CAREER: A New Theory of Social Choice for More than Two Alternatives: Combining Economics, Statistics, and Computation

PI: Lirong Xia, July 19, 2014

Project summary

Classical social choice theory (a.k.a. group decision making) sets an economic foundation for a group of agents to make a joint decision in political domains, especially political elections. In the Internet era, social choice arises naturally as a fundamental problem in many new applications including non-electoral preference aggregation, e.g. opinion polling, multi-agent systems [51], and e-commerce which includes meta-search engines [41], recommender systems [49], and crowdsourcing [67]. These new applications are beyond the scope of classical social choice theory because of three features: (1) **More than two alternatives:** the number of alternatives can be thousands or even millions. (2) **Truth-revealing:** agents' preferences are viewed as data generated from a statistical model given the true state of the world, and the main goal of social choice is to reveal the ground truth by aggregating agents' preferences. (3) **Partial preferences:** agents may not be able to compare some alternatives.

Addressing challenges in these new applications requires a combination of economics, statistics, and computation. However, most previous work overlooked at least one of these aspects. Moreover, there is a lack of principled and systematic framework for discovering new mechanisms.

The broad goal of my research career, situated within the emerging area of *economics and computation*, is to **establish a new theory of social choice to choose or discover application-specific mechanisms for more than two alternatives by combining economics, statistics, and computation**. In pursuit of this goal, I propose to initiate studies in three directions for long-term research: (i) Truth-revealing social choice for more than two alternatives; (ii) Social choice for partial preferences; and (iii) Statistical and machine learning frameworks for new mechanisms.

Intellectual merit: The combination of economics, statistics, and computation is the most distinctive feature and the main conceptual merit of the proposed research. This stands in sharp contrast to all previous work, which either focused on the economic aspect as in classical social choice, the statistical aspect as in statistics, machine learning, and political science, or the computational aspect as in *computational social choice* [25, 22].

Success of the proposed research will also create new opportunities and techniques for applying social choice theory to other research fields, especially multi-agent systems and e-commerce. Additionally, the theoretical analysis and computational techniques developed along the third direction will advance the state of the art of statistics and machine learning.

Broader impacts: Broadly, success of the proposed research will not only theoretically justify democracy and the wisdom of crowds, but also practically improve the usability and performance of social choice mechanisms in the new applications.

I am serving on an international research board of the *Trustworthy Random Sample Elections project* led by David Chaum and Deborah Hurley. This non-profit project aims at promoting democracy by developing a novel trustworthy voting system, and I expect the proposed research to have a significant impact on the theory and practice of this system. All codes in this project will be open source and be made available to different applications such as the RPI student leader elections. Another important part of the proposal is to disseminate the knowledge and spirit of classical, truth-revealing, and computational social choice to K-12 students by participating in the newly launched Rensselaer Science Ambassador program. The proposal also includes a rigorous plan on various educational and outreach activities with a focus on diversity, including supervising graduate and undergraduate students, developing curriculum, chairing workshops (M-PREF series) and conferences (including AMMA-15), and presenting tutorials.

Keywords: group decision making; social choice theory; multi-agent systems; e-commerce

1 Introduction

Classical social choice theory (a.k.a. group decision making) sets an economic foundation for a group of agents to make a joint decision in political domains, especially political elections. The primary approach of classical social choice has been to identify desired economic properties (*axioms*) to evaluate and choose manually-designed mechanisms. For two alternatives, e.g. electing a Democratic candidate or a Republican candidate, the majority rule is optimal w.r.t. economics, statistics, and computation. For more than two, such as three, alternatives, no social choice mechanism is optimal according to the Nobel-prize-winning *Arrow's impossibility theorem* [6].

In the Internet era, social choice arises naturally as a fundamental problem in many applications including non-electoral preference aggregation, e.g. opinion polling, multi-agent systems [51], and e-commerce which includes meta-search engines [41], recommender systems [49], and crowdsourcing [67]. These new applications are beyond the scope of classical social choice theory, because they often have the following three features:

1. **More than two alternatives:** the number of alternatives can be thousands or even millions.
2. **Truth-revealing:** agents' preferences are viewed as data generated from a statistical model given the true state of the world, and the main goal of social choice is to reveal the ground truth by aggregating agents' preferences.
3. **Partial preferences:** agents may not be able to compare some alternatives.

For example, when opinion polling is used to rank hotels, the number of hotels is often more than two; the goal is to reveal the ground truth ranking over the hotels; and an agent may not be able to compare an expensive and comfortable hotel to a cheap but less comfortable hotel. As another example, when a meta-search engine is used to rank webpages, there are often thousands or millions of alternatives (webpages); the goal is to reveal the ground truth ranking w.r.t. the input keywords; and preferences of an agent (search engine) are often partial, for example when the keyword is "Amazon", a search engine may not have enough information to compare the webpage about the Amazon rainforest and the webpage about the Amazon company.

Addressing challenges in these new applications requires a combination of economics, statistics, and computation. However, most previous work overlooked at least one of the economic, statistical, and computational aspects. Moreover, there is no principled and systematic framework for discovering new social choice mechanisms. I believe that computer scientists are in an advantageous position to address these challenges because computational techniques, especially machine learning, are useful in discovering new mechanisms. And considering that there is a wide range of applications of social choice in AI and e-commerce, computer science community as a whole can benefit significantly from these studies.

The broad goal of my research career, situated within the emerging area of *economics and computation*, is to **establish a new theory of social choice to choose or discover application-specific mechanisms for more than two alternatives by combining economics, statistics, and computation**. In pursuit of this goal, I propose to initiate studies in three distinct yet related directions to set the foundations for long-term research:

(i) Truth-revealing social choice for more than two alternatives.

(ii) Social choice for partial preferences.

(iii) Statistical and machine learning frameworks for new mechanisms.

My expertise and institutional context. My previous work focused on classical and computational social choice, including the truth-revealing aspects [34, 107, 102, 9, 95, 11, 10, 12, 98, 13] and an important subcase of social choice with partial preferences called *combinatorial voting* [104, 105, 106, 60, 103, 33, 107, 101, 108, 35, 32, 62], including a book chapter in preparation with Prof. Jérôme Lang [61]. My work on social choice has been reported by Duke CS News, Duke Economics News, Harvard SEAS News, and RPI news [1, 2, 3, 4, 5]. My Ph.D. dissertation titled "Computational Social Choice: Strategic and Combinatorial Aspects" won the Duke CS dissertation award and was recommended for the ACM Dissertation Award.

RPI is an ideal place to carry out my proposed research, as well as educational and outreach activities (more details in Section 5). There are many opportunities for joint work on social choice, especially with the Social Cognitive Networks Academic Research Center led by Prof. Boleslaw Szymanski. RPI's strong support of undergraduate research and the Rensselaer Science Ambassador outreach program will also contribute to the success of the proposed activities.

1.1 Intellectual Merit

The combination of economics, statistics, and computation in guiding the evaluation and discovery of social choice mechanisms for new applications is the most distinctive feature and the main conceptual merit of the proposed research. This stands in sharp contrast to all previous work, which either focused on the economic aspect as in classical social choice, the statistical aspect as in statistics, machine learning, and political science, or the computational aspect as in *computational social choice* [25, 22].

In the new social choice applications, the notion of *agents* has been extended from humans to software agents, search engines, and other sources of preferences. For example, in multi-agent systems, software agents vote to schedule the next meeting for users [51]; in meta-search engines, rankings over webpages by different search engines are merged to produce an aggregated ranking [41]; in recommender systems, rankings over items according to various criteria are combined to determine an aggregated ranking for recommendation [49]; in crowdsourcing, noisy answers to the same question from online labors are aggregated to produce a more accurate answer [67].

Addressing challenges of social choice in these new applications requires a combination of economics, statistics, and computation. Economic considerations, especially satisfiability of axioms, are desired. This is because:

(a) For new applications with strong societal contexts, e.g. opinion polling, it is important to ensure democracy and respect agents' preferences, which are often evaluated by the satisfaction of various axioms.

(b) For new applications with less societal contexts, for example meta-search engines, many social choice axioms are still desired because they represent sensible regulations on the mechanism. One example is the *monotonicity* axiom, which requires that moving an alternative up in an agent's ranking never hurts the alternative in the output of the mechanism.

The truth-revealing feature of the new applications immediately justifies the importance of and desire for superior statistical properties. Meanwhile, computation becomes a major issue as the number of alternatives increases. Hence it is important that the mechanisms have superior computational properties as well.

