SIGecom Winter Meeting 2024 Highlights

AGHAHEYBAT MAMMADOV
Penn State University
and
EMILY RYU
Cornell University
and
ROBERTO SAITTO
Stanford University

Aghaheybat Mammadov *is a PhD candidate in Economics at Penn State University, advised by Nima Haghpanah and Hadi Hosseini. Within the Economics Department and the AI Lab (College of IST), he studies problems at the intersection of Economics and AI, focusing on resource allocation and matching markets through analytic, algorithmic, and quantitative perspectives. He holds a BSc in Economics from ADA University and an MA in Economics from Bilkent University.*

Emily Ryu *is a third year PhD student in Computer Science at Cornell University, advised by Éva Tardos and Jon Kleinberg. Her research interests span algorithmic game theory, combinatorial optimization, and networks, particularly with more realistic models of behavioral and cognitive constraints. Before Cornell, she graduated from Princeton University with a B.A. in Chemistry and minors in applied math and computer science.*

Roberto Saitto *is a fourth-year PhD student in Economics at Stanford University, advised by Paul Milgrom and Matt Jackson. His research interests are economic theory and market design, with a particular focus on simplicity in mechanism design.*

The fourth annual ACM SIGecom Winter Meeting took place on February 15, 2024. Organized by Sigal Oren and Ran Shorrer, this year's meeting brought together researchers from economics, computer science, and adjacent fields to focus behavioral models. The virtual meeting took the form of a workshop including talks and presentations from leading experts on various directions, as well as a fireside chat on getting into the research space and exciting opportunities that lie ahead. We share some highlights from the 2024 Winter Meeting, and additional insights from follow-up interviews with the speakers.

1. INTRODUCTORY TALKS

1.1 Jon Kleinberg - Behavioral Agents in Algorithmic Environments

Consider the everyday scenarios where algorithms play a pivotal role: from recommending movies or social media content based on a user's digital footprint, to aiding hiring committees in job or school applications. In both the online and offline realms, individuals are mapped into vectors of features, which are then processed by decision-makers–algorithms, humans, or both–to predict outcomes. Moreover, algorithms can step beyond mere tools to become partners. This dynamic is evident in generative AI assistants, semi-autonomous vehicles, and even medical diagnostic tools. **Jon Kleinberg** (Cornell University) posed a crucial question in the first introductory talk of the 2024 SIGecom Winter Meeting: To what extent do algorithms need to incorporate behavioral models to truly enhance the welfare of users across these diverse environments? How can we ensure that these intelligent systems do more good than harm? Jon's exploration into the intersection of computational economics and human behavior makes us to rethink the role of algorithms in shaping our decisions and futures.

One model of agents with behavioral biases addresses how agents make decisions over time with a *present bias*, such as procrastination or task abandonment [Kleinberg et al. 2016]. Agents navigate a graph from a starting node s to a terminal node t, aiming to maximize rewards while minimizing costs. However, due to present bias, costs of immediate actions are perceived as higher, scaled by a factor $b > 1$. This bias often leads to agents abandoning their intended paths. To mitigate this, algorithms can restrict certain choices, effectively narrowing down the decision paths and helping agents commit to more optimal routes. This approach illustrates how strategic limitations in choice architecture can assist agents in overcoming inherent biases for better decision-making outcomes.

Another model examines how behavioral biases affect decisions in the consumption of social media feeds [Kleinberg et al. 2023]. A user consumes a content feed of items, and the main decision is whether to continue to the next item. A user has two sets of preferences which are referred to as System 1 and System 2. System 1 is impulsive and acts first, while System 2 uses long-range planning after System 1 chooses not to act. This model questions the reliability of session length as a metric for user welfare. Longer engagement might suggest higher utility if System 1 dominates, but this is not always the true story if System 2 is taken into account. Variability in user's impulsivity can distort this metric, as long sessions could be driven by either genuine interest or mere impulsiveness. This model underscores the need for careful interpretation of user data and design choices in digital platforms to truly enhance user experience and welfare.

The rest of the talk was on an environment where algorithms operate as *partners*. Jon discussed the design of algorithms that remain within the comprehension limits of human partners [McIlroy-Young et al. 2020]. He highlighted "Maia", a chess engine trained on human games, designed to mimic human moves at specific player rankings, known as Elo ratings, offering a seamless transition if a human needs to take over. This contrasts with more sophisticated engines like Leela and Stockfish, which can confound human partners with complex moves. In a unique chess experiment featuring teams composed of both humans and engines with no inter-team

communication, teams often faltered when humans couldn't understand or follow the engines' advanced strategies. However, a modified version of Leela named "PartnerBot" was designed to select moves that were intelligible to both human players and Maia, enhancing team performance. This experiment underscores the importance of transparency and intelligibility in collaborative environments where algorithms and humans interact, ensuring smoother hand-offs and more effective teamwork.

