Developing and Coordinating Autonomous Agents for Efficient Electricity Markets

by

Andrew Raymond Trueman Perrault

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Department of Computer Science
University of Toronto

Abstract

Developing and Coordinating Autonomous Agents for Efficient Electricity Markets

Andrew Raymond Trueman Perrault
Doctor of Philosophy
Department of Computer Science
University of Toronto
2018

Whether for environmental, conservation, efficiency, or economic reasons, developing next generation electric power infrastructure is critical. Temporally relevant, granular data from smart meters provide new opportunities for data-driven management of the power grid. New developments—for example, electricity markets with multiple suppliers, the integration of renewable power sources into the system, and spikier demand patterns due to, say, electric vehicles—create new challenges for efficient grid operation. Computer science is uniquely positioned to assist with increasingly sophisticated techniques for handling and learning from large amounts of data. The methods of game theory and multi-agent systems provide a natural framework for modeling the competing incentives of electricity market participants. This thesis focuses on the use of learning, optimization, mechanism design, and preference elicitation methods to coordinate electricity demand and supply while respecting the incentives of market participants. Specifically, we propose an approach where an autonomous agent acts on behalf of each household, coordinating with inhabitants to relay information and make decisions on their behalf about electricity consumption. We focus on three problems that arise in developing such agents: (i) how to coordinate consumers' electricity use, (ii) how to share the costs of consumption among households (via their agents), and (iii) how to gather consumption preference data from consumers.

Chapters 3 and 4 focus on different aspects of the first two problems. Both use a matching markets approach. In Chapter 3, we focus on the impact of demand smoothness and peaks on the supplier's cost, and in Chapter 4, on the impact of predictability. In both chapters, we develop new cost sharing schemes that are resilient to certain forms of strategic behavior on the part of the agents and that achieve strong performance in experiments.

Chapter 5 studies the third problem. Motivated by control of heating and cooling systems, we present a new approach to preference elicitation, where the cost and accuracy of query responses is dependent on the user's familiarity with the conditions specified in the query. We show that despite the theoretical difficulty in this setting, we can build solvers that perform well in practice.

Acknowledgements

Thank you to my supervisor, Craig Boutilier, for inspiring my research and guiding an abstract vision into something concrete.

To my committee, for their invaluable "on the ground" support at tough moments, as well as their outside perspective.

To my parents, for endless hours of reading and editing.

To my colleagues, collaborators and friends, especially Joanna Drummond, Elizabeth Greville, Marek Janicki, Aleksandr Kazachkov, Omer Lev, Nisarg Shah, Jake Snell, Tyrone Strangway, and Rory Tulk.

Contents

Chapter 1

Introduction

Electricity markets and electricity distribution are the source of many intriguing technical problems for artificial intelligence (AI). Electricity is notable from an economic perspective because it fails to behave like an ideal commodity [Kirschen and Strbac, 2004]. It is delivered through a physical network governed by the laws of physics: the electrical *grid*. Demand and supply must be balanced continuously to keep the grid functioning (*load balancing*)—failure results in a long and expensive recovery process. Most load balancing is done outside of markets because of the speed of reaction required, on the order of seconds. Because supply is pooled through the grid, individuals cannot buy from a particular supplier directly. Storage is also very expensive. These elements combine to make electricity markets challenging to operate efficiently, in the economic sense. Our research and intuitions about how free markets lead to high-quality outcomes are difficult to apply in a setting that is sufficiently non-standard.

The Nobel laureate Alvin Roth has written extensively on the challenge of defining good rules for interactions in markets [Roth, 2002]. He argues that no market operates completely removed from human design or conventions, and that it should be part of the role of economists in society to adjust the rules of markets to optimize outcomes. This area of study is called *market design* or *design economics*. While designing these rules is highly dependent on economic theory, there are often key computational challenges in the operation of the market itself. *Designed markets* frequently do away with the spontaneity of *organic* ones where market participants choose their own partner with whom to interact. Instead, participants submit "bids" (which may be non-monetary) to a central authority, which "assigns" transactions to the market participants by running an algorithm. Such markets are referred to as *matching markets*, since they often match buyers to sellers to maximize economic efficiency or some related objective. More broadly, they coordinate the actions of the market participants.

Electricity markets in particular are intensely designed. This design is driven by technical constraints, such as limitations in monitoring technology. For example, with traditional electricity meters, it is impossible to have a pricing scheme that depends on the time of consumption, as these meters do not record when electricity use occurs, only total usage. These markets are an interesting object of study, and Chapters 3 and 4 explore the problem of improving their efficiency from both the economic and computational perspectives.

As we propose more complex market rules that increase efficiency, we have to be careful to not disrupt the desirable properties of existing, simple markets. Here two key issues arise. First, the more information an agent reports to the market, the more opportunity the agent has to manipulate the

outcome of the market by reporting information falsely, which we call *misreporting*. Second, what is good for the group of agents is not necessarily good for each individual agent—a rational, self-interested agent will take advantage of opportunities to improve its own outcome, potentially by making *side deals* with other agents. These actions may in turn reduce the welfare of the group. Economists call this the problem of *stability*. A stable allocation is one where no agent can benefit by making side deals or taking actions other than those prescribed by the market. Both of these issues are well understood in the abstract, but often require special treatment in any specific market.

In Chapter 3, we apply a matching markets perspective to electricity exchange. Our use of matching markets is motivated by their application to supply chains (e.g., Chen and Roma [2010]), where they are used to increase economic efficiency while respecting global constraints on outcomes. Electricity generator cost functions often have a complex structure, including features such as minimum and maximum production levels, multiple generation sources with varying costs, and ramp constraints that constrain production adjustments over time. Due to this complexity, *social welfare* optimal matchings (i.e., those that maximize the aggregate utility of the market participants) are usually not stable. We develop two novel cost-sharing schemes. The first is based on a standard solution concept in cooperative games, *Shapley value*, and achieves the positive fairness properties of the Shapley value, but is computationally expensive in domains with thousands of agents. The second is based on a new notion of *similarity-based envy freeness*, which captures many essential properties of Shapley pricing, but scales to large numbers of consumers. We show that while both of these schemes achieve a high degree of stability in practice, we can increase the stability of the outcomes by sacrificing small amounts of social welfare.

Chapter 4 develops a cooperative game-theoretic mechanism for coordinating electricity use across consumers, particularly with respect to *predictability* of consumption. There is an incentive misalignment between the traditional fixed-rate tariffs faced by consumers and the two-stage market faced by electricity suppliers. Electricity suppliers try to purchase as much electricity as they can in advance because they pay a lower rate for such purchases [Team, 2011]. The cost incurred by energy suppliers thus depends on their ability to predict future consumption, but consumers typically have little incentive to consume predictably. We extend the work of Robu et al. [2017], who developed *prediction-of-use (POU) games* in consultation with a utility to address this problem. We show that POU games have significant shortcomings because they do not allow consumers to coordinate their consumption choices, resulting in outcomes with lower social welfare than under a fixed-rate tariff. We extend the model by introducing multiple profiles, addressing this weakness, and we call our extension *multiple-profile prediction-of-use (MPOU) games*. The introduction of multiple profiles per user creates new incentive issues that we address with a new technique called *separating functions*, which are a generic solution to the problem of incentivizing a user to take an action which has a partially observable outcome. In our experimental results, we build test instances using utility models learned from real electricity use data. MPOU games deliver an increase of about 5% in social welfare over fixed-rate tariffs.

While market design may not appear at first to be a natural topic for AI research, there are reasons why AI has embraced it more than have other areas of computer science. The main reason is AI's interest in multi-agent interaction, particularly between humans and artificial autonomous agents or between groups of artificial agents. Modeling those cooperative and competitive interactions is close in spirit to economic exchanges. In addition, AI is a center for research on problems with exponential and higher complexity. Many of the core problems studied in the early days of AI, intended to mimic human

capabilities, are highly challenging computationally.[1] This thesis connects market design with AI even more directly, through the use of human-representing autonomous agents in electricity interactions.

1.1 The Role of Artificial Autonomous Agents

It becomes intuitively obvious when studying market design in the electricity context that many improvements are in conflict with consumers' attentional limits. To increase efficiency, one needs to make interactions more complex. For example, in the presence of perfectly rational agents with infinite attention, dynamic electricity pricing would be more efficient than fixed-rate pricing from an economic perspective—dynamic pricing allows the balance of supply and demand to affect prices, resulting in a better allocation. However, in practice, dynamic prices are hard for consumers to manage. They need to pay attention to price changes throughout the day and shift their behavior to react to them. Because electricity makes up a small fraction of most people's spending, their incentive to shift in reaction to dynamic prices is quite limited.

We can address this problem by introducing autonomous agents that represent consumers. If each consumer has an agent that understands her preferences and can take actions on her behalf, we can reap the benefits of more sophisticated market mechanisms without paying the attentional costs. Introducing autonomous agents has an additional advantage. Even if consumers had infinite attention, they would also require complete knowledge of their preferences to respond rationally (in the economic sense) to incentives, as well as the computational capacity to optimize their behavior. The introduction of a consumer-representing autonomous agent allows us to break this problem into two parts: market design with rational agents (the focus of Chapters 3 and 4) and how an autonomous agent can learn the kind of preference information it needs from humans (who may not be rational). The latter is the focus of Chapter 5.

When learning consumer preferences, it is not sufficient to observe them and reproduce their behavior. Doing so would reproduce the attentional constraints of people in the learned behavior of agents. Instead, we need to query users directly. The problem of learning user preferences through querying is *preference elicitation (PE)*, a longstanding problem in AI. PE is connected to market design because the central market mechanism can be viewed as an autonomous agent that is acting on behalf of the market participants (see, for example, Sandholm and Boutilier [2006] and Parkes [2005]).

1.2 Experiential Elicitation

Research in PE has primarily focused on making infrequent decisions with high impact, but studying frequent, low-impact decisions raises new questions. Beyond the new technical challenge that arises from the problem structure, the psychology of the user changes. Dual process theory [Kahneman, 2011] states that humans have two separate decision-making systems: one for fast and frequent decisions (*System 1*) and another for slow and infrequent ones (*System 2*). This split has significant consequences for PE because people are less able to reason about their System 1 decisions because logical reasoning is a part of System 2.

[1] Market design has drawn in another field that specializes in hard computational problems: operations research, where the motivation came from optimizing industrial processes rather than mimicking human ones. Market design has also had interactions with the theoretical CS community, expert at proving properties of algorithms, and this has created overlap between research in theoretical CS and AI.

Systems that support System 1 decisions should take into account the *context* in which queries are asked. For example, people should have better access to their heating and cooling preferences about "days that 20°C and the electricity price is \$0.1 per kWh" on a day where the temperature is in fact 20°C and the price is \$0.1 per kWh than on a day when it is 0°C and the price is \$0.01 per kWh. In the former case, a person can appeal directly to System 1 to figure out what decision they would make, whereas in the latter, they are appealing to System 2 to reason about System 1. Similarly, in the case of the smartphone personal assistant, it is easier to answer a question about the relevance of being shown certain information in the current context than it is to consider the value of that information in a hypothetical scenario. At a high level, this hypothesis is no more complicated than the idea that people can answer questions about their immediate circumstances both more easily and accurately than they can answer hypothetical questions.

Integrating this hypothesis into PE systems requires that the systems take into account the current context when deciding what queries to ask. We call this approach *experiential elicitation (EE)* after Hui and Boutilier [2008], who apply a similar concept in an interface customization setting. EE presents new challenges to PE because of its temporal component. Traditional PE systems have a discrete querying phase, where the system does all of its querying before any control actions are taken. This approach does not work in EE because it may be better to wait to ask a query until a more appropriate context is reached. Thus, the problem of what control actions to take on behalf of the user and what queries to ask are interlinked. In extreme cases, it may even be optimal to take a control action that is believed to be suboptimal because it opens up the possibility of posing more useful queries.

It is important to note that EE can be useful even in the case where users can answer any query equally well from any context because it can reason about the value of delaying queries. First, delaying a querying about a particular context can result in a cost savings if that context never occurs. Second, if asking each query has a cost and people discount costs paid in the future, it is better to ask a query later than it is to ask it now. For example, suppose a user's heating and cooling preferences vary seasonally. The optimal querying policy is to delay learning about the user's preferences in each season until that season arrives.

The main contribution of Chapter 5 is to develop and study EE. We present an abstract model of EE and situate it relative to other important problems in AI. We show that our model of EE is reducible to a partially observable Markov decision process, but the reduction is extremely expensive for the kinds of query response and query cost functions that interest us (these functions determine the accuracy and cost of queries depending on the current state). We then develop EE for the home heating and cooling case, introduce a new query type that we find is particularly suited to that setting: the *relative value* query, which asks users to compare their reward between two states. We develop a *Gaussian process (GP)* based preference aggregation approach and show that GPs have natural synergies with relative value queries. We test the approach in a synthetic environment and show that it achieves higher user utility than natural baselines.

1.3 Why Study Electricity in AI?

Before continuing with the technical narrative of the thesis, we situate our research into the context of the transformation that is occurring in electricity generation, distribution and consumption systems.

Driven by the threat of climate change, many governments have instituted greenhouse gas (GHG)

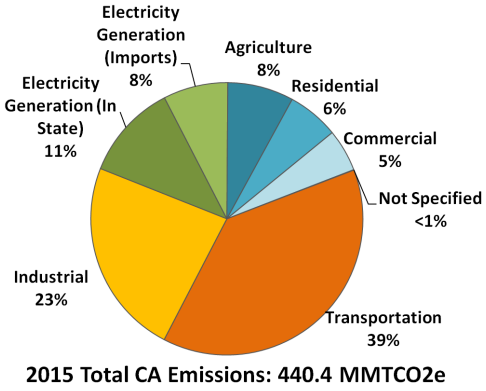


Figure 1.1: California GHG emissions by sector. Prepared by the California Air Resources Board: `https://www.arb.ca.gov/cc/inventory/data/data.htm`. MMTCO2e means million metric tons of carbon dioxide equivalent, i.e., GHG with atmospheric impact equal to that of a million metric tons of carbon dioxide.

emission targets. For example, the Global Warming Solutions Act of 2006[2] was passed in California, instituting GHG emissions targets of 1990 levels by 2020 and 80% below 1990 levels by 2050. Williams et al. [2012] study the problem of how to achieve such reductions, and they conclude that converting energy use to electricity is critical. By *electrifying*, deep emissions cuts can be achieved by *decarbonizing* electricity generation, replacing fossil fuel burning plants with generation technologies that emit less GHG. Figure 1.1 shows GHG emissions by sector in California.

Decarbonizing electricity generation is a difficult problem that is the subject of current debate. Williams et al. present four different decarbonization scenarios that achieve the reduction targets, but have different impacts. Making extensive use of renewable generation, e.g., solar and wind, is an attractive option because it is the least complicated. Because its output cannot be adjusted quickly, nuclear generation requires an end use for electricity generated beyond what is demanded, such as a large export market. It also raises safety concerns. Carbon capture storage is far from commercialization and may have unintended environmental impacts due to CO_2 injection. Yet renewable generation is unreliable, requiring large investments in storage and transmission infrastructure, as well as more capacity, according to the model of Williams et al. Additionally, Williams et al. find it impossible to surpass 74% renewable penetration given the constraints of their analysis.

One way to surpass the limits of the Williams et al. model is to increase the ability of demand to follow supply, so-called *demand response*. Their model assumes a limited form of demand response (see Section 2.4.2 for more detail). By building autonomous systems that represent and act for consumers, we can increase responsiveness (by reducing attentional costs and increasing computational capacity and speed of decision making) while maintaining user control (rather than allowing remote control). Thus, our objectives are motivated in a non-technical context as well as a technical one.

There is a business case as well. The average US household spends about $110 on electricity each month,[3] a number which will increase as energy consumption shifts to electricity. In addition, consumers are buying more home automation devices, about $10 billion each year,[4] with major tech companies such

[2] `https://www.arb.ca.gov/cc/docs/ab32text.pdf`

[3] According to the U.S. Energy Information Administration: `https://www.eia.gov/electricity/sales_revenue_price/`.

[4] According to NextMarket Insights.

as Google and Amazon competing for dominance in the market. These devices provide a foundation for deploying autonomous systems into households.

1.4 Outline

Following this introduction, Chapter 2 provides technical background and reviews related work in the following areas:

- Cooperative and non-cooperative games,
- Market design,
- Preference elicitation and mechanism design,
- Computer science and the electrical grid.

Chapters 3 and 4 focus on different aspects of the market design problem. Chapter 3 focuses on the impact of demand "shape" on the supplier's cost of supplying electricity, e.g., demand smoothness and peaks. The major contributions of Chapter 3 are:

- We develop a tractable market model for matching consumers to producers while capturing many of the complexities of electricity production and consumption (but not those pertaining to predictability of demand).
- We explore the stability properties of this model under various cost-sharing schemes.
- We develop two payment algorithms that exhibit high stability and fairness in experiments, while allowing tradeoffs between social welfare and stability.

Chapter 4 focuses on the impact of demand predictability on the supplier's costs. We extend prior work in prediction-of-use (POU) games, which were proposed by Robu et al. [2017] to optimize coordination when demand predictability affects costs. We show that the POU games model has a weakness and extend it to address that weakness. The major contributions of Chapter 4 are:

- We extend POU games to support multiple user consumption profiles, i.e., different ways of consuming electricity that each have a different value.
- We show that the extension remains convex.
- We introduce a new incentive problem that emerges from the partial observability of a user's realized demand profile and address that problem using separating functions.
- We experimentally test POU and MPOU games using utility models learned from electricity use data. We show that the social welfare of MPOU games is greater than that of the fixed-rate tariff, which is greater than that of POU games.

Chapter 5 focuses on the problem of learning user preferences, through experiential elicitation (EE), where the cost and accuracy of queries is dependent on the user's familiarity with the states the system is querying about. The major contributions of Chapter 5 are:

- We introduce and motivate the EE approach in the HVAC control setting.

- We theoretically connect optimal EE to other, well-known, problems in AI.
- We study the interplay (and synergies) of *relative value queries (RVQs)*, which are natural in HVAC settings, with a Gaussian-process (GP) model of preference prediction.
- We develop a system for EE in a smart-home HVAC setting using RVQs and GPs, and analyze it empirically using a combination of real and synthetic data, showing that it accrues higher reward than natural baselines.

Chapter 6 concludes and discusses future work.

Chapter 2

Background

This chapter presents technical background and an overview of related work on four key topics: cooperative and non-cooperative game theory, market design, preference elicitation and mechanism design, and electricity and the smart grid. These topics are important in all the remaining chapters. Work that is more narrowly related to the topics in specific chapters is covered in the chapters themselves.

2.1 Game Theory

We begin with a brief overview of non-cooperative games. We then spend the rest of the section on cooperative games, which is the main game-theoretic formalism employed in Chapters 3 and 4.

2.1.1 Non-Cooperative Game Theory

Game theory analyzes interactions between strategic agents. *Non-cooperative game theory* was developed first in the late 19th and early 20th century, where it appears in the theory of economic competition, and as its own field after von Neumann's 1928 paper "On the Theory of Games of Strategy" [von Neumann, 1928]. The key difference between cooperative and non-cooperative games is that enforceable agreements, e.g., contracts, are available in the former, but not the latter. We can see this difference in the example of the prisoner's dilemma, considering it first as a non-cooperative game and then as a cooperative game, and showing how the outcome changes.

The classical non-cooperative *prisoner's dilemma* game is as follows.[1] There are two prisoners, each suspected of a crime. Each prisoner can either stay silent or betray the other. The prosecutor offers each a deal:

- If one betrays and the other stays silent, the betrayer will go free and the one who stayed silent will serve a five-year term.
- If neither betrays, each will serve a year.
- If both betray, each will serve two years.

[1]The prisoner's dilemma was created by Merrill Flood and Melvin Dresher in 1950 and formalized by Albert Tucker, who was born in Oshawa and was an undergraduate at University of Toronto.

		Prisoner 2	
		Stay silent	Betray
Prisoner 1	Stay silent	$-1, -1$	$-5, 0$
	Betray	$0, -5$	$-2, -2$

Table 2.1: A prisoner's dilemma game.

The payoffs of the game are summarized in *normal form* in Table 2.1. Prisoner 1 is the *row player*, whose action choice is represented by the row she chooses. Similarly, prisoner 2 is the *column player*. Both players choose their action simultaneously, without being able to communicate or view any outside information, and we read the payoff from the cell of the matrix, with the row player's payoff occurring first. The game is played a single time (a *one-shot* game).

Definition 2.1. A *normal-form game* $\langle N, \boldsymbol{S} = \{S_i : i \in N\}, \boldsymbol{U} = \{U_i : i \in N\}\rangle$ is a set of n players N, a set of strategies available to each player S_i and a utility function for each player $U_i : S_1 \times \ldots \times S_n \to \mathbb{R}$ that determines her payoff depending on the strategies played.

The best outcome for the group (i.e., both prisoners) would be for both prisoners to stay silent. This outcome maximizes the *social welfare* or *efficiency* of the game by achieving an aggregate payoff of -2, which is the highest available.

Definition 2.2. A *(pure) strategy profile* $\vec{s}$ for a normal-form game $\langle N, \boldsymbol{S}, \boldsymbol{U}\rangle$ is an element of $S_1 \times \ldots \times S_n$.

Mixed strategy profiles (vs. pure strategy profiles) allow the player to play any distribution over their strategies, instead of a single strategy. They are quite important in non-cooperative games, but do not appear in this thesis.

Definition 2.3. Given a normal-form game $\langle N, \boldsymbol{S}, \boldsymbol{U}\rangle$ and a strategy profile $\vec{s}$ for that game, $SW(\vec{s})$ is the *social welfare* of $\vec{s}$:

$$SW(\vec{s}) = \sum_{i \in N} U_i(\vec{s}). \tag{2.1}$$

Non-cooperative game theory predicts that each prisoner will always choose the betray action no matter what the other prisoner does:

- If the other prisoner stay silent, staying silent receives a payoff of -1 and betraying receives a payoff of 0.
- If the other prisoner betrays, staying silent receives a payoff of -5 and betraying receives a payoff of -2.

Betraying is a *dominant strategy* in this game. In a single instance of the game, each prisoner maximizes their payoff by betraying.

Definition 2.4. Given a normal-form game $\langle N, \boldsymbol{S}, \boldsymbol{U}\rangle$, strategy s_i is a dominant strategy for player i if, for all strategy profiles $\vec{s}_{-i}$ that do not include i and all $s'_i \in S_i$,

$$U_i((\vec{s}_{-i}, s_i)) \geqslant U_i((\vec{s}_{-i}, s'_i)). \tag{2.2}$$

Definition 2.5. Given a normal-form game $\langle N, \boldsymbol{S}, \boldsymbol{U} \rangle$, strategy profile $\vec{s}$ is a dominant strategy equilibrium if $\vec{s}$ consists of only dominant strategies.

Dominant strategies do not exist in all games. The *Nash equilibrium* is a more general *solution concept*, i.e., a rule that predicts which actions the players will choose. A Nash equilibrium is strategy for each player such that no player has incentive to change her strategy even if she knew the strategies chosen by the other players.

Definition 2.6. Given a normal-form game $\langle N, \boldsymbol{S}, \boldsymbol{U} \rangle$, a strategy profile $\vec{s}$ is a *Nash equilibrium* if, for all players $i \in N$ and all strategies $s_i \in S_i$,

$$U_i((\vec{s})) \geqslant U_i((\vec{s}_{-i}, s_i)). \tag{2.3}$$

In the prisoner's dilemma game, both players choosing the betray action is a Nash equilibrium as well as dominant strategy. In the case where a dominant strategy or strategies exist for a particular player, only these strategies will appear in Nash equilibria for that player. Thus, Nash equilibria generalize dominant strategy equilibria.

The *price of anarchy* [Koutsoupias and Papadimitriou, 1999] and the *price of stability* [Schulz and Moses, 2003] characterize the social welfare of the best and worst Nash equilibria relative to the social welfare maximizing outcome. These are important in Chapter 3.

The price of anarchy is the ratio of the social welfare maximizing outcome to the social welfare of the Nash equilibrium with lowest social welfare. It can be thought of as the worst-case social cost of self-interested behavior.

Definition 2.7. Given a normal-form game $\langle N, \boldsymbol{S}, \boldsymbol{U} \rangle$, the *(pure) price of anarchy* is the ratio

$$\frac{\max_{\vec{s} \in S_1 \times \ldots \times S_n} SW(\vec{s})}{\min_{\vec{s} \in \text{NASH}} SW(\vec{s})} \tag{2.4}$$

where NASH is the set of Nash equilibria of the the game.

In practice, the numerator and the denominator of the price of anarchy and price of stability are interchanged when the payoffs of the game are negative, following the convention of the approximation algorithms. This aligns with the intuition that a high price is not desirable. In the examples in Chapter 3, the payoffs of the game will be positive. The price of anarchy of the prisoner's dilemma game, which has negative payoffs, is 2.

The price of stability is the same as the price of anarchy except the ratio is taken with respect to the Nash equilibrium with the highest social welfare, rather than the lowest. It can be thought of as the cost of self-interested behavior in the case where there is a coordinator or system operator, who can propose the best Nash equilibrium to the players.

Definition 2.8. Given a normal-form game $\langle N, \boldsymbol{S}, \boldsymbol{U} \rangle$, the *(pure) price of stability* is the ratio

$$\frac{\max_{\vec{s} \in S_1 \times \ldots \times S_n} SW(\vec{s})}{\max_{\vec{s} \in \text{NASH}} SW(\vec{s})} \tag{2.5}$$

where NASH is the set of Nash equilibria of the the game.

Since the prisoner's dilemma has a single Nash equilibrium, its price of stability is equal to its price of anarchy.

2.1.2 Cooperative Game Theory

Cooperative game theory [Osborne and Rubinstein, 1994] is the branch of game theory which analyzes multi-agent interactions where agents can enforce agreements among themselves. The prisoner's dilemma changes dramatically if the prisoners have access to enforceable contracts and an enforcement authority and the ability to make payments to each other. It is then in their mutual interest to form a contract where both prisoners commit to staying silent and are subject to a penalty for betraying. If the penalty is large enough, neither prisoner will betray the other and the maximum social welfare will be achieved. Such a contract, if obstruction of justice is the objective, might be illegal in the real world—we will be taking a more abstract perspective.

The assumption that the agents can make payments to each other is called *transferable utility*. Cooperative game theory and mechanism design, which is the subject of Section 2.3.1, both have variants where utility is non-transferable that are quite different. We assume transferable utility throughout this thesis.

We also assume *quasilinear utility*. Quasilinear utility means that the *net utility*, or *payoff*, accrued by an agent is equal to their utility for the outcome plus the value of the payment. This means that their utility function is measured in terms of the units user for payments, and there are no "wealth effects," such as agents having decreasing marginal utility for each additional unit of payment they accrue.

It is important to note that enforceable contracts are not available in all interactions. For example, adversarial interactions between countries were a driving force behind research in game theory in the 20th century. The enforceability of international contracts is limited, especially in a military context. Contracts are also often forbidden is sports and games (in the ordinary sense). Predetermining the outcome of a match, which can yield financial profits in the betting markets, is illegal.

In a cooperative game, we are not required to model the players as having individual action choices. Instead, each player has to decide which *coalition* of players to join. The coalition receives a payoff that depends only on the players it contains, and the players in the coalition have to decide how to divide that payoff among themselves.

Definition 2.9. A *cooperative game* is a tuple $\langle N, v \rangle$ where N is a set of agents and $v : 2^N \to \mathbb{R}$ is the *characteristic function*, mapping from each possible coalition to the payoff that coalition would receive.

The standard cooperative games model assumes that each coalition requires every member to agree contractually to take the action that maximizes the aggregate utility of the coalition. This contract can impose a large fine on agents who do not. The model is thus unaffected by the particular actions available to each agent except as they affect the characteristic value of each coalition. The only strategic decision that agents face is the choice of coalition to join. We note that the set of coalitions includes the singleton coalitions—choosing not to cooperate is allowed by the game. There are reasons why this assumption may be problematic in practice. For example, if the action taken by an agent is unobservable or only partially observable, the standard contract may fail to incentivize the agent to take the action agreed to by contract. We discuss this in more detail in Chapter 4.

We can formally model the prisoner's dilemma as a cooperative game. We have two agents, n_1 and n_2, representing Prisoner 1 and Prisoner 2. The characteristic function is:

- $v(\{n_1\}) = v(\{n_2\}) = -2$
- $v(\{n_1, n_2\}) = -2$

In the case where the agents cooperate, they both stay silent. When the agents do not cooperate, each receives the payoff they would have received had they betrayed, as betraying is a dominant strategy.[2] This game is *superadditive* because if we merge any two coalitions, the value of the resulting coalition is greater than or equal to the sum of the values of the merged coalitions, i.e., $v(\{n_1\}) + v(\{n_2\}) = -4 \leqslant v(\{n_1, n_2\}) = -2$. Intuitively, superadditive games are those where there are no anti-synergies among the players.

Definition 2.10. A cooperative game is *superadditive* if the characteristic function v is superadditive, i.e., v satisfies $v(S) + v(T) \leqslant v(S \cup T)$ for any disjoint sets of agents S and T.

A *coalition structure* is a set of disjoint coalitions that contains all agents. *Coalition structure generation* is the problem of calculating the optimal coalition structure given a characteristic function. It is NP-complete in general [Sandholm *et al.*, 1999], but this is unnecessary in the case of superadditive games, where the *grand coalition* of all agents maximizes social welfare. In this thesis and the remainder of this chapter, we only consider superadditive games.

Definition 2.11. The *social welfare* of a coalition structure CS is the sum of the characteristic function values of each coalition in CS.

Observation 2.1. *If a cooperative game is superadditive, the grand coalition has social welfare that is equal to or greater than that of any other coalition structure.*

In the case where the grand coalition maximizes social welfare, the question becomes how to divide the value accrued by the grand coalition among its members. This is the central question in the history of cooperative game theory. In cooperative games with transferable utility, an *allocation* is a payment from to the coalition to each of its members. Note that an allocation may be negative.

Definition 2.12. An *allocation* is a function $t : N \to \mathbb{R}$ that describes a payment from the coalition to each of its members.

A *solution concept* is a formal rule for predicting how a game will be played. There are quite a few different solution concepts in cooperative games, and we will cover only those which are necessary for work that appears later.

Perhaps the most important property for our allocations to satisfy is *efficiency*. Efficient allocations are those which distribute the entire accrued value to the agents. Because we are assuming utility is transferable, we do not need to distinguish between monetary and non-monetary utility.

Definition 2.13. An allocation is *efficient* if $\sum_{i \in N} t(i) = v(N)$.

2.1.3 Core Allocations

The first solution concept we will introduce is the *core* [Gillies, 1959].[3] The core is the set of efficient allocations that satisfy a strong stability property: the payment to any set of agents is greater than or equal to the value they would receive by forming a coalition.

[2] It is easy to come up with the characteristic function in this game because each agent has a dominant strategy if the agents do not cooperate. It can be tricky in the general case. A common strategy is to assume that each agent receives their worst possible payoff. Cooperative games are generalized by partition function games [Thrall and Lucas, 1963], where the payoff each coalition receives may depend on the structure of the other coalitions.

[3] Donald Gillies was born in Toronto and was an undergraduate at University of Toronto.

Definition 2.14. An allocation is *core-stable* if $\sum_{i \in C} t(i) \geqslant v(C)$ for any $C \subset N$.

Definition 2.15. A *core allocation* is an allocation that is efficient and core-stable. The *core* is the set of all such allocations.

Intuitively, core stability means that no set of agents would benefit by *defecting* from (i.e., leaving) the grand coalition to form their own. Because the core is defined by a set of weak linear inequalities, it is closed and convex.

Observation 2.2. *The set of core allocations is convex and closed under union and intersection.*

In the prisoner's dilemma game of the previous section, the core consists of all allocations where $t(n_1) \in [-2, 0]$ and $t(n_2) = -2 - t(n_1)$.

Unfortunately, the core may not exist and is hard to compute NP-complete [Conitzer and Sandholm, 2006]). Core allocations always exist in two-player superadditive games. Here is an example of a three-player superadditive game where there are no core allocations.

Example 2.1. Let there be three players n_1, n_2 and n_3. $v(\{1,2\}) = v(\{1,3\}) = v(\{2,3\}) = v(\{1,2,3\}) = 1$. $v(\{1\}) = v(\{2\}) = v(\{3\}) = 0$. This game has no core allocation.

The problem that occurs in this game is that any two players can cooperate to achieve the entire value of the grand coalition. When the game is played, suppose a coalition of n_1 and n_2 forms. n_3 can offer a larger split of the payoff to the agent receiving less in the coalition (w.l.o.g., n_1). n_1 is receiving $1/2$ at most, so n_3 needs to offer $1/2 + \epsilon$ to convince n_1 to defect. However, this process will repeat infinitely: after n_1 and n_3 form a coalition, n_2 can offer $1/2 + \epsilon$ to n_3, which will be larger than n_3's current payoff of $1/2 - \epsilon$.

It is worth thinking about core non-existence in superadditive games from a philosophical perspective. In a superadditive game, agents have a strong incentive to collaborate in the abstract—doing so increases the aggregate utility of the group. When core-stable allocations exist, they seem like a reasonable description of the outcomes that are likely to result from playing the game, given that the players are rational, have full information about the game, have unbounded computation and are able to defect from their assigned group with no cost.

When core-stable allocations fail to exist, it is less clear what should happen. We can view the situation analogously to the prisoner's dilemma in non-cooperative games. In the prisoner's dilemma, the agents have a "global" incentive to cooperate, but without contracts, they cannot overcome their "local" incentive to defect. In a superadditive game without a core allocation, agents have a global incentive to cooperate, but even the existence of contracts is not enough to overcome the local incentive to defect, for at least some of the agents. Nevertheless, it is not in any agent's interest to be stuck in an endless cycle of defections and therefore for no deal to be reached.

There are a few solution concepts that can be used in the case when core existence is not guaranteed. In the remainder of this chapter, we discuss concepts that relax the core by challenging the assumptions about the capabilities of the agents. In the next section, we describe the Shapley value, a completely different approach to "solving" cooperative games that avoids the non-existence problem, but achieves a lower standard of stability.

The standard core does not take into account the cost required to find a group of agents with which it would be beneficial to form a coalition. We can consider a constant "defection cost", which will increase the number of stable allocations. Formally, *strong-ϵ core* allocations violate core stability by at most a

constant ϵ, i.e., there does not exist a group of agents with an incentive of more than ϵ to defect [Shapley and Shubik, 1966].

Definition 2.16. An allocation is *ϵ-core-stable* if $\sum_{i \in C} t(i) \geqslant v(C) + \epsilon$ for any $C \subset N$.

Definition 2.17. A *strong-ϵ core* allocation is an allocation that is efficient and ϵ-core stable.

There is always a strong-ϵ core allocation for sufficiently large ϵ. An allocation that is in the strong ϵ-core for the smallest ϵ that yields a non-empty set is a *least-core* allocation [Maschler *et al.*, 1979].

Another avenue for relaxation limits the complexity of defections considered by the model. Allocations could fail to be core because there is a very large group with an incentive to defect, but it might be difficult in practice for the group to identify itself and/or coordinate its defection. *Nash-stable* allocations are those where no individual agent has an incentive to change coalitions. Nash stability is a subtler concept than core stability and its relaxations because it is defined relative to a particular *defection model*: what payoff does an agent who defects to another coalition receive? We use Nash stability in Chapter 3 where core-stable payments may not exist. In particular, we measure the *maximum incentive to defect* of an allocation, which is equivalent to the minimum ϵ for which a particular allocation is *ϵ-Nash-stable*.

2.1.4 The Shapley Value

The *Shapley value* [Shapley, 1953] is an alternative solution concept that addresses some of the weaknesses of the core, but lacks strong stability guarantees. One way to motivate the Shapley value is through the "unfairness" of core allocations. Consider the following $2n + 1$ player superadditive game. The payoff of any coalition is the minimum of the number of its even- and odd-numbered players, with players numbered from 0 to $2n$. In the core of this game, no even-numbered player can receive a positive payoff, regardless of the value of n. This occurs because there is always an "extra" even-numbered player contributing nothing to the payoff of the odd-numbered players, who will thus accept an arbitrarily low allocation, preventing any other even-numbered player from receiving an allocation greater than that.

A way around the core's acute sensitivity to scarcity is to think about the contributions of each agent locally rather than globally. Suppose we pick a coalition at random from the power set of agents in the previous example. This coalition either has: (i) more odd-numbered agents, (ii) more even-numbered agents or (iii) the same number of each. Without loss of generality, we will neglect the last case because we assume that n is large. Because there are more even-numbered agents, (ii) is slightly more common than (i). If we add an odd-numbered agent to the coalition, its value will increase by one in case (i). If we add an even-numbered agent to a coalition, its value will increase by one in case (ii). Thus, the average contribution of even-numbered agents is only slightly less than odd-numbered agents: each should get a payoff of around 1/2.

The Shapley value formalizes this reasoning. There are a few equivalent ways to define average contribution—we will use *join orders*. We imagine the grand coalition forming by having agents join one by one in a random order: a join order. For each such join order, we record the change in the coalition's value when agent n_i joins. The Shapley value for agent n_i is her average contribution over all join orders (or permutations of agents).

Definition 2.18. The *Shapley value* of agent n_i in coalition C is

$$s_C(i) = \sum_{S \subseteq C \setminus \{i\}} \frac{|S|!(|C| - |S| - 1)!}{|N|!}(v(S \cup \{i\}) - v(S))$$

The Shapley value of prisoner's dilemma game is -1 for each prisoner. For the game with no core in Example 2.1 it is 1/3 for each player, an intuitively reasonable outcome.

The Shapley value is expensive to compute exactly because there are factorially many join orders. However, it can be approximated efficiently by sampling over join orders in the most natural way. For each join order, calculate the marginal contribution of each agent in that join order and average them across the orders. The resulting estimator is unbiased [Castro *et al.*, 2009].

2.1.5 Convex Cooperative Games

Convex cooperative games are a subclass of cooperative games with particularly desirable computational properties [Shapley, 1971]. A game is convex if its characteristic value function is *supermodular*. Intuitively, this means that, when an agent joins a coalition, the value it contributes never decreases as more agents are added to that coalition.

Definition 2.19. A cooperative game $\langle N, v \rangle$ is *convex* if $v(T \cup \{i\}) - v(T) \geqslant v(S \cup \{i\}) - v(S)$, for all $i \in N$, $S \subseteq T \subseteq N \setminus \{i\}$.

Convex games have several useful properties (all due to [Shapley, 1971]). First, they are trivially superadditive. Second, they always have core allocations. Third, the set of core allocations has a specific geometry. For any join order, the allocation formed by the vector of marginal contributions of agents in that join order is a *vertex of the core*. Because the core is convex, any positive weighting of vertices yields a vector that is also in the core. Thus, the Shapley value, which is the average over all join orders, is a core allocation, and any approximation of the Shapley value consisting of the average over a subset of join orders is a core allocation.

Theorem 2.1 (Shapley [1971])**.** *Let $\langle N, v \rangle$ be a convex cooperative game. Let π be a join order over N. Then, the allocation t, where $t(i)$ is the marginal contribution of agent i in π, is a core allocation.*

Proof. **Efficiency**. The sum of marginal contributions, taken with respect to join order π, is equal to the characteristic value of the grand coalition. Therefore, t is efficient.

Core stability. Let $C \subseteq N$ be an arbitrary set of agents. We want to show that $\sum_{i \in C} t(i) \geqslant v(C)$. To do this, consider the marginal contribution each agent makes to the characteristic value of C, taken with respect to join order π (ignoring the agents that are not in C). Because $\langle N, v \rangle$ is a convex game and $C \subseteq N$, each agent's marginal contribution to C w.r.t. π is weakly less than its marginal contribution to N w.r.t. π. Because the sum of the marginal contributions to C w.r.t. π is equal to $v(C)$, $\sum_{i \in C} t(i) \geqslant v(C)$. □

Corollary 2.1. *Let $\langle N, v \rangle$ be a convex cooperative game. The Shapley value of the game is a core allocation.*

Proof. The Shapley value is the average of the marginal contributions over all join orders. Because each marginal contribution allocation is a core allocation and the core is convex, any weighted average of marginal contribution allocations is a core allocation. Therefore the Shapley value is a core allocation. □

We make use of convex cooperative games in Chapter 4.