In this proposal, I plan to focus on the following research questions in the three directions.

(i) **Truth-revealing social choice for more than two alternatives.** What is the truth-revealing ability of existing mechanisms for more than two alternatives? How can we numerically measure the truth-revealing ability of a mechanism for truthful agents and strategic agents?

(ii) **Social choice for partial preferences.** How can we extend existing social choice mechanisms and axioms to partial preferences? What are the axiomatic, statistical, and computational properties of these mechanisms?

(iii) **Statistical and machine learning frameworks for new mechanisms.** Can we establish principled and systematic frameworks to discover new social choice mechanisms given a statistical model of agents' preferences? Can we use machine learning to discover new social choice mechanisms with desired axioms?

The first and second directions aim at extending the scope of classical and computational social choice theory to new applications. Therefore, success of the proposed research will significantly enrich existing social choice theory. Practically, this will create new opportunities and techniques to develop and apply social choice theory and further deepen relationships between social choice and other research fields, especially AI and e-commerce.

The third direction is related to but technically significantly different from *automated mechanism design* [28, 29]. In (automated) mechanism design, the goal is to design the "rule of the game" to control the equilibrium outcome. In the proposed research, we want to design a social choice mechanism with desired axiomatic, statistical, and computational properties. Since the design problem cannot be handled by existing techniques in statistics and machine learning, I expect that the theoretical analysis and computational techniques developed along this direction will advance the state of the art in statistics and machine learning as well.

As such, I believe that the proposed research in this project naturally builds a promising path towards the broad goal of my research career, which is a new theory of social choice to choose or discover application-specific mechanisms for more than two alternatives by combining economics, statistics, and computation.

1.2 Broader Impacts

Success of the proposed research will directly benefit most social choice applications. For applications that aggregate preferences of humans, including elections, opinion polling, and some e-commerce applications, success of the proposed research will not only theoretically justify democracy and the wisdom of crowds, but also practically allow agents to better express their preferences and help to choose or discover application-specific social choice mechanisms.

For example, I am currently serving on the research board of the *Trustworthy Random Sample Elections project* led by David Chaum and Deborah Hurley. This non-profit project aims at promoting democracy by developing a novel trustworthy voting system using randomization [24], and an important application is opinion polling. The research board consists of a strong international team of professors from Harvard, Cornell, Concordia University, University of Birmingham, etc. A prototype of the system has been developed and under test. I believe that the success of the proposed research will have a significant impact on the theory and practice of this system.

Success of the proposed frameworks, statistical models, and learning techniques will also improve the performance of social choice systems with software agents or other abstract notions of agents, e.g. meta-search engines and multi-agent systems. These will be further supported by the open software packages and platforms developed for the proposed research (see Section 4).

This proposal also includes a plan on educational and outreach activities, including supervising graduate and undergraduate students, developing curriculum, and participating in professional activities. I will also actively participate in the newly launched Rensselaer Science Ambassador program to disseminate the knowledge and spirit of classical, truth-revealing, and computational social choice to K-12 students.

2 Background and Related Work

We start with some general notation in social choice. A set of m alternatives is denoted by $\mathcal{A} = \{a_1, \ldots, a_m\}$. Let $\mathcal{W}(\mathcal{A})$ denote the set of all *weak orders*, namely transitive and total binary relations, over $\mathcal{A}$. Agents' preferences are often assumed to be weak orders over all alternatives, and each agent j reports a weak order $V_j \in \mathcal{W}(\mathcal{A})$ to represent her preferences, which might be different from her true preferences. The collection of agents' reported preferences is called a *profile*, denoted by $P = (V_1, \ldots, V_n)$. Let $\mathcal{O}$ denote an *outcome space*. A *social choice mechanism* r is a function that maps each profile to an outcome, that is, $r : \mathcal{W}(\mathcal{A})^n \to \mathcal{O}$. Most previous work focused on three types of outcome space: (1) $\mathcal{O} = \mathcal{A}$, where the mechanism chooses a single winner and is often called a *resolute* rule; (2) $\mathcal{O} = 2^{\mathcal{A}} - \emptyset$, where the mechanism chooses multiple winners and is often called an *irresolute* rule; and (3) $\mathcal{O} = \mathcal{W}(\mathcal{A})$, where the mechanism chooses a weak order and is often called a *social welfare function*.

2.1 Classical social choice theory

Below are two examples of commonly-studied social choice mechanisms where agents' preferences are *linear orders* over all alternatives, namely transitive, antisymmetry, and total binary relations, denoted by $\mathcal{L}(\mathcal{A})$. Many other mechanisms have been invented and used in practice, see [21] for more examples.

Example 1. A *positional scoring rule* is characterized by a scoring vector $\vec{s} = (s_1, \ldots, s_m)$ with $s_1 \geq s_2 \geq \cdots \geq s_m$. For any alternative c and any linear order V, we let $\vec{s}(V, c) = s_j$, where j is the rank of c in V. Given a profile P, the irresolute version of positional scoring rule chooses all alternatives c with the maximum $\sum_{V \in P} \vec{s}(V, c)$; the resolute version chooses a single alternative by further applying a tie-breaking mechanism; and the social welfare function version ranks the alternatives w.r.t. their scores, and uses a tie-breaking mechanism when necessary. Popular special cases include *plurality* with scoring vector $(1, 0, \ldots, 0)$, which is currently used in state-wise presidential elections in the US, and *Borda* with scoring vector $(m-1, m-2, \ldots, 0)$.

Example 2. The social welfare function version of the *Kemeny rule* selects a ranking with the minimum Kendall-tau distance based on the number of pairwise orderings flipped from the profile P. That is, the output is a ranking W that minimizes $\sum_{V \in P} \text{Kendall}(V, W)$. The resolution version selects the alternative at the top of the winning ranking, and the irresolute version selects all alternatives that are ranked at the top of rankings with the minimum Kendall-tau distance from P.

We now recall some commonly studied axioms for resolute voting rules.

Example 3. A resolute voting rule r satisfies
• *anonymity*, if r is insensitive to permutations over agents;
• *Condorcet criterion*, if for any profile P where a Condorcet winner exists, it must be the winner. The Condorcet winner is the alternative that beats all other alternatives in pairwise elections.
• *reinforcement*, if for all profiles P_1 and P_2 with $r(P_1) = r(P_2)$, we have $r(P_1 \cup P_2) = r(P_1)$.

For three or more alternatives, all positional scoring rules satisfy anonymity and reinforcement, but none of them satisfy Condorcet criterion; the Kemeny rule satisfies anonymity and Condorcet criterion, but does not satisfy reinforcement.

2.2 Statistical models and maximum likelihood estimators

A *parametric model* $\mathcal{M} = (\Theta, \mathcal{S}, \Pr)$ is composed of three parts: a *parameter space* Θ, a *sample space* $\mathcal{S}$, and a set of probability distributions over $\mathcal{S}$, denoted by $\{\Pr_\theta(\cdot) : \theta \in \Theta\}$. Below are two commonly-studied parametric models in social choice that generate independent and identically distributed (i.i.d.) linear orders over alternatives. In the first model (Mallows model) the parameter space is discrete and is composed of all linear orders over alternatives. The probability of generating a linear order V decreases exponentially in the Kendall-tau distance between V and the ground truth parameter.

Example 4. Given $0 < \varphi < 1$, the *Mallows model [66] with fixed dispersion* φ is defined as follows: the parameter space Θ is $\mathcal{L}(\mathcal{A})$; the sample space $\mathcal{S}$ is $\mathcal{L}(\mathcal{A})^n$; for any $V, \theta \in \mathcal{L}(\mathcal{A})$, $\Pr_\theta(V) = \frac{1}{Z}\varphi^{\text{Kendall}(V,\theta)}$, where Z is the normalization factor. □

In the second class of models (random utility models), the parameter space is continuous. Each alternative has a random subjective inherent utility for each agent that guides her ranking.