1.2 Ori Heffetz - EBRD and Deferred Acceptance

In his intriguing talk, **Ori Heffetz** (Cornell University) shed light on a puzzling phenomenon: Why do individuals misreport their preferences even when participating in strategy-proof mechanisms such as Deferred Acceptance (DA)? Despite the theoretical assurance that participants in DA—such as students in school admissions and residents in hospital-resident matching—have no incentive to manipulate their preferences, empirical evidence suggests otherwise. Participants appear to make choices that are counter-intuitive or dominated. Ori proposed that incorporating *expectations-based reference-dependent (EBRD) preferences* might be the key to understanding and rectifying this inconsistency in behavior [Dreyfuss et al. 2022].

The literature offers various explanations for this phenomenon, ranging from a lack of understanding that DA is indeed strategy-proof [Li 2017] to misconceptions about the mechanism itself [Gonczarowski et al. 2023]. Another proposal is that altruism or a preference for higher-ranked choices might drive these behaviors [Meisner 2023]. Ori argued that applicants might be influenced not just by the schools they attend but also by expectations-based reference points, known as EBRD preferences. This model incorporates not only traditional utility from consumption but also the utility derived from "news"—the variance between new and old beliefs about consumption. The framework effectively explains why applicants might distort or omit their true preferences, often due to the fear of potential disappointment.

This theory is supported by Ori's incentivized experiment, in which he analyzed four configurations of matching environments. These configurations varied depending on whether students were on the proposing or receiving side, and whether they decided which schools to apply to statically at the beginning of the mechanism or dynamically during its operation. Ori showed that in all of these matching variants, except when students receive offers from schools dynamically (where the "news" effect is absent), the EBRD model comprehensively explained the non-intuitive behaviors or misrepresentations by applicants, starkly contrasting with traditional preference theories. This insight challenges us to rethink how we design and interpret the results of matching mechanisms, highlighting the complex interplay between expectations and decision-making.

2. FIRESIDE CHAT WITH NOAM NISAN AND ALVIN ROTH

The 2024 Winter Meeting featured a dynamic Q&A session, the Fireside Chat. In this event, two renowned professors, **Noam Nisan** (Hebrew University), a distinguished computer scientist specializing in computational complexity theory and algorithmic game theory, and **Alvin Roth** (Stanford University), a prominent economist and a Nobel Prize winner especially known for his work in market design, engaged with an audience of graduate students and established scholars. Noam and Alvin offered valuable insights into their PhD journeys and academic careers. They explored the challenges they faced, the learning opportunities they encountered, and how they have grown within the academic community.

Could you describe what a typical week looked like for you when you were a PhD student?

Noam. I participated in an extremely rigorous PhD program at Stanford University. The experience was highly demanding for each week, providing a deep dive into advanced research methodologies and critical thinking, which significantly honed my analytical and academic skills.

Alvin. I met with my advisor every week. The meeting lasted an hour, with the initial half dedicated to discussing the reasons behind the limited progress on my projects. The latter half focused on my advisor recommending specific papers for review, emphasizing the importance of understanding their content to aid in my research endeavors. This structure balanced addressing challenges while also providing constructive guidance for moving forward academically.

What are your thoughts on the advice that selecting a good advisor is more important than choosing a specific research topic for a student?

Noam and Alvin. Sharing a common interest with your advisor holds significant importance. Initially, one might not have a clearly defined research question, but through ongoing discussions and shared interests, a focused and refined question can emerge. This collaborative approach facilitates a deeper exploration of topics and fosters a productive mentorship dynamic, guiding the research process towards meaningful outcomes.

How often do you talk to your students?

Alvin. I schedule two coffee meetings each week in the morning, valuing these as a foundation for intellectual development. This approach is based on the belief that engaging conversations are a crucial source of intellectual insights, fostering an environment where innovative ideas and solutions can emerge from collaborative discussions.

How do you approach learning and developing your expertise to effectively contribute to solving problems?

Noam. I approach my work in two distinct methods. Initially, I dive deeply into the theoretical aspects of mechanism design, starting with foundational textbooks such as the Mas-Colell book to grasp the underlying concepts thoroughly. This

"original mode" focuses on building a strong theoretical foundation. Conversely, the second approach emphasizes practical application: setting aside theory to experiment with solutions directly in real-world scenarios. This mode prioritizes action and experimentation to see how concepts perform outside of theoretical frameworks, fostering a balance between theoretical understanding and practical applicability.