2.1.6 The Duality Between Cooperative Games and Markets

Shapley and Shubik [1969] motivates the idea of using cooperative game theory to centrally coordinate markets to achieve efficient, stable outcomes. They observe that *exchange economies*, which are a simple model of agents trading commodities, can be represented as a cooperative game, which they call a *market game*. They show that the market game always has core allocations, and these core allocations correspond exactly to the equilibria of the exchange economy from the economics perspective.

An exchange economy is a market where all agents are consumers. Each agent starts with an initial endowment of resources and has a utility function over any allocation of those resources. These utility functions are often continuous and concave, which is required in order to guarantee the existence of an equilibrium.

Definition 2.20. A *(pure) exchange economy* $\langle N_e, G, a, U\rangle$ consists of a set of agents N_e, G which is the set of admissible *commodity allocations* (the nonnegative orthant of a space with dimension equal to the number of commodities in the market), $a : N_e \to G$, an initial endowment of commodities to each player, and $U = \{U_i : G \to \mathbb{R}\}$, a utility function for each player.

This kind of exchange economy is "pure" in the sense that agents are not able to produce new units of the commodities.

Solution concepts for exchange economies are defined similarly to those in cooperative games: they are allocations satisfying a set of constraints. Unlike cooperative games, the allocation of goods and the payment received by each agent are dealt with separately in the equilibrium, but the payment is determined by the allocation and the price function of the economy. Here the prices will be linear, i.e., there is a constant price for each good and the payment is the sum of prices paid.

A competitive equilibrium is an allocation of commodities plus a price for each commodity that satisfies several conditions. The allocation should be *feasible*, i.e., no commodities are created or destroyed relative to the initial endowments, and it should be *envy-free* in the sense that each agent receives an allocation that maximizes their utility at the current prices.[4]

Definition 2.21. A *competitive equilibrium* $\langle p, x\rangle$ of an exchange economy $\langle N_e, G, a, U\rangle$ consists of a price function $p : G \to \mathbb{R}$ and an allocation $x : N_e \to G$ of commodities to agents, satisfying the following properties:

- x is *feasible*: $\sum_{i\in N_e} a(i) = \sum_{i\in N_e} x(i)$.
- x is *envy-free* under p: for any agent, for any $y \in G$,

$$U_i(x(i)) - p(x(i) - a(i)) \geqslant U_i(y) - p(y - a(i)) \tag{2.6}$$

If we assume that the utility functions are continuous and concave, a competitive equilibrium can be found by maximizing social welfare across the set of feasible allocations, which is a convex optimization problem. The prices are the shadow prices of the optimization, which exist because the constraints are linear. We will see that the payments under the competitive equilibrium are core payments in the naturally induced cooperative game.

[4]Note that this is a slightly different definition of envy-freeness from the one used in Chapter 3.

A *market game* is a cooperative game that models an exchange economy. The "coalitions" are groups of agents that trade *only* with each other and the characteristic value of a group of agents is the maximum sum of the valuation functions of those agents over all possible trades. Note that because agents in a coalition can always decide not to trade with each other, the game is superadditive.

Definition 2.22. A cooperative game $\langle N_c, v\rangle$ is a *market game* of an exchange economy $\langle N_e, G, a, U\rangle$ if $\langle N_c, v\rangle$ satisfies:

- The set of agents is the same, i.e., $N_c = N_e$.
- For all $C \subseteq N_c$, $v(C) = \max_x \sum_{i\in C} U_i(x(i))$ subject to $\sum_{i\in C} x(i) = \sum_{i\in C} a(i)$.

Shapley and Shubik [1969] show that the payoff vector of a competitive equilibrium of an exchange economy is a core allocation of the market game.

Theorem 2.2 (Shapley and Shubik [1969])**.** *Let $\langle N_e, G, a, U\rangle$ be an exchange economy and $\langle N_c, v\rangle$ be a market game. There there exists a competitive equilibrium $\langle p, x\rangle$ such that $t(i) = U_i(x(i)) - p(x(i) - a(i))$ is a core allocation of $\langle N_c, v\rangle$.*

Proof. Let $y : N_e \to G$ be a feasible allocation in the exchange economy that achieves the value $v(N_c)$ in the market game. Because y is the result of a convex optimization with linear constraints, there exists a linear price function q such that, for each $i \in N_e$, the expression $U_i(x(i)) - p(x(i) - a(i))$ is maximized when $x(i) = y(i)$ and $p = q$. Thus, $\langle y, q\rangle$ is a competitive equilibrium of $\langle N_e, G, a, U\rangle$.

We will show that $t(i) = U_i(y(i)) - q(y(i) - x(i))$ is a core allocation of $\langle N_c, v\rangle$. Remark that $\sum_{i\in N_c} t(i) = \sum_{i\in N_c} U_i(y(i))$ because y is feasible. Thus, t is efficient.

Let C be an arbitrary subset of N_c, and let y_C be an allocation in the exchange economy that achieves the value $v(C)$. Because y maximizes $U_i(x(i)) - q(x(i) - a(i))$, it follows that $t(i) \geqslant U_i(y_C(i)) - p(y_C(i) - a(i))$ for all $i \in C$. Summing over $i \in C$ yields $\sum_{i\in C} t(i) \geqslant \sum_{i\in C} U_i(y_C(i)) - p(y_C(i) - a(i)) = \sum_{i\in C} U_i(y_C(i))$ because y_C is a feasible allocation for C. Thus, t is core-stable. □

This correspondence suggests that the core captures key properties of a competitive equilibrium. We use this observation to motivate modeling markets as cooperative games, which is advantageous because shadow prices of the optimization often fail to exist.

2.2 Market Design

In this section, we provide a brief overview of work in market design in both economics and computer science. Market design is particularly useful in markets with *constraints*, *complementaries* and *externalities*. By constraints, we refer to allocations that are infeasible in an unusual way. For example, the output of coal-based plants can only be adjusted over a timespan of hours. Thus, they can only serve *demand profiles*, i.e., vectors of consumption over time, that are sufficiently smooth. Complementarities and externalities occur when the allocation to one agent affects the allocation or utility of another, in a positive or negative way, respectively. These do not arise in the electricity markets we study, but complementarities play a large role in the medical resident matching problem described in the next section.

Without market design, the presence of these features causes self-interested behavior among consumers to result in inefficient economic outcomes. The complexities and the economics of generation cause these features to arise in electricity markets.

We are particularly interested in exploiting the fact that consumers have preferences over different consumption bundles in ways that are not captured by current pricing schemes. This is the main focus of Chapters 3 and 4. Consumers of electricity are generally willing to shift or reduce their demand to a certain extent if they are sufficiently compensated. This makes cooperation valuable. Under a cooperative game theoretic approach, coalitions can offer subsidies to individuals to behave in a way that fits with the desired consumption of the rest of the coalition. These subsidies may result in prices that appear similar to time-of-use or critical peak prices, but they are fully adaptable to circumstances (i.e., they need not be set in advance). In addition, agents with off-peak demand pay significantly less because their demand helps flatten the aggregate demand profile of the coalition.

Section 2.2.1 provides an overview of market design's goals and history via Roth [2002]. Then, Section 2.2.2 discusses contributions to market design by computer scientists.

2.2.1 Market Design in Economics

Roth [2002] provides an overview of successes in the area of market design, particularly those in the 1990s. The technical models are not described in this section in detail because they are not relevant to the main content of the thesis, but it is worth noting that they are instances of cooperative games with non-transferable utility. Because utility is non-transferable, they use different solution concepts from those introduced so far, but they have the same goals of maximizing social welfare and achieving stability, and similar computational difficulties.

Market design, according to Roth, describes the imposition of rules on environments where agents are exchanging goods or services. The development of successful market rules has occurred "organically" (i.e., without central supervision or coordination) in many areas, particularly those where money can be used to facilitate transactions and/or goods can be inexpensively stored and transferred. There have also been several notable successes in imposing external rules of exchange in situations where the organically developed systems had major shortcomings. These scenarios often have some of the following features: (i) money can't be used; (ii) the use or exchange of goods or services is subject to constraints, externalities or complementaries; and (iii) the scale of the market is very large.

Roth's area of primary focus is the National Resident Matching Program (NRMP), where medical residents are assigned to residency programs at hospitals. Roth and Peranson [1999] formalize the NRMP problem as a two-sided, many-to-one, *stable matching* problem. Each hospital residency program is matched to one or more residents subject to feasibility constraints, primarily capacity. The goal is to find a stable matching that satisfies the constraints. They use a form of Nash stability: no resident and program would both prefer to be matched to each other over their current assignment. The resulting problem is NP-complete [Ronn, 1990].

This NRMP has all three of the properties that make market design valuable, according to Roth. The constraints on the market are dependencies between the residents that are accepted at each program. The acceptance of a resident in one program may reduce the capacity of another. For example, while most programs are one year long, there are some programs that have both one- and two-year variants. Thus, each resident enrolled in the two-year variant of a program reduces the capacity of the one-year subprogram by one. There are also parity constraints on the number of residents accepted by each

TABLE I
Stable and Unstable (Centralized) Mechanisms

Market	Stable	Still in use (halted unraveling)
American medical markets		
NRMP	yes	yes (new design in '98)
Medical Specialties	yes	yes (about 30 markets)
British Regional Medical Markets		
Edinburgh ('69)	yes	yes
Cardiff	yes	yes
Birmingham	no	no
Edinburgh ('67)	no	no
Newcastle	no	no
Sheffield	no	no
Cambridge	no	yes
London Hospital	no	yes
Other healthcare markets		
Dental Residencies	yes	yes
Osteopaths (<'94)	no	no
Osteopaths (≥'94)	yes	yes
Pharmacists	yes	yes
Other markets and matching processes		
Canadian Lawyers	yes	yes (except in British Columbia since 1996)
Sororities	yes (at equilibrium)	yes

Figure 2.1: The history of designed markets from the 1990s from Roth [2002], including whether each had survived (up to the publication of the paper in 2002) and whether each is stable.

program, and there is a complementarity in the form of residents in couples (e.g., spouses), where the preferences of one member of a couple depends on the other's assignment. The NRMP matched around 20,000 residents per year at the time of the redesign and now matches more than 30,000 [Program, 2013].

The system Roth and Peranson developed to match residents to programs (henceforth, RP) has several features that distinguish it from the previous techniques used to address the problem. First, it has a game-theoretic guarantee of stability, meaning that agents in the market cannot benefit by making side deals after the mechanism has made its assignment (at least in the case of defections of pairs of agents). Figure 2.1 reproduces Roth's comparison of designed markets up until 2002, whether they were stable and whether they survived. With the exception of the Cambridge and London Hospital markets, only stable markets survived. This is a key result in market design and has led to making stability a key design goal in most designed markets.

Second, RP is motivated on a theoretical level by simpler models. While the NRMP is too complex to reason about in its full generality, theoretical results for simpler models can be surprisingly predictive of how the market works in practice. Roth draws an analogy with bridge building: it benefits from both scientific (i.e., theoretical) and engineering (i.e., empirical) understanding.

Third, RP was empirically tested for *strategy-proofness*, that is, how much agents can benefit by strategically misreporting their preferences. One of the major issues that led to the construction of RP was a perception that the existing mechanisms were easy to manipulate, and these and various related concerns have been constant throughout the history of the NRMP. Since strategic issues were seen as very important, Roth designed the system to take them into account. It has been satisfactory and enduring.

2.2.2 Market Design in Computer Science

Computer science has had an impact on market design in three main problem areas: stable matching, *kidney exchange*, and *combinatorial auctions*. All three problems can be modeled as cooperative games,

but are not always referred to as such in the literature. Stable matching and kidney exchange are cooperative games with non-transferable utility, but they have developed their own language that mostly avoids the use of explicit concepts from cooperative game theory. Combinatorial auctions fall under the cooperative game-theoretic umbrella because they involve economic interactions where contracts are present, but the cooperative game theoretic framework is less useful because there is usually only a single seller or single buyer.

All three of these problems have in common that they involve significant computational challenges on the side of the market coordinator, which is why they are attractive problems in computer science. Some, particularly combinatorial auctions, also present computational challenges on the side of the market participant.

The main contributions from computer science to the stable matching problem are algorithmic improvements or extensions to RP. RP is incomplete—it may fail to find a stable matching even if one exists. It is possible to use general solvers for NP-complete problems, such as constraint programming, integer programming and satisfiability to achieve completeness, usually at the cost of significantly slower computation. Work from this perspective includes work by the author (Drummond et al. [2015] and Perrault et al. [2016]) and others [Biró *et al.*, 2014; Gent and Prosser, 2002; Prosser, 2014; Manlove *et al.*, 2007; Gent *et al.*, 2001; Unsworth and Prosser, 2005]. Another contribution from computer science is to extend the market model, allowing for additional constraints, complementarities or externalities. Many of the above papers do this as well, as it is often made easier by using a complete solver, but there is also work to extend RP, such as Goto et al. [2016].

Kidney exchange is an example of a market design problem where much of the initial implementation was done in computer science. The setting for kidney exchange [Abraham *et al.*, 2007] is the following: a group of patients need kidney transplants. Each is paired with a willing kidney donor with whom they are not compatible. The goal is to transplant as many kidneys as possible while respecting the constraints that (i) for each patient who receives a kidney, their donor donates a kidney and (ii) there are no donation cycles longer than a certain length, which is usually two to four patients. The latter condition enforces a type of stability: the cycles should be short enough that all the transplants can happen simultaneously, which prevents any donor from backing out after the person they are paired with has received a kidney. Kidney exchange is less game-theoretic than the other problems discussed in this section: there are no incentive conditions and no stability conditions other than the short chains constraint. The major difficulties are computational, i.e., developing fast algorithms that find a lot of matches, and extending the model to integrate more realistic features of the problem, such as donation failures [Dickerson *et al.*, 2013].

Combinatorial auctions are a type of auction where bidders' utility functions for receiving sets of items may be arbitrarily complex. The real-world example that motivates a lot of work in the area is *wireless spectrum auctions* where buyers' utility functions often depend on how much contiguous spectrum they receive. Because of the complexity of the bidders' utility functions, the problem of optimizing the set of items that each bidder receives with respect to the received bids is a challenge in itself [Rothkopf *et al.*, 1998]. The other key aspect of combinatorial auctions is how the bidders communicate their utility functions to the auctioneer. We will discuss the preference elicitation aspect of combinatorial auctions in Section 2.3.3 because of its relevance to Chapter 5.

2.3 Eliciting Preferences of Market Participants

Eliciting the preferences (or utility functions) of market participants has a different role in each of the three problems discussed in the previous section. In the standard formulation of kidney exchange, there are no preferences at all, only donor compatibilities measured by a medical test. In stable matching, preferences have a simple form: each agent on each side of the market ranks the agents on the other side. Preference elicitation in combinatorial auctions (CAs) is the most involved: preferences may be exponentially sized in the number of items and are usually elicited incrementally over a period of time (which may be days in a large auction) for that reason.

Preference elicitation in electricity markets is most similar to CAs, but differs in a few key ways. Like CAs, preferences for electricity consumption are complex, but the source of complexity is different. CAs have synergies and anti-synergies between items and sets of items. Electricity consumption preferences are context-dependent and coupled across time, e.g., deferring consumption in one time period due to high prices affects your preferences in another time period. Another key difference is that CAs have most often been studied in the context of business interactions between companies that are willing to spend time and money to understand their preferences to increase their profits. Individuals consume electricity in a way that is mostly based on intuition. The median US household spends about 2% of its income on electricity and thus is only willing to spend a limited amount of attention on electricity optimization.

To understand strategic considerations in preference elicitation, a brief background on the problem of *mechanism design* is useful. In mechanism design, agents have unknown preferences and the goal is to maximize an objective function that depends on the unknown preferences. Critically, mechanism design studies the incentives of agents to misreport their preferences in order to achieve a better outcome. CAs and all auctions are all instances of mechanism design.

In the first of the four subsections of this section, we provide an overview of mechanism design. In the second subsection, we present the critical Vickrey-Clarke-Groves (VCG) mechanism, which provides a framework under which agents are incentivized to report their preferences truthfully. This allows the preference elicitation problem to be separated from the market design problem, under certain conditions.

The remaining two subsections exploit the VCG result to focus on preference elicitation, liberated from strategic concerns. In the third subsection, we study the question of query type in preference elicitation: what form of queries should the elicitor ask the agents? We use the context of combinatorial auctions as a concrete example, but many query types are transferable to other preference elicitation settings.

In the fourth and final subsection, we discuss the question of query strategy: given the choice of query type, what specific queries should the elicitor ask the agents?

2.3.1 Mechanism Design

We first present an overview of *mechanism design* [Shoham and Leyton-Brown, 2008], a generic framework for modeling auctions and other strategic interaction where a *principal* decides an outcome and payments given the reported preferences of a set of agents N. Each agent $i \in N$ has a *type* θ_i, which typically represents the agent's utility function, known only to the agent. The principal specifies a function y (the so-called *mechanism*) that maps from the vector of reported types to an *outcome* and a payment to each agent. This function is revealed to all agents before they choose which type to report—the agents

may report strategically, i.e., a type that is not the same as their true type. After the types are revealed, the mechanism is executed, yielding an outcome and a payment to be made to each agent. The goal of the principal is to find a y such that each agent is incentivized to report their true type and the outcome maximizes the principal's objective—for example, to maximize the aggregate utility of the agents. The goal of each agent is to maximize her payoff, which is her utility for the outcome plus the payment she receives (agents are assumed to have quasilinear utility).

Definition 2.23. A *mechanism design problem* $\langle N, O, \Theta, \{\theta_i\}, U\rangle$ consists of a set of n agents N, a set of outcomes O, a set of types Θ, a type for each agent θ_i, and a utility function $U : \Theta \to (O \to \mathbb{R})$ specifying the payoff of each outcome to an agent of each type.

Definition 2.24. A *mechanism* $y = \langle f, t\rangle$ consists of a *social choice function* $f : \Theta^n \to O$, mapping from type reports to outcomes, and a payment function $t : \Theta^n \to \mathbb{R}^N$, mapping from type reports to payments to each agent.

The mechanism design interactions proceed as follows:

1. The principal specifies a mechanism y, and it is revealed to all agents.
2. Each $i \in N$ reports a type $\hat{\theta}_i$, yielding a *type profile* $\hat{\theta} = \{\hat{\theta}_1, \ldots, \hat{\theta}_n\}$.
3. The mechanism $y = \langle f, t\rangle$ is executed, yielding outcome $f(\hat{\theta})$ and payments $t(\hat{\theta})$.
4. Each agent i receives utility according to her payoff function:

$$U(\theta_i)(f(\hat{\theta})) + t(\hat{\theta})(i). \tag{2.7}$$

Formally, the principal's objective is to choose y to induce an outcome o that maximizes the sum of agent utilities under their true types:

$$\sum_{i \in N} U(\theta_i)(o). \tag{2.8}$$

The principal may also care about the revenue they receive, i.e., the sum of payments, but this thesis does not consider that case.

Mechanism design is a generic framework that includes any kind of auction, including combinatorial auctions. Stable matching and kidney exchange can also be formulated as mechanism design problems, but because there is no transferable utility, the model differs somewhat from the one we present in this subsection. As in cooperative game theory, we restrict our focus to the transferable utility case because the technical chapters allow payments.

The key difference between mechanism design and the market design problems of the previous chapter is that mechanism design emphasizes that the agents report their preferences to the principal and may not report truthfully if it is in their interest not to do so. A mechanism that incentivizes the agents to report truthfully is called *incentive-compatible* or *strategy-proof* or simply *truthful*.

The strongest form of incentive-compatibility is *dominant strategy incentive compatibility (DSIC)*. Recall from the prisoner's dilemma example in Section 2.1.2 that an agent has a dominant strategy if it maximizes her utility to follow that strategy, no matter what the other agent do. DSIC makes reporting truthfully a dominant strategy for each agent.

Definition 2.25. A mechanism $y = \langle f, t \rangle$ is *dominant strategy incentive compatible* for a mechanism design problem $\langle N, O, \Theta, \{\theta_i\}, U \rangle$ if, for each agent $i \in N$, any set of reported types by the other agents $\hat{\theta}_{-i}$, and any profile $\theta'_i \in \Theta$:

$$U(\theta_i)(f((\theta_i, \hat{\theta}_{-i}))) + t((\theta_i, \hat{\theta}_{-i}))(i) \geqslant U(\theta_i)(f((\theta'_i, \hat{\theta}_{-i}))) + t(\theta'_i, \hat{\theta}_{-i}))(i) \tag{2.9}$$

The left side of the equation represents agent i's net utility when she reports her type truthfully and the right side represents her net utility when she reports θ'_i, potentially untruthfully.

As is the case with core stability in cooperative games, there are various relaxations of DSIC. One example is when truthfulness is a Nash equilibrium—the net utility of agent i is maximized if she reports truthfully, given that all other agents do.

We present an example of a DSIC mechanism for concreteness and to motivate the VCG mechanism (because VCG generalizes this mechanism). Consider a *single-item auction*. There is a set of agents N competing for a single item that the principal is selling. The type space consists of the potential valuations of the item, which we will allow to be any real number.

Definition 2.26. A *single-item auction* $\langle N, U \rangle$ consists of a set of n agents N and a utility function $U = \{U_i : \mathbb{R}\}$ for each agent representing the value of the item to that agent. If an agent does not receive the item, they receive zero utility.

A DSIC mechanism for single-item auctions is the *second-price auction* formalized by Vickrey [1961].

Definition 2.27. A *second-price auction* $\langle f, t \rangle$ is a mechanism for a single-item auction $\langle N, U \rangle$. f assigns the item to the agent with highest reported value—the *winner*. $t(i)$ is zero except for the winner, who pays the *second-highest* reported valuation.

Theorem 2.3 (Vickrey [1961])**.** *A second-price auction is DSIC for a single-item auction.*

Proof. Consider an arbitrary agent i who reports her valuation truthfully. We will show that i cannot increase her net utility by reporting any other valuation.

- Suppose i is the winner. Notice that i's net utility is non-negative because her payment is less than her valuation.

  If she increases her valuation, the winner of the auction will not change and her payment will not change—thus, her net utility is unaffected.

  If she decreases her reported value, either she will remain the winner or another agent j will become the winner. If she remains the winner, her net utility is unaffected, as in the previous case. If a new winner j is selected, i's net utility becomes zero, which is less than it was when she was the winner.

- Suppose i is not the winner. i's net utility is zero—she does not receive the item, but also does not make a payment. If she decreases her reported valuation, her net utility is unaffected because she will never become the winner and her payment will remain zero.

  If she increases her reported valuation, she will either win the item or not. If she wins the item, her net utility is negative because she is paying more than the item is worth to her. If she does not win the item, her net utility is unaffected.

□

We can say that the second-price auction maximizes social welfare in a single-item auction, assuming that the agents follow their dominant strategies.

Definition 2.28. The social welfare of an outcome $o \in O$ is $\sum_{i \in N} U(\theta_i)(o)$.

If agents report truthfully, the second-price auction assigns the item to the agent with highest valuation, maximizing social welfare.

Note that the second-price auction does not say anything about how the agent's utilities are elicited by the auctioneer. The simplest procedure would be to ask all of the bidders what their utility for the item is, which is called a *sealed-bid* auction. Incremental elicitation procedures can achieve the same outcome with less information. Because the only pieces of information the second-price auction needs are the utilities of the highest and second-highest bidders—the values of the other bidders have no impact on the outcome or the payments—an incremental procedure can reduce the amount of information communicated between the bidders and the auctioneer significantly.

An example of a common incremental procedure is the *ascending (English) auction*. The ascending price auction begins with a minimum price. Then, iteratively:

1. Request a bid at the current price.
2. If no bid is received, terminate the procedure and assign the item to the last bidder.
3. If a bid is received, increment the price.

The outcome of this auction approaches the second-price auction as the bid increment approaches zero. In the case of a zero bid increment, the price increases until there are only two agents that are still bidding, the highest bidder and the second-highest bidder. When the auction terminates, the highest bidder receives the item and pays the value of the second highest bidder.

It is important to note that the second-price auction, while highly resilient to misreporting by individuals, is somewhat vulnerable to misreporting by small groups. In particular, the top bidder can pay the second highest bidder to lower her report. More generally, the top bidder benefits by any other bidder lowering her report. This issue will return in the next section.

2.3.2 The Vickrey-Clarke-Groves Mechanism

The second-price auction can be extended to a DSIC mechanism for a generic mechanism design problem that yields the social welfare optimal outcome if all agents report truthfully: the *Vickrey-Clarke-Groves (VCG)* mechanism. This is a critical result that we make use of in the following two subsections. It allows us to view the preference elicitation problem separately. It does not matter how we elicit agent's utility functions because they will be incentivized to respond truthfully as long as we run VCG.

The VCG mechanism can be motivated by considering the mechanics of the second-price auction. Each agent in the second-price auction pays exactly the externality she imposes on the social welfare of the other agents, i.e., her payment to the principal is equal to the other agents' aggregate utility loss as a result of her participation in the mechanism. For agents who do not receive an item, this is zero, i.e., if the agent were removed from the auction the result would be unaffected. The winner imposes an externality on the agent who would have received the item had the winner not received the item.

The amount of this externality is equal to the second-place agent's valuation, which is exactly what the winner pays.

Definition 2.29. The *Vickrey-Clarke-Groves (VCG)* mechanism $y = \langle f, t \rangle$ is a mechanism for a mechanism design problem $\langle N, O, \Theta, \{\theta_i\}, U \rangle$. The social choice function f maximizes social welfare with respect to the reported types:

$$f(\hat{\theta}) = \arg\max_{o \in O} \sum_{i \in N} U(\hat{\theta}_i)(o) \tag{2.10}$$

The payment function t pays to each agent i the aggregate utility of the other agents plus an arbitrary function $h_i(\hat{\theta}_{-i})$ (the *pivot rule*) that depends only on the reports of the other agents.

$$t(\hat{\theta})(i) = \sum_{j \in N : j \neq i} U(\hat{\theta}_j)(f(\hat{\theta})) + h_i(\hat{\theta}_{-i}) \tag{2.11}$$

Remark that if we set $h(\hat{\theta}_{-i})$ to be zero, the mechanism always *pays* the agents, which is unusual. We can remedy this by using the *Clarke pivot rule*, an adjustment to the payments that makes them equal to the externality each agent imposes on the others as a result of their participation.

Definition 2.30. The *Clarke pivot rule* for VCG is

$$h_i(\hat{\theta}_{-i}) = -\max_{o \in O} \sum_{j \in N : j \neq i} U(\hat{\theta}_j)(o) \tag{2.12}$$

Under the Clarke pivot rule, remark that for single-item auctions, the VCG mechanism is exactly equivalent to the second-price auction. In general, the Clarke pivot rule ensures that the mechanism's revenue is non-negative.

Observation 2.3. *Given a mechanism design problem, running the VCG mechanism with the Clarke pivot rule yields a non-positive payment to each agent.*

In addition, VCG payments with the Clarke pivot rule satisfy the condition that the agent is better off having participated in the mechanism rather than abstaining from reporting a type. This condition is known as *individual rationality.*

Observation 2.4. *Given a mechanism design problem, the VCG mechanism with the Clarke pivot rule satisfies individual rationality.*

We can show that VCG is DSIC by remarking that, ignoring the pivot rule for now, net utility received by each agent is exactly equal to the social welfare of the selected outcome. Then, because the pivot rule depends only on the reports of the other agents, it does not affect agent i's incentives.

Theorem 2.4 (Vickrey [1961], Clarke [1971] Groves [1973])**.** *The VCG mechanism is DSIC for a mechanism design problem* $\langle N, O, \Theta, \{\theta_i\}, U \rangle$.

The VCG mechanism maximizes social welfare in a mechanism design setting assuming that agents follow their dominant strategies.

The VCG mechanism has significant consequences from a preference elicitation perspective. If the principal announces that he will run the VCG mechanism, it does not matter how the preference information is gathered—the agents will have no incentive to misreport. The principal has to be careful

to gather *all* of the necessary information to run VCG because missing some of it can give agents an incentive to misreport. Here is an example.

Example 2.2. Consider a mechanism design problem with two agents and two outcomes, o and o'. The agents' utility functions are $U_1(o) = 0$, $U_1(o') = 1$, $U_2(o) = 2$, and $U_2(o') = 10$. The principal announces that she will run the following mechanism:

1. Suppose she assumes that $U_1(o) = 0$ and $U_1(o') = 1$. She will query agent 2 for the value $U_2(o) = 2$. All values that she has not queried or assumed (i.e., just $U_2(o')$), she will assume are 0.
2. She will execute VCG with the Clarke pivot rule on the results.

If agent 1 reports truthfully, she will receive a net utility of $U_2(o)+U_1(o)-U_1(o') = 2$ because outcome o will be selected. If she instead reports $U_2(o) = 0$, she will receive a net utility of $U_2(o')+U_1(o')-U_1(o') = 10$.

VCG is computationally no harder than maximizing social welfare, excepting for the arbitrary complexity of the pivot rule. Under the Clarke pivot rule, we have to maximize social welfare once for each agent, which may or may not be a problem depending on the complexity of that calculation and the number of agents. It is worth noting that Facebook uses the VCG mechanism in its ad auctions.

As is the case with the second-price auction, VCG is vulnerable to collusion among agents. Because each agent's payment is dependent on the reports of the other agents, collusion among agents can reduce all payments to the mechanism to zero.

Example 2.3. Consider a mechanism design problem with three agents and three outcomes. The utility functions are given in the table below.

		Outcome		
		o_1	o_2	o_3
Agent	1	3	2	1
	2	2	1	3
	3	1	3	2

Running VCG with the Clarke pivot rule yields an arbitrary outcome—say o_1. Agent 1 pays 2, agent 2 pays 1 and agent 3 pays 0. If agent 2 reduces $U_2(o_3)$ to 1 and agent 3 reduces $U_3(o_2)$ to 2, then no agents make any non-zero payments while achieving the same outcome. Agents thus have a strong incentive to collude if they can gather enough information about the other agents. Collusion is strictly beneficial for all of the agents in the example above in the case that agent 3 receives a payment of at least ϵ from agent 1 or 2.

2.3.3 Preference Elicitation and Query Types

Using the VCG mechanism, we can treat the preference elicitation problem separately because agents are incentivized to report truthfully. We focus on the preference elicitation problem in this and the next subsection. In this subsection, we discuss the choice of query type: what is the form of the questions the principal should ask the agents? In the next section, we discuss query strategy: given a choice of query type, what specific queries should the principal ask?

We will use the CA domain as a concrete example to discuss query types, but many of the query types are transferable to other settings. We begin with a brief overview of the CA problem. Solving a

CA consists of solving two related problems. The first is *winner determination*: given a CA instance, allocate the items in a way that maximizes the aggregate utility of the bidders, i.e., maximize social welfare. The second is to elicit the utility functions from the agents, a necessary prerequisite to winner determination.

We will begin with a formal definition of CAs [Cramton *et al.*, 2005].

Definition 2.31. A *combinatorial auction (CA)* is a tuple $\langle N, M, U \rangle$ where N is a set of n agents, M is a set of m distinct, indivisible items, and $U = \{U_i : 2^M \to \mathbb{R}\}$ is a utility function for each agent.

In order to formalize the winner determination problem, we need to specify the model of an allocation.

Definition 2.32. An *allocation* $a : N \to 2^M$ is a mapping from each player to a set of items she receives. An allocation a is *feasible* if it allocates each item to at most one player: for each item j, $|a^{-1}(j)| \leqslant 1$. The *winner determination problem* is to find a feasible allocation that maximizes social welfare, where social welfare is the aggregate utility of the agents.

The set of feasible allocations is our outcome space from a mechanism design perspective.

Like most other interesting problems in market design, winner determination is hard. There is an exponential number of allocations and no structure to the utility functions in the worst case. It can be shown to be NP-complete by reduction to weighted set packing.

Theorem 2.5 (Rothkopf et al. [1998])**.** *The winner determination problem is NP-complete.*

NP-completeness is not necessarily a problem in practice. CAs can be formulated as integer programs in a straightforward way, allowing many of the spectrum auctions around the world to be run on an ordinary desktop computer. The exact form of the integer program depends on the way the utility functions are communicated to the auctioneer, the so-called bidding language.

Sandholm and Boutilier [2006] provide an overview of the preference elicitation problem in CAs. They remark that many auctions can be viewed as instances of a framework (quoting directly):

1. Let C_t denote information the elicitor has regarding bidder valuation functions after iteration t of the elicitation process. C_0 reflects any prior information available to the auctioneer.

2. Given C_t, either (a) terminate the process, and determine an allocation and payments; or (b) choose a set of (one or more) queries Q_t to ask (one or more) bidders.

3. Update C_t given response(s) to query set Q_t to form C_{t+1}, and repeat.

Note that the ascending auction described in Section 2.3.1 is an instance of this framework.

The choice of query type is critical. If the kinds of queries that may be asked are not limited, we could query the entire valuation functions in a single query. There is a general trade-off between the amount of information a query provides about the user's valuation function and how difficult that query is to answer—the *cognitive cost*. The authors describe four types of queries: demand, value, order, and bound-approximation, which we present in roughly decreasing order of cognitive cost.

Demand queries ask a buyer which bundle she would purchase given a particular set of prices, which may be linear or non-linear with respect to the items. Almost all of the procedures for CAs that are used in practice use demand queries. Blumrosen and Nisan [2010] study the computational properties of demand queries and find that demand queries achieve the best possible approximation ratio of social

welfare of any of the query types under the constraint of polynomially-sized communication. They also show that demand queries can efficiently simulate a number of other query types.

On the other hand, Blumrosen and Nisan [2010] remark that it is important to consider the additional burden that demand queries put on bidders compared to other query types. A few studies in the experimental economics literature [Scheffel *et al.*, 2012; Bichler *et al.*, 2013] that show that people struggle to receive a high value from demand query-based CAs in a lab setting. In particular, they focus on a limited search space prior to the auction and have a hard time adjusting when prices become high enough that they should switch their target bundles. In large-scale CAs, the bidders often have to solve a hard optimization problem to determine their most preferred bundle at a particular price.

The most prevalent alternative to the demand query is the *value query*, where the elicitor may ask the buyer directly for the value of any bundle. In earlier work in CAs, before demand queries became the predominant choice in practice, value queries were evaluated on their own merits. Hudson and Sandholm [2004] show that making queries about bidder-bundle pairs that may have high value or are likely to be matched to a bidder outperforms random queries in experiments. They show that any deterministic evaluation policy using value queries can be tricked in the sense that an instance can be constructed where they must make all possible queries, whereas the random query policy will save queries in expectation on every instance. Recent work on value queries, such as Brero et al. [2017], emphasize how value query-based approaches can be superior to those based on demand queries. They use a machine learning-based approach that we will return to in the next section.

We conclude by briefly describing two less-studied query types. *Bound-approximation* queries ask an agent to tighten its lower and upper bound on the value of a bundle. These queries were motivated by the model where each agent runs an anytime algorithm to determine its value for each bundle [Hudson and Sandholm, 2004] and is effective in that setting.

Order queries ask a buyer which of two bundles they prefer. These result in qualitative preferences (rankings) since the value of bundle can never be fully determined. Order queries have the advantage of being particularly cognitively cheap because they have yes/no answers. Hudson and Sandholm [2004] find that, in a setting where the principal has access to both order and value queries, it is worth mixing them only if order queries cost 10% or less than value queries, and otherwise it is better to use value queries only.

2.3.4 Query Strategies in Preference Elicitation

The *query strategy* describes what queries to ask the agents given the principal's current information, i.e., step 2 of Sandholm and Boutilier's [2006] framework. We restrict our attention to strategies for value queries in this section, both for simplicity and because they are the most studied query type from this perspective. In many cases, the strategies are applicable to other query types as well. For a concrete setting, we refer to the analysis done by Charlin et al. [2012], which studies the problem of matching reviewers to papers for a conference from an active learning perspective. While not a CA, it is very similar from this perspective, and it provides the most complete analysis of query strategy selection. We will refer to work on CAs as well. The choice of query strategy will be of particular importance in Chapter 5 where we introduce a new query type: the relative value query.

It is important to note that most deployed procedures use a demand-query-based meta-strategy that is a generalization of the ascending auction to multiple items [Leyton-Brown *et al.*, 2017; Ausubel *et al.*, 2006; Ausubel, 2004]. At each iteration of the elicitation, all agents respond with what bundles they

would purchase at the current posted prices. The elicitor then raises the prices on the bundles that are demanded by multiple agents, until a feasible allocation is found. This section focuses on alternatives to that approach which may place less cognitive burden on the agents.

In Charlin et al.'s model, each reviewer has a suitability for each paper, and the objective is to maximize an aggregate suitability score subject to the constraints that each paper must have at least a minimum number of reviewers and that no reviewer may be assigned more than a certain number of papers to review. They assume that all reviewers report truthfully to the mechanism and use an integer program formulation for the optimization.

How does this setting compare to a CA? The major simplifying assumption is that agent utility functions are assumed to be linear, i.e., the value of a bundle is equal to the sum of the values of its items. This is a major simplifying assumption that, in particular, allows Charlin et al. to easily apply machine learning to predict the value of items based on the queries asked so far. In the standard CA setting, machine learning is more difficult to apply. It is easy to make predictions, but difficult to apply the predictions to the winner determination problem without querying the machine learning algorithm separately for each bundle. Brero et al. [2017] present an approach that overcomes these limitations—through clever use of support vector regression as the prediction algorithm, they solve the winner determination problem efficiently.

Other than the assumption of linear utility functions, the differences between the paper matching setting and CAs have little impact. The most obvious difference is that no payments are allowed, but since the agents are assumed to be truthful, payments are not needed to align the bidder's individual net utilities with social welfare. They also use a different feasibility constraint—"each paper must be assigned to at least a certain number of reviewers and each reviewer must be assigned at most a certain number of papers"—rather than the CA feasibility of "each item must be assigned to at most one agent". This change in constraints has little impact on elicitation with value queries.

The ideal myopic strategy for elicitation would be to query the reviewer-paper pair that has the highest *expected value of information*—the expected improvement in the objective function relative to the value of the current best matching. However, the authors report that this is too expensive to compute, and they instead test several heuristic querying schemes.

They present two naive strategies that do not take into account the current matching. One is to query the suitability score with highest uncertainty in their prediction model, and the other queries the score with highest estimated mean that has not yet been queried. Then, they present three more sophisticated strategies that take into account the current matching. The simplest of these is to query the reviewer-paper pair with the highest estimated score that is currently matched together in the suitability-maximizing matching. The gist of the strategy—to take the solution that looks the best and query some part of it that maximizes an objective—is sometimes referred to as *current solution* elicitation after Boutilier et al. [2006]. The aforementioned Brero et al. [2017] uses a similar strategy in CAs: at each step, they query every agent-bundle pair that is currently assigned by winner determination that has not yet been queried. In Charlin et al., this strategy has the weakness that the queried reviewer-paper pair may have very low uncertainty.

Empirical Results

Charlin et al. run an empirical analysis of the different query strategies on several different data sets, including paper bidding data from a conference. The authors found that the two naive strategies, un-

certainty sampling and querying the reviewer-paper pair with highest estimated score, are outperformed by querying uniformly at random. The remaining three strategies, which take into account the current matching, outperform random querying in all of their tested domains, indicating that taking the current match into account is critical. It is worth noting, however, that even the best performing strategy outperforms random querying by 5% at most.

Sandholm and Hudson [2004] find that taking the current solution into account in query selection has a large impact in a CA domain, without a machine learning model. They compare the effectiveness of randomly querying an *allocatable* bundle-agent pair, i.e., a bundle-agent pair that appears in a potential solution to the winner determination problem for some completion of the preferences, to querying a random bundle-agent pair. In their experimental setting (synthetic CA problems), they find the random query policy asks around 90% of *all* queries to find a provably optimal matching. Restricting the random queries to allocatable bundle-agent pairs causes a dramatic reduction in the number of queries needed: it requires between 60% and 15% of all queries, depending on the number of items in the auction. Sandholm and Hudson also perform a theoretical analysis on the impact of restricting queries to allocatables. They find that it may increase the number of queries required to find a provably optimal matching, compared to the random strategy, but this increase is by no more than a factor of 2.

It seems likely that the relatively small impact of sophisticated query strategies in Charlin et al.'s setting is explained by the success of the collaborative filtering model. Even with random queries as input, it quickly generates high quality predictions, yielding suitability scores that are hard to beat at the level of query strategy sophistication the authors consider.