Example 5. In a *random utility model* [90], each alternative a_i is characterized by a utility distribution $\mu_i(\cdot|\vec{\gamma}_i)$ parameterized by a vector $\vec{\gamma}_i$. The model is defined as follows: the parameter space Θ is composed of all combinations of parameters for alternatives $\vec{\gamma} = (\vec{\gamma}_1, \ldots, \vec{\gamma}_m)$, which is often a continuous subspace of a finite-dimensional Euclidean space; the sample space $\mathcal{S}$ is $\mathcal{L}(\mathcal{A})^n$; given any parameter $\vec{\gamma} = (\vec{\gamma}_1, \ldots, \vec{\gamma}_m)$, an agent generates a ranking in the following way: she independently samples a random utility U_i for each alternative a_i with probability $\mu_i(\cdot|\vec{\gamma}_i)$, then ranks the alternatives according to their U_i's, such that she prefers a_{i_1} to a_{i_2} if and only if $U_{i_1} > U_{i_2}$. For most random utility models, no closed-form formula for $\Pr_{\vec{\gamma}}(\cdot)$ is known. □

Well-known special cases of random utility models include the Plackett-Luce model [84, 65] and the Bradley-Terry-Luce model [20, 65], which have been widely applied in econometrics [68] and machine learning, especially *learning to rank* [64].

In statistics, an *estimator* of a parametric model $\mathcal{M} = (\Theta, \mathcal{S}, \Pr)$ is a function $T : \mathcal{S} \to \Theta$ that outputs a parameter for each data. Given a parametric model $\mathcal{M}$, a *maximum likelihood estimator* is a function $T_{\text{MLE}} : \mathcal{S} \to \Theta$ such that for any $P \in \mathcal{S}$, $T_{\text{MLE}}(P)$ is a parameter that maximizes the likelihood of the data. That is, $T_{\text{MLE}}(P) \in \arg\max_{\theta\in\Theta} \Pr_\theta(P)$. For example, the maximum likelihood estimator of the Mallows model is equivalent to the social welfare function version of the Kemeny rule (Example 2).

One of the most desired statistical properties for an estimator w.r.t. a parametric model is *consistency*, which states that as the data size goes to infinity, the output of the estimator is the same as the ground truth with probability 1.

2.3 Truth-revealing social choice for two alternatives

The study of truth-revealing social choice mechanisms can be dated back to the Condorcet Jury Theorem in the 18th century.

The Condorcet Jury Theorem [27]. Fix $p > 1/2$. Suppose there are two alternatives, one of them is "correct", and each agent has an independent probability of p to be correct. Then, the probability for the majority aggregation of agents' votes to be correct converges to 1 as the number of agents goes to infinity.

In statistical terms, the Condorcet Jury Theorem states that the majority rule is consistent w.r.t. a natural parametric model for two alternatives. The Condorcet Jury Theorem is of central importance in Political Science as it "*lays, among other things, the foundations of the ideology of the democratic regime*" [80], but it only received due attention in the mid-20th century, after the rediscovery of Condorcet's manuscript by Black [16]. Since then, numerous works in economics and political science extended the theorem to correlated agents [77, 89, 17, 55, 56, 57, 38, 37], heterogeneous agents [50, 77, 70, 78, 80, 14, 52, 15], and strategic agents [8, 43, 69, 44, 45, 71, 92, 39, 48, 82]. However, most of these extensions focused on **two alternatives** and the techniques cannot be generalized to three or more alternatives.

Only a few works extended the Condorcet Jury Theorem beyond two alternatives. Young [109] extended Condorcet's probabilistic model and observed that a maximum likelihood inference coincides the Kemeny rule [53]. Lam and Suen [58] extended the theorem for the majority rule with more than two alternatives. Chierichetti and Kleinberg [26] extended the theorem for the plurality rule when agents receive private signals and follow some protocols to report preferences. The work by Caragiannis et al. [23] can be viewed as proving the theorem for various commonly-studied mechanisms w.r.t. the Mallows model.

2.4 Computational social choice

Computational social choice studies computational aspects of, and also applies computational thinkings to social choice problems. Below we briefly recall three representative topics in computational social choice. Other major topics can be found in recent surveys [25, 22].

Preference elicitation studies how to elicit agents' preference by as few queries as possible. Elicitation algorithms have been proposed and analyzed under the minimax-regret principles when no probabilistic information about agents' preferences is available [18], under the information maximization principle when probabilistic information is available [11], and to compute the outcome of given social choice mechanisms [31].

Statistical approaches in computational social choice have mainly focused on viewing agents' votes as i.i.d. samples from a parametric model and computing their maximum likelihood estimators [30, 34, 107, 102, 9, 23]. Only a few works went beyond maximum likelihood estimators. Young [109] proposed to select an alternative with the largest probability to be ranked in the top, and this idea has been extended by Procaccia et al. [87] and Elkind and Shah [42]. There are two critical limitations of previous work. First, none of them systematically studied the satisfaction of axioms of the proposed mechanisms. Second, none of them provided a principled and systematic way to discover new mechanisms when the parameter space of the model is different from the designated outcome space of the mechanism, which is often the case in new social choice applications.

Proposed research in Section 3.3 aims at addressing these limitations.

Combinatorial voting is an important subcase of social choice with partial preferences. In combinatorial voting, the number of alternatives is exponentially large, and each alternative is characterized by its values on multiple issues. Previous research has focused on using a compact language to represent agents' preferences as structured partial orders [19, 59, 85], and the design and analysis of new mechanisms w.r.t. axiomatic and computational properties [104, 105, 106, 60, 103, 33, 107, 101, 108, 35, 32, 62]. See [22, 61] for recent surveys. However, social choice for general partial preferences is still a challenging open question, which will be the main topic of Section 3.2.

3 Proposed Research

3.1 Truth-Revealing Social Choice for More than Two Alternatives

While the axiomatic and computational properties of existing social choice mechanisms for more than two alternatives are well-understood, little is known about their truth-revealing properties. As a first step, I plan to understand whether the Condorcet Jury Theorem holds for existing social choice mechanisms for more than two alternatives.

Research Question 1. *Can we extend the Condorcet Jury Theorem to more than two alternatives for heterogeneous, correlated, and strategic agents?*

We recall that the original Condorcet Jury Theorem proved the consistency of the majority rule w.r.t. a parametric model (Section 2.2). Therefore, in statistical terms RQ 1 asks whether a given social choice mechanism is consistent w.r.t. a given parametric model. Instead of conducting case-by-case studies, I plan to extend the framework and techniques proposed in an ongoing work of mine [98], where I introduced a general class of mechanisms defined as follows.

Fixing the number of alternatives m, a *generalized outcome scoring rule* is defined by two functions (f, g), such that f maps each weak order V to a vector $f(V)$ in K-dimensional Euclidean space for some fixed K, and g selects an outcome based on the *ordering* of the K components of $\sum_{V \in P} f(V)$ for the input profile P.

Generalized outcome scoring rules are quite general, because they include the resolute version, irresolute version, and social welfare function version of many commonly-studied social choice mechanisms, including positional scoring rules and Kemeny. For homogeneous, independent, and truthful agents, I obtained a general necessary and sufficient condition for a generalized outcome scoring rule to be consistent w.r.t. a parametric model [98].

I plan to leverage the results for homogeneous, independent, and truthful agents to the setting with heterogeneous, correlated, and strategic agents. The latter setting is more realistic, especially when agents interact in social networks. Meanwhile, it is also much more challenging, because first, it is not clear how to model the correlation among agents' preferences, and second, agents' strategic behavior makes the output hard to predict—agents' reported preferences now depend not only their true preferences, but also on reported preferences of other agents.

For two alternatives, Dietrich and List [38] modeled the correlation among agents' preferences by a simple two-layer Bayesian network [81]. I plan to explore the more realistic case of more than two alternatives by modeling agents' preferences by a general and more complicated Bayesian network, and then investigate the consistency of social choice mechanisms for truthful and for strategic agents. A promising first step is to incorporate social networks into the Bayesian network to model the correlation among agents' preferences.

Only knowing whether the Condorcet Jury Theorem holds (equivalently, the consistency of a mechanism w.r.t. a parametric model) may not be sufficiently discriminative as a basis for choosing mechanisms, because it is possible that many mechanisms are consistent w.r.t. the same parametric model [23]. Consequently, we need to adopt more discriminative, and hopefully numerical, measures to evaluate and compare various mechanisms. This is our next research question.

Research Question 2. *How can we numerically measure the truth-revealing ability of a social choice mechanism w.r.t. a parametric model?*

To answer this question, I plan to bring in statistical considerations and start with two natural statistical measures: the *frequentist risk* and the *Bayesian risk*. Given a parametric model, a mechanism, and a ground truth parameter θ, the frequentist risk is the probability for the outcome of the mechanism to be θ, when the profile is generated from the model given θ. The Bayesian risk is calculated by further taking an expectation over a prior distribution over the parameters. I plan to characterize both types of risk for commonly-studied mechanisms and parametric models, especially the Mallows model and the random utility models.