Alvin. At all times, delving into books and consulting with field experts, such as doctors for the kidney exchange problem, is essential. This approach of dynamic engagement and exchanging ideas with specialists enhances the depth of understanding and brings new perspectives to the forefront. It underscores the importance of interdisciplinary communication in refining research questions.

As a final insight, could you provide an elaboration on a real-life problem in the field of mechanism design?

Noam. During my tenure with Google Ads, I observed that despite advertisers having budgets, they often opted for simple models over complex ones. This realization highlighted a disconnect between the sophisticated models developed in theory and the practical needs of advertisers. The simplicity and applicability of models, rather than their theoretical sophistication, were more aligned with the advertisers' objectives, showing the importance of developing solutions that are both effective and user-friendly for real-world applications.

Alvin. Over time, the dynamics within the field of kidney exchange have evolved significantly. Initially, the focus was on individual practitioners and patients. However, as the system has grown and developed, the involvement of hospitals and kidney exchange program directors has introduced strategic considerations into the process. This shift reflects a broader and more complex approach to organizing kidney exchanges, highlighting the need for strategic planning and coordination at higher organizational levels.

2.1 Further Q&A with Noam Nisan

After the meeting, Noam was kind enough to share some additional thoughts ranging from his own research interests to general directions within the EconCS subcommunity, and other advice for researchers in the space.



What can machine learning and human behavior tell us about each other? And what other sorts of cross-pollination (e.g. human learning? machine behavior?) should we be thinking about?

For a long time, we've thought that machine learning can teach us about human learning—that by uncovering or recreating the cleverness of human reasoning by developing algorithms and mechanisms, we can understand human intelligence more. I agree with this, but with a twist: I think machine learning helps us understand how much *less* spectacularly we should think of ourselves.

Sometimes when people work with ChatGPT and get annoyed by foolish errors, what really strikes me is that these errors are exactly the same kinds of things that

humans do. One of my favorite examples: if you ask ChatGPT to solve a simple math problem that it's not smart enough to do, it'll give you an answer where every line looks like it's trying to develop something sort of reasonable, but if you really look at the whole thing it's complete nonsense. And it always reminds me exactly of the exams of students that don't know the material, but you can see that they're trying their best to write things that sort of seem correct to get partial credit. So I think that to a very large extent, one of the things we can learn from ML is how humans behave in ways that are *not* so clever.

We humans think of ourselves as being very clever, say by playing chess, but now that we have machines that can play chess perhaps we see that chess does not represent the highest level of intelligence. Maybe object recognition is a higher level of intelligence–and once we have machines that can do that, we'll move onto solving math problems, or cleaning a room, or some other task entirely. But every time we get a machine that can do something new, one way to look at it is as demystifying what humans do, and maybe thinking of ourselves a little less highly.

So would you say that there's this overall perception that humans are somehow "better" than AI and machines, but you're trying to challenge this notion?

First of all, if you're not assuming some sort of God or dualism or other principle, humans are also biological, psychological machines, right? So the question is: will we be able, in the near future, to build machines that are better than us in various tasks? Nobody can know for sure, but I don't see any reason why not, sooner or later. And I think that's fine—maybe at a certain point, we'll just have to accept that our worth as human beings is not because of our intelligence; maybe it's our morality, or decision-making, or something else, but not just pure intelligence.

So that's a philosophical point of view. But practically speaking, yes, I do believe that computers will eventually outperform humans. Right now everyone's talking about LLMs; maybe next year we'll all be using something else entirely. I don't see any major reason to believe that in 5, or 10, or 20 years, machines won't be better than humans at proving mathematical theorems–like now, there's no human that can outrun a car, and so be it. Maybe our generation was lucky that we could find gainful improvement by proving theorems, and that may not be the case 30 years from now.

That's definitely a fascinating perspective on the future of AI. To take it back to the (recent) past, you've also done some work on blockchains and cryptocurrencies. Do you view this line of work as entirely separate from your work on behavioral models, or is there some insight that might be gained at the intersection of the two?

It's still disjoint, but related in a sense. Blockchains are a technical means to the goal of finding a mechanism that works without a centralized authority, and there are various computer science, economic, and engineering challenges in building them, but their overarching reason—what's really interesting about them—is the social visions that they allow. This could be similar to the introduction of the Limited Liability Company a few hundred years ago, which represented a different

form for humans to organize themselves in a way that turned out to be extremely socially important. Maybe blockchains will bring about another type of revolution into self-organizing, self-operating, complex networks. From this point of view, I'm interested in the social aspects of these networks, and of course social aspects are related to human behavior.