The results of Charlin et al., Brero et al. and Hudson and Sandholm indicate that preference prediction has the largest impact on solution quality, at least in domains where preferences are structured enough to leverage at the individual and/or the group level (i.e., when collaborative filtering is informative). Using iterative querying with a heuristic query strategy that takes the current proposed solution into account may have some effect, particularly (i) when the number of allowed queries is low or queries are expensive; (ii) when the prediction model is not that sophisticated; or (iii) when collaborative filtering is not used in prediction. None of these papers tests a very sophisticated query strategy such as estimating the expected value of information. Brero et al.'s results suggest that, at least for their domains, the benefit of such a strategy is concentrated in settings where the allowed number of queries is low.

Truthfulness Under Preference Prediction

Preference elicitation approaches that use preference prediction are not truthful in general. Recall that mechanisms must gather all of the necessary information to run VCG in order to guarantee truthfulness. There are at least two ways that preference prediction can fail to do this. Consider, for example, a simple augmentation of current solution elicitation with preference prediction, which is quite similar to that of Brero et al.

1. At each elicitation step, the proposed solution is the optimal solution w.r.t. to the predictions of the value of each bundle. The algorithm queries an unknown aspect of the proposed solution and updates the predictions.

2. Update the proposed solution. If it changes, repeat. Otherwise, terminate.

This algorithm has two obstacles that prevent truthfulness. First, it may fail to sufficiently explore the space and terminate prematurely. Sometimes agents can exploit this by misreporting in a way that causes other agents to receive queries that are unhelpful to them. Second, the prediction for a bundle may not match the query response, incentivizing the agent to misreport his preferences in order to make the prediction for each bundle match its true value as closely as possible.

One way to address the former is to allow users to "push" information to the elicitor, i.e., answer queries that have not been asked, which is discussed by Sandholm and Boutilier. A scheme that allows pushing can overcome the early termination problem because the agents can continue to provide information to the mechanism until they are satisfied with the outcome.

The latter can sometimes be overcome by applying a filter to the prediction algorithm to ensure that the predictions match the query responses, if doing so does not interfere with tractability of the optimization (for example, this is not possible in Brero et al.'s model).

The incentive properties of VCG are extremely sensitive to the mechanism selecting an outcome that does not maximize social welfare—each agent may have an incentive to defect that is as large as the difference in social welfare between the selected outcome and the outcome that maximizes social welfare. This is the situation that occurs when the latter problem is solved, but the former is not.

Observation 2.5. *Let $\langle N, O, \Theta, \{\theta_i\}, U\rangle$ be a mechanism design problem. Consider a mechanism that selects outcome o' and uses standard VCG payments with the Clarke pivot rule. Given a fixed report $\hat{\theta}_{-i}$ of the other agents, agent i's incentive to misreport is bounded above by*

$$U(\theta_i)(o^*) + \sum_{j \in N : j \neq i} U(\hat{\theta}_j)(o^*) - (U(\theta_i)(o') + \sum_{j \in N : j \neq i} U(\hat{\theta}_j)(o')) \tag{2.13}$$

where o^* is the social welfare maximizing outcome.

A similar bound can be established in the case where neither problem is solved. In unpublished work, Brero et al. prove that, in the CA case, the incentive to misreport is bounded by the largest error in the prediction algorithm. In particular, if the prediction algorithm predicts the value of every *reported* bundle within ϵ, the incentive to misreport is bounded by

$$SW(a^*_U) - SW(a^*_{\tilde{U}}) + \epsilon n \tag{2.14}$$

where $SW(a^*_U)$ is the social welfare of the true optimal allocation and $SW(a^*_{\tilde{U}})$ is the social welfare of the optimal allocation under the query responses with preference prediction applied. Note that this is the same as Observation 2.5, but with an additional ϵn term added. This bound will likely not be very effective in practice because of the multiplicative impact of the error in the prediction algorithm.

In Chapter 5, we will study preference elicitation in an electricity management setting. The choices we make are informed by past work. In particular, we use a preference prediction model where the expected value of information of a query can be computed by sampling. However, our setting differs in a key detail: queries have costs. We find that a sophisticated query strategy has quite a large effect because avoiding queries that yield little information (i.e., less than their cost) produces dramatic savings.

2.4 The Smart Grid and the Role of Computer Science

The traditional role of the electrical grid is to deliver power to consumers at minimum cost with only limited knowledge of how transmission is happening in real time or what the end uses are. The transmission and distribution is treated as a pipe—generators load electricity in at one end and consumers take it out at the other. Agents transacting with the grid are paid a flat rate for electricity they contribute and charged a flat rate for electricity they consume, irrespective of the actions of other agents. In pipe model, the primary challenge in grid operation is of an engineering nature: ensure that there is enough power is available to meet demand, but not so much that the limits of grid operation are exceeded. This view has been challenged by several recent developments: the availability of improved sensors and several societal, political and economic factors that make the traditional view unattractive.

Sensing infrastructure that can monitor the actions of agents on short timescales has become available, transmitting that information to a control center in real time. These sensors allow the system operator to take a more nuanced view of how distribution is occurring and understand how the behavior of consumers and producers is affecting costs. The most important of these sensors for trading purposes is the *smart meter*, which allows the tracking of inputs and outputs to the system in real time. In the pipe model, demand is exogenous. With sensors, various aspects of consumer behavior can be observed, hence, influenced by prices.

In addition to the availability of sensors, several environmental factors have created societal incentives to achieve more responsive control of the grid. Reducing emissions from electricity-producing technologies has become a central strategy to mitigate to climate change. Reduction of carbon dioxide emissions depends on integrating renewable sources of energy into the grid, which can only be made to conform to consumer demand patterns in a limited way. Thus, the level of consumer response to price signals plays a large role in the cost of integrating renewables. Beyond the climate change context, research on the environmental effects of energy extraction for certain sources of electricity provides another reason to integrate renewables.

From a political perspective, the traditional grid system relies on large infrastructure projects that are built and paid for over many years. These projects require significant commitment on behalf of the public because of the debt they create. Infrastructure projects have become difficult to fund in general in recent years, and it is certainly true that they are subject to extensive public debate because of their cost. Adding smart technology to the grid can be done in small batches for low cost, particularly in the form of *microgrids*. Microgrids are autonomous subgrids that are able to detach from the main grid. Much of smart grid technology can be implemented and tested on microgrids without the expense of rebuilding the entire grid.

Thus far, utility companies themselves have been the primary drivers of innovation in smart grid technology, possibly because deregulation in many countries has given them an opportunity to profit from these changes. The most direct approach to increasing efficiency in electrical grids is to use modern technology to assist the grid in performing its traditional role: i) better sensing equipment allows for quicker fault detection and recovery; and ii) actuators prevent problems by adjusting equipment to keep operation within tolerances, or, in the case of a fault, isolating the affected area. Early visions of the smart grid focused on the potential for self-healing and remote operation. More recent perspectives see the opportunity for greater interactivity that is provided by sensing infrastructure. Farhangi [2010] states that the smart grid "empowers consumers to interact with the energy management system to adjust their energy use and reduce their energy costs," increasing the value generated by transacting

with the smart grid.

In this section, we discuss three papers relating to the smart grid: the aforementioned Farhangi [2010], the Williams et al. paper [2012] introduced in the previous chapter, and Ramchurn et al. [2012], a paper focusing on applications of AI to the smart grid.

2.4.1 The Path of the Smart Grid

Farhangi [2010] provides a comprehensive overview of the smart grid: the motivation behind the smart grid, changes to electrical grids so far and predictions of what is coming next. Many shortcomings of the current grid design and implementation are addressed, the three that receive the most attention are: (i) high peak energy usage, (ii) energy lost between generation and use due to the distribution system, and (iii) the spread of failures.

High peak energy usage is a feature of the current grid system that has received significant attention. The oft-cited statistic is that 20% of the grid's capacity exists to meet peak demand only, i.e., it is in use only 5% of the time. The reason is that the demand on the grid has high variance and it is not very responsive to generation costs. Thus, reducing peak demand would result in huge cost savings by allowing the grid's generation capacity to be reduced. Thus, any new program or technology is scrutinized for the effect it might have on peak demand.

Demand response is the industry's term for programs that aim to increase the responsiveness of consumers to changes in generation costs, a concept that far predates the smart grid. Demand response focuses on reducing demand when cost savings are high, which occurs particularly at peak times. Studies have shown that when consumers are motivated to reduce peak demand (e.g., in the case of a crisis, such as systemic electricity shortages) very large reductions can be achieved. However, attempts to reduce usage at peak times under normal circumstances have not been greatly successful. We outline the general types of demand response strategies below. One of the major hopes of the smart grid is that the presence of increased information and communication infrastructure will allow for the development of more effective demand response strategies. It is worth noting, however, that the type of infrastructure that advanced demand response requires is a relatively low priority for the utility companies that are building the smart grid. Initial smart grid construction may focus on sensors and actuators that are used by the utility companies to achieve better control of transmission (long-distance) and distribution (short-distance) infrastructure, rather than developing consumer-facing technology.

Existing demand response interventions have generally fallen under two categories: economic and social. Economic interventions offer financial incentive to consumers to reduce consumption under certain circumstances. The two most prominent are *time-of-use pricing* and *critical peak pricing*. Time-of-use pricing sets an energy cost that varies based on the time of day and potentially the time of the year. Critical peak pricing allows generators to declare critical peak events when the grid is heavily stressed, and increase prices dramatically for a short period of time. One of the major challenges faced by economic interventions is that the variable cost of electricity production is relatively low so, for many consumers, prices have to be increased dramatically to impact their consumption decisions.

Social interventions use social pressure as the means to influence behavior. The most prominent form is comparison with other consumers. By showing a consumer that they consume more than others, they can be motivated to change their behavior. These interventions take on a variety of forms, from annotations on electricity bills to full-on gamification of electricity usage. These interventions are inexpensive to implement, but have only had a modest effect in practice.

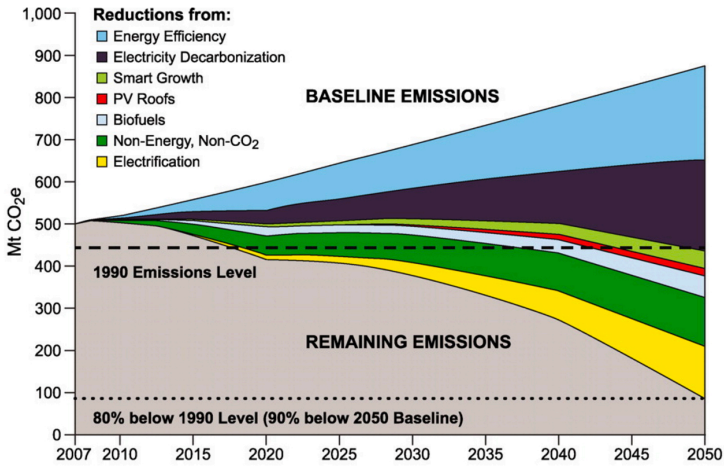


Figure 2.2: The breakdown of emissions reductions from Williams et al.'s [2012] model.

Farhangi's second major focus is deficiencies in the electricity transmission and distribution systems. Much of the North American infrastructure has not been updated in the last half-century and it suffers from many inefficiencies: 8% of electricity production is lost between generation and the end-user largely due to transmission and distribution inefficiencies. In addition, it is highly prone to failure: 90% of all power outages and disturbances are caused by the distribution network. The smart grid has great potential to increase its efficiency by allowing the flow of more fine-grained and up-to-date information throughout the system, rather than allowing sensing only at the end points and requiring site visits to acquire more information.

2.4.2 The Technology Path to Deep Greenhouse Gas Emissions Cuts

Williams et al. [2012] study how to achieve targets to cut greenhouse gas (GHG) emissions by 80% by 2050 in the state of California. They build a physical and economic model of California, including transmission and distribution infrastructure and resource constraints, and attempt to minimize the economic impact required to achieve the target. Their most notable result is that decarbonizing generation and increasing energy efficiency will not suffice—it is necessary to electrify sectors such as transportation.

Figure 2.2 shows projected sources of GHG emissions reductions. The largest reducer is *energy efficiency*—the construction of more energy efficient appliances, vehicles, and industrial processes. Energy efficiency has increased dramatically in the history of electricity. For example, LED bulbs consume about 15% of incandescent bulbs, while providing the same brightness. Unlike other GHG reduction approaches, energy efficiency is often driven by economics. For example, while LED bulbs are more expensive up front, total lifetime cost (including operation) is only about 20% of incandescent bulbs. Because of its attractive economic properties, energy efficiency is considered the most desirable form of GHG emission reduction. Williams et al. calls for a 1.3% improvement to energy efficiency per year, which is an unprecedented level that matches the level of improvement achieved during the California energy crisis of 2000–2001. There is little debate that there is an enormous potential for emissions reductions through energy efficiency, but their goal is at the very top end of what is considered achievable by other studies.

The second largest reducer is electricity decarbonization. For this portion of the analysis, they build

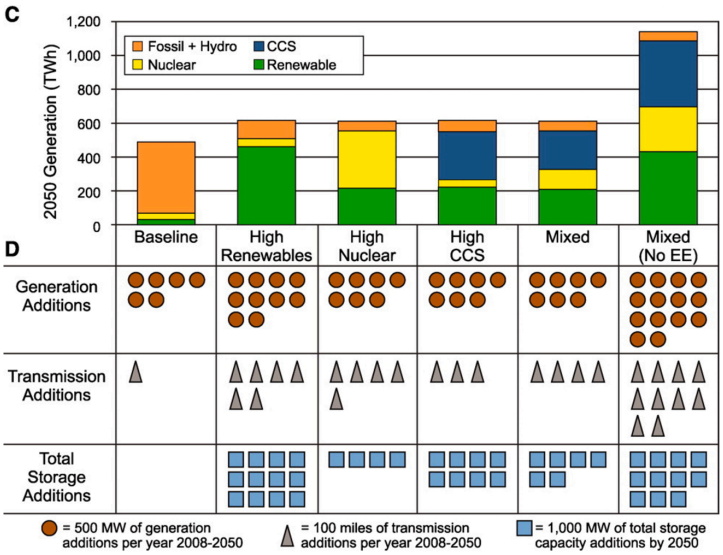


Figure 2.3: Williams et al.'s summary of generation decarbonization scenarios. CCS is carbon capture and storage and EE refers to energy efficiency. The rightmost bar "Mixed (No EE)" is included to demonstrate the impact of energy efficiency in their analysis.

several different decarbonization scenarios, reflecting the uncertainty over which low-carbon forms of electricity will dominate: renewables, nuclear or carbon capture and storage (CCS). CCS refers to the process of capturing CO_2 from the atmosphere and injecting it underground. They do not attempt to evaluate which scenario is superior, and remark that despite the renewable case requiring more generation, transmission and storage, it does not require CO_2 transmission and storage as in the case of CCS or fuel cycle facilities as nuclear does. Their main finding on this dimension is the infeasibility of increasing renewable penetration beyond 74% of total generation while respecting the constraints of their model. This is despite the fact that they make extensive technical assumptions that benefit renewable generation, such as perfect generation forecasting and smart charging of electric vehicles. This result is a subject of intense debate.

As Figure 2.2 shows, these two strategies combined are enough to reduce GHG emissions to 1990 levels, but no further (despite generous assumptions in each case). Most of the rest of the reductions originate from non-energy, non-CO_2 uses, such as those from cement and agriculture, and electrification, particularly of heating and industrial processes. This was a major result at the time the paper was published and is still under debate. These two reduction strategies, which are smaller, but still substantial, had received much less research attention.

Critically for the approach of this thesis, Williams et al. pay little attention to demand response as a source of GHG reductions beyond the load shifting that results from smart charging of electric vehicles. They indicate that they consider demand response to be subsumed under energy efficiency in their model, but the difference is important, particularly in the renewable generation case, because of the ability of demand response to counteract the unreliability of renewable generation. Their model indicates that agents that represent consumers could have significant impact on generation cost.

2.4.3 Artificial Intelligence and the Smart Grid

Ramchurn et al. [2012] suggest that there are many relationships between core tasks in the smart grid and well-studied problems in AI, particularly multi-agent systems. They envision autonomous agents working on behalf of market participants, and that these agents will be faced with the communication and coordination problems that occur in multi-agent systems. They outline several topics that they view as the most interesting from an AI point of view and present them as research challenges to the community. The authors focus on three main topics: coordination of electricity consumers, coordination of small producers, and self-healing networks.

Their first major topic is coordination among electricity-consuming agents. As an example of the risk of uncoordinated behavior, they present the case of "peak-shaving" schemes, i.e., methods to reduce demand at peak times. They contend that current economic approaches to reducing peak demand would not work well in practice, even if agents were more responsive, because they inherently *synchronize* agent behavior, which in turn can cause a detrimental effect on grid infrastructure. For example, if prices are increased between 12 and 1 pm, there will be a peak of energy usage at 11:59 am and 1:01 pm. In addition, they predict that the increased prevalence of electric vehicles will worsen this problem by significantly increasing the peaks of household consumption. They predict that, in order to address these issues, household agents must coordinate, and the behavior of these autonomous systems will have to be analyzed as a group in order to avoid causing problems at an aggregate level.

The second problem they focus on is coordination of small electricity producers. They foresee that many current consumers will be able to generate a small amount of electricity, e.g., from solar panels, wind or electric vehicle batteries, that they will want to sell back to the grid. However, because the electricity generation is on a small scale and unpredictable, it will be necessary for these agents to form groups. Collectively, these groups will be able to produce electricity at a scale and degree of reliability that can match that of traditional generators and thus will be able to produce value for the grid. These groups then have to deal with the problem of coordinating their behavior and dividing the benefits of their cooperation amongst themselves. These are problems of cooperative game theory.

Their third topic is self-healing grids. They foresee every grid component being represented by an autonomous agent. These agents present their sensor outputs and their action space to other parts of the grid. These parts of the grid then coordinate and determine the best joint action, forming a decentralized control system. This approach is very different from the predominant engineering approach where all information and decisions go through a central control center.

The systems proposed by Ramchurn et al. [2012] are decentralized. They model market participants as agents that run their own algorithms and look at the global performance of the resulting system. There is limited game-theoretic consideration because the system designer can choose the algorithms and the behavior differs between market participants only because the actor each represents reports different preferences. They generally assume that actors are not strategic in preference reporting and instead report their true preferences. In cases where agents have the opportunity to defect, the authors propose a trust system, where agents that do not operate honestly are excluded from the group over time, which differs from the game-theoretic approach of trying to incentivize truthfulness even in a one-shot game.

Chapter 3

Approximately Stable Pricing for Coordinated Purchasing of Electricity

In this chapter, we study the problem of coordinating and sharing costs among electricity consumers under complex generation cost functions, which are affected by the "shape" of demand, i.e., its peaks and smoothness. We approach this problem as a matching market, using the tools of cooperative game theory, developing two novel cost-sharing schemes: one based on Shapley values that is "fair," but computationally intensive; and one based on *similarity-based envy freeness*, which captures many of the essential properties of Shapley pricing, but scales to large numbers of consumers. Empirical results show these schemes achieve a high degree of stability in practice and can be made more stable by sacrificing small amounts ($< 2\%$) of social welfare. Some of the work in this chapter originates in Perrault and Boutilier [2015].

3.1 Introduction

Coordinating the group purchase and consumption of electricity can offer significant benefits in terms of economic efficiency, predictability, and fairness. These benefits emerge for several reasons. The first reason is that consumers who are able to shift their loads away from periods of high consumption can be compensated by others for the inconvenience or discomfort of doing so, with the resulting flatter demand profiles reducing overall cost of generation. The increasing penetration of renewable generation makes the issue of demand smoothness particularly pronounced. The most prominent sources of renewable generation, wind and solar, produce significantly more electricity during the day than at night. Thus, the load net of renewables decreases quickly in the morning and increases quickly in the evening. This effect is summarized in Figure 3.1 from Denholm et al. [2015], who remark that the load net of renewables during daytime hours decreased from 2012 to 2013 in California and projected that the trend was likely to continue. An updated figure [Reinhold, 2016] (Figure 3.2), demonstrates that Denholm et al. were correct and that the decrease appeared largely due to solar (at least on the day that he analyzed).

The steep decline and rapid increase of demand caused by renewables put a lot of pressure on

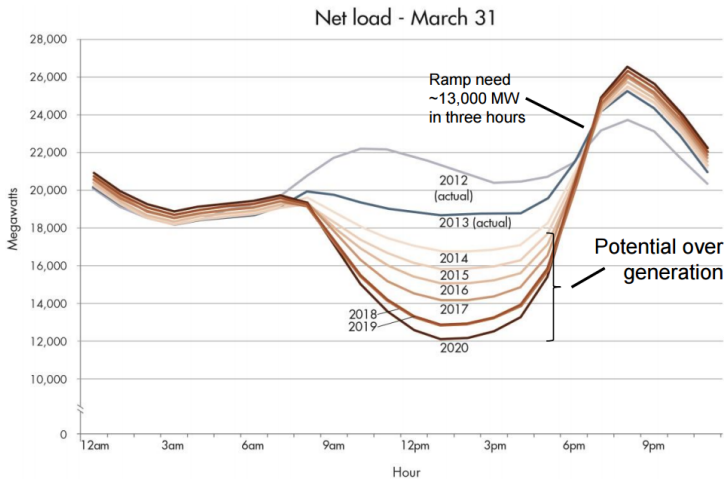


Figure 3.1: The original "duck curve" from Denholm et al. [2015], projecting steep ramping requirements for conventional generation.

conventional generation sources, increasing the cost of generating electricity. This issue is one of the factors that makes it difficult to increase renewable penetration past a certain threshold, which is a focus of debate in the literature (see Section 2.4.2). Coordinating electricity use can increase the ability of demand to match supply and thus reduce the consequences of this effect.

The second reason is economic: the formation of groups of consumers can increase competition on the supply side of the market. Rassenti et al. [2003] found that group purchasing in the electricity setting resulted in prices that are more responsive to market conditions, decreasing the ability of generators to exercise market power. The real-world appeal of this idea was demonstrated in The Big Switch [Waddams *et al.*, 2014], a UK program where 30,000 households agreed to have their demand auctioned collectively to the lowest-bidding provider. Since then, similar programs have been launched in Australia, Scotland and Ireland.[1]

The third benefit is that groups can predict aggregate consumption levels more reliably than individuals and incentivize their members to consume at predicted levels [Robu *et al.*, 2017]. This can have the consequence of reducing generation cost due to better predictions of demand by generators themselves. We do not study this issue in this chapter, but it is the focus of Chapter 4.

In this chapter, we develop a model to facilitate group-level coordination using the tools of cooperative game theory to determine mutually acceptable cost sharing among users in a group. This is a market design problem which we approach from a cooperative game-theoretic perspective. We address the standard concerns of social welfare and stability, but we do not consider the issue of elicitation. Instead, we can run a standard elicitation protocol of the style outlined in Section 2.3.3, i.e., using either demand or values queries. We take this approach because achieving high social welfare and stability is quite difficult in this setting. Our objective is to develop techniques that will scale to thousands of users—on a scale similar to previous group buying projects in electricity.

Electricity markets present a technical challenge due to their size and complexity. Maximizing social welfare in electricity markets requires optimizing the production of electricity by each producer. The producers' generation cost functions can be quite complicated. A typical electricity producer sells

[1] Australia: `https://www.onebigswitch.com.au/`, Scotland: `https://www.onebigswitch.co.uk/` and Ireland: `http://www.onebigswitch.ie/`.

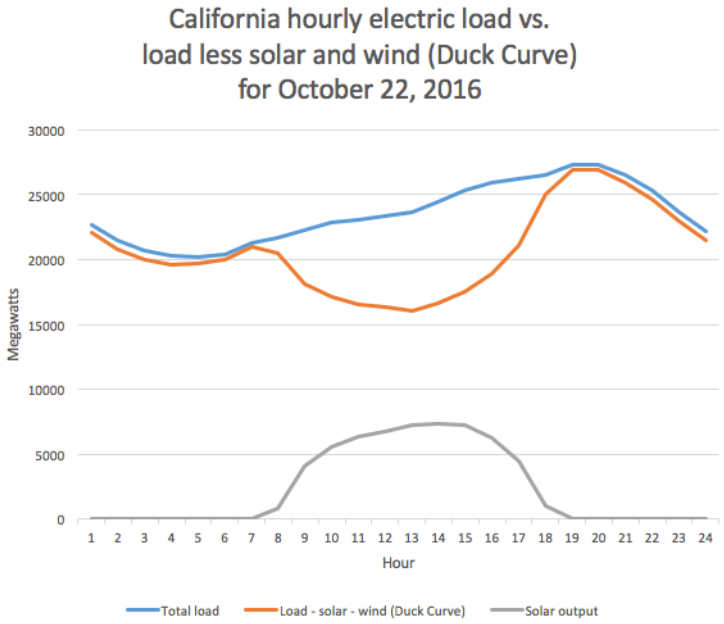


Figure 3.2: An updated "duck curve" for October 22, 2016 [Reinhold, 2016]. The gray line shows solar generation separately.

electricity from a mixture of generation sources, which have different properties. At a high level, we can divide these sources into two groups: *dispatchable* and *non-dispatchable* generation. Dispatchable generation can be controlled by the operator in response to market needs. In California, these sources are primarily natural gas, hydroelectric and coal. Most renewable energy is non-dispatchable. Nuclear power is somewhat dispatchable, but reaction times are slow—it might take 24 hours to adjust the output level. The producer's problem is to optimize the control of dispatchable generation subject to current demand and current generation from non-dispatchable sources. In the general case, this problem is a stochastic optimization due to uncertainty about demand and renewable production. In this chapter, we assume that demand and non-dispatchable generation are perfectly predictable. This simplifying assumption allows us to focus solely on the impact of demand "shape" in the dispatch optimization.

Dispatchable sources have several interesting features [Kirschen and Strbac, 2004].

- Minimum and maximum production levels: coal and nuclear plants have energy generating processes that must be operated within specific limits. Shutting down and starting up such plants also incurs a cost.

- Limitations on *ramping*: the rate of adjustment to output over time is physically constrained.

- Multiple *layers* of generation with different associated costs: there may be an inexpensive *base layer* that is slow to adjust and an expensive but easily adjustable *tracking layer*.

On the other side of social welfare optimization, consumers may be willing to shift their consumption if they are sufficiently compensated. We take a black box view of consumer utility functions and assume that each consumer provides us with a finite set of consumption *profiles*: a vector of electricity consumption over time and the utility that would be accrued by using that profile. In Chapter 5, we begin to address this problem by developing an elicitation approach for HVAC control.

The complexity of the social welfare function poses challenges from a game-theoretic perspective. Outcomes that maximize social welfare may not support stability, either from a coalitional perspective (i.e., core-stability) or from a purely strategic perspective (i.e., Nash stability). In fact, we show that the social welfare of Nash-stable allocations may be arbitrarily worse than the best unstable outcome. Thus, there is a natural tradeoff between stability and social welfare. We explore this tradeoff using two different cost-sharing schemes, one based on Shapley values, and the other based on a new notion of *similarity-based envy freeness*. We show that small sacrifices in social welfare can provide large gains in stability. Furthermore, our similarity-based envy-free cost-sharing scheme, while not as conceptually simple as the Shapley scheme, achieves greater stability and has significantly better computational properties.

The main contributions of this chapter are:

- We develop a tractable market model for matching consumers to producers while reflecting many of the complexities of electricity production and consumption.
- We explore the stability properties of this model under various cost-sharing schemes.
- We develop two payment algorithms that exhibit high stability and fairness, while allowing tradeoffs between social welfare and stability.

In Section 3.2, we describe the basic market model and related work. Section 3.3 describes the form of producer price functions we use and how social welfare can be optimized as a mixed integer program, Section 3.4 addresses stability in the model and describes two payment models. Section 3.5 describes the model of consumer electricity used in the experiments, and the experiments in Section 3.6 demonstrate the efficacy of our new payment algorithms.

3.2 Setting

To build a market model, we need to specify the utility functions of the consumers and the producers. We take the perspective that each consumer has a finite number of discrete *demand profiles*, where each profile reflects an "acceptable" consumption pattern (electricity use per period in kilowatt-hours (kWh)) along with a value for consuming according to that demand profile. These values indicate her preferences for consumption and her degree of potential flexibility.

Each producer has a *producer price function* (PPF). The price function is an arbitrary function that captures the *prices* posted by the producer for serving demand of different shapes. It represents the outcomes of the generation dispatch problem. We make the simplifying assumption that producers have regulatory profit margins which are included in their price function—they do not act strategically to increase their profit. In Section 3.3, we develop a model of PPFs that aims to capture many of the features of real-world price functions.

Given these utility models, our goal is find a matching of consumers to producers and consumers to demand profiles such that social welfare is maximized *and* stability is high. We use stability concepts that are developed for this particular setting, which are described in Section 3.4. We optimize social welfare using a *mixed-integer program (MIP)*. Stability is difficult to optimize directly—we instead take an ad hoc approach to maximizing it.

Formally, we represent a market as $\langle N, M, T, \{\Pi_i : i \in N\}, \{V_i : i \in N\}, \{\alpha_i : i \in N\}, \{P_j : j \in M\}\rangle$. Let N be a set of n consumers and let M be a set of m electricity producers that each control a set

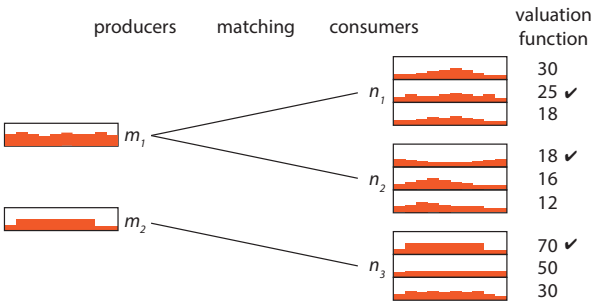


Figure 3.3: The market model. The demand profiles chosen by the consumers depend on the prices charged by the producers. Each producer has a posted price function, which is not shown here.

of generation facilities. We assume T time periods, representing, for example, the hours in a day or week. Each consumer i has a non-empty set of *demand profiles* Π_i, where each profile $\boldsymbol{\pi} \in \Pi_i \subset \mathbb{R}^T$ reflects an "acceptable" consumption pattern (electricity use per period in kilowatt-hours (kWh)) for i. Each consumer has a *valuation function* $V_i : \Pi_i \to \mathbb{R}$ indicating her value (in dollars) for each of her demand profiles. Such profiles may be explicitly elicited or estimated using past consumption data. This is an abstraction of reality, of course, because consumers' valuations derive from the actions that use the available electricity, not from the electricity itself.

A *matching* μ maps each consumer i to a producer $\mu(i)$, from whom she purchases electricity, and demand profile $\mu^p(i) \in \Pi_i$, indicating her consumption. In this matching, i pays the price per unit *posted* by producer $\mu(i)$, which may depend on the aggregate demand of all agents matched to $\mu(i)$. In order to satisfy individual rationality, we include a *null producer* which represents any consumer's best outside option—we can imagine that some consumers may have access to a source of electricity outside the main market, e.g., if they are connected to a nearby source of renewable electricity via private infrastructure. The value i being matched to null is α_i. Figure 3.3 shows a diagram of an instance of the model with two producers and three consumers, each with three demand profiles. The total demand that is served by each producer is the sum of the demands across the profiles chosen by the consumers matched to that producer.

Each producer j has a *price function* $P_j : \mathbb{R}^T \to \mathbb{R}$ that maps total demand in each time period to a price. We treat P_j as exogenous—it represents j's *posted prices*. We assume that P_j captures the fundamental features of generation costs (see next subsection), and that producers are not strategic, instead they simply recover their costs. The null producer has a fixed zero price—the net value α_i accrued to the consumer includes any price charged.

The *social welfare* of matching μ is the net utility realized by all consumers acting on the demand profiles selected by μ, i.e., the sum of the consumers' valuations for the selected profiles minus the sum of the corresponding producers' prices:[2]

$$SW(\mu) = \sum_{i \in N} V_i(\mu^p(i)) - \sum_{j \in M} P_j \left(\sum_{i' \in \mu^{-1}(j)} \mu^p(i') \right) \tag{3.1}$$

Lu and Boutilier [2012], henceforth LB, prove that social welfare maximization in LB is NP-hard via reduction to knapsack. Theorem 3.1 shows that maximizing social welfare in an instance of LB can be

[2]Note that we consider the producer's profit to be part of the cost of generating electricity.

reduced to maximizing social welfare in our electricity market model with no increase in problem size. This shows that maximizing social welfare in our model is NP-hard even for the small class of PPFs used in the reduction.

LB is inspired by both matching markets in supply chains and online daily deal providers such as Groupon and Living Social. A set of buyers is each looking to buy an item from one of a set of sellers. The items sold by the sellers are only partial substitutes—thus, each buyer has a potentially different valuation for the item sold by each seller. In addition, the price of each seller decreases as the number of buyers matched to that seller increases. In particular, there is a set of *discount thresholds* and prices decrease every time the number of buyers surpasses a threshold.

Formally, an instance of LB is represented as $\langle N_{LB}, M_{LB}, \{V_i^{(LB)} : i \in N_{LB}\}, \{\delta_j : j \in M_{LB}\}\rangle$. N_{LB} is a set of buyers, and M_{LB} is set of sellers. $V_i^{(LB)} : M_{LB} \to \mathbb{R}$ represents buyer i's value for the product of each seller. $\delta_j = (\tau_j, p_j)$ represents the discount schedule for seller j: $\tau_j = [\tau_j^{(1)}, \ldots, \tau_j^{(d)}]$ is a vector of discount thresholds, and $p_j = [p_j^{(0)}, \ldots, p_j^{(d)}]$ is a vector of positive prices. If the number of buyers assigned to j is at least $\tau_j^{(k)}$ but less than $\tau_j^{(k+1)}$, vendor j sells the item at price $p_j^{(k)}$. Formally, the social welfare of a matching μ in LB is

$$SW_{LB}(\mu) = \sum_{i \in N_{LB}} V_i(\mu(i)) - \sum_{j \in M_{LB}} |\mu(j)| p_j(|\mu(j)|). \tag{3.2}$$

where $p_j(q)$ is the price that producer j should charge to q buyers according to δ_j.

Theorem 3.1. *Maximizing social welfare in LB can be reduced to maximizing social welfare in our model.*

Proof. Consider an instance of LB $\langle N_{LB}, M_{LB}, \{V_i^{(LB)} : i \in N_{LB}\}, \{\delta_j : j \in M_{LB}\}\rangle$. We want to construct an instance of our electricity market model $\langle N, M, T, \{\Pi_i : i \in N\}, \{V_i : i \in N\}, \{\alpha_i : i \in N\}, \{P_j : j \in M\}\rangle$ that represents this LB instance.

We let $N = N_{LB}$ and $M = M_{LB}$. We assign a different time period to each agent, i.e., let $T = N$. Each agent $i \in N$ has a single demand profile that demands a single unit of electricity in time period i and 0 otherwise.

Because our consumer valuation functions do not allow for different values for different producers, we instead represent these values in the prices charged to the consumers. Thus, we let $V_i = 0$ and $\alpha_i = 0$.

P_j needs to charge the correct discounted price as well as "reimbursing" each buyer for the utility they should receive by buying from seller j. To recover the number of buyers matched to j, we sum the demand profiles matched to j. We define $P_j(\boldsymbol{x})$ as follows, where $\boldsymbol{x} \in R^T$ is the total demand matched to j:

$$P_j(\boldsymbol{x}) = \left(\sum_{t \in T} \boldsymbol{x}_t\right) p_j \left(\sum_{t \in T} \boldsymbol{x}_t\right) - \sum_{t \in T} V_t^{(LB)}(j) I[\boldsymbol{x}_t > 0] \tag{3.3}$$

Recall that agent i's demand only exists at time t. Under this price function, the producer's price function for a particular demand vector is exactly what it would be if it were matched to the same set of buyers in LB. Likewise, the social welfare for a particular matching is identical. Formally, suppose we have a matching from consumers to producers in the electricity market model. No matter which producer consumer i is matched to, her valuation for her assigned profile is zero. Thus, the left term of the social welfare expression (3.1) is zero. The right term is given by 3.3: it is the per unit (potentially

discounted) price times the number of units sold minus, for each producer j, the sum of the consumer values for being matched to j for the consumers that are matched to j. This is exactly equal to the social welfare expression for LB for μ (3.2). □

Related Work

Assignment games and matching markets have been extensively studied using different stability concepts and pricing models [Shapley and Shubik, 1971; Gale and Shapley, 1962; Demange *et al.*, 1986]. Research in real-world markets has largely focused on revenue maximization for monopolistic sellers, though strategic aspects are sometimes considered. The literature on group buying, summarized in [Anand and Aron, 2003; Chen and Roma, 2010], considers the value of offering discounts to groups of buyers who purchase items in bulk. Several group buying models are similar to ours.

Our work extends that of Lu and Boutilier [2012], where the focus is on a more restrictive model of buyer preferences (unit demand, only the supplier affects utility) and seller price functions (volume discounts). Similarly to them, we focus on the strategic behavior of buyers and treat seller prices as exogenous (strategic behavior of sellers was later investigated by Meir et al. [2014]). Two similar group buying models are those of Anand and Aron [2003] and Chen et al. [2007]. Both have seller prices that are affected by the amount purchased, but neither allow for sufficiently complex price functions to model electricity generation. Anand and Aron focuses on a single vendor and does not consider buyer coordination, while Chen et al. uses a multi-stage auction mechanism.

In the AI literature, the process of finding an optimal seller for a group of fully cooperative buyers has been studied [Sarne and Kraus, 2005; Manisterski *et al.*, 2008]. In the context of electricity, group purchasing has been suggested as a way of reducing seller uncertainty about stochastic buyer demands [Robu *et al.*, 2017], an aspect which we do not consider here, but which is the focus of Chapter 4.

3.3 Producer Price Functions (PPFs)

Deciding the optimal output levels of a group of generation facilities in order to meet demand is complex and has been studied extensively [Kirschen and Strbac, 2004]. We focus on three of its most important features. (i) Generation facilities have limited *ramp rate*—the amount by which they can change their output from one time period to the next. Ramp rates of different generation facilities vary radically (e.g., demand tracking plants such as natural gas can ramp up or down in half an hour, whereas nuclear plants make take 24 hours). (ii) Different kinds of generation have different variable costs (e.g., most renewables have low variable cost, while natural gas has high variable cost). (iii) Finally, certain kinds of plants (e.g., coal) have high costs when run below a certain level, which imposes considerable wear on the components. Shutting down these plants also incurs costs.

To capture these features, we model each producer as follows. It has a *base layer* that has low generation costs, but a low ramp rate, and the base layer is expensive to take below a certain level of generation in any time period (the *minimum economic generation level (MEGL)*). It also has a *tracking layer* that can be adjusted rapidly or shut off entirely, which has high generation costs and limited capacity.

We provide an overview of the form of *producer price functions (PPFs)* that we use below. For producer j's base layer, let $c_j^{(l)}$ be the price per kWh, $d_j^{(l)+}$ be the capacity (kWh), $d_j^{(l)-}$ be the MEGL (kWh), and r_j be the maximum ramp rate between periods (kWh). Let s_j be the *shutdown cost* (in

Symbol	Meaning
P_j	Price function. $P_j = BC_j + RC_j + SC_j$
$d_j^{(l)-}$	MEGL for the base layer
$d_j^{(l)+}$	Maximum capacity of the base layer
$c_j^{(l)}$	Per unit price for electricity provided through the base layer
r_j	Maximum ramp rate of the base layer
s_j	Shutdown cost that is incurred when base layer generation is below the MEGL
$d_j^{(h)+}$	Maximum capacity of the tracking layer
$c_j^{(h)}$	Per unit price for electricity provided through the tracking layer
BC_j	Base cost function
RC_j	Ramp cost function
SC_j	Shutdown cost function

Figure 3.4: Table of notation for producer cost function for producer j.

dollars) that is incurred when demand is reduced below the MEGL. Let $c_j^{(h)}$ be the price of the tracking layer per kWh and $d_j^{(h)+}$ be its capacity (kWh). The notation is summarized in Figure 3.4.