Once we have a good understanding of the measures for the truth-revealing ability, we can set out to measure the effect of agents' strategic behavior.

Research Question 3. *How can we numerically measure the effect of agents' strategic behavior given a social choice mechanism and a parametric model?*

The question of measuring the effect of agents' strategic behavior in general is often answered by applying *game theory* [46] and study the *price of anarchy* of the game [54, 76]. Given a quality metric of the outcomes, the price of anarchy of a game is the optimal quality of outcomes divided by the worst quality of the equilibrium outcomes, e.g. *Nash Equilibria* [75], of the game. However, for social choice problems, there is no well-accepted quality metric.

Having truth-revealing as a desired property in mind, it is natural to take the truth-revealing ability developed for RQ 2 as the quality metric, which leads to the following definition based on the frequentist risk. A similar definition can be obtained based on the Bayesian risk as well. Given a social choice mechanism r, a parametric model $\mathcal{M}$, and the number of agents n, the *price of sophistication*, denoted by $\text{PoSn}_n(r, \mathcal{M})$, is defined as:

$$\text{PoSn}_n(r, \mathcal{M}) = \max_{\theta \in \Theta, r^*} \frac{\Pr_{P_n \sim \mathcal{M}}(r(P_n) = \theta)}{\Pr_{P_n \sim \mathcal{M}}(r^*(P_n) = \theta)}$$

In this definition, P_n is a randomly generated profile of n agents from $\mathcal{M}$ given θ, and r^* outputs an equilibrium of the game. In words, the price of sophistication is the worst-case ratio of the frequentist risk for truthful agents over the frequentist risk of the equilibrium outcome, where the worst case is taken for all ground truth θ and all ways to choose an equilibrium. Unlike the price of anarchy, surprisingly, it is possible that the price of sophistication is smaller than 1. This implies that in such situations we should encourage agents' strategic behavior, because the outcome of the mechanisms will better reveal the ground truth. An example of such situations was implicitly illustrated in the calculations by Wit [92] for the majority rule w.r.t. a simple parametric model for two alternatives. Yet the price of sophistication for mechanisms and parametric models for more than two alternatives is unknown.

I plan to initiate the study of the price of sophistication for commonly-studied social choice mechanisms and parametric models by leveraging my experience in analyzing equilibrium of social choice games, studying the ordinal price of anarchy for social choice games [100, 108, 97], and applying asymptotic theory to social choice [99, 10, 12, 13].

3.2 Social Choice for Partial Preferences

Most previous work in social choice assumed that agents are able to compare all pairs of alternatives. However, this assumption is often too strong and impractical in many new social choice applications with more than two alternatives, due to the following two reasons.

1. **Alternatives are incomparable.** For example, an agent may not be able to compare an expensive and comfortable hotel to a cheap but less comfortable hotel.

2. **Lack of information.** The agent may not have enough information to compare some alternatives. For example, a search engine may not have enough information to compare a webpage about the Amazon rainforest and a webpage about the Amazon company.

The rationality behind partial preferences has been justified and studied at the individual-agent level by Aumann [7] and many other economists, yet little is known about making social choice based on partial preferences. The goal of this section is to initiate a social choice theory for partial preferences combining considerations in economics, statistics, and computation.

Considerations in Economics. As a first step, I plan to focus on the following problem to understand natural extensions of existing mechanisms and axioms. Answers to this question will provide a basis for other questions in this section.

Research Question 4. *How can we extend existing social choice mechanisms and axioms to partial preferences? What axioms do these mechanisms satisfy?*

For general partial preferences, some preliminary results have been obtained by AI researchers [83, 72], but the answer to RQ 4 is still largely open. To proceed, I plan to start with exploring extensions of my previous work on combinatorial voting [105, 60] to general partial preferences.

Considerations in Statistics. To understand statistical properties of mechanisms for partial preferences, we need natural parametric models that generate partial preferences to model agents' preferences. However, there are only a few parametric models that generate pairwise comparisons, e.g. the Bradley-Terry-Luce model [20, 65], which are very special cases of partial preferences. The main challenge is the lack of natural parametric models for general partial preferences.

Research Question 5. *What are natural statistical models for partial preferences?*

I plan to start with extending existing models that generate linear orders or weak orders to partial preferences. Recently myself and colleages extended Mallows model and Condorcet's model to partial preferences, and investigated axiomatic and computational properties of their maximum likelihood estimators [107, 102]. I plan to continue working on other possible extensions of these models as well as other popular parametric models, especially the random utility models.

Considerations in computation. I plan to focus on preference elicitation for partial preferences in this project, because this is a natural first question when agents have a different type of preferences from that of classical social choice. Investigating other important topics in computational social choice for partial preferences are natural subjects for future research.

Research Question 6. *How can we design efficient elicitation algorithms for mechanisms for partial preferences?*

When no probabilistic information is available, we are interested in robust elicitation algorithms based on minimax-regret principle as done by Boutilier [18] in a series of work for linear orders. When probabilistic information is available, especially given a parametric model for partial preferences in RQ 5, we can follow the information maximization principle recently explored by myself and colleagues [11]. I also plan to characterize the communication complexity of mechanisms developed in RQ 4, as Conitzer and Sandholm [31] did for linear orders.

3.3 Statistical and Machine Learning Frameworks for New Mechanisms

In this subsection, I will outline research questions for building two principled and systematic frameworks to discover new application-specific social choice mechanisms based on statistical decision theory and machine learning, respectively.

3.3.1 A framework based on statistical decision theory

Since it is impossible to directly design a mechanism with designated economic, statistical, and computational properties, a more realistic and practical approach is to establish a general and statistically justified framework, then discover new mechanisms using the framework and analyze their axiomatic and computational properties.

A first natural idea is to use a standard statistical estimator, especially a maximum likelihood estimator, as the social choice mechanism. However, standard statistical estimators are applicable only if the parameter space of the model is the same as the designated outcome space of the mechanism, which is seldom the case for new social choice applications.

For example, suppose we know that agents' preferences are generated by the Mallows model, where the parameter space consists of all linear orders over alternatives, and we want to design a new social choice mechanism to select a single alternative as the winner. In this case we cannot simply use an maximum likelihood estimator because its outcome is never an alternative. The naïve approach of choosing the top-ranked alternative in the linear order with the maximum likelihood has been criticized by Young [109].

For the very specific case where the parameter space consists of rankings and the outcome space consists of alternatives,Young [109] proposed to take a Bayesian approach and choose the alternative with the largest marginal posterior probability to be ranked in the top. However, for general cases it is not clear how such marginal probability is defined. Specifically, when agents' preferences are generated by a random utility model, where the parameter space is continuous, Young's idea does not work anymore. Hence, we have the following question.

Research Question 7. *How can we discover mechanisms with a statistical model where the parameter space is different from the designated outcome space?*

I plan to start with a novel statistical decision theory framework recently proposed by myself and colleagues in an ongoing work [13], where we introduced a loss function to connect the parameter space and the outcome space. More precisely, a *statistical decision theory framework for social choice* is a tuple $\mathcal{F} = (\mathcal{M}, \mathcal{O}, L)$, where $\mathcal{M}$ is a parametric model, $\mathcal{O}$ is the designated outcome space of the social choice mechanism, and $L : \Theta \times \mathcal{O} \to \mathbb{R}$ is a loss function.

Given a statistical decision theory framework for social choice, we can choose a mechanism to minimize *frequentist expected loss* or *Bayesian expected loss* [13]. We showed that this framework is very flexible, because it naturally generalizes many previous approaches including maximum likelihood estimators and Young's marginal probability approach [109, 87, 42]. We gave examples of new mechanisms obtained under this framework and showed that they satisfy unique combinations of axiomatic, statistical, and computational properties. Therefore, I believe that this is a fruitful and promising framework, and is worth pursuing further.

Research Question 8. *What are the axiomatic, statistical, and computational properties of the new mechanisms obtained in the statistical decision theory framework for social choice?*

There are many natural configurations of the framework I plan to explore, each uniquely defines a new social choice mechanism. To name a few, the parametric model can be Mallows model or any random utility model; the outcome can be a single alternative, a set of alternative with fixed/unfixed size, or a ranking over alternatives; the loss function can be the 0-1 loss function or other smooth loss functions; and the new mechanism can be the minimizer of the frequentist expected loss or the Bayesian expected loss.

