that we observe \( n \) samples \( z_1, \ldots, z_n \) with the self-selection bias described in the previous section from \( P \in \mathcal{D} \), where \( \mathcal{D} \) is the family of product measures over \([0, 1]^d\). Then there exists an algorithm to estimate \( P \) the samples with self-selection bias with error \( \varepsilon \) in Levy distance\(^3\) and has sample and time complexity \( O((1/\varepsilon)^d) \).

As we already explained the above theorem can be applied to learning from auction data which is a very fundamental problem in econometrics [Athey and Haile 2002]. For a detailed presentation of Informal Theorem 3.2 we refer to [Cherapanamjeri et al. 2022a]. As we show in [Cherapanamjeri et al. 2022a] the exponential dependence on \( d \) is necessary for this problem and can only be avoided if we restrict our attention to estimating the probability distribution only on a subset of its support. This can be formulated by changing the distance metric from the Levy distance to one that only measures the difference in a large subset of the support. In that case our sample and time complexity become \( \text{poly}(1/\varepsilon) \).

---

\(^3\)Levy distance is very similar to the Kolmogorov distance but allows for an \( \varepsilon \) error in the x-axis of the cumulative distribution functions.
3.1 Beyond Density Estimation

The main focus of this letter is on density estimation, but similar formulations can be provided in other data analysis tasks like regression and classification. In linear regression, for example, we have access to samples of the form \((x, y)\) where \(y = w^T x + \varepsilon\) and our goal is to estimate \(w\). Truncation or self-selection in this setting applies to the dependent set of variables \(y\) and we observe \((x, h(y))\) instead of \((x, y)\). In this setting the assumption on the family of probability distributions \(D\) applies to the distribution of the random noise \(\varepsilon\). Similar formulations can be done for classification problems as well. For a precise formulation and results on regression and classification problems with systematic bias we refer to [Daskalakis et al. 2019; Ilyas et al. 2020; Daskalakis et al. 2020; Daskalakis et al. 2021; Cherapanamjeri et al. 2022b].

4. OPEN PROBLEMS

As we mentioned in the previous section there are a lot of results for density estimation, regression, and classification problems under truncation or self-selection bias. An interesting direction that is still not well-explored is how truncation or self-selection bias affects the learning algorithms in online or dynamic environments. A first step in this direction has been taken in [Plevrakis 2021], but there are many interesting problems in this area that are still open: (1) Are there online learning or bandit algorithms that are robust to truncation or self-selection bias? For self-selection bias this is closely related with the problem of imitation learning; (2) Can we effectively control dynamical systems for which we cannot observe their state at every time step?

Beyond truncation and self-selection the natural question that arises is the following: Can we provide a characterization of the distribution families \(D\) and the functions \(H\) for which we can solve the density estimation problem with systematic bias?

REFERENCES


ACM SIGecom Exchanges, Vol. 20, No. 1, July 2022, Pages 55–63


Lee, A. 1914. Table of the gaussian “tail” functions; when the “tail” is larger than the body. *Biometrika*.


Fairness and Equity in Resource Allocation and Decision-Making: An Annotated Reading List

FAIDRA MONACHOU
Stanford University
and
ANA-ANDREEA STOICA
Columbia University

Fairness and equity considerations in the allocation of social goods and the development of algorithmic systems pose new challenges for decision-makers and interesting questions for the EC community. We overview a list of papers that point towards emerging directions in this research area.

Categories and Subject Descriptors: F.2 [Theory of Computation]: Analysis of Algorithms and Problem Complexity; J.4 [Computer Applications]: Social and Behavioral Sciences—Economics

General Terms: Algorithms, Machine Learning, Economics, Applied Modeling

Additional Key Words and Phrases: Fairness, Equity, Resource Allocation, Decision-making

Improving fairness and equity of decision-making systems used in the allocation of social goods is an important priority in a wide range of domains. For example, how can a government fairly and equitably allocate scarce medical resources to citizens? What does it mean for a machine learning algorithm used for loan allocation to be fair? How can an online matching platform for freelance workers ensure equal access to employment opportunity for all?