But I don't think we've reached a level of understanding of human aspects of behavior to then understand the social implications yet. Right now, we're focused on the technical questions of how to build blockchains, and are basically operating off intuitive assumptions of how people behave. We don't have nice theory or empirical results yet. But I believe this is definitely a direction that science should advance towards–at the end of the day, social structures depend on the behavior of the humans inside.

What's one problem within this space that you would love to see more researchers from the EC subcommunity work on?

I would be interested in people trying to develop the theory of tokenomics: an economic theory of these types of blockchain networks, and how they should behave. A lot of research has been done on the details of how all these blockchains differ in their abstractions and implementations, how you can maybe cheat a bit even with a small minority of faulty nodes, how you can get more computational efficiency, and so on. But I'm more interested in supposing that we've reached a situation where these decentralized networks work—now, what do they do? How do we understand their dynamics? There's economic and design questions like token pricing, voting, fees, and other system constraints. In general, these are complex systems, and I would like to see more work that tries to formalize and capture and understand the high-level aspects of the operation of these systems, under the abstraction that they work as intended.

And what would you say to the blockchain skeptics out there—any interesting non-blockchain behavioral questions you have in mind?

First, they're very right to be skeptical. I think it's a very big question whether blockchains will succeed—I would not be at all surprised if everything completely fails tomorrow, nor would I be surprised if in 10 years from now blockchains run half the world. So I really don't know what will happen—there are many possibilities that we could see.

With regard to general behavioral questions, this is an area that I believe is very important, but it's challenging for me to highlight specific questions. My tendency is generally to try to simplify things until I get a question that I can handle, but with humans, sometimes you can oversimplify and lose essence. This makes behavioral questions very tricky, and I have to admit, less aligned with my research approach. I hope that people who have this touch of combining psychology, computer science, and economics do more work in this area, but I really am not able to suggest concrete directions yet.

In general, how do you find balance between trying to gain expertise in a single focused area versus staying open to exploring a broad range of new areas (like behavioral modeling)?

That's a good question, especially as computer science is rapidly growing and changing. Back when I was a PhD student in the 80's, you could basically know all of theoretical computer science—I would say a senior grad student could feasibly go to STOC or FOCS and understand maybe 80% of the papers there.

That's not the case today anymore. It was certainly a fun time to grow up scientifically–I tried to go deep in one area and wide in general, because that's the standard recommendation. But I'm afraid that today it's difficult to be wide in general–so how wide do you go?

It's difficult to say, but you do need to constrain yourself to *something*, after all. I think you should try to find topics that will be important for the future on one hand, and on the other hand where we can make some meaningful progress today. Think about building a time machine: perhaps one of the most important questions for the future, but we currently have no idea how to do anything useful towards this direction. Easier said than done of course, but I think every scientist, every generation, faces this question.

So what's next for you and your research agenda?

I really like the blockchain world, and I'm trying to understand more stuff there. The nice thing about being a tenured professor is that understanding doesn't necessarily have to be the same thing as writing papers at this point—for me, I really like going to talks to get the basic ideas, and then if I'm interested in completing the picture, I try to figure things out myself by reading the details of the paper.

Another thing I'm interested in is how machine learning relates to algorithmic game theory, like in marketplace dynamics with an ML component—how does regret minimization work in markets? how do autobidders operate?–and other such questions combining AGT and markets with learning.

Finally, what's one interesting, non-CS fact about you?

Well, of course one never wants to share the *most* interesting answers to this question... but I suppose I could let some people here know that I have been doing Israeli folk dance for over 40 years!

3. SPOTLIGHT TALKS AND INTERVIEWS

3.1 Modibo Camara - Computationally Tractable Choice



In the first spotlight talk, **Modibo Camara** (Stanford GSB) presented a paper that studies choice when agents face computational constraints [Camara 2022]. Agents have only a limited amount of time to make their choices, and making good decisions takes time. In particular, always making the optimal choice may be too time consuming. For these reasons, agents often rely on behavioral heuristics to make their choices. An important example is choice bracketing—partitioning choices, and then optimizing within each element of the partition.

Are such behavioral heuristics a necessary response to computational constraints? Modibo gives a two-folded affirmative answer to this question, using concepts from theoretical computer science to define tractability. A choice rule is *tractable* if it can be implemented in polynomial time by some algorithm. First, if a choice rule that is consistent with the expected utility axioms is tractable, then it is observationally equivalent to some form of choice bracketing. Second, an expected utility maximizer who is constrained to use tractable choice rules can be better off by using heuristics that would not appear rational to an outside observer.

More precisely, consider an agent who has to specify a choice rule for a set of menus, which satisfies some richness conditions. Each menu consists of lotteries over eventually-zero sequences of rational numbers. A product menu is the Cartesian product of its *partial menus*, defined as the (marginal) lotteries that the agent may choose for each entry of the sequence.