3.3.2 A framework based on machine learning

Most existing social choice mechanisms were designed manually. Therefore, it is natural to ask whether we can use machine learning to shift the burden of design from humans to computers.

Social choice mechanisms can be naturally viewed as classifiers that take a profile as input and output an element in the outcome space. Taking a supervised learning approach, Procaccia et al. [86] assumed that the training data consist of multiple (P, o) pairs, where P is a profile and o is the "correct" outcome, and characterized the sample complexity of learning positional scoring rules. A natural question is whether we can learn a different class of social choice mechanisms. For example, given the training data, we may want to learn a mechanism that satisfies Condorcet criterion, which is not satisfied by any positional scoring rule.

The common approach in supervised learning is to compute an optimal classifier to minimize an objective function, which is often chosen to be the *empirical risk*—the probability to misclassify the training data, or *structural risk*—the empirical risk plus has a regularization term to control the complexity of the classifier and prevent overfitting. However, our problem of learning a social choice mechanism with designated axiomatic properties cannot be tackled by existing supervised learning techniques, because none of the satisfiability of axioms can be modeled as a regularization term in the objective function. Therefore, our main challenge is the following.

Research Question 9. *How can we learn a social choice mechanism that satisfies designated axioms?*

I plan to explore two directions discussed in a recent position paper of mine [95]. First, we can learn a mechanism from a general hypothesis class of social choice mechanisms, and then analyze its axiomatic properties. Due to the elegant mathematical structures of the *generalized scoring rules* we introduced earlier [99] and their close relationship to machine learning [96], I believe that generalized scoring rules are promising candidates for the hypothesis class.

The second direction is motivated by structure risk minimization. I plan to incorporate approximate satisfiability of axioms into the data and the objective function. The high-level idea is to automatically generate new training data as the result of enforcing axioms, then add a penalty term to the objective function for misclassifications of such new data. This is related to but technically significantly different from the work by Dütting et al. [40] on automated mechanism design, where the objective is to learn the payment rule that makes the mechanism approximately strategy-proof.

As an example of the second direction, for the Condorcet criterion, we shall add some pairs (P, c) as positive examples to the training data, where P is a profile with the Condorcet winner c; and each misclassification of such an example contributes $\alpha > 0$ to the objective function. The main challenge is how to effectively generate new training data from given axioms, for which I plan to explore active learning techniques [88].

4 Broader impacts: Open Software Packages and Social Choice Platforms

While most of the proposed research is theoretical, I also have a rigorous plan to develop open software packages and platforms for the proposed parametric models, frameworks, and mechanisms, and test them on synthetic data as well as real-world data.

Datasets: The main source of data is Preflib (www.preflib.org), which offers a comprehensive collection of datasets, including political elections, ratings of figure skaters' performances, and product rating. Agents' preferences are also in rich format, including rankings over all alternatives, ranking over a subset of alternatives (which is a special case of partial preferences), and ratings.

Python Packages: I plan to continue working with students to develop efficient implementations of the proposed research in Python, and make them open online. An evolving version has already been put up on GitHub: https://github.com/zmjjmz/prefpy. This plan works well with RPI's strong support on undergraduate research (see the next section for more details) and the Rensselaer Center for Open Source Software.

Practical Social Choice Platforms: As a direct beneficiary of the proposed research and the Python packages, open online platforms mainly serving non-electoral social choice applications will be built by my students and I in parallel with the Chaum/Hurley Trustworthy Random Sample Elections project. We plan to design a convenient interface for users to express partial preferences, as well as to develop a smart phone application for better preference elicitation (RQ 6). Data collected by these platforms will be made open for research.

Applications: RPI is an ideal place to deploy and test the proposed platforms. Historically, RPI students have a strong spirit of democracy. For example, one of the most important annual events at RPI is the election of the student leader called the *Grand Marshall* since 1866. Students are also highly autonomous in many student organizations. I am currently in touch with RPI's student union to explore possibilities of adopting novel social choice mechanisms, computer-aided preference elicitation systems, and online voting platforms in various activities, for example the election of the Grand Marshall and other decision-making scenarios in student organizations.

5 Broader impacts: Education and Outreach Activities

Supervising Ph.D. students: I enjoy working with students and view it as a vital part of my academic career. I have supported an incoming Ph.D. student Ms. Erika Mackin to attend the 2014 workshop on computational social choice, and helped her prepare and present a poster. I will continue putting a high priority on inspiring and encouraging female and minority students to pursue a research career in academia or industry.

Undergraduate research projects: I have been actively attracting undergraduate students to participate in research in the interdisciplinary area of economics and computation, which is jointly supported by RPI's undergraduate research program. In the past semester, I supervised two talented undergraduate students: Mr. William Schneider, and Mr. Zach Jablons, who took my course on computational social choice and became interested in the topic, to take a first step in extending existing mechanisms to partial preferences (RQ 4) and building Python packages. William just graduated in the summer and will join TripAdvisor soon. Zach will continue working on this project. Another new student Mr. Ethan Gertler will join my research group in the next semester to work on game theoretic aspects of social choice (RQ 1 and 3).

Outreach to K-12 students: I plan to actively participate in the Rensselaer Science Ambassadors program. This program aims at promoting science among K-12 students by selecting RPI undergraduate students as science ambassadors, helping them develop presentational skills, create scientific presentations, and finally give presentations at participating schools. I plan to sponsor one science ambassador and work closely with him/her on presentations and interactions with K-12 students on topics classical and computational social choice. This, I believe, will promote the spirit and practice of democracy to students at early ages.

Curriculum development: In the last year, I have developed and taught a multi-disciplinary graduate-level course titled "Computational Social Choice" to help new Ph.D. students and senior undergraduate students develop general methodologies for independent research in computer science in the context of computational social choice. In the first half of the course I gave lectures on important topics in the field, and the second half was seminar-like: students presented selected papers and actively participated in discussions. A research project was also required for each student.

The course was designed under the principle of *situated learning* [63], by putting students in simulated scenarios of doing research. I pretended to be their academic supervisors, and setup regular individual meetings to go through the following common lifecycle in computer science research: literature review, critical thinking, brainstorming, developing ideas and results, paper writing, and presentations. This differs from a common course-work project not only on the level of engagement of the instructor, but also, more importantly, on the instructor's mindset of guiding students to discover and solve new research problems in the long run. Students were mostly evaluated by their performance and efforts as researchers rather than mere learners.

The course attracted both graduate and senior undergraduate students from various departments including CS, economics, ISE, and math, and I am delighted and encouraged by a high average overall score of 4.75/5.0. Some students continued working on their projects after the class towards publications. I plan to teach the course in the next 5 years and further improve it along the direction of situated learning. For example, in the next offering of the course in 2014 fall, I will integrate the review experience into the course by applying a novel peer review mechanism [93] combined with a truncated-mean mechanism [47] for students to evaluate each other's work.

I am planning to develop two undergraduate courses. One is an introductory junior-level course on economics and computation, which aligns well with the proposed undergraduate research and outreach activities. The other is a senior-level course called "Computational Social Processes" that I will develop jointly with Prof. Anshelevich and Prof. Magdon-Ismail, to offer an introduction to the computational and algorithmic aspects of social processes. The course will cover a wide range of topics to illustrate the diverse challenges in algorithmic social process analysis including social networks and social choice theory.

Professional activities: I am currently co-charing the 8th multidisciplinary workshop on preference handling (M-PREF-14), and will co-chair the 2015 conference on Auctions, Market Mechanisms and Their Applications (AMMA). I will continue contributing to organizational activities in this area, including submitting a proposal for M-PREF-15 and actively participating in the organization of the computational social choice workshop in 2016, which is one of the most important venues in the community. I have also given tutorials on computational social choice at ACM EC-12, IJCAI-13, and WINE-13, and a winter school in Singapore in 2013. I am developing a new tutorial on truth-revealing social choice for AAAI-15. These outreach activities provide opportunities to inspire and encourage new researchers to the field.

6 Collaborations, Evaluation, and Timeline

Collaborations within RPI: I plan to collaborate with Prof. Elliot Anshelevich on game-theoretic aspects (RQ 3), Prof. Malik Magdon-Ismail on the machine learning approach (RQ 9), Prof. Sibel Adali on modeling heterogenous and correlated agents using social networks (RQ 1), Prof. Mukkai Krishnamoorthy and Prof. Boleslaw Szymanski on extending basic social choice theory to partial preferences (RQ 4), and Prof. Bulent Yener on security concerns in the social choice platforms (Section 4).