While many different disciplines have studied these problems, we argue that they lie at the very center of the economics and computation field, as algorithms constitute a central mechanism in many traditional decision-making systems. Whether algorithms are in the role of assisting in resource allocation at scale, diagnosing differences in access, or governing new environments in which resources are distributed (e.g. online platforms), fairness questions lie at the core of their design and implementation. Thus, these challenging questions pose a unique opportunity, not only to uncover untapped research insights utilizing tools from mechanism design, algorithms, machine learning, and optimization, but most importantly to collectively contribute potential research-grounded solutions to emerging societal problems.

In an attempt to highlight some of the most promising directions for future research, in this article we offer a list of papers that could serve as a useful starting point for the interested readers in the EC community. We acknowledge that the list below is far from exhaustive and fully representative, since due to space constraints, we omitted many important and closely related papers. Nevertheless, we hope it serves as a useful starting point for the interested readers in the EC community, highlighting some of the most promising directions for future research.

Authors’ addresses: monachou@stanford.edu,astoica@cs.columbia.edu

ACM SIGecom Exchanges, Vol. 20, No. 1, July 2022, Pages 64–66
In his seminal paper, Phelps, often credited together with K. Arrow, introduces statistical discrimination: the theory that, even in the absence of prejudice, discrimination can arise due to uncertainty about individuals’ true characteristics.

This classic paper introduces discrimination as coordination failure, i.e., when groups of ex ante identical agents choose different equilibrium strategies. Its simple equilibrium model is the basis of multiple subsequent works until today.

A staple of the fairness research literature, this paper presents a theoretical analysis exploring the trade-offs between three main statistical definitions of group fairness, showing that not all can co-exist when the underlying data is not completely unbiased.

This paper proposes an alternative definition of fairness to demographic parity that shows a better alignment between objectives and diversity considerations. Formulated as equality of opportunity, this definition bridges notions of equality and fairness and opens up avenues of research in designing mechanisms that equalize chances of obtaining resources across different groups.

This paper explores the connections between fair classification and social welfare maximization, pointing towards how “more fair” classifiers can worsen welfare outcomes for all social groups.

This paper proposes a generalized conceptual method for defining fairness through a causality criterion, generalizing from observational methods that define fairness. Through its practical distinction between protected attributes and their proxies, this paper opens up avenues of interdisciplinary research for establishing interventions that remove discrimination through causal pathways.

Following upon the concept of price of fairness introduced by the same authors, this paper considers fairness-efficiency trade-offs that a decision-maker faces in the allocation of scarce resources under the α-fairness scheme.

This paper considers the problem of allocating goods to sequential arriving agents with varying levels of need in an efficient and equitable manner, contributing to less explored areas such as dynamic fairness.


This paper presents a series of limitations to observational definitions of fairness through the lens of social welfare and theories of justice, and shifts the focus towards hidden concepts of within-group heterogeneity and merit-based inequity.


As EC researchers are puzzled with contradicting notions of fairness and the legal limitations of their proposed technical solutions, this essay offers a law perspective to how data-driven algorithmic techniques can lead to disparities and highlights open questions in the intersection of law and computation.
Economics and Computation Meets Cognitive Biases: A (Biased) Annotated Reading List

SIGAL OREN
Ben-Gurion University of the Negev

This is an annotated reading list on papers in the intersection of economics and computation and behavioral economics.

General Terms: Economics, Theory
Additional Key Words and Phrases: Behavioral Economics, Mechanism Design, Present Bias

A recently growing line of works aims to bridge the gap between Economics and Computation and Behavioral Economics. Works in this space often come in one of these two flavors: (1) using tools from theoretical computer science to devise and study general models of cognitive biases and (2) considering players with cognitive biases in well-studied settings from Algorithmic Game Theory. This reading list is composed of some typical papers for each of these flavors as well as related papers from Behavioral Economics. The list is intended to serve as a starting point for researchers interested in this line of work and is not comprehensive.

First, we consider works about planning for the future when the agents exhibit some planning-related biases. Such biases include present bias, sunk-cost bias and projection bias. These biases are well studied in the behavioral economics literature in experimental and empirical settings. However, the theoretical models suggested to capture them are often quite specific to the setting. This line of work demonstrates that by harnessing tools from theoretical computer science and graph theory, we could obtain a much more general model and use it to ask and answer questions otherwise impossible. Our focus on this list will be on present bias.

The second type of works we list takes settings well-studied in Algorithmic Game Theory and aims to make them more applicable to real-life by considering some cognitive biases that the players exhibit. While planning-related biases, we previously discussed, have negative implications, here, there are settings in which taking the biases into account may have positive implications. In this reading list, we focus on works of this type situated in the field of algorithmic mechanism design.