Within this framework, rational and tractable choice rules that satisfy a symmetry condition can be rationalized by additively separable Bernoulli utility functions. This implies that, for product menus, such choice rules are observationally equivalent to narrow choice bracketing—that is, optimizing over partial menus separately. More generally, a choice rule is rational and tractable if and only if it is observationally equivalent to *dynamic choice bracketing*, which allows the relevant brackets to change in the process of making the choice.

For the final result, consider an expected utility maximizing agent who can only choose tractable choice rules. Surprisingly, her Bernoulli utility function may be such that the following is true: There is no rationalizable tractable choice rule that guarantees the agent at least one-half of a menu's optimal value, while there are non-rationalizable tractable choice rules which do. This result has a particularly interesting takeaway: Once we accept the idea that real-world agents have limited computational capacity, behavioral heuristics that may appear irrational from the perspective of an outside observer can actually be more effective than fully rationalizable choice rules.

3.1.1 *Interview with Modibo Camara.*

In an interview after the Winter meeting, Modibo Camara talked about the process that led to his paper and his perspective on future research in the field.

How did you get interested in the theory of simplicity, and more specif-

ically how did you come up with the idea for your paper?

When I started my Ph.D., my initial impression was that economic models often assume that people can do things that are too complicated. As a consequence, some of the recommendations of these models felt unrealistic and not very applicable. This made me think that taking complexity seriously may be of great importance for economic theory.

The research process that eventually led to my job market paper started during my first year. I was not fully satisfied with the axiomatic approach to decision theory: I believed that, if complexity were to be taken into account, then an agent may be better off by choosing simpler behaviors, that—despite violating the consistency requirements of expected utility theory—are approximately optimal. In particular, I had a feeling that when considering large numbers of decisions, a complexity constrained agent had to rely on some sort of narrow bracketing. I formalized this idea with my paper.

What other topic are you working on that you find the most exciting?

Related to bounded rationality and computer science, one topic that a collaborator and I have been thinking about recently is Artificial Intelligence (AI) alignment. One way to think about AI is as a tool to overcome humans' bounded rationality. From that perspective, if humans did not find computation taxing, there would be no need for AI. But there is still a computational bottleneck. In order to delegate a decision to the AI, the human needs to communicate her preferences to the AI, either directly or indirectly (e.g., through past choices). Studying the complexity of such communication seems important for identifying settings in which AI can be effective, as well as settings in which AI may never be fully aligned with human preferences.

More broadly, what topics at the intersection of computer science and bounded rationality in economic theory do you think are most promising for future research?

One interesting topic for future research is recommender systems. These are algorithms used by many online platforms, and their purpose is to resolve market failures that arise because users are boundedly rational. Take Netflix, for example. One of the main services that Netflix provides is that it uses behavioral data to filter a massive catalog of movies and television series into personalized recommendations for its users. Without this algorithm, users that struggle to process Netflix's massive catalog would likely end up making significantly worse choices. I think recommender systems represent a promising application area for work on bounded rationality, since they drive a huge amount of consumption online and are fundamentally a response to users' cognitive limitations. Optimistically, the EconCS community should be able to explain why these systems exist and provide guidance on how to design them.

3.2 Ryan Oprea - What Makes a Rule Complex

In the second spotlight talk, **Ryan Oprea** (University of California, Santa Barbara) presented experimental work which investigates what makes rules complex for human beings [Oprea 2020].

The paper aims to make a first step into ranking decision tasks and predicting which sub-optimal but less costly algorithms, such as heuristics, will be favored by humans. To this end, it considers the following experimental setting to measure subjective cognitive costs of performing algorithms directly and understanding what makes a rule complex for humans. Agents are asked to implement some rule—a sequence of choices as a function of a sequence of events. For example, an agent may be asked: Pick x until you see a, then switch to y. They get paid if and only if the task is performed correctly. Then, the experimenters elicit their willingness to pay to avoid to be asked to implement it again in the future. Which rules are more complex? Why?

In order to answer these questions, rules need to be systematically comparable. To this end, the paper represents rules as automata. That is, each rule is decomposed into a set of states and transitions: Each state is identified with an action to take, and is graphically represented as a circle. Each transition is identified with an event that triggers a shift from a state to another, and is graphically represented as a connecting arrow.

This approach reveals several recurring patterns of complexity. While agents make few mistakes, they have high willingness to pay to avoid to implement the proposed rules. Moreover, agents' willingness to pay varies significantly with the number of states of a given rule. Consistently with intuition, the data show that costs are increasing in the number of states and transitions. Interestingly, agents do not efficiently represent algorithms to themselves: In particular, they fail to see that some states and transitions may be redundant.