International Collaborations: I have collaborated frequently with Prof. Jérôme Lang (Paris Dauphine University) on a wide range of topics in combinatorial voting [104, 105, 106, 103, 33, 107, 108, 35, 62, 61], which are important subcases of partial preferences, including a book chapter in progress [61]. I was offered a paid visiting professor position at University of Paris Dauphine hosted by Prof. Jérôme Lang and Prof. Jérôme Monnot (the position was typically offered to established researchers in the field), and plan to spend one month in Paris in 2015 to collaborate with them on a wide range of topics especially on social choice for partial preferences in Section 3.2.

I have also collaborated frequently with Prof. Toby Walsh (UNSW and NICTA) on a wide range of topics in computational social choice [36, 73, 91, 74]. I plan to visit Prof. Toby Walsh for one month in 2016 to continue our collaborations especially on RQ 6. Since Prof. Toby Walsh is the creator and manager of the Preflib datasets, I also expect to collaborate with him on the Python packages and social choice platforms proposed in Section 4.

Evaluation. The proposed research, if successful, will result in multiple publications at prestigious peer-reviewed conferences in computer science, especially artificial intelligence, multi-agent systems, and machine learning, including AAAI, AAMAS, ACM EC, CP, ICML, IJCAI, KR, NIPS, and UAI. I also plan to publish papers at prestigious journals in artificial intelligence and economics, including Artificial Intelligence Journal, Journal of AI Research, ACM Transactions on Economics and Computation, Social Choice and Welfare, and Journal of Economic Research.

In the long run, the success of the proposed research will be evaluated by the social choice mechanisms discovered, evaluated, and deployed using the proposed principles and frameworks. More precisely, successful research in Section 3.1 will result in (1) a list of existing social choice mechanism where the Condorcet Jury Theorem can be extended (for RQ 1), (2) frequentist risk and Bayesian risk of these mechanisms w.r.t. popular parametric models (for RQ 2), and (3) the price of sophistication of these mechanisms (for RQ 3). Successful research in Section 3.2 will lead to (1) formally-defined extensions of axioms, extensions of social choice mechanisms, and their satisfiability w.r.t. the extended axioms for partial preferences (for RQ 4), (2) new parametric models for partial preferences (for RQ 5), and (3) efficient preference elicitation algorithms (for RQ 6). Successful research in Section 3.3 will lead to two principled and systematic frameworks to discover new mechanisms, and a comprehensive table of axiomatic, statistical, and computational properties of some new mechanisms obtained in the new frameworks (for RQ 7–9).

Timeline. In the first three years I will mainly focus on truth-revealing social choice in Section 3.1 plus extensions of existing social choice mechanisms and axioms to partial preferences (RQ 4) and new statistical models for partial preferences (RQ 5), because these are the basis for other research questions. In the 3rd, 4th, and 5th years I plan to focus on computational social choice for partial preferences, especially on elicitation algorithms (RQ 6) and frameworks for new mechanisms discussed in Section 3.3. A brief timeline is listed in Table 1.

Months 0 -12: Extension of Condorcet Jury Theorem to > 2 alternatives (RQ 1)
Months 13 - 36: Measuring the truth-revealing ability (RQ 2)
Months 13 - 36: Measuring the effect of agents' strategic behavior (RQ 3)
Months 0 - 24: Extending mechanisms and axioms to partial preferences (RQ 4)
Months 0 - 36: Statistical models for partial preferences (RQ 5)
Months 25 - 60: Elicitation algorithms for partial preferences (RQ 6)
Months 25 - 60: New mechanisms for parameter space $\neq$ outcome space (RQ 7)
Months 25 - 60: A framework based on statistical decision theoretic (RQ 8)
Months 25 - 60: Learning mechanisms with designated axioms (RQ 9)

Table 1: A brief timeline of proposed research.

Prior NSF Support. In 2011–2013 I was supported by NSF under Grant #1136996 to the Computing Research Association for the CIFellows Project, for conducting postdoctoral research at Harvard University. Together with colleagues, I have made significant progress in computational social choice and machine learning. The work was published at top-tier conferences in AI and machine learning, including AAAI, AAMAS, ACM EC, CP, ICML, KR, NIPS, UAI. Among other things, I took an important step forward towards better fraud detections in elections by studying the computation of the margin of victory [94]; we obtained new results in using computational complexity to protect elections [79, 91]; we provided approximation and constraint-satisfaction viewpoints of combinatorial voting [32, 62]; and we designed several new computational methods to tackle statistical inferences for general random utility models, which were not known to be computationally tractable [9, 11, 10, 12].

References

[1] https://www.cs.duke.edu/people/profiles/individual/2009_lirong_xia/, 2009.

[2] http://econ.duke.edu/news/archive/2010/12/09/once-master-s-students-now-ph-d-job-market-candidates-this-year, 2010.

[3] http://www.seas.harvard.edu/news/2012/11/counting-votes-precinct-and-web, 2012.

[4] http://econ.duke.edu/news/archive/2013/12/26/duke-alum-and-rpi-professor-on-the-new-msec-program, 2013.

[5] http://news.rpi.edu/content/2014/04/16/expert-social-choice-lirong-xia-joins-rensselaer-department-computer-science, 2014.

[6] Kenneth Arrow. *Social choice and individual values*. New Haven: Cowles Foundation, 2nd edition, 1963. 1st edition 1951.

[7] Robert J. Aumann. Utility Theory without the Completeness Axiom. *Econometrica*, 30(3): 445–462, 1962.

[8] David Austen-Smith and Jeffrey S. Banks. Information Aggregation, Rationality, and the Condorcet Jury Theorem. *The American Political Science Review*, 90(1):34–45, 1996.

[9] Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. Random utility theory for social choice. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, pages 126–134, Lake Tahoe, NV, USA, 2012.

[10] Hossein Azari Soufiani, William Chen, David C. Parkes, and Lirong Xia. Generalized method-of-moments for rank aggregation. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, Lake Tahoe, NV, USA, 2013.

[11] Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. Preference Elicitation For General Random Utility Models. In *Proceedings of Uncertainty in Artificial Intelligence (UAI)*, Bellevue, Washington, USA, 2013.

[12] Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. Computing Parametric Ranking Models via Rank-Breaking. In *Proceedings of the 31st International Conference on Machine Learning*, Beijing. China, 2014.

[13] Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. *Statistical Decision Theory Approaches to Social Choice*. Working paper, cs.rpi.edu//~xial/Files/SDT.pdf, 2014.

[14] Daniel Berend and Jacob Paroush. When is Condorcet's Jury Theorem valid? *Social Choice and Welfare*, 15(4):481–488, 1998.

[15] Daniel Berend and Luba Sapir. Monotonicity in Condorcet's Jury Theorem with dependent voters. *Social Choice and Welfare*, 28(3):507–528, 2007.

[16] Duncan Black. *The Theory of Committees and Elections*. Cambridge University Press, 1958.

[17] Philip J. Boland, Frank Proschan, and Y. L. Tong. Modelling Dependence in Simple and Indirect Majority Systems. *Journal of Applied Probability*, 26(1):81–88, 1989.

[18] Craig Boutilier. Computational Decision Support: Regret-based Models for Optimization and Preference Elicitation. In P. H. Crowley and T. R. Zentall, editors, *Comparative Decision Making: Analysis and Support Across Disciplines and Applications*. Oxford University Press, 2013.

[19] Craig Boutilier, Ronen Brafman, Carmel Domshlak, Holger Hoos, and David Poole. CP-nets: A tool for representing and reasoning with conditional ceteris paribus statements. *Journal of Artificial Intelligence Research*, 21:135–191, 2004.

[20] Ralph Allan Bradley and Milton E. Terry. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

[21] Steven J. Brams and Peter C. Fishburn. Voting procedures. In Kenneth J. Arrow, A.K. Sen, and Kotaro Suzumura, editors, *Handbook of Social Choice and Welfare*, volume 1, pages 175–236. North Holland, 2002.

[22] Felix Brandt, Vincent Conitzer, and Ulle Endriss. Computational social choice. In G. Weiss, editor, *Multiagent Systems*. MIT Press, 2013.

[23] Ioannis Caragiannis, Ariel Procaccia, and Nisarg Shah. When do noisy votes reveal the truth? In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, Philadelphia, PA, 2013.