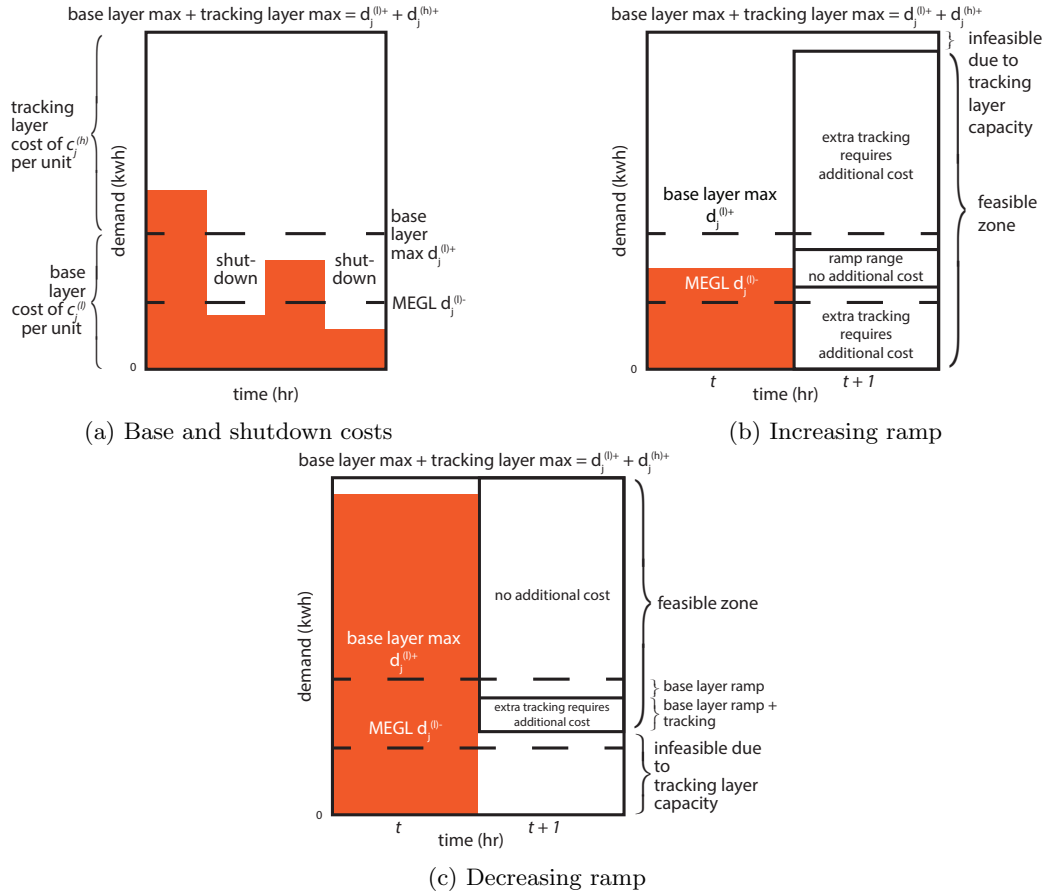
- If demand is smooth and does not exceed the maximum base layer capacity or fall short of the MEGL, only base layer costs are incurred. Formally, if demand in every period is in the interval $[d_j^{(l)-}, d_j^{(l)+}]$, and the largest period-to-period change in demand does not exceed r_j, the unit price is $c_j^{(l)}$. Demand that exceeds the base layer capacity will be met using the tracking layer, if capacity is available, at a price $c_j^{(h)}$. In Figure 3.5a, the total cost to meet demand in the first time period is $c_j^{(l)} d_j^{(h)+}$ plus $c_j^{(h)}$ times the amount of demand that exceeds $d_j^{(h)+}$.

- A shutdown cost is charged if demand in a period is less than the MEGL and the demand in the previous period was greater than the MEGL. If demand in the previous period is greater than $d_j^{(l)-}$ and demand in the current period is less than $d_j^{(l)-}$, the shutdown cost of s_j is charged. In Figure 3.5a, a shutdown occurs in the second period.

- If there is a large increase in demand between two periods, the first r_j units of the increase are met using the base layer at price $c_j^{(l)}$, and the remaining units of the increase are met using the tracking layer at price $c_j^{(h)}$. Figure 3.5b shows the ramp costs that are incurred at time $t+1$ given a moderate demand at time t. Note that the base or tracking layer may have insufficient capacity, which would result in a demand profile that cannot be served, i.e., it has infinite cost.

- If there is a large decrease in demand from period to period, an additional fee of $c_j^{(h)} - c_j^{(l)}$ per unit of decrease exceeding r_j is charged, which represents the cost of meeting the necessary amount of the previous period's demand using the tracking layer. Figure 3.5c shows the ramp costs incurred at time $t+1$ given a high demand at time t.

Our PPFs have the Markov property: the price paid in any period depends only on demand in that period and in the previous one, which makes them easy to compute. It is also gives a lower bound on the cost of meeting the demand by optimizing base and tracking layer production levels in each time period.



(a) Base and shutdown costs

(b) Increasing ramp

(c) Decreasing ramp

Figure 3.5: Diagrams showing how the various components of the producer price function are calculated.

Every cost incurred in the price function must be also be incurred by any solution that satisfies the generation constraints, but the price function may underestimate costs (e.g., it assumes that base layer ramping can be performed within two time periods). Our general approach for optimizing social welfare and stability can be applied to a variety of PPFs—the form of the PPF may be application-dependent.

The key requirement for our approach is that PPFs can be easily represented as constraints in a MIP. PPFs need not be Markov to satisfy this requirement. In the electricity domain, we could represent more complex PPFs as Markov by augmenting the state vector. For example, if there is a requirement that a production source be shutdown a certain percentage of the time, we can add that statistic to the state directly.

If we want to perform an optimization to determine the value of a PPF, we could split the optimization into two parts and iterate, repeating until convergence:

- Maximize social welfare subject to an PPF approximation.
- Improve the PPF representation around the social welfare maximum.

This procedure increases computational cost, but that may not be an issue, particularly in the case where we compute payments separately from social welfare maximization, so that we only need to maximize social welfare once.

We now formalize our model of PPFs. The price P_j charged by the producer is the sum of three non-negative functions: the base cost, ramp cost and shutdown cost.

Let $BC_j(x)$ represent the base cost of meeting demand at level x for producer j, assuming no additional ramping costs. Demand is preferentially met using the base layer, and the tracking layer is used only if necessary.

$$BC_j(x) = \begin{cases} c_j^{(l)} x & \text{if } 0 \leqslant x \leqslant d_j^{(l)+} \\ c_j^{(l)} d_j^{(l)+} + c_j^{(h)} \left(x - d_j^{(l)+} \right) & \text{if } d_j^{(l)+} < x \leqslant d_j^{(l)+} + d_j^{(h)+} \\ \infty & \text{if } x > d_j^{(l)+} + d_j^{(h)+} \end{cases} \tag{3.4}$$

Let $RC_j(x_1, x_2)$ represent the cost of ramping from generation level x_1 to generation level x_2. The ramp cost is zero if both demand levels are on the tracking layer or they differ by no more than the ramp rate of the base layer. Otherwise, the ramp cost is the cost of using the tracking layer to cover the amount of change exceeding the ramp rate by substituting it for generation that would ordinarily be served using the base layer. This substitution may occur in the generation schedule of the current period or that of the previous period. The amount of additional ramp capacity required may not exceed the total unused capacity of the tracking layer in the relevant (current or previous) period. This gives rise to the following ramp cost function.

$$RC_j(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 \in \left(d_j^{(l)+}, d_j^{(l)+} + d_j^{(h)+} \right] \wedge x_2 \in \left(d_j^{(l)+}, d_j^{(l)+} + d_j^{(h)+} \right] \\ 0 & \text{if } x_1 \in \left[0, d_j^{(l)+} + d_j^{(h)+} \right] \wedge x_2 \in \left[0, d_j^{(l)+} + d_j^{(h)+} \right] \wedge \\ & \quad |x_2 - x_1| \leqslant r_j \\ (c_j^{(h)} - c_j^{(l)})(x_2 - x_1 - r_j) & \text{if } x_1 \in \left[0, d_j^{(l)+} \right] \wedge x_2 \in \left[0, d_j^{(l)+} \right] \wedge x_2 - x_1 - r_j \in \left(0, d_j^{(h)+} \right] \\ (c_j^{(h)} - c_j^{(l)})(x_1 - x_2 - r_j) & \text{if } x_1 \in \left[0, d_j^{(l)+} \right] \wedge x_2 \in \left[0, d_j^{(l)+} \right] \wedge x_1 - x_2 - r_j \in \left(0, d_j^{(h)+} \right] \\ (c_j^{(h)} - c_j^{(l)}) \max \left(d_j^{(l)+} - x_1 - r_j, 0 \right) & \text{if } x_1 \in \left[0, d_j^{(l)+} \right] \wedge x_2 \in \left(d_j^{(l)+}, d_j^{(l)+} + d_j^{(h)+} \right] \wedge \\ & \quad x_2 - x_1 - r_j \in \left(0, d_j^{(h)+} \right] \\ (c_j^{(h)} - c_j^{(l)}) \max \left(d_j^{(l)+} - x_2 - r_j, 0 \right) & \text{if } x_1 \in \left(d_j^{(l)+}, d_j^{(l)+} + d_j^{(h)+} \right] \wedge x_2 \in \left[0, d_j^{(l)+} \right] \wedge \\ & \quad x_1 - x_2 - r_j \in \left(0, d_j^{(h)+} \right] \\ \infty & \text{otherwise} \end{cases} \tag{3.5}$$

The last component component required is the shutdown cost $SC_j(x_1, x_2)$. The shutdown cost is charged if the previous demand level is above the MEGL and the current demand is below it:

$$SC_j(x_1, x_2) = \begin{cases} s_j & \text{if } x_1 \geqslant d_j^{(l)-} \wedge x_2 < d_j^{(l)-} \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

Note that this formulation represents a lower bound on a model that calculates the cost by optimizing over the use of the base and tracking layer directly. This is because every cost incurred in the estimation must be also be incurred by any solution that satisfies the generation constraints, but we allow several constraints to be violated. For example,

- In the case of a large ramp downward, the amount of power that is required to be transferred from

Symbol	Meaning
$y_{i,j,k}$	Binary variable indicating whether consumer i is matched to producer j and using profile k
$w_{j,t}^{BC}$	Continuous variable indicating the base layer cost for producer j in time period t
$I_{j,t}^{RC}$	Auxiliary binary variable used in the calculation of the ramp cost
$w_{j,t}^{RC}$	Continuous variable indicating the amount of ramp beyond the ramp rate required
$I_{j,t}^{SC}$	Binary variable indicating whether production is above or below MEGL
$J_{j,t}^{SC}$	Auxiliary binary variable used in the calculation of the shutdown cost
$\Phi(\cdot)$	Abbreviation used to represent the constraints of a particular cost component

Figure 3.6: Table of variables for MIP optimization for producer j in time period t.

the base layer to the tracking layer in the previous period could be large enough that it forces the base layer generation to decrease below the MEGL.

- Infeasible ramps may be performed because we assume that the base layer can ramp to any level below its maximum capacity within two time periods.

3.3.1 Mixed Integer Program Encoding

We can develop a MIP encoding of the PPF that we use for social welfare optimization. Figure 3.6 summarizes the notation in this section. It is important to note that, because the price function is the sum of three non-negative functions on the demand of each consecutive pair of time intervals, social welfare optimization attempts to minimize the posted prices—we will use this "pressure" to simplify the required constraints.

The first component function is the base cost function $BC_j(x_{j,t})$. This function can be written as:

$$BC_j(x_{j,t}) = c_j^{(l)} \min\left(x_{j,t}, d_j^{(l)+}\right) + c_j^{(h)} \max\left(0, x_{j,t} - d_j^{(l)+}\right) \tag{3.7}$$

plus the condition that $x_{j,t}$ is less than the capacity of the generation system, $x_{j,t} \leqslant d_j^{(l)+} + d_j^{(h)+}$. This is represented in the social welfare optimization as the following linear constraints $\Phi(BC_j, t)$, introducing one continuous variable (we use I and J for auxiliary variables that are intended to take on binary values, even if they are later relaxed to be continuous, and w for continuous variables):

$$\begin{aligned} BC_j(x_{j,t}) &= w_{j,t}^{(BC)} \\ w_{j,t}^{(BC)} &\geqslant c_j^{(l)} x_{j,t} \\ w_{j,t}^{(BC)} &\geqslant c_j^{(h)} x_{j,t} - d_j^{(l)+}\left(c_j^{(h)} - c_j^{(l)}\right) \\ x_{j,t} &\leqslant d_j^{(l)+} + d_j^{(h)+} \end{aligned} \tag{3.8}$$

The second component function is the ramping cost between two periods, $RC_j(x_{j,t}, x_{j,t+1})$. The ramping function can be written as:

$$RC_j(x_{j,t}, x_{j,t+1}) = \left(c_j^{(h)} - c_j^{(l)}\right) \max\left(\min\left(x_{j,t+1}, d_j^{(l)+}\right) - x_{j,t} - r_j, \min\left(x_{j,t}, d_j^{(l)+}\right) - x_{j,t+1} - r_j, 0\right) \tag{3.9}$$

plus the constraint that $|x_{j,t} - x_{j,t+1}| - r_j \leqslant d_j^{(h)+}$. To write these in the MIP, we introduce a binary

variable $I_{j,t}^{(RC)}$ which is 1 if and only if the base layer of producer j is saturated at time t, i.e., $x_{j,t} \geqslant d_j^{(l)+}$. We denote the following constraints as $\Phi(RC_j, t, t+1)$:

$$\begin{aligned}
RC_j(x_{j,t}, x_{j,t+1}) &= \left(c_j^{(h)} - c_j^{(l)}\right) w_{j,t}^{(RC)} \\
I_{j,t}^{(RC)} &\in \{0,1\} \\
I_{j,t+1}^{(RC)} &\in \{0,1\} \\
w_{j,t}^{(RC)} &\geqslant x_{j,t} - x_{j,t+1} - r_j - U I_{j,t}^{(RC)} \\
w_{j,t}^{(RC)} &\geqslant d_j^{(l)+} - x_{j,t+1} - r_j - U\left(1 - I_{j,t}^{(RC)}\right) \\
w_{j,t}^{(RC)} &\geqslant x_{j,t+1} - x_{j,t} - r_j - U I_{j,t+1}^{(RC)} \\
w_{j,t}^{(RC)} &\geqslant d_j^{(l)+} - x_{j,t} - r_j - U\left(1 - I_{j,t+1}^{(RC)}\right) \\
w_{j,t}^{(RC)} &\geqslant 0 \\
x_{j,t} &\leqslant d_j^{(l)+} + U\left(1 - I_{j,t}^{(RC)}\right) \\
d_j^{(l)+} &\leqslant x_{j,t} + U I_{j,t}^{(RC)} \\
x_{j,t} - x_{j,t+1} - r_j &\leqslant d_j^{(h)+} \\
x_{j,t+1} - x_{j,t} - r_j &\leqslant d_j^{(h)+}
\end{aligned} \tag{3.10}$$

where U is large enough that U is greater than

$$\max(|x_{j,t} - x_{j,t+1} - r_j|, |x_{j,t+1} - x_{j,t} - r_j|, |d_j^{(l)+} - x_{j,t} - r_j|, |x_{j,t} - d_j^{(l)+}|) \tag{3.11}$$

for any j and t. The fourth-to-last and third-to-last constraints confirm that the minimum selected by the optimization is the true minimum. This will be needed later, but it is not necessary for the simple social welfare optimization.

The third and final component is the shutdown cost $SC_j(x_{j,t}, x_{j,t+1})$, which can be written as:

$$SC_j(x_{j,t}, x_{j,t+1}) = s_j I\left[x_{j,t} \geqslant d_j^{(l)-}\right] I\left[x_{j,t+1} < d_j^{(l)-}\right] \tag{3.12}$$

We charge the shutdown cost only when a producer shifts from being above the MEGL to below it; after incurring the shutdown cost once, the producer can remain below the MEGL for any number of consecutive periods without incurring additional penalties. This is represented by the following constraints $\Phi(SC_j, t, t+1)$ in the MIP, introducing one binary variable per time period $I_{j,t}^{(SC)}$ which

represents whether the load in period t on producer j is below the MEGL, i.e., $x_{j,t} \leqslant d_j^{(l)-}$:

$$\begin{aligned} SC_j(x_{j,t}, x_{j,t+1}) &= s_j J_{j,t}^{(SC)} \\ I_{j,t}^{(SC)} &\in \{0,1\} \\ I_{j,t+1}^{(SC)} &\in \{0,1\} \\ x_{j,t} - d_j^{(l)-} &\geqslant -U I_{j,t}^{(SC)} \\ d_j^{(l)-} - x_{j,t} &\geqslant -U(1 - I_{j,t}^{(SC)}) \\ x_{j,t+1} - d_j^{(l)-} &\geqslant -U I_{j,t+1}^{(SC)} \\ d_j^{(l)-} - x_{j,t+1} &\geqslant -U(1 - I_{j,t+1}^{(SC)}) \\ J_{j,t}^{(SC)} &\geqslant I_{j,t+1}^{(SC)} - I_{j,t}^{(SC)} \\ J_{j,t}^{(SC)} &\geqslant 0 \end{aligned} \tag{3.13}$$

U again denotes a large constant, greater than $|x_{j,t} - d_j^{(l)-}|$, for any j and t.

With the components assembled, we will combine them into a single optimization problem. Our decision variables are the binary matching variables $y_{i,j,k}$, each of which is true if and only if household i is matched to producer j and household i is using demand profile $\boldsymbol{\pi}_i^{(k)}$.

$$\begin{aligned} \max_{\boldsymbol{y},\boldsymbol{w},\boldsymbol{I}} & \sum_{i\in N} \sum_{k\in\Pi_i} \sum_{j\in M} y_{i,j,k} V_i\left(\boldsymbol{\pi}_i^{(k)}\right) - \sum_{j\in M} \sum_{t\in[T]} BC_j(x_{j,t}) - \\ & \sum_{j\in M} \sum_{t\in[T-1]} (RC_j(x_{j,t}, x_{j,t+1}) + SC_j(x_{j,t}, x_{j,t+1})) - \sum_{j\in M} s_j(1 - I_{j,0}^{(SC)}) \end{aligned}$$

subject to

$$\begin{aligned} y_{i,j,k} &\in \{0,1\} && \forall i \in N, \forall j \in M, \forall k \in \Pi_i \\ \sum_{j\in M} \sum_{k\in\Pi_i} y_{i,j,k} &= 1 && \forall i \in N \\ x_{j,t} &= \sum_{i\in N} \sum_{k\in\Pi_i} \pi_{i,t}^{(k)} y_{i,j,k} && \forall j \in M, \forall t \in [T] \\ \Phi(BC_j, t) & && \forall j \in M, \forall t \in [T] \\ \Phi(RC_j, t, t+1) & && \forall j \in M, \forall t \in [T-1] \\ \Phi(SC_j, t, t+1) & && \forall j \in M, \forall t \in [T-1] \end{aligned} \tag{3.14}$$

Note that we define several abbreviations: $x_{j,t}$ represents the total demand assigned to producer j in period t and the function symbols BC_j, RC_j and SC_j, which are defined as before. These abbreviations can be eliminated by substituting the expression for each, which is specified as a constraint, in each place where they occur. The Φ symbols represent sets of constraints as defined earlier in this section. In our case, we interpret it as representing the amount of power loss incurred when power is delivered from producer j to consumer i, which we will assume is proportional to the distance between them.

The MIP has $NMT + 3MT \in O(NMT)$ binary variables and $2MT \in O(MT)$ continuous variables, and $18MT \in O(MT)$ constraints. It can be solved by relaxing the binary matching variables $y_{i,j,k}$, leaving a number of binary variables proportional to MT, and independent of the number of consumers

and profiles. As in many matching problems, most relaxed variables are integral at the optimal solution in practice (just a few consumers may have their demand split across several generators). It may be acceptable for such agents to have contracts split across generators; otherwise, LP rounding may be used. Social welfare maximization can be solved for large instances, since the number of producers (requiring integer variables) is generally very small compared to the number of consumers.

3.4 Cost Sharing and Stability Concepts

Finding a social welfare optimal matching is relatively straightforward, although somewhat involved due to the complexities of the producer price functions. More difficult is the question of appropriate cost sharing among the group of consumers. By coordinating demand to maximize social welfare, some consumers sacrifice their own utility for the benefit of the group and thus should be compensated. Various notions of stability can be used for this purpose. Recall that, given some cost-sharing scheme, stability measures the incentive for any consumer to defect, i.e., change their profile or producer. We show below that core stability is not a particularly useful concept in this setting, and we thus focus on Nash stability. To do this, we need to price defections, i.e., what does a consumer pay if they change their matching. We approach the issue of defection pricing from two perspectives: a *marginal cost defection model*, where a producer accepts any defector who pays the marginal cost they impose by defecting; and an *envy-free defection model* where a defector pays the same as any other consumer originally matched to that producer with a similar profile.

Other than stability and budget balance (i.e., all producers' costs are paid), there are several other desiderata for a cost-sharing scheme. A matching is *envy-free* if no consumer would prefer the matched pair of any other consumer. This notion requires some generalization in our model (as we discuss below). A scheme should be transparent: it should be clear why a consumer is paying what they pay, and what they can do to change what they pay. It should be deterministic and easy to describe so that outcomes do not appear arbitrary. We also desire computational scalability. Finally, we would like to be able to sacrifice social welfare to achieve these properties, especially stability, in a controllable way. As noted in the background, history shows that stability in particular is an essential component of successful matching mechanisms [Roth, 2002].

First, we analyze cost sharing under the marginal-cost defection model and find that stability is difficult to achieve. Second, we present two cost-sharing schemes under the envy-free defection model that achieve many of our desiderata.

3.4.1 Cost Sharing under the Marginal-Cost Defection Model

One difficulty in defining the cost/price of a defection is that the cost imposed by a consumer on a producer depends on the consumers already matched to that producer. We address this with two cost-sharing schemes below. Here we begin by considering *marginal cost payments*—where a consumer defecting to a new producer, or changing profile, pays the marginal cost imposed on the new producer (or existing producer)—and their stability properties.

Ideally, we would like payments to be core stable, wherein no group of consumers can benefit by defecting. Core stability is not always achievable in this game: since LB can be reduced to our model, their proof that the core may be empty (under much simpler producer cost functions) applies to our model as well. Core stability is a very strong notion that is hard to attain without sacrificing social

welfare. Fortunately, achieving it is not critical in practice if large groups of consumers cannot effectively discover, communicate and coordinate their actions.

A weaker, but more practical concept is Nash stability, which accounts for defections by single consumers. It is informative to examine the pure Nash equilibria (NE) with the worst and best social welfare.

Figure 3.1 shows our results for the *price of stability (PoS)* and *price of anarchy (PoA)* under producer price functions with various combinations of model features. The PoS results form a dichotomy. With capacity constraints, but without ramp or shutdown constraints, there is a Nash equilibria that maximizes social welfare. Allowing either ramp or shutdown constraints is sufficient to ensure the non-existence of (pure) NE. The dichotomy of PoS results can be understood in terms of the market games of Section 2.1.6. With base and tracking layer only, the social welfare optimization is convex, which guarantees that stable payments can be found. Adding ramp or shutdown constraints may violate convexity to an arbitrary degree.

Theorem 3.2. *There is no cost-sharing scheme that achieves a pure price of stability better than ∞ when producer price functions have capacity constraints and ramp constraints.*

Proof. Consider an instance with two consumers, two producers and a single time period. Each consumer has a single demand profile of 1 with value 2 and outside option value 0. Producer m_1 has an MEGL of 1.5, base layer cost 1, shutdown cost 1 and large base layer capacity. Producer m_2 has no MEGL, base layer cost 0.75 and base layer capacity 1. This instance has no pure NE. There are three feasible matchings. First, we can match both consumers to m_1: total cost is 2, so one of the consumers pays at least 1. This consumer has incentive to defect to m_2 to pay 0.75. The other matchings have one consumer matched to m_1 and the other to m_2. Because the shutdown cost is removed when the m_2-consumer moves to m_1, the net cost she imposes by defecting is 0. Thus, the consumer who is matched to m_1 must pay the entire cost imposed on both agents for the assignment to be stable, which is $3 > 2$, making defection to the null producer attractive. Thus, no matching is stable. □

Theorem 3.3. *There is a cost-sharing scheme that achieves a pure price of stability of 1 when producer price functions have only tracking layers and capacity constraints.*

Proof. Consider the social welfare optimal matching μ and suppose each producer charges each matched consumer the average price per unit times the number of units she consumes. By way of contradiction, suppose consumer n_1 benefits by defecting to m using profile π. Let the matching after defection be μ'. Let $p_{\mu(n_1)}$ and $p'_{\mu(n_1)}$ be the average price per unit for $\mu(n_1)$ before and after n_1's departure, respectively. When n_1 defects, the average unit cost for demand on $\mu(n_1)$ decreases because some demand that was previously met with the tracking layer may be met with the base layer. Thus $p'_{\mu(n_1)} \leqslant p_{\mu(n_1)}$. Since n_1 defected, $V(\mu^p(n_1)) - |\mu^p(n_1)|p_{\mu(n_1)} \leqslant V(\pi) - C_m(\mu'^{-1}(m)) + C_m(\mu^{-1}(m))$. These inequalities can be rearranged to show that social welfare before the defection is less than the social welfare after, which is a contradiction. □

Theorem 3.4. *There is no cost-sharing scheme that achieves a pure price of stability better than ∞ when producer price functions have shutdown costs and capacity constraints.*

Proof. Consider an instance with two consumers, two producers and a single time period. Suppose each consumer has a demand of 1, value of 2 for their demand and outside option of value 0. Suppose that

Feature	w/ capacity constraints	w/o capacity constraints
Shutdown costs	PoS = ∞	?
Ramp constraints	PoS = ∞	PoS=?, PoA = ∞
Tracking layer	PoS = 1, PoA = ∞	N/A
Base layer only	PoS = 1, PoA = ∞	$PoA = 1$

Table 3.1: Table of stability results for combinations of producer price function features under the marginal cost defection model.

producer m_1 has a MEGL of 1.5, base layer cost of 1, shutdown cost of 1 and a large base layer capacity. Producer m_2 has no MEGL, base layer cost of 0.75 and base layer capacity of 1. This instance has no pure Nash equilibria. There are three feasible matchings. First, we can match both consumers to producer m_1. In this case, since the total cost is 2, one of the consumers is paying $\geqslant 1$. This consumer has incentive to defect to m_2 to pay 0.75. The other feasible matchings have one consumer matched to m_1 and the other to m_2. Because the shutdown cost is removed when the consumer that is matched to m_2 moves to m_1, the net cost that customer imposes by defecting is 0. Thus, the consumer who is matched to m_1 would have to pay the entire cost imposed on both agents for the assignment to be stable, which is $3 > 2$, which makes defecting to the null producer attractive. Thus, none of the matchings are stable. □

We show that capacity constraints plus any other feature ensures an infinite price of anarchy.

Theorem 3.5. *There is no cost-sharing scheme that achieves a pure price of anarchy better than ∞ when producer price functions have ramp constraints.*

Proof. Suppose we have a two-period model with two producers with no ramp capacity, i.e., the only feasible matchings have an identical load on a particular producer in both periods. Suppose that there exists a feasible matching with social welfare > 0. If there does not exist a consumer with identical loads in both periods, the matching where every consumer is matched to the null producer is also stable. In the case where the outside options are 0, the price of anarchy is ∞. □

Theorem 3.6. *There is no cost-sharing scheme that achieves a pure price of anarchy better than ∞ when producer price functions have capacity constraints.*

Proof. Consider an instance with two producers, two consumers and a single time period. Producer m_1 has base layer cost 1.5 and capacity 1 and producer m_2 has base layer cost 0 and capacity 1. Suppose consumer n_1 has demand $\epsilon << 1$, valuation ϵ and outside option 0 and consumer n_2 has demand 1, valuation 1 and outside option 0. The social welfare optimal matching μ^* is $\mu^*(n_1) = m_1$ and $\mu^*(n_2) = m_2$ which has $SW(\mu^*) = 1 + \epsilon - 1.5\epsilon = 1 - 0.5\epsilon$. Consider the matching $\mu(n_1) = m_2$ and n_2 unmatched, and where n_1 pays the entire cost of ϵ^2. $SW(\mu) = \epsilon - \epsilon^2$. n_1 does not want to defect to m_1 because she would incur a cost of $1.5\epsilon > \epsilon^2$, nor does she want to be unmatched as $\epsilon - \epsilon^2 \geqslant 0$. n_2 does not want to defect to m_1 because her net change in utility would be $1 - 1.5 < 0$. n_2 cannot defect to m_2 because m_2 does not have enough capacity. Thus μ is stable. $SW(\mu^*)/SW(\mu) = \frac{1-0.5\epsilon}{\epsilon-\epsilon^2}$. As ϵ approaches 0, this ratio approaches ∞. □

Allowing for a base layer only without capacity constraints reduces the model to an economic exchange variant, which has a price of anarchy of 1.

Theorem 3.7. *There exists a cost-sharing scheme that achieves a price of anarchy of* 1 *when producer price functions have a base layer only with no capacity constraints.*

Proof. Consider the cost-sharing scheme where each consumer pays the number of units consumed times the price of the producer they are matched to. By using this scheme in this setting, the prices paid are not affected by the behavior of other consumers. Thus, the social welfare optimal matching is obtained by maximizing the utility of each consumer individually. Note that by using this scheme in this setting, the worst-case defection model and the envy-free defection model are equivalent. Also, the prices paid under the defection model are the same as those that would have been paid if the consumer had originally been matched to the producer they were defecting to. The social welfare optimal matching is stable because switching profiles and producers cannot increase the net utility of any consumer since this quantity is maximized in the social welfare optimal matching. Any other matching is not stable because this quantity can be increased. Thus, the price of anarchy is 1. □

While the best NE may be arbitrarily worse than the social welfare optimal matching, one might hope that the optimal matching is close to a NE in practice. We have found this is generally *not* the case, but we do not focus on that question in this work. Since marginal cost defection pricing fails to induce stability, we consider two cost-sharing schemes that assume "envy-free" defection pricing, in which a defector is treated no differently than a consumer who was originally matched to that producer. Envy-free defection pricing assumes that while producers are free to make offers to any consumer, they cannot offer a deal to a potential defector that they do not offer to other consumers. This assumption is realistic in a setting with many small consumers.

3.4.2 Shapley-Like Payments

Since the underlying problem is a cooperative game, one natural approach to cost sharing is to use the Shapley value. We consider the group of consumers matched to a single producer to be a *coalition*. The Shapley value charges each agent the average marginal cost (or benefit) they contribute to their coalition over all possible *join orders*. Formally, the Shapley value of consumer n_0 matched to producer m_0 under matching μ is:

$$s(n_0) = \alpha \sum_{S \in \mu^{-1}(m_0) \setminus \{n_0\}} P_{m_0}(\text{dem}_\mu(S \cup \{n_0\})) - P_{m_0}(\text{dem}_\mu(S)) \tag{3.15}$$

where α is a normalization constant (number of permutations) and $\text{dem}_\mu(x)$ is the total demand of the set x of consumers when using the profiles assigned under μ. Our setting is atypical because some join orders induce demand profiles that cannot feasibly be served (e.g., due to ramp or capacity constraints), which is not accounted for in the standard definition of the Shapley value. To deal with this, we use a large cost when joining a coalition causes infeasibility.[3] Since all costs must be recovered, we normalize the payments so that the total paid by these consumers matched to a producer j equals the total charged by j.

Shapley values provide a conceptually simple approach to cost sharing that captures price functions well and is "fair." However, it is computationally intractable if all agents are distinct because it requires

[3]We believe that it is reasonable to charge an agent a larger share of cost if he frequently causes infeasibility. Note that, in the general case, we cannot average over feasible join orders only because there may not be any join orders where every addition is feasible.

iterating through all permutations of agents. To overcome this in a practical optimization and pricing procedure, we sample permutations to approximate Shapley costs. In addition, Shapley payments do not explicitly aim for stability, and indeed, we see they are not inherently stable. Hence, we allow Shapley values to be adjusted $\pm 10\%$ within each generator to increase stability, though even this modification does not guarantee stability. 10% is chosen because it is a modest deviation from the true Shapley values. Ideally, we desire a method to sacrifice some social welfare to improve stability, specifically to find matchings on the Pareto frontier of social welfare and degree of stability. This is difficult, however, because we can't efficiently maximize stability: producer price functions are far from concave and do not admit good concave upper bounds.

However, we find that sampling the matchings with high social welfare allows us to gain a significant amount of stability without losing much social welfare.[4] We sample matchings in two ways: through *exclusions* and *cuts*. Both use the well-known linear constraint that precludes a particular assignment of binary variables $\{X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\}$ from being selected by an optimization:

$$\sum_{i:x_i=1} X_i - \sum_{i:x_i=0} X_i \leqslant \sum_{i \in [n]} x_i - 1 \tag{3.16}$$

Recall that the MIP formulation of social welfare maximization contains two types of binary variables: the matching variables $y_{i,j,k}$ that indicate that consumer i is matched to producer j and is using profile k, and the support variables of the PPFs, such as $I_{j,t}^{(SC)}$ which indicates whether producer j incurred a shutdown cost at time period t. The exclusions method focuses on matching variables only: each iteration is the standard social welfare maximization plus a constraint in the form of Equation 3.16 for each setting of matching variables corresponding to the matchings found in previous iterations. Empirically, this decreases social welfare only slightly. Thus, after a fixed number of iterations, we return the matching with the highest stability.

The cuts method requires that *both* the matching variables and the support variables are different from the values used in previously found matchings. A given iteration of the cuts method includes the following additional constraints:

1. The constraints used by the exclusion method.
2. For each previous iteration, a constraint of the form of Equation 3.16 over the setting of the binary PPF support variables used in that iteration. These variables are $\{I_{j,t}^{(RC)} : j \in M, t \in [T]\}$ and $\{I_{j,t}^{(SC)} : j \in M, t \in [T]\}$.

As is the case with the exclusions method, the cuts method causes only a small loss in social welfare. Thus, we return the matching with the highest stability after a fixed number of iterations.

Note that the behavior of the cuts method is highly dependent on the form of the PPFs and may not be applicable to all PPFs whereas exclusions can be applied in any matching setting. Cuts require a more drastic change to the matching, decreasing social welfare by a larger amount, but sampling more diverse areas of the matching space. These two methods are compared below.

3.4.3 Similarity-Based Envy-Free Payments

The standard notion of *envy-freeness*—that no agent would prefer to receive the outcome received by another agent–is too weak in our setting. Since demand profiles are real-valued vectors, they are generally

[4] The same techniques can be used to enumerate matchings when searching for approximate NEs with high social welfare.

unique in that no two consumers share an identical profile. To handle this, we consider a generalization, *similarity-based envy-freeness (SBEF)*, where vectors that are "close" (we use L_2-distance) are priced identically (on a per-unit basis). Specifically, we use a clustering algorithm to partition the demand profiles, and constrain the unit price for any profile in a given partition to be equal. Our experimental model uses 24 one-hour time periods. While we could use demand in each period as the feature-vector for clustering, we instead use higher level features: the average and standard deviation of the demand across all periods; the global maximum and minimum demands, and the gap between them; and the average and standard deviation of demand in 6-hour windows. These high-level features blur the boundaries between partitions, which could be misleadingly granular if demand profiles were used directly. For instance, with demand profiles, we might distinguish two partitions based on consuming more or less than x units from 1–2 PM, which might lead consumers to respond to these specific features (e.g., by shifting some tasks from 1–2 PM to 12–1 PM). Such specific responses are unlikely to have a large effect on generation cost, especially if other consumers behave likewise. By using abstract features, consumer responses tend to have a greater effect on generation cost.

We use *Ward clustering*, an agglomerative clustering algorithm, to partition demand profiles [Ward Jr, 1963]. Ward clustering proceeds as follows:

1. Initialize each data point to be its own cluster.
2. Repeat until the desired number of clusters is reached:
   (a) For each cluster A_i, calculate $\Delta(A_i)$, the sum of squared Euclidean distances from each point in the cluster to the centroid.
   (b) For each pair of clusters A_i and A_j, consider the cluster $A_i \cup A_j$ that would result from merging the two. Calculate $\Delta(A_i \cup A_j)$.
   (c) Merge the two clusters that minimize $\Delta(A_i \cup A_j) - \Delta(A_i) - \Delta(A_j)$.

Ward clustering runs in linear time.

Ward clustering constructs partition boundaries that are difficult to communicate. To address this, we approximate the resulting clusters by building a bounded-depth decision tree (using CART [Breiman *et al.*, 1984]) with easily understandable partition boundaries based on a small number of features, without sacrificing much stability. In general, the choice of partitioning scheme should support desiderata for price functions. In summary, the procedure we use for calculating SBEF payments is as follows:

1. Partition demand profiles using Ward clustering on high-level feature vectors of the demand profiles.
2. Approximate the resulting clusters with a decision tree.
3. Find a matching that maximizes stability, subject to (i) recovering all costs, and (ii) requiring a fixed unit price for profiles in any given partition.

The last step requires solving the social welfare maximizing MIP as well as an additional linear program to find payments that minimize the incentive to defect. The variables in the linear program are $\boldsymbol{p}$ where where $p_{j,A}$ is the per unit price for generator j and cluster A.

$$\begin{aligned}
&\min_{\boldsymbol{p}} w \\
&\text{subject to} \\
\sum_{i\in N} p_{\mu(i),\mu^P(i)} \sum_{t\in[T]} (\mu^P(i))_t \geqslant & \textsc{Cost}(\mu) \\
w \geqslant & V_i(\boldsymbol{\pi}_i^{(k)}) - p_{j,A(\boldsymbol{\pi}_i^{(k)})} \sum_{t\in[T]} \boldsymbol{\pi}_{i,t}^{(k)} - \\
& V_i(\mu^P(i)) + p_{\mu(i),\mu^P(i)} \sum_{t\in[T]} (\mu^P(i))_t \qquad \forall i \in N, \forall j \in M, \forall k \in \Pi_i \qquad (3.17)
\end{aligned}$$

where $\textsc{Cost}(\mu)$ is the cost of matching μ and $A(\boldsymbol{\pi}_i^{(k)})$ is the cluster assigned to profile $\boldsymbol{\pi}_i^{(k)}$. The first constraint ensures that the revenue raised covers the cost of the matching. The second makes the objective minimize the maximum incentive to defect.

SBEF ensures that consumers are indifferent to which producer they are matched if the generators have similar enough price functions.

Observation 3.1. *Assume two or more generators. Suppose we have a matching μ and a set of SBEF payments p. If at least one profile in each partition is assigned to each generator and the maximum incentive to defect is 0, all generators must offer the same unit price in each partition.*

Proof. If this were not the case, there exists some consumer that is matched to a a more expensive generator given her profile. This consumer would have an incentive to defect that is at least equal to the (positive) difference between the costs of two generators for that profile. □

The price differences between "adjacent" partitions will be "reasonable" if there are enough consumers with demand profiles in multiple partitions. When a consumer has demand profiles in two partitions, stability puts pressure on the difference in price between two partitions to be small w.r.t. the difference in their valuations. The SBEF price procedure is somewhat more conceptually complex than Shapley-like payments, but it is much more computationally efficient and it addresses envy-freeness more directly than Shapley. While the Shapley payments within a coalition may be intuitively fair, the assignment of similar profiles to particular producers by social welfare optimization may be somewhat arbitrary and result in payments that are far from envy-free. We see below that SBEF payments have much stronger stability properties in practice than Shapley.

3.5 Model of Consumer Demand

To test our algorithms, we use a model of the US residential energy market. Building characteristics are based on the 2011 Buildings Energy Data Book [D&R International, Ltd., 2012]. The building thermal model, which includes temperature, solar radiation and a miscellaneous factor, is derived from [Huang *et al.*, 1999]. Roughly, we independently sample square footage and insulation level from known US distributions. Using appliance surveys, we randomly generate appliances and load events for each appliance. We then calculate air conditioner loads for a variety of target interior temperatures. External conditions are those of July 10, 2010 in San Antonio, Texas: since most home electricity use is due to

air conditioning, hot summer days stress generation heavily and induce larger incentives for consumers who are willing to alter their behavior.

Buildings have the following features:

1. *Building floorspace* (square feet). The floorspace of a building is the main metric of size in surveys. We use it to determine the general level of electricity consumption of a building as well as to extrapolate the volume and surface area of a building for heating and cooling purposes.

2. *Building volume* (cubic feet). Volume is one important component of determining heating and cooling costs for buildings. We calculate the building volume by assuming that the ceilings are 8 ft high.

3. *Heater and air conditioner properties.*

   (a) *Heater type* (gas or electric). If heater is electric, we rely on *heater efficiency* (Btu per Watt-hour). Typical efficiency is 3.413 Btu per Watt-hour since we are only concerned with electric heaters. We assume that heaters have infinite output power.

   (b) *Air conditioner efficiency* (Btu per Watt-hour). Air conditioner efficiency at a given moment depends on outside temperature, but for simplicity, we assume that it is fixed. Typical air conditioner efficiency ratings are averages across a variety of conditions. The typical range for these average efficiency ratings is 10–14 Btu per Watt-hour. We assume that air conditioners have infinite output power.