The above results notwithstanding, finite automata may still miss important aspects of complexity. Indeed, more sophisticated automata (known as pushdown automata) can better fit the data, and even richer languages—such as Turing machines—may be needed to fully understand the costs of algorithms.

In closing, Ryan stressed how we only just started to understand how to design simpler social rules that can work better in practice. In particular, this work is a call for future research to build models of bounded rationality rooted on empirically grounded characterizations of complexity.

3.3 Gali Noti - Learning When to Advise Human Decision Makers

In the third spotlight talk, **Gali Noti** (Cornell University) presented joint work with Yiling Chen on learning when to advise human decision makers [Noti and Chen 2023]. AI-assisted human decision-making, in which a machine learning algorithm is intended to *help* rather than *replace* a human decision-maker by providing advice (in the form of predicted *risk scores*), is becoming an increasingly widespread paradigm in fields ranging from criminal justice to healthcare. Gali's work asks the following question: to best improve human decision-making, *when* should algorithms provide advice?

A natural guess is that algorithms should always provide advice–and indeed, this is generally what happens in current practice–but it turns out that this may not actually be the case. In fact, experience suggests that humans and algorithms have complementary strengths–algorithms are not perfect, and there are still areas where humans can outperform algorithms. Another empirically observed phenomenon is a *scarcity effect*, in which humans respond more strongly when advice is provided

less frequently. Taken together, these facts point towards a *responsive advising approach*, where the AI only provides the advice when it is expected to improve the decision. Is such an approach possible? And if so, does it actually help?

Using the context of predicting recidivism in pre-trial release decisions, this work considers an algorithmic assistant consisting of a risk assessment algorithm (produces a risk score) and an advising policy (decides whether or not to provide advice). The results? The algorithmic assistant learned to advise with 74.1% accuracy, and the assisted human decisions were approximately as accurate as the algorithmic predictions. In short–yes, responsive advising is both possible *and* helpful!

Other empirical findings included an illustration of the scarcity effect, as well as evidence of "learning on both sides" (by both the human and the algorithm). Finally, Gali emphasized that many richer forms of human-AI interaction remain unexplored. Moving forward, as we continue to navigate the various tradeoffs between learning from algorithms and preserving human strengths, it will become increasingly important to understand how humans and algorithms behave in collaboration with each other.

3.4 Nicole Immorlica - Data, the Fundamental Particle of Interaction



The final spotlight talk of the day was given by **Nicole Immorlica** (Microsoft Research) on data, the fundamental particle of interaction. Nicole began with the motivating observation that one of the most basic assumptions underlying much econCS research–that rational agents make decisions by maximizing expected utility according to prior distributions–is *not* actually an accurate model of human behavior. In reality, priors are difficult to formulate precisely; instead, humans tend to operate off past experiences. For large corporations and other organizations, this is often in the form of massive datasets; on the level of individual human decision-making, this is usually based on *anecdotal experiences*.

Anecdotes are a useful tool for representing non-experts (who lack sophisticated formal models of the world), and considerations such as data privacy or other constraints. In a stylized model, the state of the world is parametrized by some value unknown to the agent, and an *anecdote* is a noisy representation of the state of the world. Now, an agent receives anecdotes one at a time, and can choose to remember or forget each one. How many anecdotes does an agent need to use to form a "good" estimate of the state of the world?

If the agent has perfect memory, the Central Limit Theorem implies that the empirical mean of T anecdotes has loss $O(1/T)$. But what if the agent only has bounded memory? The simple but naive strategy of remembering only the most recent anecdotes requires $O(T)$ memory to match this loss bound of $O(1/T)$. In the busy and chaotic world we live in today, we might hope to ask: can we get away with remembering any less?

Good news for the forgetful folks: yes! It turns out a form of *strategic forgetting* can achieve this same $O(1/T)$ loss with memory only logarithmic in T. The complete specification and analysis of the algorithm are fairly technical, but also enlightening and exciting. One key insight is the notion of setting the optimal "step

size" parameter –too small, and the empirical distribution barely changes (reflecting an agent who stubbornly refuses to learn from experience); but too large, and the empirical distribution fails to converge (reflecting an agent who is too easily swayed and keeps changing her mind). This latter observation in particular suggests a rationale for *confirmation bias*, the widespread behavioral phenomenon in which an anecdote must be at least somewhat close to an agent's current beliefs in order for her to remember it. This has interesting implications–oftentimes we think of "rational" and "unbiased" as interchangeable terms, but in this case (and perhaps in others), bias may be rational after all.