[24] David Chaum. Random-Sample Elections. http://rs-elections.com/Random-Sample%20Elections.pdf, 2014.

[25] Yann Chevaleyre, Ulle Endriss, Jérôme Lang, and Nicolas Maudet. A short introduction to computational social choice. In *Proceedings of the 33rd Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM-2007)*, 2007.

[26] Flavio Chierichetti and Jon Kleinberg. Voting with Limited Information and Many Alternatives. In *Proceedings of the Twenty-third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1036–1055, Kyoto, Japan, 2012.

[27] Marquis de Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: L'Imprimerie Royale, 1785.

[28] Vincent Conitzer and Tuomas Sandholm. Complexity of mechanism design. In *Proceedings of the 18th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 103–110, Edmonton, Canada, 2002.

[29] Vincent Conitzer and Tuomas Sandholm. Self-interested automated mechanism design and implications for optimal combinatorial auctions. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, pages 132–141, New York, NY, USA, 2004. Early version appeared as a poster in the ACM Conference on Electronic Commerce, 2003, pp. 232–233.

[30] Vincent Conitzer and Tuomas Sandholm. Common voting rules as maximum likelihood estimators. In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 145–152, Edinburgh, UK, 2005.

[31] Vincent Conitzer and Tuomas Sandholm. Communication complexity of common voting rules. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, pages 78–87, Vancouver, BC, Canada, 2005.

[32] Vincent Conitzer and Lirong Xia. Paradoxes of Multiple Elections: An Approximation Approach. In *Proceedings of the 13th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 179–187, Rome, Italy, 2012.

[33] Vincent Conitzer, Jérôme Lang, and Lirong Xia. How hard is it to control sequential elections via the agenda? In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI)*, pages 103–108, Pasadena, CA, USA, 2009.

[34] Vincent Conitzer, Matthew Rognlie, and Lirong Xia. Preference functions that score rankings and maximum likelihood estimation. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI)*, pages 109–115, Pasadena, CA, USA, 2009.

[35] Vincent Conitzer, Jérôme Lang, and Lirong Xia. Hypercubewise preference aggregation in multi-issue domains. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI)*, Barcelona, Catalonia, Spain, 2011.

[36] Vincent Conitzer, Toby Walsh, and Lirong Xia. Dominating manipulations in voting with partial information. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 638–643, San Francisco, CA, USA, 2011.

[37] Franz Dietrich. The Premises of Condorcet's Jury Theorem Are Not Simultaneously Justified. *Episteme*, 5(1):56–73, 2008.

[38] Franz Dietrich and Christian List. A Model of Jury Decisions Where All Jurors Have the Same Evidence. *Synthese*, 142(2):175–202, 2004.

[39] John Duggan and Cesar Martinelli. A Bayesian Model of Voting in Juries. *Games and Economic Behavior*, 37(2):259–294, 2001.

[40] Paul Dütting, Felix Fischer, Pichayut Jirapinyo, John K. Lai, Benjamin Lubin, and David C. Parkes. Payment Rules Through Discriminant-based Classifiers. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 477–494, Valencia, Spain, 2012.

[41] Cynthia Dwork, Ravi Kumar, Moni Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th World Wide Web Conference*, pages 613–622, 2001.

[42] Edith Elkind and Nisarg Shah. How to Pick the Best Alternative Given Noisy Cyclic Preferences? In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*, 2014.

[43] Timothy Feddersen and Wolfgang Pesendorfer. Voting Behavior and Information Aggregation in Elections With Private Information. *Econometrica*, 65(5):1029–1058, 1997.

[44] Timothy Feddersen and Wolfgang Pesendorfer. Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting. *The American Political Science Review*, 92(1):23–35, 1998.

[45] Timothy Feddersen and Wolfgang Pesendorfer. Elections, information aggregation, and strategic voting. *Proceedings of the National Academy of Sciences*, 96(19):10572—10574, 1999.

[46] Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, October 1991.

[47] Florent Garcin, Lirong Xia, and Boi Faltings. How Aggregators Influence Human Rater Behavior? In *Workshop on Social Computing and User Generated Content*, 2013.

[48] Kerstin Gerlinga, Hans Peter Grünera, Alexandra Kielc, and Elisabeth Schulte. Information acquisition and decision making in committees: A survey. *European Journal of Political Economy*, 21(3):563—597, 2005.

[49] Sumit Ghosh, Manisha Mundhe, Karina Hernandez, and Sandip Sen. Voting for movies: the anatomy of a recommender system. In *Proceedings of the third annual conference on Autonomous Agents*, pages 434–435, 1999.

[50] Bernard Grofman, Guillermo Owen, and Scott L. Feld. Thirteen theorems in search of the truth. *Theory and Decision*, 15(3):261–278, 1983.

[51] Thomas Haynes, Sandip Sen, Neeraj Arora, and Rajani Nadella. An Automated Meeting Scheduling System That Utilizes User Preferences. In *Proceedings of the First International Conference on Autonomous Agents*, pages 308–315, Marina del Rey, California, USA, 1997.

[52] Satoshi Kanazawa. A brief note on a further refinement of the Condorcet Jury Theorem for heterogeneous groups. *Mathematical Social Sciences*, 35(1):69–73, 1998.

[53] John Kemeny. Mathematics without numbers. *Daedalus*, 88:575–591, 1959.

[54] Elias Koutsoupias and Christos Papadimitriou. Worst-case Equilibria. In *Proceedings of the 16th Annual Conference on Theoretical Aspects of Computer Science*, pages 404–413, Trier, Germany, 1999.

[55] Krishna K. Ladha. The Condorcet Jury Theorem, Free Speech, and Correlated Votes. *American Journal of Political Science*, 36(3):617–634, 1992.

[56] Krishna K. Ladha. Condorcet's jury theorem in light of de Finetti's theorem. *Social Choice and Welfare*, 10(1):69–85, 1993.

[57] Krishna K. Ladha. Information Pooling through majority-rule voting: Condorcet's Jury Theorem. *Journal of Economic Behavior and Organization*, 26:353–372, 1995.

[58] Louisa Lam and Ching Y Suen. Majority vote of even and odd experts in a polychotomous choice situation. *Theory and Decision*, 41(1):13–36, 1996.

[59] Jérôme Lang. Vote and aggregation in combinatorial domains with structured preferences. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1366–1371, Hyderabad, India, 2007.

[60] Jérôme Lang and Lirong Xia. Sequential composition of voting rules in multi-issue domains. *Mathematical Social Sciences*, 57(3):304–324, 2009.

[61] Jérôme Lang and Lirong Xia. Voting in Combinatorial Domains. Handbook of Computational Social Choice, chapter 9. https://sites.google.com/site/handbookcomsoc/chapters, 2014.

[62] Jérôme Lang, Jérôme Mengin, and Lirong Xia. Aggregating conditionally lexicographic preferences on multi-issue domains. In *Proceedings of the 18th International Conference on Principles and Practice of Constraint Programming (CP)*, pages 973–987, 2012.

[63] Jean Lave and Etienne Wenger. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, 1991.

[64] Tie-Yan Liu. *Learning to Rank for Information Retrieval*. Springer, 2011.

[65] Robert Duncan Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.

[66] Colin L. Mallows. Non-null ranking model. *Biometrika*, 44(1/2):114–130, 1957.

[67] Andrew Mao, Ariel D. Procaccia, and Yiling Chen. Better human computation through principled voting. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, Bellevue, WA, USA, 2013.

[68] Daniel McFadden. Conditional logit analysis of qualitative choice behavior. In *Frontiers of Econometrics*, pages 105–142, New York, NY, 1974. Academic Press.

[69] Andrew McLennan. Consequences of the Condorcet Jury Theorem for Beneficial Information Aggregation by Rational Agents. *The American Political Science Review*, 92(2): 413–418, 1998.

[70] Nicholas R. Miller. Information, Electorates, and Democracy: Some Extensions and Interpretations of the Condorcet Jury Theorem. In Grofman B. and Owen G., editors, *Information Pooling and Group Decision Making*, pages 173—192. JAI Press, 1986.

[71] Roger B. Myerson. Extended Poisson Games and the Condorcet Jury Theorem. *Games and Economic Behavior*, 25(1):111–131, 1998.

[72] Nina Narodytska and Toby Walsh. The Computational Impact of Partial Votes on Strategic Voting. In *Proceedings of the 21st European Conference on Artificial Intelligence*, 2014.

