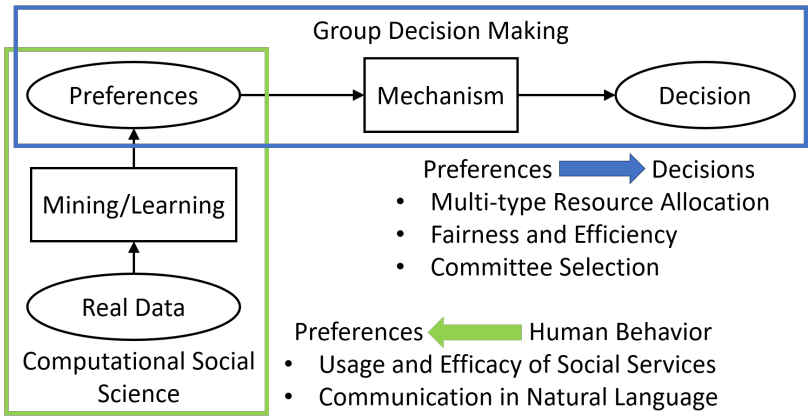
RESEARCH STATEMENT - SUJOY KUMAR SIKDAR

sujoyks@gmail.com

My research interests lie in the aggregation of information in the form of explicit or implicit preferences, in systems composed of multiple agents, and making optimal group decisions based on this information. My work revolves around problems in computational social choice, where I develop algorithms and mechanisms to make decisions for groups of agents given their preferences, with desirable economic and computational properties, and provable guarantees. I am equally concerned with a data driven approach to develop models of human preferences that are realistic, expressive, and cognitively viable. My research touches upon Artificial Intelligence, Mechanism Design, Computational Social Choice, Data Mining, Machine Learning, and Computational Social Science. Below, I briefly describe my research in group decision making and computational social science.



Group Decision Making

Fundamental societal problems often require making decisions for groups of agents over multiple attributes simultaneously, such as allocating or dividing public and private resources of multiple types, or making decisions on multiple issues simultaneously, when the agents have complex preferences over the alternatives. Consider the following examples: in a seminar class, a group of students may want to exchange research papers and time slots for presentation; in cloud computing, agents may want to allocate multiple types of resources, including CPU, memory, and storage; in determining what to serve during a multi-course dinner for a group of guests, the host may wish to have the guests vote for combinations of food items and beverages. Designing mechanisms to make optimal decisions that are fair, economically efficient, incentivize participation, and encourage agents to truthfully report preferences can become challenging in such problems because: (i) the number of alternatives grows exponentially with the number of attributes, and (ii) agents' preferences over the alternatives may have a complex combinatorial structure. These challenges often mean that there is no mechanism that "ticks all the desired boxes" when decisions must be made over multiple attributes. My work delivers a surprisingly positive message: *it is possible to design good mechanisms for multi-attribute decision making problems under mild assumptions on preferences.*

Multi-type Housing Markets. In our AAAI-17 paper [16], we study multi-type housing markets [12], where we break a long standing barrier in the literature and provide the first positive results on finding *core allocations*. In a multi-type housing market, a collection of agents each endowed with a set of indivisible items of different types, have preferences over *bundles* consisting of subsets of all items. The goal is to find a redistribution of items that best satisfies agents' preferences without exchanging money. The notion of the core [15] of the market, the set of allocations where no group of agents has incentive to deviate by exchanging their initial endowments within the group, is a particularly appealing notion of a good redistribution, as it is intuitively stable, and economically efficient. Computing core allocations for multi-type housing markets have received very little attention because the core may be empty [10]. Indeed, as Sönmez and Ünver [19] noted: *"Positive results of this section [on housing markets] no longer hold in an economy in which*

one agent can consume multiple houses or multiple types of houses." We address this problem in Sikdar, Adalı, and Xia [16], by proposing an extension of the Top Trading Cycles (TTC) mechanism [15] to multi-type housing markets which is strict core selecting, when agents' preferences are lexicographic. In our AAAI-19 paper [17], we propose a general framework for multi-type housing markets, where agents are allowed to specify sets of acceptable bundles. We propose a new graphical language to represent preferences, that generalizes previously studied preference languages, and identify a general setting where our extension of TTC is strict core selecting, and always computes allocations where every agent is allocated an acceptable bundle.