4. *Window surface area* (square feet) SA_{window}.

5. *Wall surface area* (square feet) SA_{wall}.

6. *Roof surface area* (square feet) SA_{roof}.

7. *Window shading coefficient* (dimensionless) WSC. This number represents the fraction of solar radiative heat transmitted by a window; range 0.57–0.74.

8. *Insulation levels* ($\text{ft}^2 \cdot {}^\circ\text{F} \cdot \text{hr/Btu}$). These represent the amount of heat per unit surface area carried through a particular medium per degree of difference between interior and exterior temperatures.

   (a) *Window insulation level* R_{window}. The range of insulation values will is 1.3–1.7.

   (b) *Wall insulation level* R_{wall}. The range of insulation values will is 2.5–7.

   (c) *Roof insulation level* R_{roof}. The range of insulation values will is 9–14.

9. *Temperature preference surface* (input units are time and degrees Fahrenheit, and output units are dollars). The temperature preference curve represents the cost (in dollars) of maintaining a particular temperature at a particular time; this cost may represent discomfort and/or lost revenue depending on the type of agents involved.

10. *Lighting power density* (Watts per square foot). Heating generated by lighting is an important component of heating and cooling costs. Typical power density for lights is between 1–2.5 Watts per square foot. We approximate that 75% of this power is emitted as heat.

External Conditions. The population of agents shares a common external context consisting of several factors.

1. *Temperature* (units time to degrees Fahrenheit).
2. *Solar radiation* (units time to Btu per square foot per second) $SR(t)$.

Heat Transfer Calculation. Given these numbers, we calculate the overall heat transfer. For conductive transfer, the instantaneous heat transfer is given by:

$$1.75\Delta T\left(\frac{SA_{window}}{R_{window}}+\frac{SA_{wall}}{R_{wall}}+\frac{SA_{roof}}{R_{roof}}\right) \tag{3.18}$$

where ΔT is difference between the interior and exterior building temperatures. Note that 1.75 is a corrective factor to make up for heat transfer not modeled (i.e., ventilation, foundation, and infiltration). For radiative transfer, it is:

$$0.33 * SR(t) * SA_{window} * WSC \tag{3.19}$$

where 0.33 is an estimate about the amount of time a particular window is exposed to the sun each day.

3.6 Experiments

In all experiments, we use 50 consumers, 2 producers, 4 profiles per consumer, 24 time periods, and run 50 trials for each experiment. We use a small number of consumers because Shapley values are expensive to compute—each marginal contribution calculation requires running another optimization. We are able to optimize social welfare and find SBEF payments for instances with 2500 consumers, 2 producers, 4 profiles per consumer, 24 time periods in 30 minutes on average on 12x2.6GHz, 32GB machine using CPLEX 12.51. Scalability could be increased further by: (i) using a simpler clustering algorithm and (ii) compressing the optimization by grouping similar consumers/demand profiles together. Shapley value computation could be sped up by either approximating the marginal contribution and only reoptimizing every few new agents, or by caching the optimization results. In this setting, the social welfare optimal matching has a mean social welfare of \$837.5 with standard deviation of \$138.8, or around \$16.75 per consumer.

3.6.1 Shapley-Like Payments

Figures 3.7 and 3.8 show results using Shapley payments. Each trial requires about one hour of computation, almost all of which is to approximate Shapley values. We sample 30 random join orders, a number which was determined empirically to induce convergence. Initial *maximum (over consumers) incentive to defect (MItD)* is \$17.6 on average (standard deviation \$16.4), which is 105% of the average utility per consumer. After 14 iterations, mean MItD was \$7.8 with cuts (44% of the original, 46% of per-consumer average utility) and \$13.9 with exclusions (79% of the original, 83% of per-consumer average utility). Cuts decreased MItD faster than exclusions—after two cuts, MItD decreased to \$13.8. on average, a greater reduction than 14 exclusions. (Standard deviations are large because they include the variation among instances.) Using a paired t-test, the difference between MItD using cuts vs. exclusions is statistically significant after 2 iterations ($p < 0.05$).

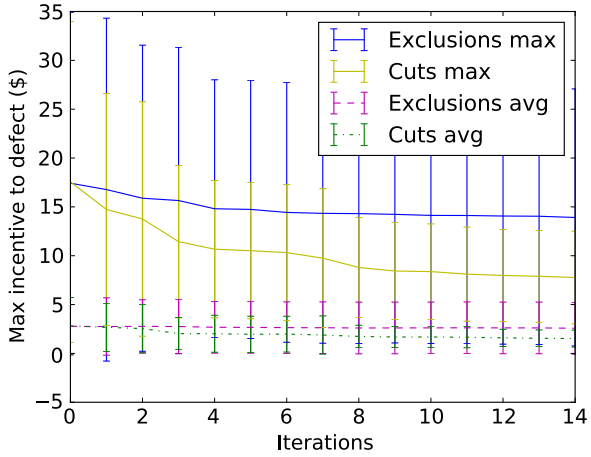


Figure 3.7: Max and average incentive to defect for maximally-stable matchings using Shapley-like payments. The corresponding social welfare is shown in Figure 3.8.

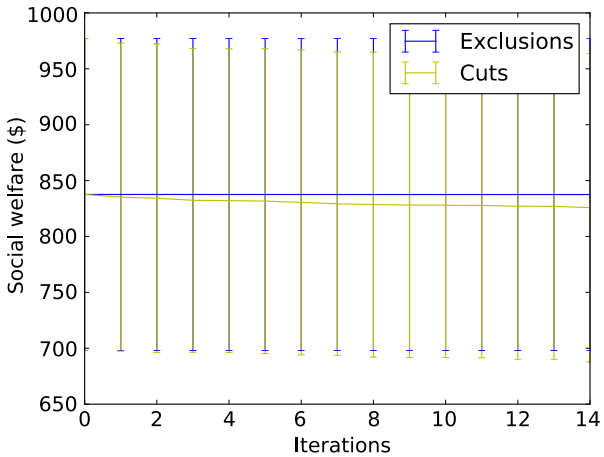


Figure 3.8: Social welfare for maximally stable matchings using Shapley-like payments.

	Avg. 12-6pm	Std. dev.	Std. dev. 12-6pm	Max
Gini importance	0.46	0.32	0.14	0.034

Table 3.2: Table of stability results for combinations of producer price function features under the marginal cost defection model.

Since the MItD is primarily influenced by agents with large demands, the average incentive to defect is also shown on Figure 3.7. For consumers with a positive incentive, (i.e., a defection that is more attractive than their assignment), the mean decreased from \$2.81 to \$2.57 with exclusions (91.3% of the original) and to \$1.52 with cuts (54% of the original), correlating with the decreases in MItD, but showing a less dramatic drop when cuts are used. The percentage of agents with positive incentive increased slightly after 14 iterations, from 50% to 52.6% with exclusions and 50.3% with cuts. This appears to be a spurious consequence of the enumeration process.

Exclusions reduced social welfare by less on average than cuts. The percentage of max social welfare under exclusions had a mean of 99.9% after 14 iterations, while cuts had a mean of 98.6%. Since exclusions enumerate every matching, exclusions enumerate those returned by cuts, but they are much slower—the first cut has greater effect than 14 exclusions.

3.6.2 Similarity-Based Envy-Free Payments

Figure 3.9 shows the effect of using different numbers of partitions and decision tree depth within SBEF payments on mean MItD. Each trial takes only a few seconds (in stark contrast to Shapley). It is important to note that these results all use the social welfare optimal matching—since stability is so high, we do not explore the trade-off between stability and social welfare (though cuts and exclusions could be used here, as above). The same enumeration techniques used for Shapley-like prices (exclusions and cuts) could also be applied here to increase stability further by trading off social welfare. We see that SBEF payments are highly stable under all tested conditions. Stability increases with the number of partitions: mean MItD is \$1.76 with two partitions (standard deviation \$2.03, 11% of per-consumer average utility) and \$0.71 with nine (standard deviation \$1.21, 4% of per-consumer average utility). Having more partitions tends to reduce potential "envy-freeness" as fewer consumers are in each. The figure also suggests that a minimum tree depth is needed for the decision-tree approximation to be as stable as the original partitioning: from one level for two partitions, up to four levels for six or more partitions. Mean incentive for customers with a positive defection incentive increases slightly with the number of partitions: \$0.21 with 2 partitions and \$0.24 with 9 partitions. The number of customers with positive defection incentive decreased from 34% with 2 partitions to 24% with 9.

We use Gini importance to assess the feature importance in the partition-approximating decision trees. Table 3.2 shows the four most important features across all instances with four partitions and a decision tree of depth 3. Because temperatures and solar radiation peak in the afternoon, the fact that afternoon consumption is the most important determinant of production price makes sense. While this cost-sharing scheme resembles time-of-use pricing, the features that affect overall cost change on the fly, dynamically reflecting their impact on generation cost.

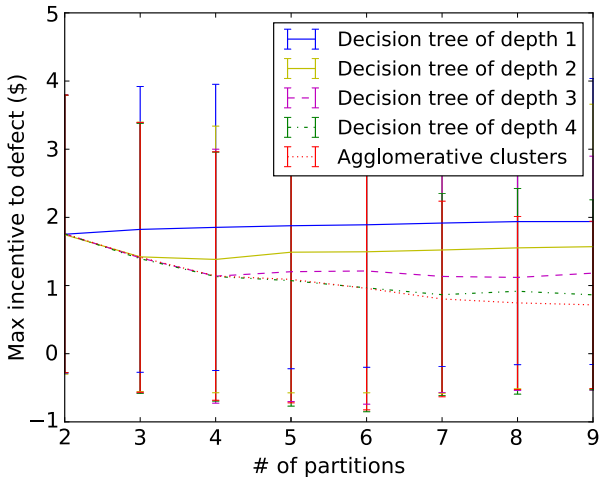


Figure 3.9: Stability of different numbers of partitions and decision tree depths approximating those partitions.

3.7 Conclusion and Future Work

We have presented a market model for matching electricity producers and consumers, which can be tractably optimized for a large number of consumers. Consumers can present multiple demand profiles, which allows the matching mechanism to offer discounts to consumers if they are willing to shift demand in a way that reduces production costs. We showed that Nash-stable matchings may not exist in settings with realistic producer price functions. In particular, the presence of capacity constraints and either shutdown or ramp constraints is enough to allow an instance to have zero Nash equilibria.

We presented two alternate cost-sharing schemes: one based on the Shapley value and a new approach called SBEF. Because stability is difficult to optimize directly, we develop ad hoc schemes for increasing stability while sacrificing social welfare. We find that in the Shapley value case, these schemes result in a significant reduction in mean MItD with only a small decrease in social welfare ($< 2\%$). However, the mean MItD is still substantial with the best ad hoc method, around 46% of the per consumer average utility. SBEF does better, achieving a mean MItD of 4-11% of average consumer utility depending on the selected parameters, which could be improved further with the methods we developed in the Shapley value case. SBEF also scales far better than Shapley.

There are several possible directions for future work.[5] The key contribution of this chapter is a framework for efficiently computing payments in a large non-convex game where even Nash stability may not be attainable. It is likely that many practical cooperative games have these properties, but games of this type have not received much attention in the literature. We show that despite the unattainability of "ideal" payments, we can define a series of desiderata and find payments that do quite well under those desiderata—in particular, by trading off a small amount of social welfare, we can gain a large amount of stability. However, guided by the message of Sandholm and Boutilier [2006], we do not study preference elicitation or incentives in preference elicitation. Note that these questions recur at the end of Chapter 4.

We would like to address these issues by making preference elicitation the main focus and dealing with the incentive issues in the context of the preference elicitation procedure, with the justification

[5]Future work that is narrowly related to each chapter is discussed in that chapter and broader related work at the end of the thesis.

that, unlike in combinatorial auctions, a very natural and low effort elicitation scheme is critical because agents have limited effort and patience to spend on interacting with the mechanism. This perspective in itself makes incentive issues less important than in the CA setting, but we would still like a procedure that has good incentive properties. We discuss two approaches to attain these properties. The first is to extend the methods of this chapter and study the questions empirically. The second is to make use of VCG payments.

The incentive properties of an elicitation scheme can be measured empirically by calculating the incentive to misreport for each query. The question then becomes when the incentive to misreport is low enough, because misreporting can have a snowballing effect: once one agent misreports, it changes the incentive properties for the other agents. This can again be measured empirically, for example, by giving each agent a stochastic function that triggers a misreport based on their myopic incentive to do so. At the end of the experiment, one can measure how much the resulting payments satisfy the desiderata raised in this section, with respect to both the true and reported utility functions. It is worth noting that misreporting effectively requires a lot of information about the other agents, and we would expect that the informational barrier would have some effect in practice but it is not clear how to model it. This is a broader question that arises in all market design problems and is mainly avoided by the existence of a mechanism that has good incentive properties even under complete information. This setting is sufficiently difficult that it is worth considering how information flows could be modeled.

Note that even with a preference elicitation scheme that is designed for convenience, there is the potential to prove guarantees on incentive compatibility. This is addressed in the Chapter 4.

The second approach to the incentive problem is to make use of the VCG mechanism. We know that if we use the VCG outcome and Clarke payments, there is no incentive to misreport. Calculating them exactly does not seem promising because it requires running the optimization once for each agent, which is intractable in a large setting. There is the possibility of approximating VCG payments and maintaining incentive compatibility, but the mechanism must use the social welfare maximizing allocation in order to access the incentive guarantee of VCG, and this chapter argues that this allocation is not desirable because of its poor stability properties. Recall that the incentive properties of VCG are extremely sensitive to the mechanism selecting an outcome that does not maximize social welfare (see Observation 2.5 in Section 2.3.4).

However, many properties of VCG are maintained under any choice of "organizing function", i.e., we can substitute some other function for social welfare and use the outcome that maximizes it instead. For example, both incentive compatibility and the non-positivity of payments under the Clarke pivot rule hold for any choice of function. Thus, there is the possibility of selecting an organizing function that includes both social welfare and stability. There are a few difficulties facing us if we were to do this.

1. The individual rationality of VCG with the Clarke pivot rule may be affected.

2. The function would likely be quite difficult to optimize exactly, which is why we take a more heuristic approach to optimization in this chapter. Unfortunately, Observation 2.5 shows that failing to optimize the function exactly can yield a massive incentive to defect for each agent.

3. Another difficulty is budget balance, i.e., that the payments of the agents sum to the price charged by the producer. Budget balance is difficult to guarantee in VCG mechanisms in general, and it is likely that empirical work would have to be done.

There is quite a lot of flexibility in VCG mechanisms through the choice of pivot rule, but, in order to maintain incentive properties, it is important that the choice of pivot rule for a particular agent is never affected by that agent's report. We return to the question of extending VCG in Chapter 6.

The second major area of future work concerns SBEF. SBEF is similar to a generalized concept of fairness in supervised learning that was introduced by Dwork et al. [2012] (henceforth, DHPRZ). They consider the following setting. Let N be a set of agents and A be a set of alternatives. The goal is to assign each user a distribution over the alternatives to minimize some loss function $L : N \times A \to \mathbb{R}$. DHPRZ define fairness in this setting by saying that a distribution over outcomes is fair if, for any two agents, the distributions they are assigned differ by at most a function of the distance between their types. They argue that classifiers should be run to minimize the loss function while respecting the fairness constraint.

DHPRZ fairness is quite conceptually similar to SBEF, except that it is a continuous version of the concept. Instead of requiring that agents that are in the same cluster pay the same price, DHPRZ fairness requires that the more similar agents are, the more similar the price they should pay.

SBEF may be applied as a more practical form of DHPRZ fairness because it is more efficient to optimize. DHPRZ fairness requires a quadratic number of constraints in the number of agents because the property must be enforced between every pair of agents. For SBEF, the optimization is to set the outcome family that is assigned to the agents in each partition. At a minimum, for example in the case where the outcomes in each partition have to be DHPRZ fair w.r.t. to the other agents in that partition, each partition can be optimized separately. In this chapter, it is even more efficient than that because we can set a price per partition and achieve fairness that way.

Chapter 4

Multiple-Profile Prediction-of-Use Games

In this chapter, we discuss an electricity market design problem driven by inefficiencies caused by consumption prediction. Our goals are (i) to coordinate the consumption of electricity across agents whose value for "consuming unpredictably" varies and (ii) to divide the costs of electricity consumption across the agents in a stable way. We approach the problem using cooperative game theory. We find we can achieve our goals by extending the techniques of Robu et al. [2017], who develop a cooperative game-theoretic model of a similar scenario, except that agents are not allowed to adjust their consumption decisions in response to those of other agents. We show that our extension maintains many of the useful properties of Robu et al.'s model, while offering a large increase in social welfare in our experiments. Giving agents the ability to change their consumption decisions raises a principal–agent problem which we address using a new technique, *separating functions*, that has many useful properties. Parts of the work in this chapter originate in Perrault and Boutilier [2017].

4.1 Introduction

In many countries, energy suppliers face a two-stage market, where suppliers purchase energy at lower rates in anticipation of future consumer demand and then reconcile supply and demand exactly by purchasing at a higher rate at the time of realization through a "balancing market" [Team, 2011]. The cost to energy suppliers is thus highly dependent on their ability to predict future consumption, but consumers typically have little incentive to consume predictably in a fixed-rate tariff environment, where consumers pay for electricity at a fixed price per kilowatt-hour (kWh). This misalignment of incentives makes the electricity market less efficient because consumers who would be willing to consume more predictably in exchange for a lower price are not given the option to do so.[1] This chapter focuses on the market design problem that arises from the misalignment of predictability incentives.

The role of uncertainty in electricity markets can be seen in Figure 4.1, which shows demand forecasts and actuals for a single day. On this day, actual demand is quite close to the predictions, but there is

[1] We talk about the misalignment problem from the perspective of suppliers (middlemen) instead of producers, in contrast to the last chapter. This is to follow Robu et al. [2017]. The underlying reason for the incentive misalignment is that producers can serve demand more efficiently (by dispatching generation more efficiently) if they have accurate forecasts, and they pass this cost structure on to the suppliers.

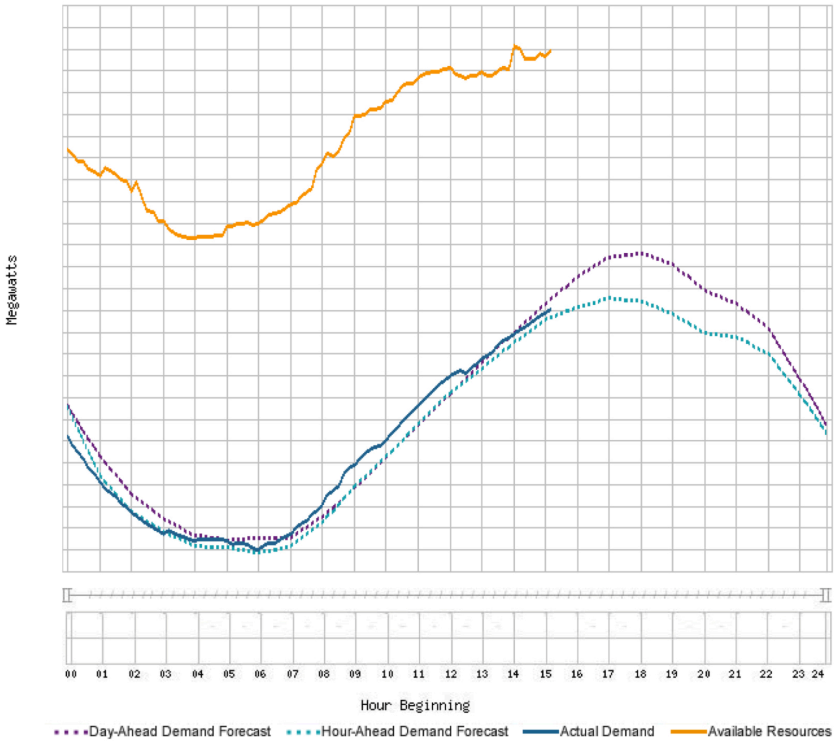


Figure 4.1: A view of electricity markets for one 24-hour period in California, taken from the California Independent System Operator, which oversees the electricity system in California.[2]

nonetheless some difference, even between actual and the hour-ahead prediction. Notice that the amount of available generation is tracking actual demand, but is about 40% higher. While the predictions appear to be quite accurate in this case, the amount of reserve indicates the level of uncertainty in the system.

Prediction-of-use games were developed by Robu et al. [2017], hereafter RVRJ, to address this misalignment of incentives. Prediction-of-use games incentivize consumers to report reasonably accurate *predictions of their own consumption*, thus offering access to their private information about the future. This has two beneficial effects. The first is to address the incentive misalignment by creating the incentive to be predictable. The second is that suppliers gain access to the consumer's private information, which may improve their predictions of consumer behavior.

RVRJ analyze a setting where flat tariffs are replaced with *prediction-of-use (POU) tariffs*, in which consumers make a payment based on both their actual consumption and the accuracy of their prediction. Similar tariffs have been deployed in practice, primarily directed at industrial consumers [Braithwait *et al.*, 2007]. RVRJ analyze the *cooperative game* induced by POU tariffs, in which consumers form *buying coalitions* that reduce (aggregate) consumption uncertainty, and find that, under normally-distributed prediction error, the game is *convex*. Convexity is a powerful property that significantly reduces the complexity of important problems in cooperative games, both analytically and computationally.

While attractive, the POU model has a significant shortcoming. Though it could be adapted to model how consumers change their consumption in reaction to price changes, consumers cannot *coordinate* their consumption choices within the POU model. A consumer's *optimal consumption profile*—a random variable representing the individual's possible behaviors or patterns of energy consumption—depends

[2]Available for current day and previous month at `https://www.caiso.com`.

on the profiles used by others. In POU games, the only consumer choice is which coalition to join—a consumer's demand is represented by a *single* prediction, reflecting just one selected (or average) consumption profile for each individual. In essence, consumers predict their behavior without knowing anything about others in the game. While the POU model can offer social welfare gains when the profiles are selected optimally, we show they can result in significant welfare loss when profile selection is uncoordinated.

We introduce *multiple-profile POU (MPOU) games*, which extend POU games to admit *multiple* consumer profiles. This allows consumers to coordinate the behaviors that change their predictions, facilitating the full realization of the benefits of the POU model. We show that MPOU games have many of the same properties that make the POU model tractable, e.g, convexity, which makes the stable distribution of the benefits of cooperation easy to compute. In addition, we show that MPOU games are individually rational and that consumer utility is monotone increasing as the number of truthfully-reported profiles increases. However, MPOU games also present a new challenge in coalitional allocation: since one can only observe an agent's (stochastic) consumption—not their underlying behavior—determining stabilizing payments for coalitional coordination requires novel techniques. We introduce *separating functions*, which incentivize agents to take a specific action in settings where actions are only *partially observable*.

We experimentally validate our techniques, using household utility functions that we learn (via structured prediction) from publicly available electricity use data. We find that the MPOU model provides a gain of 3-5% over a fixed-rate tariff across several test scenarios, while a POU tariff *without* consumer coordination can result in losses of up to 30% from a fixed-rate tariff. These experiments represent the first end-to-end study of the welfare consequences of POU tariffs.

The remainder of the chapter is organized as follows. Section 4.2 reviews cooperative games, the POU model and related work. Section 4.3 introduces MPOU games and Section 4.4 proves their convexity. Section 4.5 outlines the new class of incentive problems that arises when the mechanism designer cannot (directly) observe an agent's selected profile, and develops a general solution to that problem. Section 4.6 briefly discusses manipulation. In Section 4.7, we describe an approach for learning consumer utility models from real-world electricity usage data, and experimentally validate the value of MPOU games using these learned models in Section 4.8.

4.2 Background

We begin with basic background on POU games. For an overview of cooperative games, see Chapter 2.

4.2.1 Prediction-of-Use Games

A *prediction-of-use (POU) game* is a tuple $\langle N, \Pi, \tau \rangle$, where N is a set of agents, Π is a set of consumption profiles, and τ is a POU tariff. Each $i \in N$ uses electricity according to a *consumption profile* in Π, a normal random variable with mean μ_i and standard deviation σ_i, say, in kilowatt-hours (kWh). Let x_i denote i's realized consumption, $x_i \sim \mathcal{N}(\mu_i, \sigma_i)$. POU games will be treated as cooperative games, and the definition of coalition is given below. Agents are assumed to truthfully report their profiles to the coalition. We do not address elicitation or estimation of consumption here, except briefly in Section 4.6.

A *POU tariff* has the form $\tau = \langle p, \underline{p}, \bar{p} \rangle$, and is intended to better align the incentives of the consumer and electricity supplier, whose costs are greatly influenced by how predictable demands are. Each agent

i is asked to predict a *baseline consumption* b_i, and is charged p for each unit of x_i, plus a penalty that depends on the accuracy of their prediction: $\bar{p}$ for each unit their realized x_i exceeds the baseline, and $\underline{p}$ for each unit it falls short:

$$\psi(x_i, b_i, \tau) = \begin{cases} p_j \cdot x_i + \bar{p} \cdot (x_i - b_i) & \text{if } b_i \leqslant x_i \\ p_j \cdot x_i + \underline{p} \cdot (b_i - x_i) & \text{if } b_i > x_i \end{cases} \tag{4.1}$$

To ensure agents have no incentive to artificially inflate consumption, we require $0 \leqslant \bar{p}$ and $0 \leqslant \underline{p} \leqslant p$ [Robu *et al.*, 2017]. An agent i should report a baseline that minimizes his expected payment. RVRJ show that i does this by predicting $b^* = \mu_i + \sigma_i \Phi^{-1}(\frac{\bar{p}}{\bar{p}+\underline{p}})$, where Φ^{-1} is the inverse normal CDF. They also show that i's expected payment under the optimal baseline is $\mu_i p + \sigma_i L(\underline{p}, \bar{p})$ where $L(\underline{p}, \bar{p}) = \int_0^{\frac{\bar{p}}{\bar{p}+\underline{p}}} \Phi^{-1}(y)dy$.

To be more predictable in *aggregate*, agents may form a coalition C, where C reports its aggregate demand and is charged as if it were a single agent. C's aggregate consumption is the sum of the normal random variables corresponding to the members' profiles, itself normal with mean $\mu(C) = \sum_{i \in C} \mu_i$ and standard deviation $\sigma(C) = \sqrt{\sum_{i \in C} \sigma_i^2}$. This aggregate prediction generally has lower variance w.r.t. the mean, thus reducing total penalty payments facing C under POU tariffs (compared to members acting individually).

RVRJ analyze *ex-ante* POU games. In the ex-ante game, all agent decisions, as well as any internal transfers, or payments, are based on *expected* consumption (realized consumption plays no role). This approach is justified based on two assumptions. The first is that coalitions form at the time of consumption prediction, not at the time of consumption, and the second is that agents are risk-neutral, expected-utility maximizers. The characteristic value of coalition C is

$$v(C) = -\mu(C)p - \sigma(C)L(\underline{p}, \bar{p}) \tag{4.2}$$

and they show that the ex-ante POU game is convex.[3]

POU games are closely related to *newsvendor games* [Müller *et al.*, 2002], where a supplier must purchase inventory in advance of demand and faces a penalty for oversupply (storage costs) and undersupply (lost profit). Unlike POU games, the players are the suppliers, the demand distribution is known, and the primary object of study is the value that suppliers can gain by pooling their inventory.

In addition to POU games, others have proposed the formation of cooperatives or coalitions among electricity consumers. Rose et al. [2012] develop a similar mechanism for truthfully eliciting consumer demand. Kota et al. [2012] and Akasiadas & Chalkiadakis [2013] propose using coalitions to improve reliability and shift peak power loads. The previous chapter focuses on the formation of groups of consumers with multiple profiles to reduce peak loads. None of this work offers the theoretical guarantees of RVRJ.

Beyond electricity markets, several authors have studied the problem of group purchasing in an AI context. Lu and Boutilier [2012] study a restrictive class of buyer preferences (unit demand, only the supplier affects utility) and seller price functions (volume discounts), which have strong theoretical guarantees. Similarly, optimally matching a group of cooperative buyers to sellers has been studied [Sarne and Kraus, 2005; Manisterski *et al.*, 2008].

[3]Technically, they define the game as a *cost game* and show that the game is concave, while we use a profit game, but results from the two perspectives translate directly.

4.3 Multiple-Profile POU Games

We extend POU games by allowing agents to report *multiple profiles*, each reflecting different behaviors or consumption patterns, and each with an inherent utility or value reflecting comfort, convenience, flexibility or other factors. These profiles correspond to different discrete choices the consumer makes, e.g., what temperature to set the air conditioner at or when to do laundry or dishes. This will allow an agent, when joining or bargaining with a coalition, to trade off cost—especially the cost of predictability—with his inherent utility. These profiles are similar to the demand profiles of Chapter 3.

A *multiple-profile POU (MPOU) game* is a tuple $\langle N, \{\Pi_i\}, V, \tau \rangle$. Given set of agents N, each agent $i \in N$ has a non-empty *set of demand profiles* Π_i, where each profile $\pi_{i,k} = \langle \mu_{i,k}, \sigma_{i,k} \rangle \in \Pi_i$ reflects a consumption pattern (as in a POU model). Agent i's *valuation function* $V_i : \Pi_i \to \mathbb{R}$ indicates his value or relative preference (in dollars) for his demand profiles.[4] Admitting multiple profiles allows us to reason about an agent's response to the incentives that emerge with POU tariffs and in coalitional bargaining. Finally, τ is a POU tariff. We use the same definition of POU tariffs and agent baselines as in POU games above. Notice that the optimal baseline report for an agent is now defined *relative to the profile they use*.

As in POU games, agents are motivated to form coalitions to reduce the relative variance in their predictions. However, for a coalition C to accurately report its aggregate demand, its members must select and commit to a specific usage profile. We denote an *assignment of profiles to agents* as $A : N \to \times_{i \in N} \Pi_i$. Under such an assignment, C's consumption is normal, with mean $\mu(C, A) = \sum_{i \in C} \mu(A(i))$ and standard deviation $\sigma(C, A) = \sqrt{\sum_{i \in C} \sigma^2(A(i))}$. The aggregate value accrued by the coalition (prior to supplier payments) is the sum of its members' values: $V(C, A) = \sum_{i \in C} V_i(A(i))$.

Analogous to RVRJ's analysis of POU games, we begin by analyzing *ex-ante MPOU games*, where agents make decisions and payments before consumption is realized. The characteristic value v of a coalition C is the maximum value that coalition can achieve in expectation under full cooperation, that is, assuming an optimal profile assignment and baseline report. We thus define $v(C) = \max_A v(C, A)$, where

$$v(C, A) = V(C, A) - \mu(C, A)p - \sigma(C, A)L(\underline{p}, \bar{p}) \tag{4.3}$$

Notice that profile selection, a critical component of our MPOU models, does not arise in the POU setting.

In the following sections, we present a mechanism for MPOU games with which the grand coalition organizes the individual consumption behavior of its members (all agents in N) and the payments that flow among them. The mechanism proceeds as follows:

1. Agents report their consumption profiles to the mechanism (we assume this report is truthful).

2. The mechanism calculates an assignment A of agents to profiles that maximizes social welfare. We elaborate on this assignment optimization below.

3. The mechanism calculates an ex-ante core stable payment $t(i)$ for each agent i that is based on all agents using their assigned profiles. We address payment computation in Section 4.4.

[4] Such profiles and values may be explicitly elicited or estimated using past consumption data (see Section 4.6).

4. In Section 4.5, we find that some agents have an incentive to defect from the assigned profile. This is due to the agent's choice of profile being only partially observable by the coalition—if there is a profile with higher value than the one assigned, an agent can consume according to that profile without it being immediately obvious to the coalition. Hence, we design *separating functions* to prevent these defections. The mechanism calculates a separating function D_i for each agent with an incentive to defect from their assigned profile.

5. At realization time, each agent i receives payment $t(i)$. Each agent i with a separating function D_i receives $D_i(x_i)$, where x_i is his realized consumption.

In the MPOU model, calculating a social welfare-maximizing assignment of agents to profiles requires solving a non-convex optimization problem. We do this using a mixed integer program with an objective function given by (4.3), a binary assignment variable for each agent-profile pair, and a constraint that each agent is assigned exactly one profile. The last term of the objective is non-convex: $\sigma(C, A) = \sqrt{\sum_{i \in C} \sigma^2(A(i))}$. We replace the negative square root with a piecewise linear upper bound, which requires three binary variables per segment.

$$\max_{\boldsymbol{y}, \boldsymbol{I}, w} \sum_{i,k} y_{i,k} v_{i,k} - p \sum_{i,k} \mu_{i,k} - L(\underline{p}, \bar{p}) w$$

subject to

$$\begin{aligned}
\sum_k y_{i,k} &= 1 && \forall i \in N \\
-UI_q^{\mathrm{LB}} + \sum_{i,k} y_{i,k} \sigma_{i,k}^2 &\leqslant \alpha_q - \epsilon && \forall q \in [0, Q-1] \\
-UI_q^{\mathrm{UB}} - \sum_{i,k} y_{i,k} \sigma_{i,k}^2 &\leqslant -\alpha_{q+1} - \epsilon && \forall q \in [0, Q-1] \\
I_q^{\mathrm{LB}} + I_q^{\mathrm{UB}} - I_q &\leqslant 1 && \forall q \in [0, Q-1] \\
-w + UI_q + \frac{\sqrt{\alpha_{q+1}} - \sqrt{\alpha_q}}{\alpha_{q+1} - \alpha_q} \sum_{i,k} y_{i,k} \sigma_{i,k}^2 &\leqslant U - \sqrt{\alpha_{q+1}} + \alpha_{q+1} \frac{\sqrt{\alpha_{q+1}} - \sqrt{\alpha_q}}{\alpha_{q+1} - \alpha_q} && \forall q \in [0, Q-1] \qquad (4.4)
\end{aligned}$$

$y_{i,k}$ is a binary variable that is true if agent i is matched to profile $\pi_{i,k}$. w is an auxiliary continuous variable that represents the piecewise-linear approximation of $\sqrt{\sum_{i,k} y_{i,k} \sigma_{i,k}^2}$. $[\alpha_0, \ldots, \alpha_Q]$ is a discretization of the region of the square root function we want to approximate. For example, we can take $\alpha_0 = \min_{\boldsymbol{y}} \sum_{i,k} y_{i,k} \sigma_{i,k}^2$ and $\alpha_Q = \max_{\boldsymbol{y}} \sum_{i,k} y_{i,k} \sigma_{i,k}^2$. I_q is true if the $\alpha_q \leqslant \sum_{i,k} y_{i,k} \sigma_{i,k}^2 \leqslant \alpha_{q+1}$. The lower and upper bounds are checked separately. I_q^{LB} is true if $\alpha_q \leqslant \sum_{i,k} y_{i,k} \sigma_{i,k}^2$ and I_q^{UB} is true if $\sum_{i,k} y_{i,k} \sigma_{i,k}^2 \leqslant \alpha_{q+1}$. U is a large constant that is greater than $\alpha_Q - \alpha_0$ and greater than $\sqrt{\alpha_Q}$. As in other assignment problems, we can relax the assignment variables $y_{i,k}$: in practice, relaxed solutions are very close to integral.

4.4 Properties of MPOU Games

It is natural to ask whether, like POU games, ex-ante MPOU games are convex, since convexity simplifies the analysis of stability and fairness. We show that this is, in fact, the case. We begin with a technical lemma.

Lemma 4.1. *Let $\langle N, \Pi, \tau, V \rangle$ be an MPOU game. Let $i \in N$ and $S \subset T \subseteq N \backslash \{i\}$ and $j \in T \backslash S$. Then we have:*

$$v(S \cup \{i\}) - v(S) \leqslant v(S \cup \{i, j\}) - v(S \cup \{j\}) \tag{4.5}$$

Proof. We let $A^*(S)$ denote the assignment of profiles that maximizes the social welfare of S. In the case where there are multiple social welfare-maximizing configurations of S, we use the one with highest aggregate variance. We observe that $v(T, A^*(S)) \leqslant v(T)$ because $A^*(S)$ imposes a constraint on the behavior of S. For technical reasons, we break the proof into two cases based on whether it is more beneficial for i) i to join coalition S when S is configured to maximize $v(S \cup \{i\})$ or ii) i to join coalition $S \cup \{j\}$ when $S \cup \{j\}$ is configured to maximize $v(S \cup \{j\})$.

Case 4.1. $v(S \cup \{i\}) - v(S, A^*(S \cup \{i\})) > v(S \cup \{i, j\}, A^*(S \cup \{j\})) - v(S \cup \{j\})$

On both sides of the inequality, we are adding $\{i\}$ to a set of agents without changing the configuration of that set of agents. Thus, the inequality implies that $\{i\}$ contributes more value on the left side than on the right side. Since the amount of value that $\{i\}$ contributes depends only on the variance of the coalition that it is joining, the inequality implies that $\sigma(S \cup \{j\}, A^*(S \cup \{j\}) < \sigma(S, A^*(S \cup \{i\}))$.

Since j contributes a non-negative amount of variance, $\sigma(S, A^*(S \cup \{j\})) \leqslant \sigma(S \cup \{j\}, A^*(S \cup \{j\}))$, and likewise, $\sigma(S, A^*(S \cup \{i\})) \leqslant \sigma(S \cup \{i\}, A^*(S \cup \{i\}))$. Applying these inequalities yields $\sigma(S, A^*(S \cup \{j\})) < \sigma(S \cup \{i\}, A^*(S \cup \{i\}))$, implying:

$$v(S \cup \{j\}) - v(S, A^*(S \cup \{j\})) < v(S \cup \{i, j\}, A^*(S \cup \{i\})) - v(S \cup \{i\}) \tag{4.6}$$

Then, applying the inequalities $v(S, A^*(S \cup \{j\})) \leqslant v(S)$ and $v(S \cup \{i, j\}, A^*(S \cup \{i\})) \leqslant v(S \cup \{i, j\})$, and rearranging terms:

$$v(S \cup \{i\}) - v(S) < v(S \cup \{i, j\}) - v(S \cup \{j\}) \tag{4.7}$$

which is a stronger version of the lemma.

Case 4.2. $v(S \cup \{i\}) - v(S, A^*(S \cup \{i\})) \leqslant v(S \cup \{i, j\}, A^*(S \cup \{j\})) - v(S \cup \{j\})$

Applying the inequality $v(S, A^*(S \cup \{i\})) \leqslant v(S)$ on the left side yields:

$$v(S \cup \{i\}) - v(S) \leqslant v(S \cup \{i, j\}, A^*(S \cup \{j\})) - v(S \cup \{j\}) \tag{4.8}$$

Applying on the right side $v(S \cup \{i, j\}, A^*(S \cup \{j\})) \leqslant v(S \cup \{i, j\})$ yields the lemma:

$$v(S \cup \{i\}) - v(S) \leqslant v(S \cup \{i, j\}) - v(S \cup \{j\}) \tag{4.9}$$

□

From Lemma 4.1, we immediately obtain:

Theorem 4.1. *The ex-ante MPOU game is convex.*

Proof. If $S = T$, then $v(S \cup \{i\}) - v(S) = v(T \cup \{i\}) - v(T)$ since the welfare-maximizing configurations of S and T are the same. If $S \subset T$, we repeatedly apply Lemma 4.1 to "grow" S one agent a time, creating a series of inequalities, until we relate S and T. □

Since the ex-ante MPOU game is convex, the Shapley value is in the core, hence we can compute a core allocation by averaging the payments from any number of join orders. In our experiments, we approximate the Shapley value by sampling [Castro *et al.*, 2009].

It is important that agents are incentivized to participate in the mechanism. We show that MPOU games are *individually-rational*—no agent receives less utility than his best outside option, i.e., what he would receive if he chose not to participate in the mechanism. To achieve this, we augment any instance of the game by adding a dummy profile to each agent with value equal to that of their (best) outside option.