[73] Nina Narodytska, Toby Walsh, and Lirong Xia. Manipulation of Nanson's and Baldwin's rule. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, San Francisco, CA, USA, 2011.

[74] Nina Narodytska, Toby Walsh, and Lirong Xia. Combining voting rules together. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI)*, pages 612–617, 2012.

[75] John Nash. Non-cooperative games. *Annals of Mathematics*, 54:289–295, 1951.

[76] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, New York, NY, USA, 2007.

[77] Shmuel Nitzan and Jacob Paroush. The significance of independent decisions in uncertain dichotomous choice situations. *Theory and Decision*, 17(1):47–60, 1984.

[78] Guillermo Owen, Bernard Grofman, and Scott L. Feld. Proving a distribution-free generalization of the Condorcet Jury Theorem. *Mathematical Social Sciences*, 17(1):1–16, 1989.

[79] David C. Parkes and Lirong Xia. A Complexity-of-Strategic-Behavior Comparison between Schulze's Rule and Ranked Pairs. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 1429–1435, Toronto, Canada, 2012.

[80] Jacob Paroush. Stay away from fair coins: A Condorcet jury theorem. *Social Choice and Welfare*, 15(1):15–20, 1998.

[81] Judea Pearl. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann Publishers Inc., 1988.

[82] Bezalel Peleg and Shmuel Zamir. Extending the Condorcet Jury Theorem to a general dependent jury. *Social Choice and Welfare*, 39(1):91–125, 2012.

[83] Maria Silvia Pini, Francesca Rossi, Kristen Brent Venable, and Toby Walsh. Aggregating partially ordered preferences. *Journal of Logic and Computation*, 19:475–502, 2009.

[84] Robin L. Plackett. The analysis of permutations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 24(2):193–202, 1975.

[85] Giorgio Dalla Pozza, Maria Silvia Pini, Francesca Rossi, and K. Brent Venable. Multi-agent soft constraint aggregation via sequential voting. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, pages 172–177, Barcelona, Catalonia, Spain, 2011.

[86] Ariel D. Procaccia, Aviv Zohar, Yoni Peleg, and Jeffrey S. Rosenschein. The learnability of voting rules. *Artificial Intelligence*, 173:1133—1149, 2009.

[87] Ariel D. Procaccia, Sashank J. Reddi, and Nisarg Shah. A maximum likelihood approach for selecting sets of alternatives. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*, 2012.

[88] Burr Settles. *Active Learning*. Morgan & Claypool Publishers, 2012.

[89] Lloyd Shapley and Bernard Grofman. Optimizing group judgmental accuracy in the presence of interdependencies. *Public Choice*, 43(3):329–343, 1984.

[90] Louis Leon Thurstone. A law of comparative judgement. *Psychological Review*, 34(4): 273–286, 1927.

[91] Toby Walsh and Lirong Xia. Lot-based voting rules. In *Proceedings of the Eleventh International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 603–610, Valencia, Spain, 2012.

[92] Jörgen Wit. Rational Choice and the Condorcet Jury Theorem. *Games and Economic Behavior*, 22(2):364—376, 1998.

[93] Jens Witkowski and David C. Parkes. Peer Prediction Without a Common Prior. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 964–981, Valencia, Spain, 2012.

[94] Lirong Xia. Computing the margin of victory for various voting rules. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, pages 982–999, Valencia, Spain, 2012.

[95] Lirong Xia. Designing social choice mechanisms using machine learning. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pages 471–474, St. Paul, MN, USA, 2013.

[96] Lirong Xia. Generalized Scoring Rules: A Framework That Reconciles Borda and Condorcet. *ACM SIGecom Exchanges*, 12(1):42–48, 2013.

[97] Lirong Xia. Assigning indivisible and categorized items. To be presented at ISAIM-14 special session on Computational Social Choice, 2014.

[98] Lirong Xia. Statistical Properties of Social Choice Mechanisms. In *Proceedings of the Fifth International Workshop on Computational Social Choice*, Pittsburgh, Pennsylvania, USA, 2014.

[99] Lirong Xia and Vincent Conitzer. Generalized scoring rules and the frequency of coalitional manipulability. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, pages 109–118, Chicago, IL, USA, 2008.

[100] Lirong Xia and Vincent Conitzer. Stackelberg voting games: Computational aspects and paradoxes. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 921–926, Atlanta, GA, USA, 2010.

[101] Lirong Xia and Vincent Conitzer. Strategy-proof voting rules over multi-issue domains with restricted preferences. In *Proceedings of the Sixth Workshop on Internet and Network Economics (WINE)*, pages 402–414, Stanford, CA, USA, 2010.

[102] Lirong Xia and Vincent Conitzer. A maximum likelihood approach towards aggregating partial orders. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI)*, pages 446–451, Barcelona, Catalonia, Spain, 2011.

[103] Lirong Xia and Jérôme Lang. A dichotomy theorem on the existence of efficient or neutral sequential voting correspondences. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI)*, pages 342–347, Pasadena, CA, USA, 2009.

[104] Lirong Xia, Jérôme Lang, and Mingsheng Ying. Sequential voting rules and multiple elections paradoxes. In *Theoretical Aspects of Rationality and Knowledge (TARK)*, pages 279–288, Brussels, Belgium, 2007.

[105] Lirong Xia, Jérôme Lang, and Mingsheng Ying. Strongly decomposable voting rules on multiattribute domains. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 776–781, Vancouver, BC, Canada, 2007.

[106] Lirong Xia, Vincent Conitzer, and Jérôme Lang. Voting on multiattribute domains with cyclic preferential dependencies. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 202–207, Chicago, IL, USA, 2008.

[107] Lirong Xia, Vincent Conitzer, and Jérôme Lang. Aggregating preferences in multi-issue domains by using maximum likelihood estimators. In *Proceedings of the Ninth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 399–406, Toronto, Canada, 2010.

[108] Lirong Xia, Vincent Conitzer, and Jérôme Lang. Strategic sequential voting in multi-issue domains and multiple-election paradoxes. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, pages 179–188, San Jose, CA, USA, 2011.

[109] H. Peyton Young. Condorcet's theory of voting. *American Political Science Review*, 82: 1231–1244, 1988.

Facilities, Equipment, and Other Resources

RPI provides facilities that support regular everyday work, including high-speed wifi, library space, shared PCs. RPI library has access to articles in many major journals.

RPI computer science department offers major software packages including various scientific computing softwares. The PI has his own office space. Graduate students share offices.

Data Management Plan

The data management plan is prepared under the guidance of the CISE Directorate (http://www.nsf.gov/cise/cise_dmp.jsp). I will address the following two main questions:

1. What data are generated by the project?
2. What is the plan for managing the data?

The main data products of the projects are:

- publications based on research performed in this project;
- software written by project team members;
- experimental data collected by the proposed online open social choice platforms (see Section 4).
- metadata such as study designs, coding schemes for protocols, and software for analyzing the experimental data.

Publications. Technical reports, posters, and publications resulting from this project will be placed on my homepage and will be available for public download, which will be made valid for at least five years after the end of the project.

Software. The Python packages and social choice systems/platforms and their descriptions (see Section 4) will be made open online on my homepage as well as GitHub (https://github.com/) or Source Forge (http://sourceforge.net/). An evolving Python package for research proposed in Section 3.2 is already online at GitHub (https://github.com/zmjjmz/prefpy). These will be made valid for at least five years after the end of the project.

Experimental Data. The proposed research does not require new data. The main source of data is an open dataset called Preflib (www.preflib.org, see Section 4 for more details).

Since the proposed online open social choice platforms (Section 4) will provide free services on online preference aggregation, we expect to obtain new data on users' preferences. These data will be anonymized, stored in standard formats, and made available at my homepage and Preflib. Another potential source of data is applications of the proposed platforms at various organizations for elections, e.g. RPI student leader elections. For such case, the data will be handled in accordance with the agreements of individual organizations. We will aggressively pursue sharing as much of the relevant details of the data as possible and made them online at my homepage and Preflib. These will be made valid for at least five years after the end of the project.

I have already initiated an application for Institutional Review Board (IRB) approval and expect that many of the studies in the proposed project will be granted IRB approval under "Exempt" status, because we do not expect to store any personally identifying information on the subjects.

Experimental Metadata. All experimental metadata, that is, all data describing the experimental studies—coding schemes, use cases, details of which software versions were used, and so on—sufficient to permit replication of the work, will be treated just as experimental data. Any software written to analyze the experimental data will be treated in the same way as all the other software developed by the project.