Multi-type Resource Allocation. In our recent AAAI-20 paper [21], we study the multi-type resource allocation problem, and provide the first mechanisms which are *simultaneously fair and efficient*, when items are divisible or indivisible. We uncover a fundamental impossibility result that for the unrestricted domain of partial preferences, no mechanism can guarantee the satisfaction of desirable fairness and efficiency properties simultaneously. Notwithstanding this, our main message is positive. For divisible items, and under the natural domain restriction of CP-net preferences [3], we provide the first extensions of the well-known Random Priority [1] and Probabilistic Serial [2] mechanisms to multiple types of items, and to partial preferences, and satisfy close analogues of all of the fairness, efficiency, and strategyproofness properties of their counterparts for single types and complete preferences. We encounter a challenge in dealing with indivisible items. Unlike the single type setting, where every fractional assignment is *decomposable* into a probability distribution over discrete assignments, the same does not hold for multiple types. We propose a novel mechanism for indivisible and divisible items, matching the stronger efficiency property satisfied by PS, and the weaker fairness notion satisfied by RP, even for the unrestricted domain of partial order preferences.

Fair Division Through Information Withholding. In our recent AAAI-20 paper [9], we consider the resource allocation problem with additive valuations, and study allocations that are nearly *envy-free* in aggregate. We define a novel fairness notion based on the idea of *withholding information*. Under our notion, an agent can withhold (or hide) some of the goods in its allocated bundle, and reveal the remaining goods to the other agents. We observe that in practice, *envy-freeness can be achieved by withholding only a small number of goods overall.* Our notion is stronger that the notion of envy-freeness up to one good (EF1), a well-studied fairness notion for indivisible goods that addresses pairwise envy by the removal of at most one good. In the worst case, each pair of agents might require the (hypothetical) removal of a different good, resulting in a weak aggregate guarantee. In contrast to this worst-case guarantee, our experiments on synthetic and real-world preference data from Spliddit [7] show that existing algorithms for finding EF1 allocations withhold close-to-optimal amount of information.

Equitable Allocation of Indivisible Goods. In our IJCAI-19 paper [4], we study the *equitable* allocation of indivisible goods among agents with additive valuations. Our work is the first to consider equitability in conjunction with other well-studied notions of fairness and economic efficiency. We show that the well known Leximin algorithm always produces an allocation that satisfies Pareto optimality, and equitability up to any good, meaning that for any pair of agents, the inequity in utilities of the agents can be eliminated by the removal of any good from the agent with the higher utility. We also give a novel algorithm that guarantees Pareto optimality and equitability up to one good in pseudopolynomial time. Our experiments on real-world preference data from Spliddit [7], reveal that approximate envy-freeness, approximate equitability, and Pareto optimality can often be achieved simultaneously.

Multi-winner Elections and Committee Selection. In our AAMAS-17 paper [14], we address the following question in multi-winner elections: *How to select a committee of size k to satisfy a majority of voters, who have preferences over individual candidates?* We take the approach of extending the Condorcet criterion for single-winner elections proposed by Gehrlein [5] to multi-winner elections, which translates to the following notion of stability: every member of the committee enjoys support from a majority of voters in a head to head contest with any member outside the committee. We provide a simple template to extend Condorcet consistent rules to committee selection, whereby any single-winner Condorcet consistent rule which depends only on the tournament graph, can be extended to a multi-winner rule which depends only on the local majority graph consisting of edges between neighboring committees that differ by only one member. Additionally, we introduce a novel notion of approximate committee selection for hard to solve multi-winner rules, where we relax the constraint on the size of the committee, and instead guarantee that the total satisfaction is at least that of the optimal committee of size k. Indeed, there is often some flexibility in the size of the committee, for example: in the number of journals carried by a library, products in a company's portfolio, and members in a congressional committee. For the Maximin rule, we provide an approximation algorithm which computes a committee of size at most $2k$ that is at least as good as the optimal committee of size k.