To close, Nicole called for researchers to begin investigating a broader agenda of viewing *data* as *anecdotes* correlated with the state of the world. This framework can be extended beyond this talk's decision theory problem, to other contexts ranging from communication and persuasion games to mechanism and information design. More generally, it's time to start viewing data as the *fundamental particle* underlying much of human behavior and decision-making, and develop economic models that reflect this new paradigm shift.

3.4.1 *Interview with Nicole Immorlica.* In an interview after the meeting, Nicole gave further insight on modeling data as anecdotes, and shared some broader perspectives on behavioral research in the growing field of economics and computation.

What inspired you to start reasoning about models using data and anecdotes instead of distributions? Was it pure intellectual curiosity, motivated by a product team, or something else entirely?

I think it was mostly intellectual curiosity, but this curiosity was initiated by observations from product teams. My first paper in this space (*Incentivizing Exploration with Selective Data Disclosure*, joint work with Alex Slivkins [Immorlica et al. 2018]) was motivated by platforms like Yelp–we wanted to think about how Yelp uses reviews (which are anecdotes or data points) to persuade people to try different restaurants. Why is this a natural model? Why does Yelp actually aggregate past reviews instead of just making up arbitrary scores? Perhaps it's because that's how humans reason–we look at experiences others have had in the past. So this was the first moment where I thought this kind of framing makes sense–Yelp is using reviews because humans reason better about concrete data than they do about abstract messages.

My next paper in this framework, *Communicating with Anecdotes* [Haghtalab et al. 2022], took this further–if people are in the business of sharing facts and data points rather than beliefs and messages, how does conversation work? What happens if we can't just say anything we want (the typical *cheap talk* assumption underlying much of the economics literature), but can only communicate things that have actually happened? In this world, you can think of this as only being able to communicate *true news* via facts, evidence, and anecdotes. What beliefs then get induced? This paper was also influenced by many fascinating conversations I had with Nancy Baym, a communications scholar at Microsoft, about why people communicate, as well as some psychology literature that one of my coauthors, Markus Mobius, was reading at the time. It was fantastic to have access to people outside the EC subcommunity who think about things from different perspectives.

This paper I presented at the winter meeting was my third in this line of work. It combines the previous idea that we think about data and anecdotes with this concept of limited memory that Markus was intrigued by from a book he was reading (*Foundations of Human Memory* by Michael Jacob Kahana [Kahana 2012]).

Zooming into some technical details of your talk: Can we say anything constructive about distributions that are not necessarily single-peaked? (Using your coffee example from your talk, what if there's a "good barista" and a "bad barista" so the distribution of coffee quality is bimodal?)

All of the papers I discussed have some kind of single-peaked assumption; this serves as a focal point to narrow in on, so that the mass of the distribution is where the mean is. If you have a bimodal distribution, instead of converging, gradient descent might just keep jumping around between the two local maxima. It also depends on what exactly you're trying to learn in this case–the average quality of the coffee shop, or the quality of one barista versus the other? In the latter case, I actually do have two single-peaked distributions, just obfuscated (and if I can tell which samples are from which distribution, then we can apply the multi-dimensional extension of our algorithm).

But in general, you do need some kind of structure on the distribution–think of it as coming from the assumption that there's a ground truth of the state of the world, and observations are noisy perturbations. If the noise is zero mean, then you'll end up with a single-peaked distribution.

Have you thought about modeling other behavioral phenomena, such as rose-colored glasses, memories drifting over time, or more strongly remembering extreme memories?

For sure, I'm quite interested in looking at how cognitive constraints may lead to behavioral phenomena. I would love to assume that people are trying to optimize their true underlying utility but have limited abilities, and when they try to find the optimal algorithm subject to these constraints, it turns out to have behavioral implications. It would be cool to come up with model where the human is trying to optimize subject to being able to run an algorithm in their head, and as a feature of the best such algorithm, ends up exhibiting a particular behavior (in the talk I presented, confirmation bias arose from humans having limited memory).

Remembering extreme events is something I've talked about a bit with Nageeb Ali. You could imagine a setting where the goal is to learn the best action out of several choices (rather than a mean estimation problem), so you would love to run a multi-armed bandit algorithm, but you can't quite do so because you only have anecdotes and maybe also limited memory. What could you do instead? One thing you might want to do is to mimic UCB by remembering the best thing you've seen so far. Maybe if you have enough memory to remember something about every arm, you'd also want to remember the really terrible points so you never try them again. So interestingly, both the phenomena of remembering extreme events, or just the very positive ones, could arise from trying to mimic UCB.

You hinted at other contexts where this sort of paradigm shift from

distributions to data could be applied, like in game theory and mechanism design. What's one direction that you find particularly exciting?