Theorem 4.2. *Let G be an MPOU game where each agent has a profile $\pi_{out}^{(i)}$ with $V(\pi_{out}^{(i)}) = \theta_i$, $\sigma(\pi_{out}^{(i)}) = \mu(\pi_{out}^{(i)}) = 0$, where θ_i is the value of i's outside option. Then, G is ex-ante individually rational if core payments are used.*

Proof. Core payments exist because G is an MPOU game, hence convex. Suppose, by way of contradiction, agent i receives an expected payment less than θ_i. The stability condition of core payments requires that $t(i) \geqslant v(\{i\})$. However, this contradicts the fact that $v(\{i\}) \geqslant \theta_i$. □

4.5 Incentives in MPOU Games

MPOU games introduce a new coordination problem for coalitions that do not arise in POU games. In a fully-cooperative MPOU game, a coalition C agrees on a joint consumption profile prior to reporting its (aggregate) predicted demand. Despite this agreement, an agent $i \in C$ may have incentive to actually *use a profile that differs from the one agreed to.*

For example, consider an instance with a single agent, agent 0, who has two profiles: $\pi_{0,0} = \langle \mu_{0,0}, \sigma_{0,0} \rangle = \langle 0, 1 \rangle$ with $V_0(\pi_{0,0}) = 0$ and $\pi_{0,1} = \langle \mu_{0,1}\sigma_{0,1} \rangle = \langle 0, 10 \rangle$ with $V_0(\pi_{0,1}) = 1$. The prediction of use tariff has $p = 1$ and $L(\underline{p}, \bar{p}) = 2$. If agent 0 uses profile $\pi_{0,0}$, social welfare is $0 - 2 = -2$. If he uses $\pi_{0,1}$, social welfare is $0 - 20 = -20$. Thus, assigning agent 0 profile $\pi_{0,0}$ maximizes social welfare. However, agent 0 would increase his net utility by $V_0(\pi_{0,1}) - V_0(\pi_{0,0}) = 1$ by defecting to profile $\pi_{0,1}$.

Typically, a penalty should be imposed for such a deviation to ensure that C's welfare in maximized. Unfortunately, i's profile *cannot be directly observed.* Only his realized consumption x_i is observable, and it is related only *stochastically* to his underlying behavior (adopted profile). As such, any transfer or penalty in the coalitional allocation must depend on x_i, showing that an ex-ante analysis is insufficient for MPOU games (in stark contrast to POU games). Furthermore, since x_i is stochastic, it could have arisen from i using either profile (i.e., we have no direct signal of the i's chosen profile), which makes the design of such transfers even more difficult. Finally, the poor choice of a transfer function may compromise the convexity of the ex-ante game, undermining our ability to compute core payments.

To address these challenges, we introduce a *separating function* $D_i(x_i)$. For each agent i, D_i maps i's realized consumption to an additional *ex-post separating payment.*

Definition 4.1. D_i is a *separating function* (SF) for i under assignment A if it satisfies the *incentive* and *zero-expectation* conditions.

- **Incentive**: $\mathbb{E}_{x_i \sim A(i)}[D_i(x_i)] > \mathbb{E}_{x_i \sim \pi}[D_i(x_i)] + V_i(\pi) - V_i(A(i))$ for any $\pi \in \Pi_i$ such that $\pi \neq A(i)$.
- **Zero-expectation**: $\mathbb{E}_{x_i \sim A(i)}[D_i(x_i)] = 0$.

Intuitively, the incentive condition ensures that the agent is incentivized to use the assigned profile, and the zero-expectation condition requires that the payments introduced by the incentive condition do not affect the agent's expected payment if he uses the assigned profile. Since agents are assumed to be risk neutral, each agent's payoffs are unaffected by addition of a SF as long as the agent uses the profile assigned by the coalition. Thus, payments remain in the core after the addition of an SF.[5]

The rest of this section describes how to find SFs. We begin by showing that a weaker form of separating function can trivially be transformed into a SF.

Definition 4.2. D_i is a *weak separating function* (WSF) for i under assignment A if $\mathbb{E}_{x_i\sim A(i)}[D_i(x_i)] > \mathbb{E}_{x_i\sim\pi}[D_i(x_i)]$ for any $\pi \in \Pi_i$ such that $\pi \neq A(i)$.

Observation 4.1. *Let D_i be a WSF for i under assignment A. Then, $D_i' = w_0 D_i + w_1$ is an SF, where*

$$w_0 = \max_{\pi\in\Pi_i,\pi\neq A(i)} \frac{V_i(\pi) - V_i(A(i))}{\mathbb{E}_{x_i\sim A(i)}[D_i(x_i)] - \mathbb{E}_{x_i\sim\pi}[D_i(x_i)]} \tag{4.10}$$

and $w_1 = -\mathbb{E}_{x_i\sim A(i)}[w_0 D_i(x_i)]$.

Thus, it is sufficient to find a WSF. It is worth noting that not all SFs are WSFs. In particular, SFs can fail to satisfy the WSF condition for profiles π that have lower value than the assigned profile, i.e., $V_i(\pi) - V_i(A(i)) < 0$. These are profiles that the agent would have no incentive to defect to.

When an agent has only two profiles, this is straightforward: we let D_i be the PDF of the assigned profile minus the PDF of the unassigned profile. The proof for this statement is algebraic, using the fact that $\mathcal{N}(x;\mu_0,\sigma_0)\mathcal{N}(x;\mu_1,\sigma_1)$ has a closed form that is proportional to a normal PDF in x.

Theorem 4.3. *Let i be an agent with two profiles π_0 and π_1 and let $A(i) = \pi_0$. Then, w.l.o.g., $D_i(x_i) = \mathcal{N}(x_i;\mu_0,\sigma_0) - \mathcal{N}(x_i;\mu_1,\sigma_1)$ is a WSF for i under A.*

Proof. We show that the minimum of $\mathbb{E}_{x\sim\mathcal{N}(\mu_0,\sigma_0)}[\mathcal{N}(x;\mu_0,\sigma_0)-\mathcal{N}(x;\mu_1,\sigma_1)]-\mathbb{E}_{x\sim\mathcal{N}(\mu_1,\sigma_1)}[\mathcal{N}(x;\mu_0,\sigma_0)-\mathcal{N}(x;\mu_1,\sigma_1)]$ occurs when $\mu_1 = \mu_0$ and $\sigma_1 = \sigma_0$, and that the value of the expression at that point is positive.

We make use of the fact that $\mathcal{N}(x;\mu_1,\sigma_1)\mathcal{N}(x;\mu_2,\sigma_2)$ is a function proportional to the PDF of a normal distribution. Specifically,

$$\begin{aligned}&\mathcal{N}(x;\mu_0,\sigma_0)\mathcal{N}(x;\mu_1,\sigma_1) = \\ &\qquad \mathcal{N}\left(\mu_0;\mu_1,\sqrt{\sigma_0^2+\sigma_1^2}\right)\mathcal{N}\left(x;\frac{\sigma_0^{-2}\mu_0+\sigma_1^{-2}\mu_1}{\sigma_0^{-2}+\sigma_1^{-2}},\frac{\sigma_0^2\sigma_1^2}{\sigma_0^2+\sigma_1^2}\right)\end{aligned} \tag{4.11}$$

Then, by expanding terms and applying (4.11):

$$\begin{aligned}&\mathbb{E}_{x\sim\mathcal{N}(\mu_0,\sigma_0)}[\mathcal{N}(x;\mu_0,\sigma_0) - \mathcal{N}(x;\mu_1,\sigma_1)]- \\ &\mathbb{E}_{x\sim\mathcal{N}(\mu_1,\sigma_1)}[\mathcal{N}(x;\mu_0,\sigma_0) - \mathcal{N}(x;\mu_1,\sigma_1)] \\ &\quad = \frac{1}{2\sigma_0\sqrt{\pi}} - 2\mathcal{N}\left(\mu_1;\mu_0,\sqrt{\sigma_0^2+\sigma_1^2}\right) + \frac{1}{2\sigma_1\sqrt{\pi}}\end{aligned} \tag{4.12}$$

[5] Our use of zero-expectation payments for risk-neutral agents is mechanically similar to Cremer and McClean's [1988] revenue-optimal auction for bidders with correlated valuations.

We then minimize with respect to μ_1 and σ_1. Since the middle term is the only one that contains μ_1, we can minimize it separately:

$$-\frac{2}{\sqrt{2\pi(\sigma_0^2+\sigma_1^2)}}\exp\left(-\frac{(\mu_0-\mu_1)^2}{2(\sigma_0^2+\sigma_1^2)}\right) \tag{4.13}$$

Since the exponent (i.e., the argument of the exp function) is always non-positive, it is maximized when it is zero, i.e., $\mu_1 = \mu_0$. Making this substitution yields:

$$\frac{1}{2\sigma_0\sqrt{\pi}} - \frac{2}{\sqrt{2\pi(\sigma_0^2+\sigma_1^2)}} + \frac{1}{2\sigma_1\sqrt{\pi}} \tag{4.14}$$

Setting the derivative with respect to σ_1^2 to zero yields two real roots of $\sigma_0 = \pm\sigma_1$. The second derivative at these points is positive. Thus, it is a minimum. The value of the original expression at this point is 0 and positive otherwise. □

With more than two profiles, this approach does not always work. Instead, we can use a linear program (LP) to find the coefficients of a linear combination of the profile PDFs. Formally, denote the PDFs of the profiles as $\boldsymbol{\mathcal{N}}_i(x_i) = \langle \mathcal{N}(x_i;\mu_0,\sigma_0),\ldots,\mathcal{N}(x_i;\mu_{|\Pi_i|-1},\sigma_{|\Pi_i|-1})\rangle$, their weights as $\boldsymbol{y}_i$, and search over $\boldsymbol{y}_i \in \mathbb{R}^{|\Pi_i|}$ for a separating function of the form $D_i(x_i,\boldsymbol{y}_i) = \boldsymbol{y}_i \cdot \boldsymbol{\mathcal{N}}_i(x_i)$. We use an LP that minimizes the L_1-norm of $\boldsymbol{y}_i$ subject to $\mathbb{E}_{x_i\sim A(i)}[D_i(x_i,\boldsymbol{y}_i)] > \mathbb{E}_{x_i\sim\pi}[D_i(x_i,\boldsymbol{y}_i)]$ for all $\pi \in \Pi_i, \pi \neq A(i)$. Ideally, we would also like to minimize the variance of the separating payment, giving agents *maximal certainty* w.r.t. this payment; however, this objective is not tractable in an LP (we leave this question to future work). In our experiments below, we do, however, assess the variance of the separating payment.

A feasible $\boldsymbol{y}_i$ corresponds to a linear combination of vectors whose sum has only positive entries. We call these the *difference vectors* of D_i. While we cannot prove that a feasible $\boldsymbol{y}_i$ always exists, viewing the problem in terms of difference vectors suggests why they exist in practice:

Definition 4.3. Let $A(i)$ be π_0 (w.l.o.g.). For each profile $\pi_k \in \Pi_i$ the *difference vector* $\boldsymbol{d}_k = \mathbb{E}_{x\sim\pi_k}[\mathcal{N}(x;\pi_0,\sigma_0] - \langle \mathbb{E}_{x\sim\pi_k}[\mathcal{N}(x;\mu_1,\sigma_1)],\ldots,\mathbb{E}_{x\sim\pi_k}[\mathcal{N}(x;\mu_{|\Pi_i|-1},\sigma_{|\Pi_i|-1})]\rangle$.

Note that these vectors do not depend on $\boldsymbol{y}_i$. We can restate the LP constraints using difference vectors:

Theorem 4.4. *Let i have profiles Π_i and let A assign a profile to i. There exists $\boldsymbol{y}_i \in \mathbb{R}^{|\Pi_i|}$ that makes $D_i(x_i,\boldsymbol{y}_i)$ a WSF if and only if there is a linear combination of the difference vectors of $D_i(x_i,\boldsymbol{y}_i)$ that has only positive entries.*

Proof. First, we prove the forward direction. Let $\boldsymbol{c}$ be the coefficients of the linear combination of the difference vectors that has only positive entries, i.e., $\sum_{k\in|\Pi_i|}\boldsymbol{c}_k\boldsymbol{d}_k = \boldsymbol{b}$ where $\boldsymbol{b}$ is element-wise positive. Then, $\mathbb{E}_{x_i\sim A(i)}[D_i(x_i,\boldsymbol{c})] - \mathbb{E}_{x_i\sim\pi}[D_i(x_i,\boldsymbol{c})] = \boldsymbol{c}\boldsymbol{d}_k = \boldsymbol{b}_{k-1}$. Since $\boldsymbol{b}$ is element-wise positive, letting $\boldsymbol{y}_i = \boldsymbol{c}$ makes $D_i(x_i,\boldsymbol{y}_i)$ a separating function.

The reverse direction is also straightforward. Suppose $D_i(x_i,\boldsymbol{y}_i)$ is a separating function. Then, let $\boldsymbol{b}_{k-1} = \mathbb{E}_{x_i\sim A(i)}[D_i(x_i,\boldsymbol{c})] - \mathbb{E}_{x_i\sim\pi}[D_i(x_i,\boldsymbol{c})] = \boldsymbol{y}_i \cdot \boldsymbol{d}_k$. Thus, taking $\boldsymbol{y}_i$ as the coefficients of the linear combination of difference vectors equals $\boldsymbol{b}$, which has only positive entries. □

Corollary 4.1. *Let $\boldsymbol{d}_k$ be the difference vectors for agent i. If the difference vectors are linearly independent, a setting of $\boldsymbol{y}_i$ exists that makes $D_i(x_i,\boldsymbol{y}_i)$ a WSF.*

Proof. If the difference vectors are linearly independent, there exists a coefficient vector $\boldsymbol{c}$ that makes $\sum_{k \in |\Pi_i|} \boldsymbol{c}_k \boldsymbol{d}_k$ elementwise positive. We can take $\boldsymbol{y}_i = \boldsymbol{c}$ to satisfy the corollary. □

We generally expect a random set of vectors to be linearly independent as the set of matrices drawn from the reals with non-independent rows has Lebesgue measure zero. We have yet to encounter an instance where a separating function does not exist in our experiments. It is an open question as to whether a separating function of this form always exists.

4.6 Manipulation in MPOU Games

While we defer a thorough investigation of manipulation of MPOU games to future work, we briefly discuss a simple form of manipulation: *adding profiles to, or removing profiles from, an agent's report.* Formally, we say that an agent can *manipulate* an MPOU game if they gain expected utility by misreporting their true set of profiles. Here, we simplify the discussion by assuming that agents have a true underlying set of profiles, and we rely on the results of the previous section by assuming that each agent can be incentivized to use their assigned profile without changing their expected payoff.

Agents are not incentivized to strategically withhold information if they otherwise report truthfully. However, reporting additional untruthful profiles will benefit the agent, as long as those profiles are not assigned by the mechanism.

Theorem 4.5. *Let G be an MPOU game, let G' be identical to G except agent i reports an additional profile $\pi^{(i)}_{extra}$. Let all of i's reported profiles be truthful except $\pi^{(i)}_{extra}$ and let at least one of these conditions hold: (i) $\pi^{(i)}_{extra}$ is truthful or (ii) $\pi^{(i)}_{extra}$ is not the assigned profile. Then, agent i's payoff in G' is greater than or equal to its payoff in G if payments are used that average marginal contributions over the same join orders.*

Proof. First, we establish that i's Shapley value is greater with the additional profile. Each time agent i is added to a coalition S in a join order, agent i's marginal contribution to $v(S \cup \{i\})$ with the extra profile is greater than or equal to its contribution with its original profiles. Thus, $t_{G'}(i) \geqslant t_G(i)$.

This condition is not sufficient to ensure that i increases his payoff, which is equal to his coalitional payment minus the reported value of the assigned profile plus the true value of the assigned profile. In condition (i), the Shapley value equals the payoff value and in condition (ii), the assigned profile is the same in G and G'. Thus, i's payoff is greater or equal in G' in either case. □

Note that the theorem applies both to the Shapley value, which can be expressed as an average over marginal contributions over join orders, and to sampling-based approximations, such as the ones used in our experiments.

We outline two ways of heuristically combating manipulation by reporting additional profiles. The first is to simply limit the number of reported profiles, either by creating a cap or by charging agents per profile they report, limiting the amount agents can gain by manipulating. This approach leads to a non-truthful equilibrium, and it penalizes agents who have more complicated utility functions.

The second approach emerges from an approximation to the Shapley value that happens to remove the incentive to add additional profiles that are not selected. Recall that i's Shapley value in coalition C can be interpreted as the average marginal value that i contributes over all orders that agents join C. Computing this requires recalculating the optimal assignment of profiles before and after i joins since

the addition of i may cause a change in the optimal assignment for the other agents. Because this is computationally expensive, we approximate it by fixing agents to the profile they are assigned in the grand coalition. Formally, we let i's *Shapley value with fixed profiles* be

$$s_C(i,N) = \sum_{S \subseteq C \setminus \{i\}} \frac{|S|!(|C| - |S| - 1)!}{|N|!}(v(S \cup \{i\}, A^*(N)) - v(S, A^*(N))) \tag{4.15}$$

Recall that $v(S, A^*(N))$ is the value of coalition S under the assignment that maximizes the value of coalition N. We find the approximation is quite close to the true Shapley value in our setting. The approximation sacrifices exact convexity because it does not discriminate between agents based on how attractive their unassigned profiles are, which has the additional consequence that, as long as agents report their true profiles, they have no incentive to add additional false ones.

Theorem 4.6. *Let G be an MPOU game, let G' be identical to G except agent i reports an additional profile $\pi^{(i)}_{extra}$. Let all of i's reported profiles be truthful except $\pi^{(i)}_{extra}$. Then, agent i's payoff in G' is less than or equal to its payoff in G, if payments are used that average marginal contributions over the same join orders and fix i's profile to its assigned profile.*

Proof. Since we assume that i reports all of its profiles truthfully, the true value of $\pi^{(i)}_{extra}$ is 0. Then, either the mechanism selects $\pi^{(i)}_{extra}$ or it does not. If it does, i's payoff will be negative since it receives 0 value from $\pi^{(i)}_{extra}$, and thus, its payoff decreased because the mechanism is individually rational according to Theorem 4.2. If it does not, i's payoff is unchanged because $\pi^{(i)}_{extra}$ does not affect its payoff. □

4.7 Learning Utility Models

To empirically test the MPOU framework and our separating functions, we require consumer utility functions. As we know of no data set with such utility functions, we learn household (agent) utility models from real electricity usage data from Pecan Street Inc. [Rhodes *et al.*, 2014].[6] We define our prediction period as 4–7 pm each day, when electricity usage typically peaks in Austin, Texas, where the data was collected. We decompose utility into two parts: $V_i^{(\mu)}(w,\mu)$ describes the value an agent i derives from his mean consumption given a vector w of weather conditions; and $V_i^{(\sigma)}(\sigma,\mu)$ represents utility derived from variance in consumption behavior. Agent i's utility is $V_i(w,\mu,\sigma) = V_i^{(\mu)}(w,\mu)V_i^{(\sigma)}(\sigma,\mu)$.

Estimating $V_i^{(\mu)}$ is difficult, since we lack data for some aspects of the problem. Thus, we make some simplifying assumptions: (i) consuming 0 kWh yields value \$0; and (ii) $V_i^{(\mu)}(w,\mu)$ is concave and increasing. We learn a model for each of 25 households that have complete data (i.e., hourly consumption) from 2013–2015 (about 1100 data points per household), using select weather conditions w and mean consumption between 4–7 pm as input, and outputting value (in dollars). We use this valuation function to predict consumption by maximizing an agent's net utility under the observed price:

$$V_i^{(\mu)}(w,\mu) = z_i^{(0)}(w)\left(\mu - z_i^{(1)}(w)\right)^{z_i^{(2)}(w)} + z_i^{(3)}(w) \tag{4.16}$$

constraining $z_i^{(0)} > 0$, $z_i^{(1)} > 0$, $0 < z_i^{(2)} < 1$, $z_i^{(3)}(w) \geqslant 0$ (Figure 4.2 depicts the utility model). We use a homogeneous function to represent utility [Simon and Blume, 1994]. The term $z_i^{(3)}(w)$ has no influence on predictions: it can be viewed as inherent value due to weather, and accounts for the

[6] Publicly available at `pecanstreet.org`.

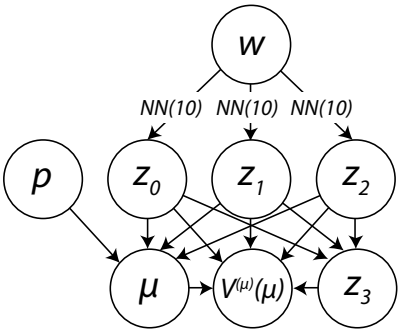


Figure 4.2: The learned valuation model. $NN(10)$ denotes a neural network with 10 hidden units.

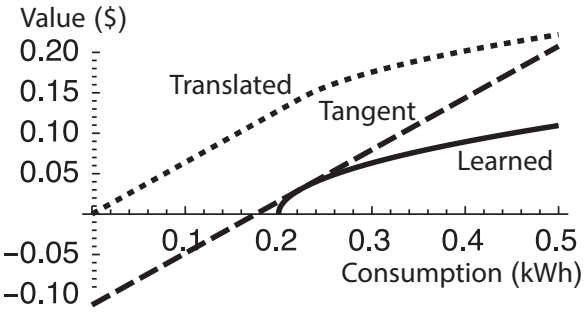


Figure 4.3: Translating the valuation function to pass through the origin

flexibility provided by the $z_i^{(1)}$ term, which may create valuations where consumption 0 yields negative value (violating our assumptions). To prevent this, we set $z_i^{(3)}(w)$ to ensure the tangent at the predicted consumption for \$0.64 (the largest price in the data set) passes through (0,0) (see Figure 4.3). When this tangent crosses the y-axis above 0, we set $z_i^{(3)}(w) = 0$ and splice in an exponential ax^b that passes through (0,0) and matches the derivative at the splice point.

For training, we use the model to predict consumption by solving the net utility maximization problem, $\max_\mu(V_i(w,\mu) - \mu p)$, yielding:

$$\hat{\mu}(w,p) = \frac{p}{z_i^{(0)}(w) z_i^{(2)}(w)}^{\frac{1}{z_i^{(2)}(w)-1}} + z_i^{(1)}(w) \tag{4.17}$$

We represent $z_i^{(0)}, z_i^{(1)}$ and $z_i^{(2)}$ in fully-connected single-layer neural networks, each with 10 hidden units and ReLU activations, and train the model with backpropagation. We implement the model in TensorFlow [Abadi *et al.*, 2015] using the squared error loss function and the Adam optimizer [Kingma and Ba, 2015]. We use Dropout [Srivastava *et al.*, 2014] with a probability of 0.7 on each hidden unit.

We split the data into 80% train and 20% test for each household. Table 4.1 compares the prediction accuracy of our model ("valuation") to (i) an unstructured neural network, and (ii) the best constant prediction for each household. The unstructured net learns a mapping from $\langle w, p\rangle$ to μ directly using 10 hidden units, without an intervening utility model.[7] The best constant prediction disregards weather and price data, and simply predicts average consumption for that household. Table 4.1 shows that the valuation model overfits somewhat, but that predictive accuracy is on par with the unstructured model. Moreover, the valuation model has lower variance in test RMSE. This shows that our constraints on the form of the valuation function are not unduly restrictive and validates the value predictions produced by these learned models. However, we believe these value functions significantly underestimate value because we lack consumption observations when the price is higher is than \$0.64.

Figure 4.4 shows the learned valuation for the 25 households. Each line represents a household's response to different weather conditions. While temperature is the most significant predictor of power usage, different households appear to exhibit sensitivity to different factors.

[7] Our other implementation choices are the same as the valuation model, except we use Dropout of 0.5.

Table 4.1: Comparison of model prediction accuracy by root-mean-square error (RMSE). We divide each household's consumption amounts by their largest observed consumption.

Model	Mean train RMSE	Std. dev. train RMSE	Mean test RMSE	Std. dev. test RMSE
Valuation	0.137	0.0168	0.148	0.0194
Unstructured	0.142	0.0226	0.144	0.0284
Constant	0.204	0.0345	0.205	0.0411

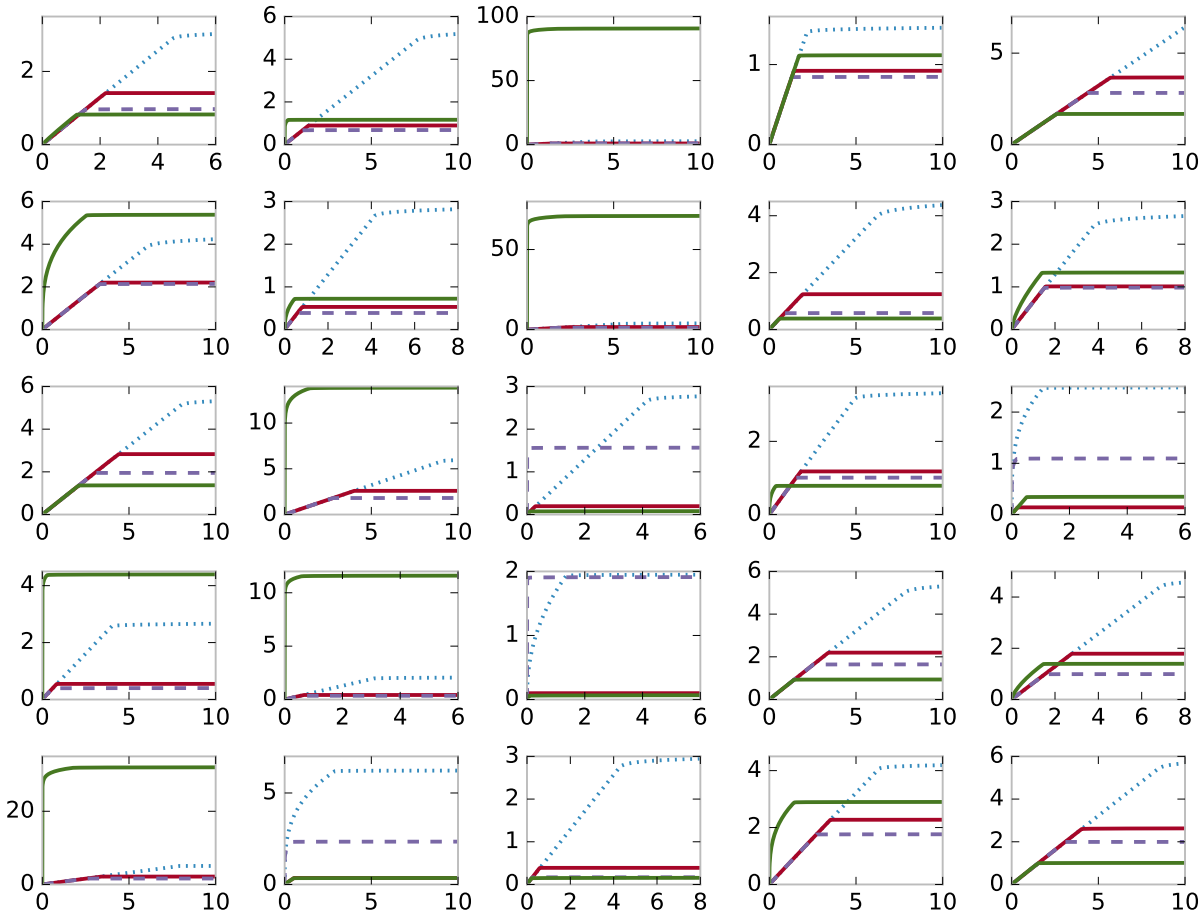


Figure 4.4: Learned value models for the 25 households with consumption mean (kWh) on the x-axis and value ($) on the y-axis. The red line represents the median weather conditions. The dotted line represents the median day with 90th percentile or higher temperature. The dashed and green lines are the same for sunshine and humidity, respectively.

Modeling Unpredictable Consumption

Unfortunately, we do not have access to electricity usage data where consumers are charged differently depending on the accuracy of their predictions. Our model of the value of unpredictable consumption is thus speculative, but uses the Pecan Street data as a starting point. We assume that each household chooses the standard deviation that maximizes its utility (since they are not being charged for σ), and that it has an optimal fraction β_i of σ/μ that does not depend on other conditions. We estimate β_i from the data by treating each data point as having an observed σ equal to the absolute error in consumption prediction made by the learned valuation model. We assume no value is gained by increasing σ above the optimal ratio, and use an exponential to represent the loss in value when σ is reduced,

$$V_i^{(\sigma)}(\sigma, \mu) = \max\left(\frac{\mu/\sigma}{\beta_i}, 1\right)^{\gamma_i}, \tag{4.18}$$

where γ_i is a constant representing i's cost for being predictable. A higher γ_i means that consumer i values variance more highly. In our experiments, we sample γ_i from the uniform distribution over the interval $[0.1, 2]$.

4.8 Experiments

We experimentally evaluate our mechanism for MPOU games. The questions we study experimentally are:

1. How important is consumer coordination under POU tariffs?
2. What is the social welfare gain from using an MPOU model vs. a flat tariff?
3. How important is an agent's choice of reported profiles?
4. What are the variances of the payments introduced by the separating functions?

4.8.1 Experimental Setup

We first describe the experimental setup: how we select agents, profiles and tariffs. For each trial, we select weather conditions w uniformly at random from the Pecan Street data. To generate agents representing households, we sample from our 25 learned household utility models, using w as input and adding a small amount of zero mean noise to the model parameters. We sample γ_i from the uniform distribution $[0.1, 2]$ for each agent i. Each data point is an average of 100 trials with 5000 sampled (household) agents, unless otherwise noted. One of the goals of our experiments is to study the consequences of different choices of reported profile. To do this, we vary the way profiles are generated. Each agent has four profiles: a *base profile* (predicted to be optimal under a flat rate tariff with rate equal to the fixed-rate p of the POU tariff), and three others reflecting reduced consumption mean or variance. The first reduces the base profile *mean* by the amount required to reduce value by u%, which we call the *profile spacing*. The second reduces *variance* to reduce value by u%. The third reduces both. For example, if the base profile has mean μ and standard deviation σ, the reported profiles would be:

1. The base profile π_0.

2. Profile π_1 with standard deviation σ and mean μ_1 such that $V_i((\mu_1, \sigma)) = (1 - \frac{u}{100})V_i((\mu, \sigma))$.
3. Profile π_2 with mean μ and standard deviation σ_2 such that $V_i((\mu, \sigma_2)) = (1 - \frac{u}{100})V_i((\mu, \sigma))$.
4. Profile π_3 has mean μ_1 and σ_2.

We vary u throughout the experiments.

To generate tariffs, we vary the amount of emphasis each tariff puts on accurate predictions vs. the amount consumed. We let the *predictivity emphasis* (PE) of a tariff w.r.t. a group of agents be the fraction of the expected total cost paid for prediction penalties when each uses his base profile. In practice, PE should be set to match the properties of the reserve power generation capacity that is available: a higher PE corresponds to more expensive reserves. A tariff is *revenue-equivalent* to another with respect to a specific set of profiles if the revenue of the two is the same for that set. All of our tariffs will be revenue-equivalent with respect to the set of base profiles. To find a revenue-equivalent tariff with a certain PE, we use a numerical solver to find a tariff of the form $\langle p, r, r \rangle$ with the appropriate total cost. Intuitively, a higher PE should result in larger benefits from POU tariffs, and we find that to be the case in our experiments.

To generate Shapley values, we sample a number of join orders equal to the logarithm of the number of agents in the instance. Shapley values were very close to linear in the standard deviation of the assigned profile. The average Shapley payment for prediction was \$0.41 per kWh of uncertainty across trials with PE 10%, and \$0.82 per kWh with PE 20%.[8] Within a single trial, the standard deviation of this ratio was less than 0.01 on average, suggesting that it is not necessary to optimize the choice of profiles every time an agent is added in a join order—it is sufficient to fix each agent's profile to the one it is assigned. We exploit this fact to run larger experiments.

4.8.2 Results

We first address the question of how important it is for agents to coordinate their consumption under a POU tariff. We define the *uncoordinated POU setting* as the scenario where agents are subject to a POU tariff, but do not coordinate their consumption behavior, i.e., each agent uses the profile that individually maximizes its net utility relative to that POU tariff. Then, as is standard in that setting, the grand coalition forms and makes the optimal baseline prediction. Figure 4.5 shows the social welfare derived by agents in the uncoordinated POU setting as a percentage of their social welfare under a revenue-equivalent fixed-rate tariff. We see that the average social welfare achieved in the uncoordinated POU setting is less than that of the fixed rate setting for all profile spacings. Individual agents react to the POU tariff by increasing their predictivity, and thus decreasing their realized value, but they do not account for the predictivity discount that results from being part of a coalition. As profile spacing increases, more agents shift away from their base profile and social welfare decreases, dropping to 70% of the fixed-tariff social welfare when spacing is 25%. These results underscore the need for some way for agents to coordinate their profile choices under POU tariffs and highlight one of the main challenges of successfully implementing POU tariffs in practice.

Next, we study the social welfare gains that can be achieved by a POU tariff when agents coordinate optimally under the MPOU framework. Figure 4.6 shows the effect of profile spacing (u) on the welfare gained by switching from a fixed-rate tariff to a revenue-equivalent POU tariff.[9] Overall welfare gains

[8] This and other tariffs in this section have $0.2 \leqslant \underline{p} = \bar{p} \leqslant 1.5$.

[9] Each instance took around 3 minutes on a single thread of 2.6 GHz Intel i7, 8 GB RAM.

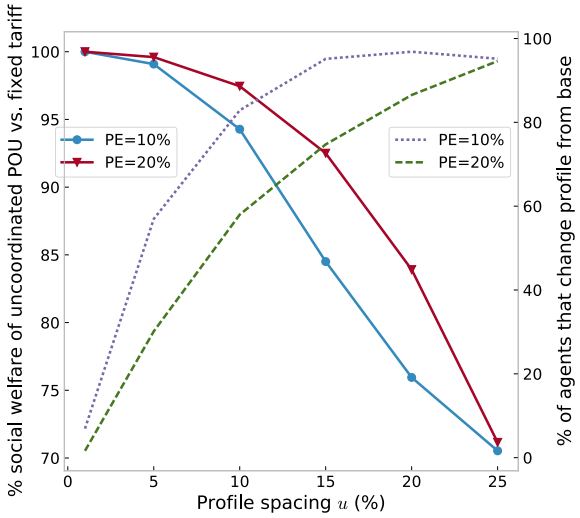


Figure 4.5: Profile spacing vs. % of social welfare of fixed-rate tariff for uncoordinated POU setting and % of agents that change profile

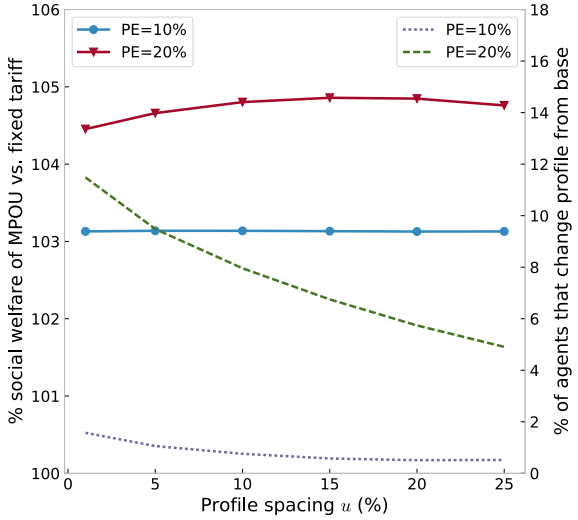


Figure 4.6: Profile spacing vs. social welfare % gain from fixed-rate tariff and % of agents that change profile

are moderate, around 3.13% for PE of 10% and 4.4–4.9% for PE of 20%. A higher PE results in a larger social welfare gain because agents only benefit from cooperating when trading off predictivity for inherent utility. Profile spacing appears to have limited impact on social welfare gain, suggesting that most of the gain is achieved by the effective reduction in fixed-rate price under a POU tariff. We note that these experiments are the first to study end-to-end social welfare gain from a POU tariff. They show that MPOU games offer a large increase in social welfare over the POU games model, from 3% at low profile spacing to 47%–50% at 25% profile spacing.

Figure 4.6 appears to indicate that personalizing profile spacing based on each agent's value for predictivity would increase social welfare further. We see this because increasing profile spacing increases welfare up to a spacing of 15% for both PE levels, but the number of agents that shift profiles decreases as spacing is increased (shown on the right-side axis). We hypothesize that welfare could be further increased if agents with higher γ spaced their profiles farther apart than those with lower.

Next, we address the question of uncertainty introduced by separating payments. Recall that while separating payments have expectation zero, they introduce additional uncertainty to agent payments. We find that the amount of uncertainty introduced is, in fact, minimal, and decreases with instance size and increased PE. Figure 4.7 shows the ratio of the standard deviation of the separating payment to the Shapley payment for predictivity. The standard deviation of the separating payment is on average 15–20% of predictivity payment for PE of 10% and 7.5–10% for PE of 20%, and increases slightly as profile spacing increases. Note that only agents that actually require a separating function are taken into account, around 1–2% of all agents for PE of 10% and 5–10% for PE of 20%, on average. More agents require separating payments as PE increases, but the uncertainty introduced by each decreases. Note that these are uncertainties for a single instance of the game, and if the game is played repeatedly (e.g., every day), the aggregate uncertainty will decrease as the independent random variables are added.

Figure 4.8 shows the same uncertainty ratio for a single large instance versus the predictivity flexibility (γ) of each agent. This instance has PE of 20%, 100,000 agents, profile spacing of 15% and takes 90 min. to solve. The ratio is shown for the 4876 agents that require separating functions. The magnitude of the

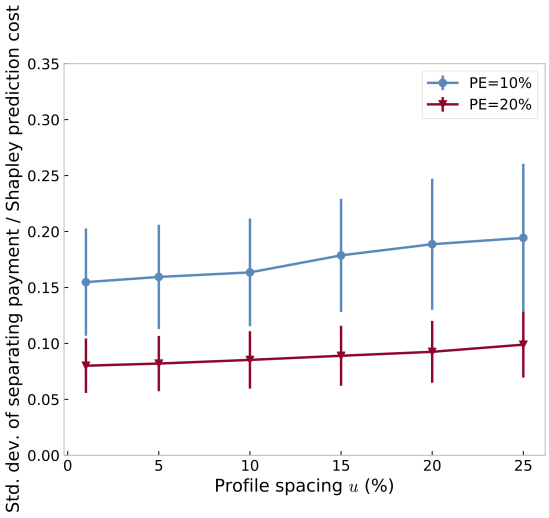


Figure 4.7: Comparison of the standard deviation of the separating function payment to the ex-ante payment for prediction accuracy. Bars show one standard deviation. 5000 agents, 100 trials

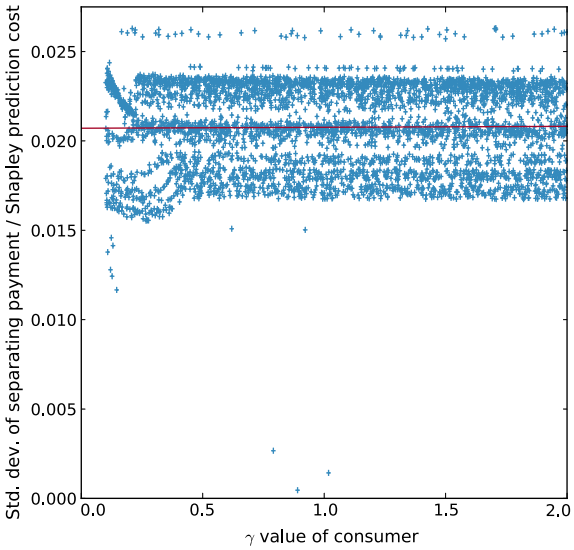


Figure 4.8: Comparison of the standard deviation of the separating function payment to the ex-ante payment for prediction accuracy

introduced uncertainty is smaller in this larger instance with an average of 2.07% (and not exceeding 3% for any agent). In addition, predictivity flexibility has little affect on the introduced uncertainty: the linear least-squares fit (red line) has slope of less than 10^{-4}.

To summarize, POU games, while well intentioned, decrease social welfare compared to fixed-rate tariffs in our experiments because they do not allow agents to coordinate their choice of profiles. MPOU games realize the promise of POU tariffs, increasing social welfare by 3–5% over fixed rate tariffs, and 3%–50% over POU games. Separating functions are a computationally efficient way of addressing the principal–agent problem that arises for some agents in MPOU games, 1–10% of agents in our experiments. For these agents, separating functions solve the principal–agent problem at the cost of adding a variance of 7.5–20% to the agent's predictivity payment. Again, the independence of the separating functions across time periods causes the separating function payment to approach 0 rapidly over longer periods of time (e.g., a month, the typical time period between electricity bills). It is worth noting that while we assume risk neutrality in our discussion of SFs, it is not necessarily the case that risk neutrality is required if the game is played sufficiently often because of the reduction in variance that results from repeated plays.