Computational Social Science

Preferences of Communities in Social Media. In an ICCCN-17 paper [8], we study the voting behavior of users on the popular online social media platform reddit, where users engage in discussions on several topics ranging from politics to science, and steer the discussion by promoting specific threads by voting. Informed by social science theory, and applying techniques in natural language processing and machine learning, we identify the key attributes of comments that affect users' preferences for certain threads of discussion, and provide key insights into the multi-attribute nature of preferences of a diverse set of communities.

Information Credibility. In a SocialCom-13 paper [18] which *won the best paper award*, we studied the problem of identifying credible information on Twitter. Our work is among the first to develop reliable metrics, and provide accurate models of information credibility for social networking data that involve multiple factors that affect credibility such as competence and reliability of sources from social ties and linguistic features. A key component to the success of our models was the development of carefully designed crowdsourced studies to annotate large scale data from Twitter to train and validate our models.

Ongoing and Future Work

Machine Learning for Group Decision Making. A particularly exciting avenue for future research is the use of *machine learning to help group decision making*. In a recent AAAI-19 paper [20], we develop a novel framework for the use of machine learning to develop practical, anytime, *exact algorithms* for computationally hard group decision making and winner determination problems for the first time. In another IJCAI-19 paper [11], we develop a learning and recommendation approach to improve the *usability* of the online voting platform OPRA [13] developed by an outstanding team of undergraduate researchers at RPI. We recommend an initial ordering of alternatives to minimize the time taken by a user to input her preference by learning user preferences and user behavior in interacting with the platform from real world data. Our work is motivated by the increasing use of online voting and preference reporting platforms such as OPRA, with considerations of practicality and scalability in mind.

Mechanism Design for Social Good. I am particularly interested in the design of fair and efficient mechanisms for the division of resources, and making public decisions for social good. In ongoing work, I consider the allocation of child protection services using real world data. The broader problem of designing of fair and efficient dynamic matching algorithms, and identifying and mitigating bias in the allocation of social services, raises several exciting possibilities for interdisciplinary work involving social science and social work. I strongly believe that there is a great opportunity to create virtuous cycles where data drives research and the development of improved mechanisms applied to serve the needs of society, using data science and artificial intelligence to incorporate large scale real-world data to assess society's needs, the availability of resources, and design mechanisms that are fair and efficient.

Computational Social Choice. My recent work on mechanism design for multi-type resource allocation and the fair and efficient resource allocation present several interesting avenues for future research. In ongoing work, I expand on my current work on multi-winner elections and the idea of relaxing the size of the selected committee to design approximation algorithms for hard to compute multi-winner voting rules. In multi-type resource allocation, often, different agencies are responsible for the allocation of different types of resources. An open problem is the role of a central authority that wishes to co-ordinate the agencies to fairly and efficiently allocate resources.

Computational Social Science. In ongoing work, I study the strategic behavior of political candidates who communicate their positions to voters over different channels of communication such as social and print media, when voters may each subscribe to different channels. I am particularly interested in realistic game theoretic models that allow voter abstention, multiple issues and political dimensions, and party loyalty from voters and candidates. I use data science and natural language processing techniques to identify the political positions of candidates from text of their communications in social and print media, and validate empirical results from our game theoretic model.

In ongoing work outlined in my dissertation, I study a question answering setting, and develop models to understand why humans give some responses to a question with higher frequency, using data from a carefully developed laboratory experiment in collaboration with experts in cognitive neuroscience, where subjects are asked to provide one word answers to general knowledge questions. Together, we identify novel and relevant natural language and semantic features relevant to provide new insights into human preference formation and decision making in question answering, and develop models inspired by work on "fast and frugal heuristic (FFH)" [6] based decision models from the psychology literature that are cognitively viable and require very little training data.