1. [O'Donoghue and Rabin 1999] - Individuals exhibiting present bias focus more on the present than on the future. According to the hyperbolic discounting model such agents will multiply the cost (or utility) they get now by a factor of $\beta$ and the the utility at any future step $t$ by a factor of $\beta^t$. This classic paper in behavioral economics studies simple planning settings (e.g., “which day of the week should I complete the review?”) for agents exhibiting present bias.

2. [Kleinberg and Oren 2014] - This paper suggests a general graph-theoretic
model for capturing the planning behavior of agents that exhibit present bias. The generality of the model enables it to capture different scenarios that previously each required a different specific model and compare them. For example, it allows a characterization of the types of scenarios in which the loss of the agent due to its bias is large.

(3) [Gravin et al. 2016] - While most papers assume that the present bias parameter of the agent is fixed, the current paper assumes that the parameter is sampled each step independently from a given distribution. The paper identifies graphs in which the loss of the agent due to its bias is bounded and constructs for each bias-distribution a graph maximizing the agent’s loss. The latter is done by identifying surprising connections to optimal pricing theory.

(4) [Albers and Kraft 2021] A high-level question in the setting of [Kleinberg and Oren 2014] is how could we make changes in the task graph in order to help the agent mitigate the effects of its bias. Previous papers on this question consider setting deadlines; this corresponds to computing a subgraph where the agent will reach the target. The current paper takes a different approach. It sets penalties on some tasks to deter the agent from completing them and proves that this approach is quite effective.

(5) [Strack and Taubinsky 2021] The last paper we mention on this topic is a bit of an outlier that offers a different perspective on planning behavior. Present bias (or as it is often termed these days *present focus*) often leads to time-inconsistent behavior in which agents keep changing their plan. The paper suggests an alternative explanation, which essentially says that we cannot distinguish between an agent that behaves inconsistently and an agent that behaves consistently but has some uncertainty regarding the future.

(6) [Kahneman et al. 1991] - This paper describes a classic experiment demonstrating and estimating the endowment effect (i.e., an individual values items more once he owns them.) The rough idea is to give half of the subjects some object (e.g., a coffee mug) and ask them how much they are willing to sell it, and compare it with the price that the other half of the subjects are willing to pay for it.

(7) [Babaioff et al. 2018] - In a combinatorial auction, Walrasian equilibrium is only guaranteed to exist if the players’ valuations are gross substitutes. The paper suggests a model of endowment effect for combinatorial auctions and shows that if we assume that the players have a (mild) endowment effect, we can considerably extend the class of valuations for which a Walrasian equilibrium exists to sub-modular valuations.

(8) [Ezra et al. 2020] - This paper considers different ways to model the endowment effect in combinatorial auctions and studies the implications of the different modeling assumptions on the class of valuations in which a Walrasian equilibrium exists. The paper highlights the need for more experimental work on the endowment effect in combinatorial auctions to understand which assumptions are more plausible.

(9) [Gneezy 2005] - The paper belongs to a line of work in behavioral economics, arguing that people facing an opportunity to increase their payoff by lying do
not always choose to lie. The paper presents experiments demonstrating that people tend to lie less when their benefit from the lie is smaller with respect to the loss of another person, even a stranger, from the lie.

(10) [Dobzinski and Oren 2022] - The paper builds on literature on lying behavior in behavioral economics to challenge a fundamental paradigm of mechanism design: bidders that could lie to increase their payoff will always do so. Based on behavioral assumptions extracted from this literature, the paper studies an auction model in which to determine whether to report their valuations truthfully, the bidders compare their gain from lying against the loss of the others. The paper asks whether an auctioneer can take advantage of bidders behaving this way to increase its revenue.

REFERENCES


Here we provide an overview of an important issue in online field experiments: spillover effects. We include a reading list for researchers in both academia and industry who are interested in this topic.

General Terms: Electricity, Commerce, Economics, Agents, Meta, Stuff
Additional Key Words and Phrases: Templates, Skeletons, Things

Field experiments typically aim to quantify how a given intervention (e.g., a new policy) affects certain outcomes in a population. With the growing popularity of online communities and marketplaces, there has been a corresponding increase in online field experiments. However, many online and offline experiments are subject to interference spillover effects. Spillover effects can take various forms: in some experiments, a user’s outcome may be affected by the treatment assignments of other subjects; in within-subject experiments, the treatment assignment that the subject receives at one stage may affect the outcome at a later stage (also referred to as carryover effects).