4.9 Conclusion and Future Work

We have introduced *multiple-profile POU (MPOU) games*, a framework for coordinating agent behavior under POU tariffs. MPOU games allow agents to express their consumption utility functions, while maintaining convexity of the basic POU model. MPOU games introduce a new class of incentive problems due to agent actions being partially observable: we introduce *separating payments* to restore proper incentives. Our experimental utility models are learned from historical electricity usage data in a novel way. Our experiments show that the social welfare gained by introducing the MPOU model (relative to a fixed-rate tariff) is non-trivial, and the POU model appears to cause significant social welfare losses

due to the lack of coordination. As a result, the social welfare gains of MPOU vs. POU are significant, strongly indicating that MPOU is vital to the practical deployment of POU tariffs. The effectiveness of the MPOU model depends both on the predictivity emphasis (PE) of reserve generation and on the value to consumers of consuming unpredictably, which are both areas where more real-world data is needed. We find that the uncertainty introduced by separating payments decreases as instance size increases, and decreases in aggregate as more iterations of the game are played. Increased PE increases the number of agents that need separating functions, but the uncertainty introduced decreases.

Interesting future directions for MPOU games remain. Better access to household utility data, especially for variance of consumption, and data about the PE of generation mixes would allow us to test social welfare gain more precisely.

We would ideally like to calculate separating functions to minimize the amount of variance they introduce to agents' payments, but this requires solving a difficult optimization problem which we do not address here. There is additionally an open question about separating function existence. While we show that the set of problems where separating functions fail to exist has measure 0, it would be useful to know if these problems can occur in practice.

Other critical aspects of the system are the ability of agents to manipulate, which we touch on only briefly, and how to elicit household utility functions. As in Chapter 3, this can be approached in an ad hoc manner, extending the analysis of Section 4.6 or through using the VCG mechanism. Note that separating functions are agnostic to the choice of mechanism, and the principal–agent problem they solve would occur in any model of this problem. We will return to the question of manipulability in the discussion of demand response models below.

A clear limitation of the MPOU games framework is the inability to handle correlations in demand across agents, e.g., due to weather. A related issue is that agents make individual predictions of global conditions, e.g., the weather, in this model, which is not useful information to the principal, who can presumably make better predictions of global conditions. The information that the principal is most interested in is how the agents would react to different global contexts. Developing the framework along these lines could yield a useful refinement. One approach to this would be to have each agent issue sets of demand profiles conditional on the realization of the global conditions. Then, social welfare maximization is performed for each realization of the global conditions, and agents are paid according to which conditions occur. One issue that would have to be addressed is that it is likely impossible to make the demand profiles completely conditionally independent. This raises the question of how badly the model performs under a low degree of correlation compared to full independence and if any tweaks can be made that improve performance in that setting.

While our discussion of POU and MPOU games has focused on electricity markets, we believe the approach may be more widely applicable in other cases where agents are contending with a scarce resource, e.g., internal allocation of computing resources across groups in a company or university.

Separating functions are worthy of future study. They provide a generic computational approach to the principal–agent problem with theoretical guarantees. Note that the separating function approach does not depend on the random variables having normal distributions or even distributions with any compact representation. The only requirement for the separating function approach is the ability to calculate the "overlap" between distributions, i.e., the expectation of the PDF of variable A taken with respect to the distribution of variable B. We can efficiently compute this by sampling as long as we can compute the PDF value at a point for at least one of the two random variables.

MPOU games are related to work in demand response, which has been the subject of several recent papers in AI [Ma *et al.*, 2016; Meir *et al.*, 2017; Ma *et al.*, 2017]. Demand response is a system where suppliers can call on consumers to reduce their consumption for a price. The idea is that demand response can act as a generation source. To do this, consumers bid on how much they are willing to reduce their consumption for what price.

The demand response problem is closely related to the incentive misalignment problem studied in this chapter. The problems address quite similar issues, but from different perspectives, and unifying the two models may be an interesting avenue of future work. To make the discussion concrete, it is worth describing the demand response model in these papers. They study a two-period interaction. In the first period, the agents report their types to the mechanism, the mechanism posts demand response prices and the agent chooses whether to prepare for demand response. In the second period, the agents choose whether to respond or not and the mechanism pays the agents based on the posted price function. The goal of the mechanism is reduce demand by a target M with probability at least τ. Specifically, in the first period:

1. The agents report their types to the mechanism. The types consist of a function describing how the agent can prepare to undertake demand response and a function that states the utility they will lose (as opportunity cost) if they reduce their consumption in the second period. The latter is stochastic and the agents will observe their realization of it in the second period.

2. The mechanism posts a payment function, i.e., for each agent, the reward or penalty to be received for reducing consumption by each amount.

3. The agents observe the payment functions and choose how to prepare for demand response.

In the second period:

1. Each agent observes their realized utility function and decides whether to undertake demand response or not, based on the incentives posted by the mechanism.

2. The mechanism pays the agent based on the posted prices.

The analysis in these papers is theoretical and studies restricted versions of the problem with different conditions on the form of how agents invest effort into responding and restrictions on the utility functions. The general idea of the theory is to show that truthful reporting is a dominant strategy, which is done through a mechanism similar to VCG with the Clarke pivot rule.

There are a few key differences between the demand response model and MPOU games, beyond the obvious (and important) one that demand response models the phase of the interaction where the agents report their types to the mechanism. The first and most critical difference is the objective of the optimization. Rather than maximize social welfare directly, demand response focuses on a reduction and reliability objective: reduce demand by a target M with probability at least τ. The perspective is that the central system knows how much demand response it needs to balance the market. This perspective is completely insensitive to cost to the central system. It is impossible to tell what the cost of the reduction will be before the mechanism is executed, and it could be that reducing the reduction target or reliability by a small amount would yield a large cost savings.

This difference in objectives greatly affects the mechanism's operation. The goal can be reformulated as selecting the set of agents who can reliably reduce their demand the most cheaply. In contrast, MPOU games' social welfare objective causes the model to interact with each agent.

The second major difference between the two models is the treatment of stability. MPOU games use a group concept of stability rooted in cooperative games whereas the demand response model considers only the incentive of individual agents to misreport. It is nevertheless clear that individuals have an incentive to form groups to bid in the demand response model. By doing so, members of group would reduce their uncertainty in the realization of their utility functions between the first and second period of the mechanism.[10]

It is possible that MPOU games could be applied directly to the demand response problem. This would require, at the least, extending them to allow for non-independent prediction errors and to address the question of truthful reporting. MPOU games are in some respects more principled than the demand response market approach, particularly because they allow an explicit tradeoff between the cost of reducing variance and the value achieved. The fact that MPOU games consider the possibility that agents would combine and enter the market as a single large agent is also an advantage.

[10]Note that although the MPOU has uncertain consumption and the demand response model has uncertain utility functions, the difference is primarily superficial.

Chapter 5

Experiential Preference Elicitation for Autonomous HVAC Systems

In this chapter, we study the problem of preference elicitation in a sequential decision-making problem, where the cost and accuracy of queries is dependent on the current state. We call this problem *experiential elicitation*. We are motivated by the problem of preference elicitation in an HVAC setting. We begin by presenting a general model of experiential elicitation that uses Markov decision processes to model the sequential decision-making problem, but is agnostic to the query type and noise and cost model. We then specialize that model to the HVAC problem, developing a new query type, the *relative value query*, which asks the user to compare the value of two states. We show that a preference prediction model based on *Gaussian processes*, generally well-suited to preference elicitation problems, can be extended to integrate the responses to relative value queries in closed form. We conclude by comparing our model to several natural baselines and show that it accrues more reward. Some parts of work in this chapter originate in Perrault and Boutilier [2018].

5.1 Introduction

In the previous two chapters, we presented techniques that improve the efficiency of electricity markets but which require extensive attention from the user: he has to report his preferences periodically to the market in the form of demand profiles. In this chapter, we focus on the problem of developing autonomous agents that represent users in these systems, reducing their attentional costs.

AI systems have two distinct components: the underlying mathematical model and the specific data that the mathematical model receives as input. The rigidity of the separation varies from problem to problem as does the way the data is acquired. In a single-agent interaction, an expert designer may control both the specification of the model and the input of data, blurring the line between the two components. Problems with features such as data that is expensive to acquire or communicate or when there is a non-expert user (who treats the model as a black box) make the separation more distinct. In these cases, the designer may explicitly model the problem of data acquisition, which may be as hard or harder than the problem solved by the AI system.

The data acquisition problem is particularly important in multi-agent systems where it is distributed, with each agent providing data about themselves or users on whose behalf they act. The agents may

have competing interests, e.g., in an auction, where each agent has an incentive to misreport in order to receive an item that they wouldn't have otherwise or to pay a lower price. These incentives have to be taken into account when acquiring data from the agents. Incentives to misreport can be just as strong in a cooperative setting. Consider an elicitation procedure where the underlying interaction between agents is a *convex cooperative game*. The agents have a strong incentive to coordinate their behavior because doing so increases their aggregate utility and it is easy to find a mutually acceptable division of the benefits of cooperation. Nevertheless, an individual may receive a larger share of the benefits of cooperation by misreporting their preferences.

In this chapter, we distinguish between the agents and the users they represent. As previously discussed regarding the electricity market design problems of Chapters 3 and 4, it does not make sense for individuals to interact in real time with the principal—too much communication is required for insufficient benefit. Instead, we envision an agent that reports profiles to the principal in each time period, using knowledge of their user's preferences, and takes the necessary coordinating actions on behalf of the user, e.g., controlling heating and cooling systems and dispatching appliances.

To represent the user, the agent has to obtain the user's preferences. This can be done in a variety of ways, which we divide into two broad categories. The first is using *revealed preference data*, in which the user's past decisions in the target (or a related) setting provide a (partial) view of their utility preferences [Samuelson, 1948; Beigman and Vohra, 2006; Ng and Russell, 2000]. The second is through *preference elicitation (PE)*, i.e., directly asking the user about their preferences for possible decisions or outcomes, using queries of various types [Chajewska *et al.*, 2000; Boutilier, 2002; Braziunas and Boutilier, 2010]. Either approach (or some combination) may be better suited to particular settings.

The revealed preference approach has the advantage of being completely passive. The only costs involved are those required to gather the requisite data, i.e., in the electricity case, sensors and a place to store the sensor data. If we have enough data about the user's past decisions, we can learn their preferences using a regression model. This is the approach of Chapter 4, where we model the decisions of users based on the Pecan Street electricity use data.

The key weakness of the revealed preference approach is that it requires that the relevant data exist in sufficient quantity. This may pose a chicken-and-egg problem. In order to gather data on how users interact with the mechanism, the mechanism must be implemented. However, because of the level of attention required to interact with the mechanism, the data gathered from observing users interact with it, without an intervening agent, would be of low quality.

One might object that, given enough low quality data, we can learn an accurate representation of the user's preferences, but there is a deeper problem: introducing a user-representing agent *changes* the user's preferences. Consider for example a case where electricity prices suddenly drop in the middle of the day, while the user is busy in a meeting at work. Without a user-representing agent, the user may be unable to take advantage of the decrease in prices. That does not mean that he would not choose to have an autonomous agent do so on his behalf.

It is worth mentioning that even in the case where relevant data exists, the user's desire for privacy may deter him from providing the data to the system.

The PE approach avoids this problem by *querying* the user about how he would act in hypothetical situations. PE has its own limitations. We will study two of the most important. First, by asking a user to introspect and analyze his preferences for specific outcomes—usually at some remove from the scenario where the data will be used—the *cognitive costs* of PE can be high. Second, PE often requires users

to imagine and compare outcomes without experiencing them directly, leading to inaccurate responses (even in cases where the user is incentivized to be truthful). For instance, an apartment selection assistant may ask queries such as "would you prefer apartment w at price x or apartment y at price z?" [Braziunas and Boutilier, 2010]. This is quite a cognitively demanding question, but the answer may not be especially predictive because people make real estate decisions based as much on intuition as on reasoning. According to the BMO Psychology of House Hunting Report,[1] prospective buyers visited an average of 10 homes over 5 months and 80% of prospective buyers knew whether a home was right for them as soon as they stepped inside.

People's reliance on "experiential" information to make decisions can be explained by *dual process theory* [Kahneman, 2011]. Dual process theory hypothesizes that people have two distinct reasoning systems: System 1 and System 2. System 1, or intuition, is a primarily subconscious reasoning system that we use to make decisions on issues with which we have extensive experience. In the example of real estate selection, we can quickly evaluate how comfortable a living space is if we can visit it in person. However, performing the same evaluation given only a description of the space is a difficult and unfamiliar task. Because we are unable to apply System 1, we appeal to System 2, the logical reasoning system. In this case, System 2 is perhaps attempting to predict System 1's reaction by deduction or analogy, imagining other spaces we have experienced. The result is significantly less accurate than System 1's output.

The PE literature underemphasizes the role of experiential information in the quality and cost of responding to queries. In particular, experiential information is important in decisions where we have significant experience, as in the real estate example above, and when we have limited cognitive effort to devote to the task. These conditions correspond with Kahneman's characterization of when System 1 engages [Kahneman, 2011].

Electricity use satisfies these conditions. We interact with electricity without much conscious attention and on a daily basis, and its cost does not represent a large budget item for most people.

In this chapter, we develop a theory of *experiential elicitation (EE)* using home temperature control as a concrete example. The name "experiential elicitation" originates in Hui and Boutilier [2008], who use a qualitative version of EE in an application customization setting. In traditional PE that does not take experiential information into account, it is possible to separate the elicitation and control phases of the interaction. In the elicitation phase, the system queries the user until it is satisfied that it understands the user's preference well enough to make the necessary decisions. In the control phase, the system takes actions on behalf of the user and does not pose additional queries. In EE, it is often impossible to separate the two phases cleanly. It may be worthwhile for the system to take control actions that put the user in a useful experiential state. In the real estate example, an EE would dispatch the user to visit various listings, but would take the cost of such visits into account.

Experiential information does not have a role in all instances of PE. Consider *security games* [Kiekintveld *et al.*, 2009]. Security games model the situation where a *defending player* has to decide on a stochastic allocation of *defense resources*, such as patrols, to a number of potential *targets*. The *attacking player* then observes the defender's allocation and chooses which target to attack. The objective of the model is to decide how the defender can best allocate its resources. This requires the defender to evaluate the *payoff* (cost) of a successful attack on each target. According to Nguyen et al. [2014]:

[1] Summarized at https://newsroom.bmo.com/2013-05-02-BMO-Psychology-of-House-Hunting-Report-Home-Buyers-Visited-an-Average-of-10-Homes-Before-Buying.

> Equilibrium strategies for defenders can be extremely sensitive to these payoffs, yet they can be extremely difficult to assess, requiring security experts to evaluate the potential impact on lives, property, and economic activity associated with specific attacks on particular targets. Even with this effort, this assessment is generally characterized by significant uncertainty.

System 1 is not very helpful in responding to these queries because even experts have limited experience with the problem. It may even hinder accurate assessment by providing the intuition that any successful attack would be devastating.

In this work, we develop a model of EE in *sequential* decision-making settings with relevant experiential information, i.e., the cost and accuracy of a user's response to a preference query depends on the state of the system. Specifically, if a user is asked to assess (e.g., compare) outcomes, both cost and response noise increase with how "closely" the user has *experienced* these outcomes, allowing different forms of distance (e.g., state similarity, recency of experience) to play a role. In this setting, there is an explicit trade-off between exploitation (optimal control given the current preference information) and exploration (having the user encounter new circumstances that may allow preference queries that can improve control).

We motivate our methods by considering a smart home agent that controls the heating, ventilation and cooling system (HVAC) system in a user's home, changing settings in response to variable electricity prices and a user's complex temperature preferences. The increasing presence of renewable sources of power generation has created highly variable pricing, and users cannot realistically adapt behavior to such price variations in real time (due to attentional costs, inability to forecast price and temperature changes, etc.). However, for an autonomous agent to adapt HVAC on a user's behalf, considerable preference information about small changes in comfort levels are needed. PE queries can be difficult to answer unless a user is actually experiencing (or has recently experienced) the conditions in question (e.g., temperature, humidity, price). As such, this a natural domain for EE.

The main contributions of this chapter are as follows:

1. We introduce and motivate the EE approach for a variety of AI systems that interact with users over time.

2. We provide a theoretical analysis connecting optimal EE to other, well-known problems in AI.

3. We study the interplay (and synergies) of *relative value queries (RVQs)*, which are natural in HVAC settings, with a preference prediction model based on *Gaussian processes (GPs)*.

4. We develop a system for EE in a smart-home HVAC setting using RVQs and GPs, and analyze it empirically using a combination of real and synthetic data, showing that it outperforms natural baselines.

The chapter is organized as follows. Section 5.2 reviews Markov decision processes, GPs, and related work in PE. Section 5.3 provides a formal model of EE and relates it to other problems in AI. Section 5.4 introduces RVQs and our GP-based model for EE. In Section 5.5, we outline a natural cost and noise model for RVQs and we present experimental results in Section 5.6.

5.2 Background

Markov Decision Processes (MDPs): MDPs [Puterman, 1994] are a generic model of sequential decision making processes in discrete time. We use MDPs as the central sequential decision-making model in this chapter. The problem of HVAC control naturally lends itself to being solved as an MDP. At each time step, the agent inputs a control action, heating or cooling the house. The interior temperature in the next time step depends on the interior and exterior temperatures in the previous step and the action that was taken.

The most important parts of the MDP model are:

- A set of states S, with a state corresponding to each possible state of the world.
- A set of actions A, which describes what actions can be taken from each state.
- A *transition function* P_{sa}, which yields the distribution of the next state given the current state and the action taken.
- A *reward function* $r(s, a)$, which tells the agent the utility (payoff) of taking a particular action from a particular state.

In a standard MDP, all properties of the model are known to the agent in advance.

Formally, we assume a fully observable MDP $\mathcal{M} = \{S, A, \{P_{sa}\}, \gamma, \beta, r\}$ with finite state set S, action set A, transition models P_{sa}, discount factor γ, initial state distribution β and reward function $r(s, a)$. The initial state distribution is the distribution over starting states of the model. The discount factor describes to what extent the agent should prefer reward in the current step compared to the next step: reward accrued i steps in the future is scaled down by γ^i. We consider infinite-horizon models.

The usual goal in MDPs to find a (deterministic) *policy*, i.e., a mapping from each state to an action, that maximizes discounted reward. Formally, a policy $\pi : S \to A$ has *value* V^π, given by: $V^\pi = \sum_{s_1 \in S} \beta(s_1) \mathbb{E}\left[\sum_{i=1}^{\infty} \gamma^{i-1} r(s_i, \pi(s_i)) | \pi\right]$, where expectation is taken over the distribution of state sequences induced by π.

An *optimal policy* π^* maximizes expected value—$\pi^* \in \operatorname{argmax}_\pi V^\pi$. An optimal policy can be computed by a variety of means [Puterman, 1994]. It can be done in polynomial time [d'Epenoux, 1963], but faster methods lacking polytime guarantees are typically used. In this chapter, we focus on *value iteration* (for technical reasons). Value iteration uses the *value function* $V : S \to \mathbb{R}$, which will converge to the expected value of executing the optimal policy from each state. The value function is initialized arbitrarily. At each step of value iteration, the value function for a particular state is updated, using the formula:

$$V(s) := \max_a \left(\sum_{s'} P_{sa}(s')(R(s, a) + \gamma V(s')) \right) \tag{5.1}$$

Value iteration converges to the optimal policy, but it may require an exponential number of steps to do so.

Gaussian Processes (GPs): GPs [Rasmussen and Williams, 2006] are universal function approximators that model the points of a function as a multivariate Gaussian distributed with known covariance, given by applying the *kernel function* to their inputs. In addition to estimating the value at any point,

GPs quantify the uncertainty in that estimate. GPs have been used previously in PE because they address several of the main desiderata of preference models: a) they can capture any utility function, b) they model uncertainty in a principled manner, c) they facilitate query selection to improve the user's objective and d) they allow for the incorporation of prior knowledge [Bonilla *et al.*, 2010]. In our EE model of HVAC control, GPs are additionally valuable because they can integrate the responses to relative value queries while still providing closed-form predictions.

A GP is represented as $\{K, \boldsymbol{x}_{\text{train}}, \boldsymbol{y}_{\text{train}}, \sigma^2\}$ with kernel function $K(\cdot,\cdot)$, training data $(\boldsymbol{x}_{\text{train}}, \boldsymbol{y}_{\text{train}})$, and Gaussian observation noise σ^2.[2] Typical kernel functions include the squared exponential $\exp[\frac{(x_0-x_1)^2}{2\ell^2}]$, where the *lengthscale* parameter ℓ governs the distance that new observations affect.

Calculating a GP's predicted mean $\boldsymbol{\mu}_{\text{test}}$ and covariance $\bar{\boldsymbol{\Sigma}}$ at any number of test points $\boldsymbol{x}_{\text{test}}$ can be done by applying the standard conditional expectation and covariance expressions for multivariate Gaussians and making use of the fact that the sum of Gaussians is Gaussian (for Gaussian observation noise). Let $\boldsymbol{\Sigma}$ denote the covariance matrix that results from stacking the test points on top of the training points:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\text{test,test}} & \boldsymbol{\Sigma}_{\text{test,train}} \\ \boldsymbol{\Sigma}_{\text{train,test}} & \boldsymbol{\Sigma}_{\text{train,train}} \end{bmatrix} \tag{5.2}$$

where $\boldsymbol{\Sigma}_{i,j} = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Then,

$$\begin{aligned} \boldsymbol{\mu}_{\text{test}} &= \boldsymbol{\Sigma}_{\text{test,train}}(\boldsymbol{\Sigma}_{\text{train,train}} + \sigma^2 I)^{-1}\boldsymbol{y}_{\text{train}} \qquad (5.3) \\ \bar{\boldsymbol{\Sigma}} &= \boldsymbol{\Sigma}_{\text{test,train}}(\boldsymbol{\Sigma}_{\text{train,train}} + \sigma^2 I)^{-1}\boldsymbol{\Sigma}_{\text{train,test}} + \boldsymbol{\Sigma}_{\text{test,test}} + \sigma^2 I \end{aligned}$$

Matrix inversions make both operations (5.3, 5.4) scale cubically with the number of training points. Since our work aims to minimize this number (i.e., user queries asked), this is not a practical obstacle, (for large data sets, DNGO can be used [Snoek *et al.*, 2015]). If data arrives arrive incrementally, block matrix operations can be used to perform a posterior update with complexity that is only quadratic in the training set size. Since test points correspond to states of the MDP, performance of the conditional expectation operation is more critical—it is linear in data set size.

While we present GPs with uniform observation noise, we will make use of the fact that the matrix $\sigma^2 I$ can be swapped with an arbitrary covariance matrix, e.g., to allow for non-identically distributed Gaussian noise.

Because of their cubic complexity in the amount of data, GPs can be difficult to profitably apply directly to supervised learning problems. One domain where they have been particularly successful is the optimization of black-box functions [Snoek *et al.*, 2012]. This setting resembles preference elicitation in a sense: a costly estimate of the function at a particular point can be obtained by querying. This problem is of special interest in supervised learning where it is often used for hyperparameter tuning.

Related Work: PE in MDPs has been previously studied by Regan and Boutilier [2009; 2011], who assume that elicitation and control are separable. Thus, they query the user until enough reward information has been acquired to solve the MDP (approximately) optimally. In our work, the ability of the user to effectively assess preference/reward queries depends on the current system state, thus inducing a tight coupling between elicitation and control.

[2] We assume a prior mean of zero for simplicity.

Shann and Seuken [2013; 2014] study preference elicitation and optimal control in an MDP-based HVAC model.[3] Their active learning model assumes a specific functional form of user utility with four parameters. The PE system is allowed a specific number of demand queries each day, i.e., queries of the form "what would your preferred temperature be, given current conditions (including price)?" There are two key differences between their work and ours. The first is that their PE system does not consider experiential information: it is impossible for the user to tell the system anything except their preferences in the current context. The second is that their end-goal is an HVAC system, whereas we use an HVAC system as a motivating example for developing EE and RVQs. As a consequence, they are focused on learning four utility parameters that are relevant to HVAC, whereas our system has the goal of learning a utility function with unknown functional form, a harder task.

PE in sequential decision making has also been studied from the perspective of "active" inverse reinforcement learning. In standard inverse reinforcement learning [Ng and Russell, 2000; Rothkopf and Dimitrakakis, 2011], the agent receives a set of demonstrations from the user and attempts to reconstruct the reward function from the demonstrations. Judah et al. [2014] extends this by allowing the user to specify additional reward information in the form of a shaping reward function. From a PE perspective, these papers can be viewed as one-shot "push" systems: the user provides as much preference information as they want to the system, which then learns as accurate a model as possible from it. Natarajan and Odom [2016] consider an iterative "pull" PE system where the agent requests demonstrations from the user. None of these systems have an experiential dimension to them.

Ideas similar to EE have appeared in past PE work. Hui and Boutilier [2008] use a qualitative version of EE in an application customization setting. In PE for stable matching, Drummond and Boutilier [2014] use a two-stage elicitation process. The first stage "hypothetical" queries are inexpensive and limited in accuracy and the second stage "interview" queries are expensive and perfectly accurate.

GPs were first used for preference learning by Chu and Ghahramani [2005]. They study the problem of learning to rank a set of alternatives, which are represented by feature vectors, given the results of a number of pairwise preference queries. Their main contribution is to develop a new likelihood function for the preference relations and a variational approach for inferring the model parameters that makes use of a Laplace approximation. Note that Chu and Ghahramani do not perform any elicitation.

Bonilla et al. [2010] extend Chu and Gharamani. First, they extend the likelihood model and inference procedure to multi-user collaborative filtering. Second, they perform elicitation by estimating the expected value of information of each query and eliciting in descending order of EVOI. They find that their method outperforms several others in well-studied PE domains.

5.3 A Model of Experiential Elicitation

In this section, we describe an abstract model for the experiential elicitation (EE) problem and draw theoretical connections to other sequential decision models.

An *EE problem instance* is $\{\mathcal{M}, R, Q, N = \{N_q\}, D = \{D_q\}, C\}$, where $\mathcal{M}$ is a fully observable MDP. Unlike a standard MDP, the true reward function r is not (initially) known to the agent. R is the *reward uncertainty model* capturing the agent's prior over possible rewards. We take a probabilistic perspective—R is a distribution over reward functions [Xu and Mannor, 2009]. We sometimes abuse

[3] Shann and Seuken [2014] use GPs as well, but for a different purpose: to predict the outdoor temperature and the electricity price.

notation and use R to refer to the support of the distribution (i.e., set of feasible reward functions). The remaining parameters specify the process available to the agent to elicit user's reward function. Q is the set of available *queries* that can be asked, while N associates a set of possible user *responses* N_q with each $q \in Q$. The *response function* D specifies a response distribution for each query. It maps from $R \times S^*$ to a distribution N_q for each $q \in Q$, defining a distribution over user responses given any possible true reward function r and history of past states. The *query cost function* $C : Q \times S^* \rightarrow \mathbb{R}^+$ specifies the cost of asking a query, given a state history.

We give a brief example of an EE model for the HVAC domain. Here the MDP would be the standard heating and cooling model, where the actions represent thermostat control. The queries might be demand queries:[4] given a particular state of the world (temperature, weather, who is home and what they are doing and wearing, the current electricity price) what would their most preferred temperature be? The response set is all possible temperatures. The response function is the user's true preferred temperature with Gaussian noise added, noise which is of higher magnitude the more distant the queried state is to the current state. The query cost function is also proportional to the distance from the queried state to the current state.

We believe that EE may be useful in other domains where decision-making is primarily subconscious. For example, an autonomous personal assistant on a smartphone may make decisions on our behalf that have significant subconscious components, such as when to schedule meetings, what information to show the user on the lock screen, and when to provide the user with reminders. It may be hard for the user to respond accurately to queries about the utility of hypothetical actions of personal assistants, whereas experiential queries may be effective in eliciting useful responses.

An agent acting in an EE instance knows all problem elements except r. In other words, it doesn't know the true reward r, but only the space and distribution R of possible rewards. This reflects the fact that an agent has incomplete and imprecise information about a user's true preferences. The agent can at each stage choose to ask queries or take actions, specifically:

1. An initial state s, drawn from β, is revealed (the MDP state is fully observable).
2. Repeat infinitely:
   (a) The agent asks the user zero or more queries. The user responds to each according to the query response function (w.r.t. r and the state history).
   (b) The agent executes some $a \in A$ and state transitions according $\{P_{sa}\}$, with the new state revealed.

The goal of the agent is to maximize expected total discounted reward w.r.t. the user's (unknown) reward function r less discounted query costs.

We make two assumptions for ease of exposition. First, MDP state transitions evolve more slowly than any "reasonable" number of queries. If this is not the case, we can model exogenous state transitions when queries are being asked or introduce time-based discounting to account for query "delays." Second, actions provide no reward information. The extension to actions with information content is straightforward.

In the remainder of this section, we connect the EE model to several related AI models. We find exploiting these connections to be difficult in the domain that we consider, but they may valuable in

[4] We consider different query types below.

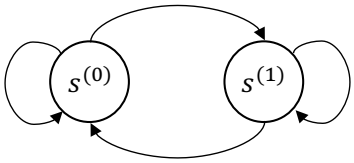


Figure 5.1: Diagram of a simple two-state MDP with deterministic transitions.

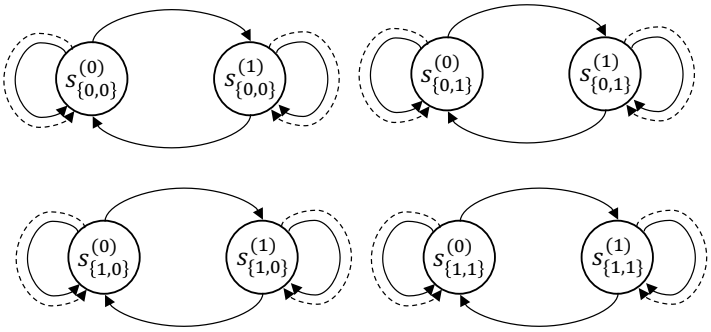


Figure 5.2: Diagram of the corresponding POMDP, where dashed lines represent query actions. There is one query action available in each state.

other settings and they are of theoretical interest.

The first connection is that an EE instance can be formulated as a *partially-observable MDP (POMDP)* [Smallwood and Sondik, 1973]. A POMDP is an MDP where the state is not directly observable. The agent instead receives an *observation* when a state transition occurs, and the distribution of observations is known for each state-action pair. Formally, a POMDP is $\{\mathcal{M}, \Omega, \{O_{sa}\}\}$, where $\mathcal{M} = \{S, A, \{P_{sa}\}, \gamma, \beta, r\}$ is an MDP, Ω is the set of observations and $\{O_{sa}\}$ is the distribution over Ω for each state action-pair.

Theorem 5.1. *Any EE instance can be formulated as a POMDP.*

Proof. The fundamental idea of the proof is to embed the set of possible rewards in the POMDP state space. This transformation is related to Poupart et al.'s [2006] method of solving Bayesian reinforcement learning problems by formulating them as POMDPs. State transitions act on S, but never change the reward embedded in the state; thus, for any distinct $r, r' \in R$, the corresponding embedded state sets $S^{(r)}$ and $S^{(r')}$ do not communicate. The reward uncertainty distribution may be discrete or continuous (necessitating a POMDP with infinite states in the latter case). The action space A is augmented by the set of queries, so the agent can ask queries or take (original) actions; queries cause no state transition. The observation function for (original) actions reflects full observability of the (original) state space (observations are the states themselves), while for queries, the observation function captures the distribution over responses.

Figures 5.1 and 5.2 give a visual example of the transformation. Figure 5.1 shows the MDP of a two-state EE model where the reward in each state is either 0 or 1. Figure 5.2 shows the equivalent POMDP representation where there is a single query action available for each state.

We begin by constructing a model without elicitation.

Formally, let S be the same as S in the MDP, except each state is duplicated $|R|$ times, with a copy corresponding to each possible reward function. We denote the set of states where r is the true reward

function as $S^{(r)}$. Consider a state $s_r \in S^{(r)}$ that corresponds to $s \in S$. $A(s_r)$ is a superset of $A(s)$.

- Transitions: for each action $a \in A(s)$ and each state $s'_r \in S'^{(r)}$ with corresponding original state s', $P_{s_r a}(s'_r) = P_{sa}(s')$. $P'_{s_r a}(s'_r)$ is zero otherwise.
- Observations: each action emits the observation $\omega_{s'}$ with probability 1, which informs the agent of the identity of the state $s' \in S$ corresponding to the state s'_r it has transitioned into.
- Rewards: the reward is $r(s_r, a) = r(s, a)$.
- Initial state distribution: β' is $\beta(s_r) = \Pr(r)\beta(s)$ where $\beta(s)$ is the initial state distribution in the MDP.

The model now matches the EE instance exactly, but there is no way for the agent to gain information about the true reward function. To allow this, we will augment the action set with "querying" actions. These actions do not cause state transitions, but they emit observations that reveal information about the true reward function. The rewards of the actions are negative and given by the query cost function.

Let $q \in Q$ be a query. Create a corresponding action a_q which is available from every state.

- Transitions: a_q causes no transitions. $P_{s_r a_q}(s_r) = 1$.
- Rewards: $r(s, a_q) = -C(q, \mathbf{s})$ where $\mathbf{s}$ is equal to the state history. We need to augment the POMDP's state space to track the necessary history for both the query cost function C and the query response function. This augmentation only affects the rewards and observations of querying actions. For example, suppose our query cost and response models depend on the shortest distance between the queried state and any state we have visited in the last 50 states. We would need to augment the state representation with a S^{50} vector representing the past 50 visited states. If the model depends on an infinite state history (as do those in Section 5.5), the POMDP requires an infinite dimensional state in the worst case.
- Observations: let Ω contain $\bigcup_q N_q$. $O_{s_r a_q} = D_q(r, \mathbf{s})$.

The agent can now perform elicitation. This concludes the construction, except for one issue: discounting. In the POMDP, discounting is applied when any action is taken (including a query), but in the EE model, discounting is only applied when a control action is taken. POMDPs can be extended so that certain actions do not cause discounting (and this does not interfere with many standard results). If we wish to make no such modification, we can augment the state space (storing the number of queries that have been asked so far) to fix the problem.

We show that an optimal policy for the POMDP is optimal for the EE instance. Observe that for a given sequence of actions and queries and a starting state distribution, the expected reward in the EE and POMDP models is identical. The actions have the same transition probabilities except that the POMDP state has the reward information embedded in it, and action trajectories can never cross from a state with a particular reward embedded to a state with a different reward embedded. The queries have the same costs associated with them and do not cause the state to change in either model. Thus, because the set of policies is the same in each model and the expected reward of each policy is the same, the optimal policy in the POMDP is the same as that in the EE.

□

Note that the POMDP belief state for an EE problem will reflect the agent's posterior over R, reflecting information captured about a user's preferences by queries and responses. POMDPs have been used to model elicitation problems in the past [Boutilier, 2002; Holloway and White, 2003]. These formulations differ from ours because they require only one state for each potential reward function.

We discuss the three main obstacles in the POMDP reduction and their impacts from a practical and theoretical perspective. The first is that the POMDP may require infinite states if R is continuous. This is not an important issue for two reasons: i) it would occur for any PE scenario where the support of R is infinite, regardless of model; and ii) the state space of POMDPs is often approximated in practice, even when it is finite, because of the high computational complexity of solving POMDPs.

The second, bigger, issue is that the POMDP may require an infinite-dimensional state space to keep track of the state history. EE, as we define it, is not Markov whereas POMDPs must be. Infinite dimensional POMDPs can be a problem because compressing the state dimension while keeping the relevant information is hard. However, it is unlikely in that any practical EE system would require unbounded state history to model responses accurately. Our definition permits unbounded state history, and we make use of it because it is elegant, but it is not necessary from a practical perspective.

The third issue is the different way discounting is handled in the two models. This is quite annoying from a practical perspective because it prevents any EE instance from being solved as a POMDP exactly without requiring an infinite-dimensional state space (unless the discount factor is 1). However, from a theoretical perspective, POMDPs having a fixed discount rate is more for ease of exposition than it is essential. From a practical perspective, we may simply ignore the difference (or tweak the discount rate, taking into account the expected number of queries that will be asked).

There is a direct connection between EE and reinforcement learning (RL) as well.

Observation 5.1. *Consider an RL problem $\langle \mathcal{M} \rangle$ that consists of an MDP $\mathcal{M}$ which is known to the agent except for the reward function, and whenever the agent transitions into a state, it receives the reward information for that state. This problem can be reduced to EE and the reduction requires increasing the number of states by a factor of $O(|S| \times |A|)$.*

Proof. Create an EE that retains the model details (states, actions, transitions, rewards, discount). We let reward uncertainty R be an uninformative prior. For each state s, we allow *value queries* for any state-action pair (s, a), which asks a "user" (representing the environment) for the reward for that pair, and the response function represents the RL (stochastic) reward for that pair. Query cost is zero if asking about the action just taken at the previous state, and infinite otherwise. To encode this query cost function, the EE state must include the previous state visited and action taken, which causes a state space blowup of $|S| \times |A|$.

In this EE instance, the only "available" query at a state is "free," so it is optimal to always ask it, giving an EE agent the same information as an RL agent. □

Lastly, suppose we ignore an EE agent's query strategy, or equivalently, make no queries available. The optimal control policy under reward uncertainty for a risk-neutral agent can be solved as an MDP.

Observation 5.2. *Given a risk-neutral agent and an MDP with uncertain reward R, its optimal policy is that of an MDP where each state-action reward is its expected reward under R. (This holds even if rewards are correlated under R).*

Proof. This follows from a simple argument using linearity of expectation. Let $R_{s,a}$ be a random variable representing the (possibly correlated) state-action reward according to the agent's current beliefs. Given a risk-neutral agent, the optimal policy π^* of an MDP with uncertain reward R satisfies

$$\pi^*(s) = \underset{a \in A(s)}{\operatorname{argmax}} \mathbb{E}\left[R_{s,a} + \gamma \sum_{s' \in S} P_{sa}(s') V^*(s') \right] \tag{5.4}$$

where

$$V^*(s) = \mathbb{E}\left[R_{s,\pi^*(s)} + \gamma \sum_{s' \in S} P_{s\pi^*(s)}(s') V^*(s') \right] \tag{5.5}$$

We can rewrite $\pi^*(s)$ using linearity of expectation and using the fact that $V^*(s)$ is already an expectation w.r.t. to R:

$$\pi^*(s) = \underset{a \in A(s)}{\operatorname{argmax}} (\mathbb{E}[R_{s,a}] + \gamma \sum_{s' \in S} P_{sa}(s') V^*(s')) \tag{5.6}$$

We can rewrite $V^*(s)$ likewise:

$$V^*(s) = \mathbb{E}[R_{s,\pi^*(s)}] + \gamma \sum_{s' \in S} P_{s\pi^*(s)}(s') V^*(s') \tag{5.7}$$

Combining the two equations, we get the standard Bellman equation for the value function, but with $R_{s,a}$ replaced with $\mathbb{E}[R_{s,a}]$:

$$V^*(s) = \max_{a \in A(s)} \left(\mathbb{E}[R_{s,a}] + \gamma \sum_{s' \in S} P_{sa}(s') V^*(s') \right) \tag{5.8}$$

Since the optimal policy is the only policy that satisfies the Bellman equation, π^* is the optimal policy of the MDP with reward replaced by its expected value. □

This is equivalent to the imprecise-reward MDP [Regan and Boutilier, 2009], but we solve it in a Bayesian manner instead of using minimax regret. We exploit this observation when defining expected value of information in Section 5.4.

5.4 EE with Relative Value Queries

A *relative value query (RVQ)* asks the user to articulate the difference in value or utility between states. An RVQ $(\boldsymbol{x}_0, \boldsymbol{x}_1)$ comprises two points in a d-dimensional state space. A response y is the user's estimate of the difference $r(\boldsymbol{x}_0) - r(\boldsymbol{x}_1)$ in their immediate rewards. We study RVQs because they naturally capture the way that users make decisions in certain domains. For example, in the HVAC setting an RVQ represents the query "in the current conditions, what electricity cost savings would be required so that raising the temperature by one degree for one hour would be acceptable?" When a user sets their HVAC system in a given context, they are "reasoning" over the set of RVQs. In this section, we examine EE systems that use RVQs.