References

[1] Abdulkadiroglu, A., and Sonmez, T. 1998. Random Serial Dictatorship and the Core from Random Endowments in House Allocation Problems. *Econometrica* 66(3):689.

[2] Bogomolnaia, A., and Moulin, H. 2001. A New Solution to the Random Assignment Problem. *Journal of Economic theory* 100(2):295–328.

[3] Boutilier, C.; Brafman, R. I.; Domshlak, C.; Hoos, H. H.; and Poole, D. 2004. Preference-Based Constrained Optimization with CP-Nets. *Computational Intelligence* 20(2):137–157.

[4] Freeman, R.; Sikdar, S.; Vaish, R.; and Xia, L. 2019. Equitable Allocations of Indivisible Goods. In *IJCAI-19*.

[5] Gehrlein, W. V. 1985. The Condorcet Criterion and Committee Selection. *Mathematical Social Sciences* 10(3):199–209.

[6] Gigerenzer, G., and Goldstein, D. G. 1996. Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Psychological Review* 103(4):650.

[7] Goldman, J., and Procaccia, A. D. 2015. Spliddit: Unleashing Fair Division Algorithms. *ACM SIGecom Exchanges* 13(2):41–46.

[8] Horne, B.; Adali, S.; and Sikdar, S. 2017. Identifying the Social Signals that Drive Online Discussions: A Case Study of Reddit Communities. In *ICCCN-17*.

[9] Hosseini, H.; Sikdar, S.; Vaish, R.; Wang, J.; and Xia, L. 2019. Fair Division Through Information Withholding. In *AAAI-20*.

[10] Konishi, H.; Quint, T.; and Wako, J. 2001. On the Shapley–Scarf Economy: The Case of Multiple Types of Indivisible Goods. *Journal of Mathematical Economics* 35(1):1–15.

[11] Li, H.; Sikdar, S.; Vaish, R.; Wang, J.; Xia, L.; and Ye, C. 2019. Minimizing Time-to-Rank: A Learning and Recommendation Approach. In *IJCAI-19*.

[12] Moulin, H. 1995. *Cooperative Microeconomics: A Game-Theoretic Introduction*. Prentice Hall.

[13] Online Preference Reporting & Aggregation. http://www.opra.io. Accessed: 2019-12-31.

[14] Sekar, S.; Sikdar, S.; and Xia, L. 2017. Condorcet Consistent Bundling with Social Choice. In *AAMAS-17*.

[15] Shapley, L., and Scarf, H. 1974. On Cores and Indivisibility. *Journal of Mathematical Economics* 1(1):23–37.

[16] Sikdar, S.; Adalı, S.; and Xia, L. 2017. Mechanism Design for Multi-type Housing Markets. In *AAAI-17*.

[17] Sikdar, S.; Adalı, S.; and Xia, L. 2019. Mechanism Design for Multi-type Housing Markets with Acceptable Bundles. In *AAAI-19*.

[18] Sikdar, S.; Kang, B.; ODonovan, J.; Höllerer, T.; and Adah, S. 2013. Understanding Information Credibility on Twitter. In *2013 International Conference on Social Computing*, 19–24. IEEE.

[19] Sönmez, T., and Ünver, M. U. 2011. Matching, Allocation, and Exchange of Discrete Resources. In Benhabib, J.; Bisin, A.; and Jackson, M. O., eds., *Handbook of Social Economics*. North-Holland. chapter 17, 781–852.

[20] Wang, J.; Sikdar, S.; Shepherd, T.; Zhao, Z.; Jiang, C.; and Xia, L. 2019. Practical Algorithms for STV and Ranked Pairs with Parallel Universes Tiebreaking. In *AAAI-19*.

[21] Wang, H.; Sikdar, S.; Guo, X.; Xia, L.; Cao, Y.; and Wang, H. 2020. Multi-type Resource Allocation with PartialPreferences. In *AAAI-20*.