I'm starting to try to understand the impact of LLMs on how we write our economic models and games. Here, I think of LLMs as both content creators and content aggregators. In the former aspect, there's this whole theme of generating the content (data points) that we see. How does the existence of these AI machines that can contribute to this environment influence what data we create? If the LLM is creating data and also competing with me for attention, does this influence how I create data myself?

Regarding the other concept as an aggregator, so far I've been operating under the basic assumption that I can see and communicate and remember a data point. But with a summarization tool, I can remember *combinations* of data points instead. If I can remember arbitrary combinations, then I get back to standard models of distributions, but what if there are constraints on the combinations? Further, what if the LLM wants to influence humans to think a certain way? On one hand, LLMs can support human learning by providing combinations of data points; on the other hand, these combinations may be biased by the LLM. More broadly, how do LLMs affect how people learn (especially if incentives are misaligned)?

More generally, what is your process for coming up with interesting research problems?

I think everyone has their own process, and what inspires people may be different. For me, I'm much more into the modeling aspect than the theorem-proving aspect, so I try to look for something I don't really understand and figure out why it's happening. In this instance: people often make anomalous, seemingly sub-optimal decisions. Why? Maybe it's because they have limited memory. Overall, my research is a tool for helping me understand the world around me.

There's a big gap between figuring out what you're interested in the world and what questions to ask. Another thing I've been interested in for a while but haven't formed the right questions around is the concept of identity. What is identity? How does it matter to me that I identify as a woman, as both an economist and a computer scientist, as a Microsoft employee, as a member of my family? How do identities change? How do conflicts between my perception of my own identity and others' perceptions of me influence my actions? There's this quote from a Walt Whitman poem: "Do I contradict myself? Very well then I contradict myself, (I am large, I contain multitudes.)" This is another hint that we do things that may seem absurd from a rationality perspective, but maybe that's because we are not single rational actors! Even a single individual is composed of many forces pulling in different directions–how can we use this notion of identity to think about how we act as a result?

So to get ideas, I read poetry. I read related works (Rachel Kranton and George Akerlof have some great stuff). I talk to people outside of the subfield. Then I start trying to write down models.

What's one random fun fact about you?

I really value breadth of experience, both academically and personally. For example, just pre-pandemic I took a poetry course at MIT because I'm interested in

engaging with fields beyond STEM. In fact, one of my favorite classes in undergrad was German film post-World War II!

REFERENCES

Camara, M. K. 2022. Computationally tractable choice. In *Proceedings of the 23rd ACM Conference on Economics and Computation*. EC '22. Association for Computing Machinery, New York, NY, USA, 28.

Dreyfuss, B., Glicksohn, O., Heffetz, O., and Romm, A. 2022. Deferred acceptance with news utility. Tech. rep., National Bureau of Economic Research.

Gonczarowski, Y. A., Heffetz, O., and Thomas, C. 2023. Strategyproofness-exposing mechanism descriptions. Tech. rep., National Bureau of Economic Research.

Haghtalab, N., Immorlica, N., Lucier, B., Mobius, M., and Mohan, D. 2022. Communicating with anecdotes. *arXiv preprint arXiv:2205.13461*.

Immorlica, N., Mao, J., Slivkins, A., and Wu, Z. S. 2018. Incentivizing exploration with selective data disclosure. *arXiv preprint arXiv:1811.06026*.

Kahana, M. J. 2012. *Foundations of human memory*. OUP USA.

Kleinberg, J., Mullainathan, S., and Raghavan, M. 2023. The challenge of understanding what users want: Inconsistent preferences and engagement optimization. *Management Science 0,* 0, null.

Kleinberg, J., Oren, S., and Raghavan, M. 2016. Planning problems for sophisticated agents with present bias. In *Proceedings of the 2016 ACM Conference on Economics and Computation*. EC '16. Association for Computing Machinery, New York, NY, USA, 343–360.

Li, S. 2017. Obviously strategy-proof mechanisms. *American Economic Review 107,* 11, 3257–3287.

McIlroy-Young, R., Sen, S., Kleinberg, J., and Anderson, A. 2020. Aligning superhuman ai with human behavior: Chess as a model system. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD '20. Association for Computing Machinery, New York, NY, USA, 1677–1687.

Meisner, V. 2023. Report-dependent utility and strategy-proofness. *Management Science 69,* 5, 2733–2745.

Noti, G. and Chen, Y. 2023. Learning when to advise human decision makers. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, E. Elkind, Ed. International Joint Conferences on Artificial Intelligence Organization, 3038–3048. Main Track.

Oprea, R. 2020. What makes a rule complex? *American Economic Review 110,* 12 (December), 3913–51.