With the presence of spillover effects, the conventional way of randomizing samples may be problematic. Intuitively, spillover effects mean that the outcome of an observation is affected not only by their own treatment assignment but also by the treatment assignments of other observations. The existence of spillover effects violates the stable unit treatment value assumption (SUTVA), the standard assumption in causal inference. For example, the existence of social contagion provides empirical evidence of the spillover effect: if one person is assigned a treatment, their family, friends, or acquaintances may also indirectly receive this treatment.

The following papers discuss how to detect or take into account spillover effects when we design experiments or analyze experimental data:


This paper presents an overview of the design and analysis of online field experiments. It covers representative studies from both economics and computer science.

Using a randomized experiment, this study presents evidence of social influence in social networks, suggesting that the existence of spillover effects may challenge the SUTVA when we analyze experiments on social networks.


This paper designs an approach to tackle with spillover effects on large-scale social networks by creating clusters on social networks and performing randomization on the cluster level. This approach helps reduce bias when we estimate the treatment effect with the presence of spillover effects.


This paper proposes a novel approach for detecting spillover effects in social networks by simultaneously running individual (Bernoulli) and cluster level randomized experiments and comparing the resulting estimates.


This paper aims to address the interference on two-sided markets. The authors also compare cluster randomization versus Bernoulli randomization and find that the latter greatly reduces bias in estimating treatment effects.


This paper discusses the optimal design and analysis of switchback experiments. The authors also discuss how to account for carryover effects: the previous treatment assignment to one unit may affect the unit’s future outcome.

By relaxing SUTVA, this paper proposes a general framework, named “exposure mapping”, to analyze experimental data with the presence of spillover effects.


This paper proposes regression adjustment estimators to reduce bias in experiment settings that violate SUTVA. The paper also proposes to consider the modeling of spillover effects as a feature engineering problem, which can further increase the precision of estimating treatment effects.


This paper discusses practices, challenges, and pitfalls in evaluating results of online controlled experiments from an industry perspective (LinkedIn).

Please note that this list is far from exhaustive and there are other related and exciting papers that are not included.
Puzzle: Communicating to Plan Noam Nisan’s 60th Birthday Workshop

VINCENT CONITZER
Duke University and University of Oxford

Please send solutions to the author by e-mail, with the title of this puzzle in the subject header. The best solution will be published in the next issue of SIGecom Exchanges, provided that this solution is of sufficiently high quality. Quality is judged by the author, taking into account at least soundness, completeness, and clarity of exposition. (Incidentally, there is another birthday puzzle for which we still need a solution [1]!) This is a puzzle in honor of Noam Nisan’s 60th birthday and the June 2022 workshop associated with it (and perhaps also a bit in honor of one of its distinguished attendees, Hervé Moulin). This workshop was held at the Hebrew University of Jerusalem in Israel.

Michael, Moshe, and Shahar—i.e., a constant number of organizers—are planning the workshop for Noam’s 60th birthday, and are trying to predict who, out of $n$ people, will attend. Whether a person wants to attend is a function of who else attends. “The more the merrier,” so for each person $i$, if $i$ would attend when $S$ is the set of other attendees, and $S \subseteq S'$, then $i$ would attend when $S'$ is the set of other attendees. Let $S_i$ be the set of sets $S$ of other people for which $i$ would attend (so, $S_i$ is upward closed).

To split the work, the organizers partition the set of $n$ people among themselves. Subsequently, each of them figures out, for every player $i$ in his own part, what $S_i$ is. (Note that each organizer thus still needs to think about how much “his” people like the people in the other parts. But each organizer knows $S_i$ only for people $i$ in his own part.) At this point, the organizers, who of course want the workshop to be successful, must communicate with each other to find the largest possible set of people $S^*$ that can consistently attend (i.e., the largest set with the property such that every person in it will attend given that everyone else in the set attends; i.e., for each $i \in S^*$, we have $S^* \setminus \{i\} \in S_i$, and $S^*$ is the largest set with that property).

Up to a constant factor, how many bits of communication do the organizers need to figure this out?

REFERENCES

[1] I thank Shahar Dobzinski for helpful feedback.

Authors’ addresses: conitzer@cs.cmu.edu