We compare RVQs to a few other query types in the HVAC domain. Value queries would ask a user

directly for their utility (in dollars) for a particular state. This is quite difficult to answer as there is no need to have a value function to make HVAC decisions—instead, the user is faced with a set of actions and needs to choose among them. Following that intuition, we could ask users what their preferred action would be at a particular state (this is one possible interpretation of a demand query). One disadvantage of this query strategy is that requires the user to reason about the cost of the action they choose. It is cleaner to separate the raw reward of a state and the cost of getting to that state. Another disadvantage is that there is stochasticity in the model, so the information we receive is comparing the expected values of two distributions. Another idea for interpreting demand queries is to ask the user what their preferred interior temperature would be for a particular state. The difficulty here is that the user's response is completely removed from the cost of getting to that temperature.

A first key observation is that the MDP embedded in an EE instance can be solved optimally using only RVQs. Having only difference information for all state pairs means that state rewards are known only up to an additive factor. However, by the equivalence of utility functions (in this case, the MDP value function) under positive affine transformation, it immediately follows that adding a constant to the reward function does not change the optimal policy.

Observation 5.3. *An MDP can be solved optimally given only information about the differences in rewards between a collection of state-action pairs.*

Proof. Knowing the difference in reward between all state-action pairs is equivalent to knowing the reward function up to an additive factor. Thus, it suffices to show that adding c to the reward of each state-action pair does not change the optimal policy.

The optimal policy π^* of an MDP is a policy that satisfies

$$\pi^*(s) = \underset{a \in A(s)}{\operatorname{argmax}} \left(r(s,a) + \gamma \sum_{s' \in S} P_{sa}(s') V^*(s') \right) \tag{5.9}$$

where $V^*(s) = r(s, \pi^*(s)) + \gamma \sum_{s' \in S} P_{s\pi^*(s)}(s') V^*(s')$.

Consider the effect on Equation 5.9 of replacing $r(s,a)$ with $r'(s,a) = r(s,a) + c$. Each $V^*(s)$ will increase by $\sum_t c\gamma^t$, where t is the number of steps remaining in the MDP. Because each $V^*(s)$ increases by the same amount, $\pi^*(s)$ remains the same. □

To exploit RVQs effectively, we need a learning method that, given a set of responses to RVQs, can estimate the rewards (and uncertainty) of all states. GPs can be adapted to this purpose while maintaining the critical properties that drive their usefulness for PE, in particular, the ability to estimate the expected value of information of a query.

We show that the GP posterior update has a closed form in the case of RVQs. Since an RVQ represents the difference of two Gaussian random variables (RVs) with known covariance, the affine property of multivariate Gaussian RVs ensures they are closed under linear transformations.

Observation 5.4. *Let $\mathcal{X}$ be an RV distributed as $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{c}$ a constant M-dimensional vector, and $\boldsymbol{B}$ a constant $M \times N$ matrix. Then $\boldsymbol{c} + \boldsymbol{B}\mathcal{X} \sim \mathcal{N}(\boldsymbol{c} + \boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^T)$.*

Proof. The proof uses characteristic functions: a random variable Y of dimension N with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ is normally distributed *if and only if* for any constant vector $\boldsymbol{t}$ of dimension $N \times 1$,

$$\mathbb{E}[\exp(i\boldsymbol{t}^T Y)] = \exp(i\boldsymbol{t}^T\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}). \tag{5.10}$$

(i is the imaginary unit). We will show that $\mathbb{E}[\exp(i\boldsymbol{t}^T(\boldsymbol{c}+\boldsymbol{B}\mathcal{X})]$ has the required form.

$$\mathbb{E}[\exp(i\boldsymbol{t}^T(\boldsymbol{c}+\boldsymbol{B}\mathcal{X})] = \exp(i\boldsymbol{t}^T\boldsymbol{c})\mathbb{E}[\exp(i\boldsymbol{t}^T\boldsymbol{B}\mathcal{X})] \tag{5.11}$$

$$= \exp(i\boldsymbol{t}^T\boldsymbol{c})\mathbb{E}[\exp(i\boldsymbol{\alpha}^T\mathcal{X})] \tag{5.12}$$

where we let $\boldsymbol{\alpha}^T = t^T\boldsymbol{B}$. Because $\boldsymbol{\alpha} = \boldsymbol{B}^T\boldsymbol{t}$ is a constant vector of dimension $N \times 1$,

$$\mathbb{E}[\exp(i\boldsymbol{t}^T(\boldsymbol{c}+\boldsymbol{B}\mathcal{X})] = \exp(i\boldsymbol{t}^T\boldsymbol{c})\exp(i\boldsymbol{\alpha}^T\mu - \frac{1}{2}\boldsymbol{s}^T\Sigma\boldsymbol{\alpha}) \tag{5.13}$$

$$= \exp(i\boldsymbol{t}^T(\boldsymbol{c}+\boldsymbol{B}\mu) - \frac{1}{2}\boldsymbol{t}^T\boldsymbol{B}\Sigma\boldsymbol{B}^T\boldsymbol{t}). \tag{5.14}$$

□

Suppose our training data consists of RVQ responses.[5] Our n training points have form $(\boldsymbol{x}_0, \boldsymbol{x}_1, y)$, where $(\boldsymbol{x}_0, \boldsymbol{x}_1)$ is the RVQ (two states) and y its (noisy) response, i.e., estimation of reward difference. Let $\boldsymbol{X}_0$ (resp., $\boldsymbol{X}_1$) be the matrix resulting from stacking the $\boldsymbol{x}_0$ (resp., $\boldsymbol{x}_1$) vectors, and let $\boldsymbol{X}_{\text{test}}$ be the set of m test points, arranged as a $m \times d$ matrix. To compute the posterior of the GP at $\boldsymbol{X}_{\text{test}}$ (states at which we wish to predict reward), we construct an affine transformation matrix $\boldsymbol{B}$. Let $\boldsymbol{X}$ be the data points stacked as $\boldsymbol{X} = \left[\begin{smallmatrix}\boldsymbol{X}_{\text{test}}\\ \boldsymbol{X}_0\\ \boldsymbol{X}_1\end{smallmatrix}\right]$, a matrix of size $(m+2n) \times d$. Define

$$\boldsymbol{B} = \begin{bmatrix}\boldsymbol{I}_{m,m} & \boldsymbol{0}_{m,2n}\\ \boldsymbol{0}_{n,m} & \boldsymbol{B}_{\text{diff}}\end{bmatrix} \tag{5.15}$$

where $\boldsymbol{B}_{\text{diff}}$ is $[\boldsymbol{I}_{n,n} - \boldsymbol{I}_{n,n}]$, $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{0}$ is the zero matrix. ($\boldsymbol{B}_{\text{diff}}$ is $n \times 2n$ and $\boldsymbol{B}$ is $(m+n) \times (m+2n)$.) $\boldsymbol{B}$ transforms the training rows into differences without affecting the test points.

In the absence of observations, the distribution of $\mathcal{X}$—the RV that represents the distribution over $\boldsymbol{X}$— is just the prior zero-mean and covariance $\boldsymbol{\Sigma}$ Gaussian, where $\boldsymbol{\Sigma}$ results from applying our chosen kernel. Applying Observation 5.4 using $\boldsymbol{B}$ yields an RV that relates the training and test points: a Gaussian with covariance $\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^T$ and mean $\boldsymbol{B}\boldsymbol{0} = \boldsymbol{0}$. This Gaussian has dimension $m+n$ (cf. the original $m+2n$)—we now have a *single* dimension for each training example. The usual conditional GP expectation and variance (Equations 5.3, 5.4) in the transformed space yields the posterior at the test points.

As in a standard GP model, we can also incorporate observation noise. Suppose that each observation y has i.i.d. Gaussian noise with variance σ^2. We model this with $\frac{\sigma^2}{2}$ noise on the underlying Gaussian, which results in observation noise of σ^2 on the difference. Non-i.i.d. noise is handled by replacing $\sigma^2 I$ with our chosen noise matrix.

Using this procedure for posterior updates, we can estimate *(myopic) expected value of information (EVOI)* of a query using a GP, adapted to the MDP setting. The estimate is myopic because it does not account for future queries. The myopic EVOI is the amount our expected reward increases as a result of asking a particular query, according to the current preference predictions of the GP. In our experiments, we use the EVOI to determine whether a query is expected to increase our discounted reward by more than the cost of the query. In theory, we could use the EVOI directly to determine what query provides the highest increase in discounted reward net of query cost, but this would require some sophistication

[5] This method can be modified to accommodate any combination of training points as long as each represents the result of applying a linear transformation. This includes ordinary value queries.

due to the size of the set of RVQs.

Suppose we want to estimate the EVOI of an RVQ q. Let R be our current reward uncertainty and $R_{(q,y)}$ be the posterior after asking q and receiving response y. Let Y be an RV representing the RVQ response according to R. Applying Observation 5.2, let V^* be the optimal policy value under belief R and let $V^*_{q=y}$ be the optimal policy value under belief $R_{(q,y)}$.

Definition 5.1. The *(myopic) expected value of information (EVOI)* of a query q is $\mathbb{E}_Y[V^*_{(q,y)} - V^*]$.

In practice, we compute the expectation by sampling query responses from R and averaging the changes in policy value. We use the following procedure:

1. Sample n query responses for q from the GP posterior.
2. For each query response y:
   (a) Add y to the set of query responses (i.e., treat it as if it were a query that we actually asked the user).
   (b) Calculate the value of the optimal policy under the new set of query responses and save it.
   (c) Remove y from the set of query responses.
3. Return the optimal policy value, averaged over all the responses, minus the original optimal policy value.

Random forest model: As a point of comparison, we also define a simple *random forest (RF)* model [Breiman, 2001; Friedman *et al.*, 2009] for EE with RVQs. We choose an RF because of its excellent performance with small amounts of training data.

RFs consist of many decision trees, each trained on a sample of the data. To make a prediction, the outputs of all of the decisions trees are averaged. Formally, for each decision tree:

1. Draw a sample from the data uniformly at random, with replacement.
2. Fit a decision tree to the sample by recursively performing the following splitting operation on each terminal node of the tree that is larger than the minimum node size (i.e., the number of data points which terminate at that node of the tree) until no such nodes exist:
   (a) Sample a subset of data dimensions uniformly at random.
   (b) Split the node into two new terminals, using the dimension and split point that makes the numbers of data points that terminate on each new node as close to equal as possible.

Prediction is performed by averaging the outputs of all trees with equal weight. We use the scikit-learn [Pedregosa *et al.*, 2011] implementation of RFs, which uses a sample size equal to the size of training data (but sampled with replacement) and uses all data dimensions when searching for the best split.

We train the RF using each training example twice, $(\boldsymbol{x}_0, \boldsymbol{x}_1, y)$ and its "reversed" RVQ as $(\boldsymbol{x}_1, \boldsymbol{x}_0, -y)$, so the model is informed of the "opposite" prediction. At test time, we predict a response to q' by averaging the model response to q' and the negative response to the reversal of q'. To evaluate reward at a set of states, we formulate each as an RVQ that compares it to the state most observed in the training data (recall that by Observation 5.4, only relative values affect the MDP solution). Unlike the GP, this model is incapable of evaluating EVOI of queries.

5.5 Query Response and Cost Model

We develop a model for users' responses to queries, i.e., how accurate and how costly such responses are. We are motivated by the HVAC setting, where RVQs compare states representing the internal and external temperature, humidity, time, etc. Our response and cost models should satisfy the following criteria:

1. They should align with the concept of *just noticeable difference (JND)* [Gescheider, 2013]. An RVQ comparing two states that differ only in internal temperature is more difficult to answer if the difference is small. Such queries are difficult (i.e., cognitively costly and error prone).

2. It should be harder to compare states that are different in several ways rather than few, e.g., it is easier to compare two states that differ only in internal temperature than states that differ in internal and external temperature.

3. It should be easier to answer queries about states that are similar to the current state, or to one that was visited recently. For example, it is easier to compare two "summer states" in the summer than it is in the winter.

4. The mere act of answering a query should be expensive because it is disruptive. Thus, there is a lower bound on query response costs.

Query response model: We use the following query response model:

$$D(\boldsymbol{x}_{q0}, \boldsymbol{x}_{q1}) = \mathcal{N}(r(\boldsymbol{x}_{q0}) - r(\boldsymbol{x}_{q1}), \sigma_{noise}) \tag{5.16}$$

$$\sigma_{\text{noise}} = c_{\text{f\_n}} + c_{\text{d\_n}}(||\boldsymbol{x}_{q0} - \boldsymbol{x}_{q1}||_1 + \min_{0 \leqslant i \leqslant t}(1+\delta)^{t-i}(||\boldsymbol{x}_{q0} - \boldsymbol{x}_i||_1 + ||\boldsymbol{x}_{q1} - \boldsymbol{x}_i||_1)) \tag{5.17}$$

where $\boldsymbol{x}_{q0}$ and $\boldsymbol{x}_{q1}$ are the queried states, $\boldsymbol{x}_i$ is the sequence of past/visited states and δ is the user's discount factor. ($\mathcal{N}(\mu, \sigma)$ is a normal random variable with mean μ and variance σ.)

Modern psychophysics [Gescheider, 2013] assumes that sensor noise is the cause of JND. The **f**ixed **n**oise constant $c_{\text{f\_n}}$ ensures that noise is non-zero.

The second and third desiderata are satisfied by using the L_1-distances between states, which captures both the number of (state feature) differences and their magnitude, weighted by the **d**istance **n**oise constant $c_{\text{d\_n}}$. We capture the temporal component (i.e., less noise queried states are close to recently visited states) by discounting distances to previously visited states and using the state with least discounted distance. The literature generally embraces the notion that sensors are well-calibrated, thus, we use an unbiased response model.

Our query cost model has the same form, with one addition. We augment the model to reflect that queries near the JND are more costly by adding a term that decreases with the distance between $\boldsymbol{x}_{q0}$ and $\boldsymbol{x}_{q1}$:

$$\begin{aligned} C(\boldsymbol{x}_{q0}, \boldsymbol{x}_{q1}) =& c_{\text{f\_c}} + c_{\text{d\_c}}(||\boldsymbol{x}_{q0} - \boldsymbol{x}_{q1}||_1 + \\ & \min_{0 \leqslant i \leqslant t}(1+\delta)^{t-i}(||\boldsymbol{x}_{q0} - \boldsymbol{x}_i||_1 + ||\boldsymbol{x}_{q1} - \boldsymbol{x}_i||_1)) + \frac{c_{\text{JND}}}{1 + e^{||\boldsymbol{x}_{q0} - \boldsymbol{x}_{q1}||_1}} \end{aligned} \tag{5.18}$$

5.6 Experiments

Our experiments assess: (i) how the GP-based approach to EE compares to several natural baselines; and (ii) how filtering prospective queries by EVOI affects both the quality of the learned policy and overall utility accrued. We first describe our experimental setup. Our MDP model is derived from HVAC and temperature data from Pecan Street Inc. [Rhodes *et al.*, 2014], the same data used in Chapter 4. The states of the MDP contain the time, date, exterior and interior temperature. We discretize time into hours and temperatures into 20 buckets between 5 and 45°C. There are 3.5 million states. There are 10 HVAC control actions (5 heating, 5 cooling). Control response depends on properties of the HVAC system and thermal properties of the house in question, which are selected randomly from a range of realistic values. Action costs are determined by electricity prices in Austin, TX (where the data was collected). The user's reward for state s has the form:

$$r(s) = \exp\left[-h\frac{(\text{INT}(s) - (\text{INT}^* + \text{BIAS}(\text{EXT}(s) - \text{INT}^*)))^2}{w}\right] \tag{5.19}$$

where INT* is the user's most preferred internal temperature, INT(s) and EXT(s) are the internal and external temperatures in state s, and BIAS represents the degree to which a user's preference for internal temperature is affected by external temperature. Parameters h and w control the height (strength) and width (breadth) of a user's utility function. INT*, w, h and BIAS are drawn from $\mathcal{U}(18, 22)$, $\mathcal{U}(0, 20)$, $\mathcal{U}(0, 3)$, and $\mathcal{U}(0, 0.4)$, respectively. All users discount exponentially at a rate of 1%. Query response/cost models use: $c_{\text{f\_n}} = c_{\text{f\_c}} = c_{\text{d\_c}} = c_{\text{d\_n}} = 0.05$ and $c_{\text{JND}} = 2.5$. We run 100 instances of EE, each for 1000 hours, with weather sampled from the Pecan Street data.

Note that while performance is affected by the choice of user utility function, the model can accommodate any form of utility function, as long as it depends only on attributes that are included in the state vector. This is in contrast to previous work in learning HVAC utility functions such as Shann and Seuken [2013].

We select the action for the current time step by solving an MDP, representing the next 24 hours, using value iteration. For EVOI estimation of queries, we use 10 response samples and a longer lookahead of 50 hours.

Query strategies: In each query strategy, a single candidate query is proposed at each time step. The strategies differ in how they decide whether to ask a query and how they integrate responses. A query compares (i) a sampled *next state* from transitions induced by the current best action, and (ii) a sampled next state using the second best action. If the states are identical, we re-sample. This meta-strategy is effective because (i) it provides information that is immediately relevant, and (ii) it asks queries that tend to be close to the current state. More complex methods could increase performance at the cost of additional computation (e.g., using lookahead or approximate solutions to the induced POMDP).

We compare the performance of five query strategies: two that are GP-based, one that is RF-based, and two that are baselines.

- GP_ALWAYS uses a GP-based preference prediction model. It does not calculate EVOI and always asks the proposed query.
- GP_EVOI also uses a GP-based model, but only asks the proposed query if its EVOI is greater than its cost.

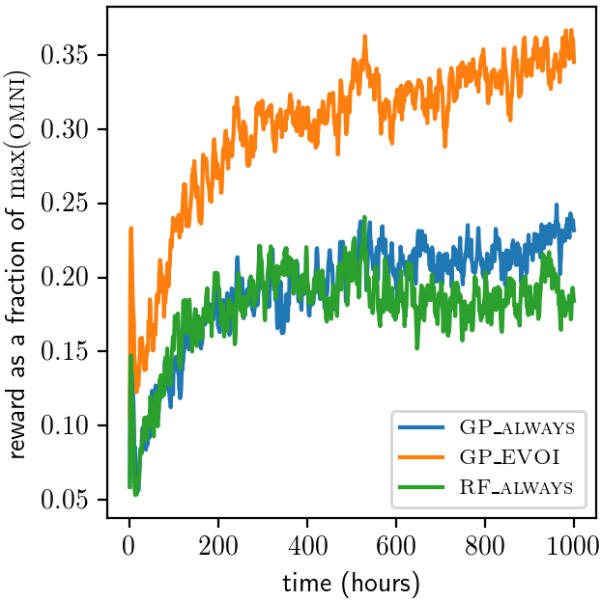


Figure 5.3: Reward accrued vs. time.

- RF_ALWAYS, uses a random forest model that always asks the proposed query (the RF model does not calculate EVOI).
- OMNI is optimally omniscient, and takes the action of the optimal policy in the underlying (known-reward) MDP without querying. It provides an upper bound on performance, which cannot be realized practically.
- NULL always takes the null action, i.e., does nothing.

Both GP models use the squared exponential kernel with lengthscale of 0.1°C. The random forest uses 25 decision trees.

Computing: We run all experiments on Intel "Broadwell" CPUs at 2.1 GHz. At time 1000—when all methods have their largest set of responses (which imposes maximal demands on GP computation)—all three (non-baseline) strategies take 0.1 s (on average) to select the next control action. GP_ALWAYS and RF_ALWAYS do not estimate EVOI and take 0.15 s to select a query, whereas GP_EVOI takes 2.0 s. The strategies also differ in the time required to evaluate state rewards given the current query set: GP_ALWAYS takes 0.7 s, RF_ALWAYS 0.3 s, and GP_EVOI takes 0.1 s—GP_EVOI is more efficient since it asks far fewer queries (see below).

Results: We first analyze total reward, including query cost, accrued by each strategy. Figure 5.3 shows (average) reward accrued over time for the three core strategies. Because each user has different utility height and width, we normalize results w.r.t. the maximum policy value, max(OMNI), observed in that instance by the OMNI baseline. This ensures users with large utility values do not dominate the data.

OMNI and NULL perform at a constant level of 44.6% and 18.8%, respectively, of max(OMNI), and are not shown in Figure 5.3. GP_EVOI outperforms GP_ALWAYS and RF_ALWAYS at each point (each performs similarly to NULL). These differences are statistically significant from iteration 90 onward (with $p = 0.05$). We will show that GP_EVOI's success is due to its ability to achieve control policy quality

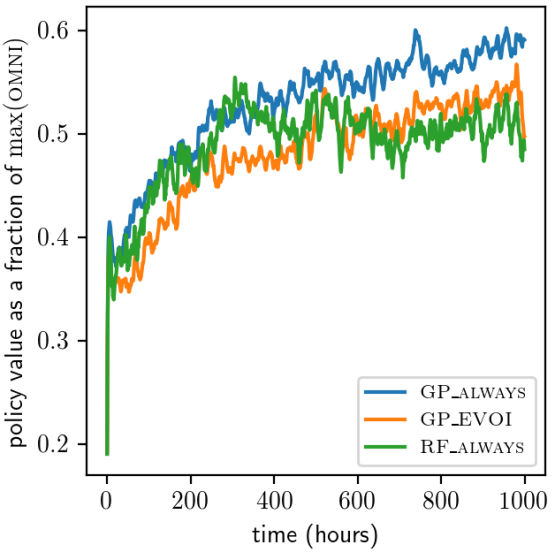


Figure 5.4: Policy value vs. time

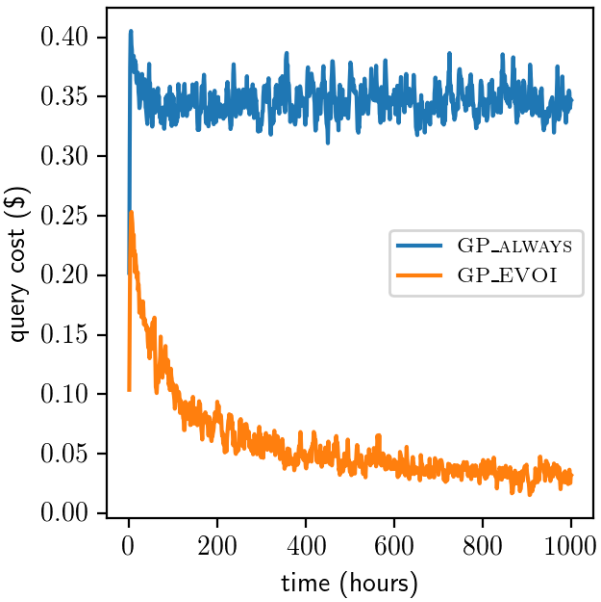


Figure 5.5: Query cost vs. time.

similar to the other methods, while incurring much lower query cost. This shows that the ability to estimate query EVOI is invaluable in this domain.

We see that training under this model takes hundreds of queries to complete. 1000 hours represents around 6 weeks of training time. One way of reducing training time would be to integrate collaborative filtering techniques—building a more accurate initial model for each household by using the results of queries to other households with similar features.

We next analyze the quality of the policy learned by each strategy. Figure 5.4 shows (ground truth) expected value of the current (optimal) control policy (as specified by Observation 5.2, assuming no more queries are asked) evaluated under the *true reward function*. OMNI and NULL (not shown) perform at a constant level of 80% and 1.7%, respectively, of max(OMNI). The three main strategies perform similarly w.r.t. policy value, with GP_ALWAYS having a slight advantage over GP_EVOI. This is in part because GP_ALWAYS (and RF_ALWAYS) receives a new query response every step, while GP_EVOI asks many fewer queries: 43.8 ($\sigma = 9.7$) queries on average by time 100, 121.1 ($\sigma = 25.8$) by time 500 and 179.6 ($\sigma = 39.4$) by time 1000.

The query numbers above lead to significant query cost savings for GP_EVOI. Figure 5.5 shows the query cost accrued by GP_ALWAYS and GP_EVOI (RF_ALWAYS is not shown since its query cost is similar to GP_ALWAYS; OMNI and NULL are not shown since they ask no queries). Low query cost is a major factor in GP_EVOI's strong performance. By time 100, GP_EVOI has incurred total query cost of \$15.4 vs. \$35 for GP_ALWAYS. By time 500 and 1000, GP_EVOI's costs are \$40.2 and \$58.4, resp. (vs. \$173 and \$348 for GP_ALWAYS). Average cost per (asked) query for GP_EVOI is similar to that of GP_ALWAYS, \$0.32 vs. \$0.35, showing that GP_EVOI's advantage lies largely in asking queries with higher EVOI rather than lower cost.

We ran additional experiments to study (i) how the strategies respond to rapidly changing context; and (ii) whether GP_EVOI correctly determines that queries in a previously visited context have low value after spending a lot of time in a different context. To achieve this, we use a setup that is the same as our other experiments except it has 1500 time steps and two different contexts. The first 500 time steps are consecutive hours starting in either Dec., Jan. or Feb. of our data set. The next 500 are consecutive hours in Jun., Jul. or Aug. The last 500 return to the same time steps as the first. In Austin, TX. the average temperature difference between the two contexts is around 15 °C.

Figures 5.6, 5.7 and 5.8 show the results. We observe that the context change has substantial effects. In policy quality terms (Figure 5.7), the first context change results in a drop of around 17% on average, which is similar for GP_ALWAYS, GP_EVOI and RF_ALWAYS. This is because the query strategies lack information about the reward function in the "summer" (second) context and not because the optimal policy has a radically different value: OMNI achieves an average policy quality of 76% and 74% of max(OMNI) in the winter and summer contexts, respectively.

Figure 5.8 shows that GP_ALWAYS and RF_ALWAYS suffer increases in query cost at each of the context changes.[6] When the change happens, each of the strategies has a state history that is not very useful for the new context, resulting in higher query costs and lower accuracy.

GP_EVOI's query cost also increases at the first context change. This increase in query costs is not due solely to having a less useful state history because there is no corresponding increase at the second context change. This shows that GP_EVOI correctly recognizes that its model has become more uncertain because of the change in context and querying has thus become more valuable. At the second

[6]Recall that RF_ALWAYS's query cost is nearly identical to GP_ALWAYS's.

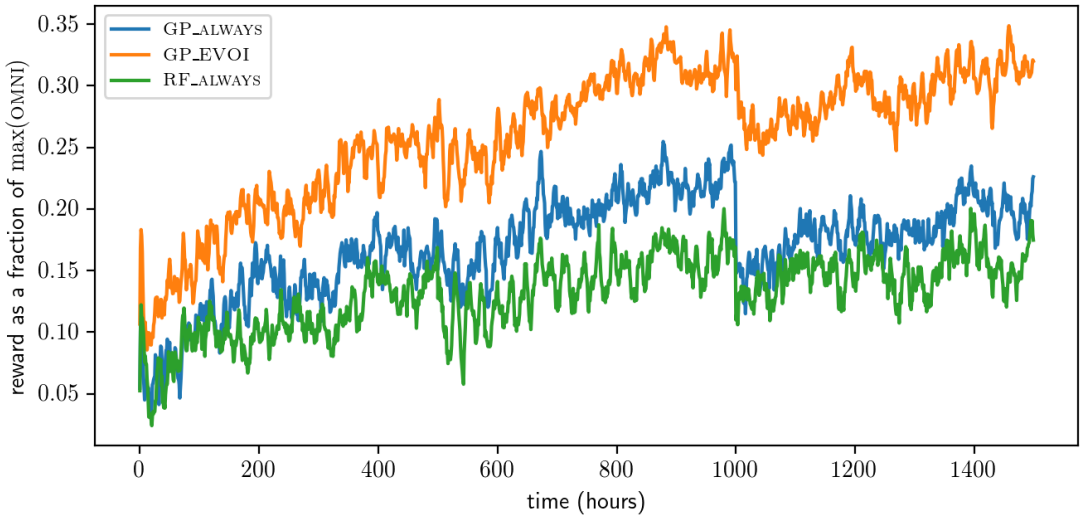


Figure 5.6: Reward accrued vs. time.

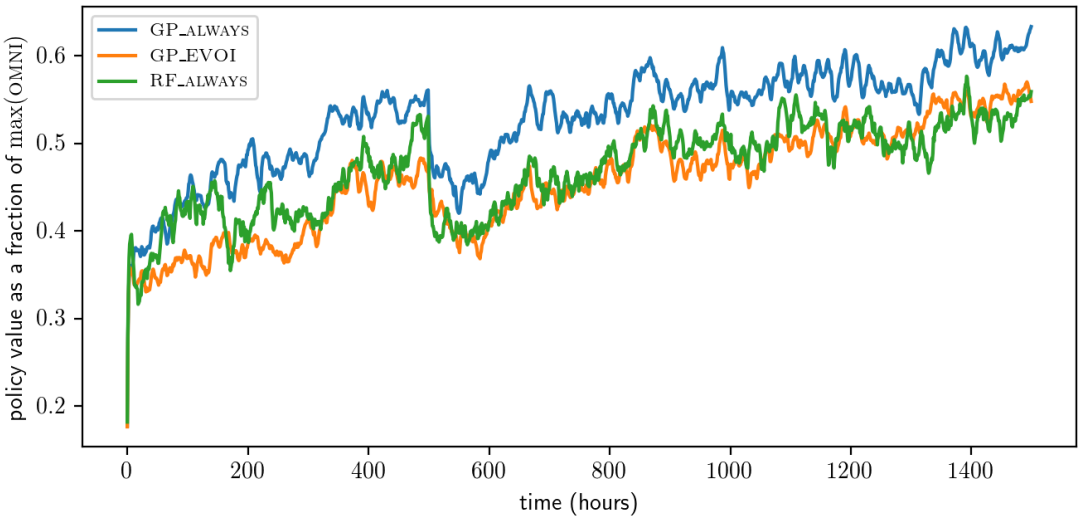


Figure 5.7: Policy value vs. time.

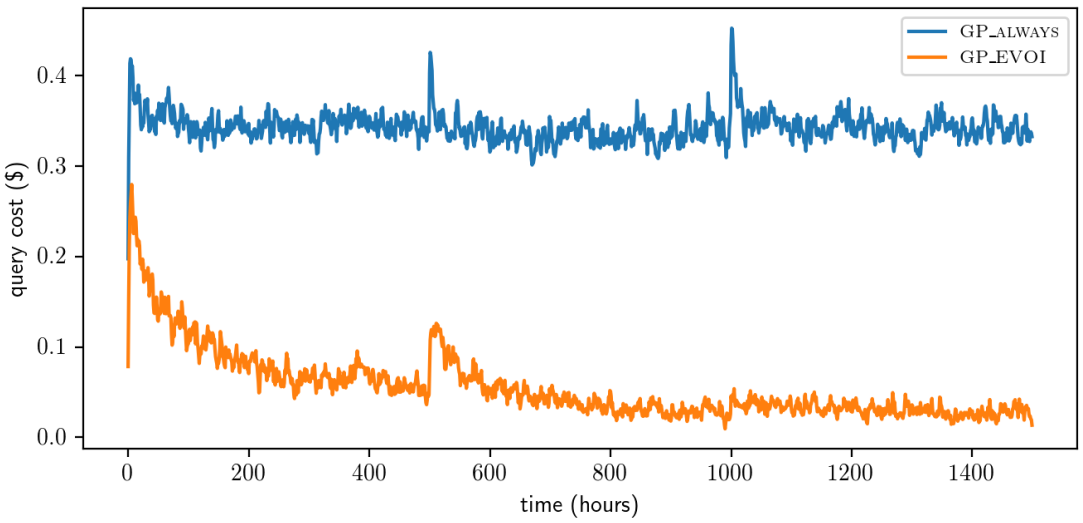


Figure 5.8: Query cost vs. time.

change, GP_EVOI assesses that it already has a high quality utility model for the new context and does not increase its querying.

The difference in how GP_EVOI handles the context changes compared to the other strategies indicates that the benefit of query EVOI estimation increases with a rapidly changing context. For example, in a model where the context rotated every 10 steps, GP_EVOI could concentrate its queries in the later periods of each context, when more state history is available. Meanwhile, the strategies without EVOI estimation would constantly suffer higher query costs due to the lack of relevant state history.

The relative performance of the three strategies is quite similar to the original experiments. GP_EVOI lags behind GP_ALWAYS in policy quality, but asks many fewer queries. The context change slightly increases the number of queries asked by GP_EVOI: 46.4 ($\sigma = 9.8$) by time 100, 132.9 ($\sigma = 19$) by time 500, 204 ($\sigma = 23.9$) by time 1000 and 254 by time 1500 ($\sigma = 36.5$). In terms of reward accrued, GP_EVOI significantly outperforms the other two strategies starting after the first 90 hours.

5.7 Conclusion and Future Work

We have introduced *experiential elicitation*, a framework for preference elicitation in settings where a user's ability to (easily) answer queries is context dependent. Our model of experiential elicitation has tight connections to POMDPs, RL and uncertain-reward MDPs. We studied a new query type for GPs, the *relative value* query, that is well-suited to the home HVAC domain. We showed that GPs naturally accommodate RVQs and offer effective EVOI computation, which is critical for trading off query cost vs. value. Our experiments show that GP-based elicitation using EVOI outperforms other natural baselines in the HVAC setting.

Interesting future directions for EE remain. Most obvious is the significant gap between the performance of our approach and that of the optimal omniscient algorithm. While this gap cannot be closed entirely, a more sophisticated query strategy may have significant effect. In particular, the model allows multiple queries per time step, which our strategy does not make use of. We might achieve better per-

formance by asking many speculative queries at the start of the process. Another improvement would be to make better use of EVOI estimations in query selection. To minimize the increase in computational requirements, we could use a heuristic to come up with a short list of potential queries and then evaluate myopic EVOI for each and then ask the query that offers the most EVOI net of query cost.

Our figures do not fully measure the performance of the algorithm under changing background conditions. We would expect that GP_EVOI should recognize when conditions are shifting into a region where it does not understand user preferences well, and it should respond by asking more queries. Unfortunately, because this requires a large shift in weather, it requires long experiment running times to observe.

General EE problems may be much harder than the ones we encounter in this chapter. The HVAC domain is well-behaved, with limited incentive to deviate from a simple control strategy to seek out better queries. In the worst case, the EE system may be required to make a deviation of unbounded cost from the optimal policy to receive any reward at all. Consider the following example.

Example 5.1. Let $S = \{s_0, \ldots, s_n, s_{elicit}\}$ be the set of states. Let R be that one of $\{s_0, \ldots, s_n\}$, selected uniformly at random, that provides a reward of x, and the others give a reward of zero. The agent can transition between any of the states deterministically. Transitioning between any of the s_i is free, but transitioning to s_{elicit} costs y. A query revealing which of the s_i contains the reward may only be asked from s_{elicit} and it has zero cost and perfect accuracy. The initial state is uniformly distributed over the s_i.

The optimal control policy in this setting without querying is to randomly traverse the s_i. It yields a discounted reward of

$$\sum_{t=0}^{\infty} \gamma^t \frac{x}{n} \tag{5.20}$$

Suppose the following holds:

$$y\gamma \leqslant \frac{x}{n} + \sum_{t=2}^{\infty} x\gamma^t \tag{5.21}$$

The optimal policy is go to s_{elicit} and query once, and then spend the rest of time in the node that yields reward. This yields a reward of

$$\frac{x}{n} - y\gamma + \sum_{t=2}^{\infty} x\gamma^t \tag{5.22}$$

(If the assumption does not hold, the non-querying policy is optimal.)

Let both n and x approach infinity, but n does so faster (so that $\frac{x}{n}$ approaches 0). The reward of the optimal policy without querying approaches 0. Under these conditions, the reward of the optimal querying policy is

$$-y\gamma + \frac{x\gamma^2}{1-\gamma} \tag{5.23}$$

The term on the right-hand side approaches infinity. Let y approach infinity, but asymptotically slower than x. Then, the above approaches infinity. Because y approaches infinity, achieving the optimal policy

requires a deviation of unbounded cost from the optimal policy without querying.

Implementation choices for EE and utility functions that violate quasi-linearity: We discuss two potential issues with our EE model. The first is that RVQs are difficult for users to answer because they require numerical answers. We can adapt our model to address this criticism by replacing RVQs with binary or Likert scale queries, e.g., "Do you prefer the current state to a state that is 1°C degree cooler at a cost of \$0.10 (and what is the strength of your preference on a scale of 1 to 5)?" Chu and Gharamani [2005] and Bonilla et al. [2010] provide variational inference procedures for preference prediction using GPs with binary queries, which could be directly substituted for our exact GP prediction model. Extending their models to Likert scale queries is a potential avenue of future work.

The second issue is that the form of utility functions we require excludes users who violate quasi-linearity because of budget effects, wealth effects (i.e., diminishing returns for money) or non-monetary motivations (e.g., environmentalism). Solving an MDP subject to a budget has been studied previously [Altman, 1999; Boutilier and Lu, 2016]. Since our approach does not depend on the specific algorithm that is used to solve the MDP, a budgeted MDP could be directly substituted into the model. Non-monetary violations do not necessarily pose an obstacle unless they do not have monetary equivalents. The MDP framework may or may not be extensible to quasi-linearity violations depending on the specific form of the violation.

Concept drift: EE is related to the problem of learning under so-called *concept drift*. After describing a generic online classification problem, Widmer and Kubat [1996] write:

> A difficult problem in such a learning scenario is that the concepts of interest may depend on some hidden context. Mild weather means different things in Siberia and in Central Africa; Beatles fans had a different idea of a fashionable haircut than the Depeche Mode generation. Or consider weather prediction rules, which may vary radically depending on the season. Changes in the hidden context can induce more or less radical changes in the target concepts, producing what is generally known as *concept drift* in the literature (e.g., Schlimmer and Granger [1986]).

The problem is in some sense similar to the one faced by the EE system, which must separate features into those which are the main drivers of utility (e.g., internal temperature), those which provide important context (e.g., the month of the year may determine whether the heating system is active or not) and those which are completely irrelevant. Over time, the EE system should learn when the context has changed in a relevant way and know to return to the user with additional queries at that point. This could be used as a PE-based approach to concept drift. When the supervised learning model is uncertain in the current context, it can request more samples.

Chapter 6

Conclusions and Future Work

We begin with a summary of the main contributions of the thesis and then discuss future work.

6.1 Summary

This thesis approaches the problem of developing AI systems for efficient electricity markets from two different angles. In Chapters 3 and 4, we build systems that solve the problem of coordinating a group of electricity-consuming agents as a matching problem. The market has two sides: the sellers of electricity and the buyers. Each buyer has a set of demand profiles, each with a different value, that represent different electricity consumption plans. The objective is to maximize the social welfare, subject to stability constraints, by matching each buyer to a seller and each buyer to a demand profile. The chapters explore different elements of the problem.

In Chapter 3, we focus on the impact of demand "shape" on generator costs. When sellers are assigned by the mechanism to a group of buyers, they are faced with the problem of how to serve that demand in a way that minimizes their costs (the so-called dispatch problem). The way demand changes over time can affect cost. For example, "smoother" demand is cheaper to serve, ceteris paribus, because it requires less ramping (adjusting the output) of generators between time periods. Generator cost functions have a variety of features, including shutdown costs and different layers of generation with different properties. From a technical perspective, the cost functions we consider present the challenge that stability is hard to achieve. We show that even weak definitions of stability are not achievable in the worst case. Thus, we focus on a balanced approach, trading off a small amount of social welfare for increased stability.

The major contributions of Chapter 3 are:

- We develop a tractable market model for matching consumers to producers while capturing many of the complexities of electricity production and consumption (but not those pertaining to predictability of demand).
- We explore of the stability properties of this model under various cost-sharing schemes.
- We develop two payment algorithms that exhibit high stability and fairness in experiments, while allowing tradeoffs between social welfare and stability.

Chapter 4 focuses on a different aspect of generator cost functions: the impact of demand predictability on cost. In Chapter 3, we assumed that demand was deterministic. The addition of stochasticity requires that generators maintain reserve capacity to be drawn on when demand is unexpectedly high. Robu et al. [2017] develop prediction-of-use (POU) tariffs to model this situation. They show that the cooperative game induced by POU tariffs are convex, which makes computation of core-stable payments inexpensive. We perform the first end-to-end test of POU tariffs and find that they can decrease social welfare because they introduce a new coordination problem among electricity consumers without presenting a means for solving it.

To that end, we develop multiple-profile prediction-of-use (MPOU) games. MPOU games bring the "demand profiles" concept from the previous chapter to a setting with uncertainty, allowing each consumer to report a set of demand random variables to the system, with an associated value for each. We show that convexity is maintained in this model, and that, empirically, MPOU games achieve a higher social welfare than fixed-rate tariffs. However, introducing demand profiles in a setting where the choice is not fully observable creates a new principal-agent problem: consumers may choose not to select the demand profile that is assigned by the system. We address this problem through separating functions. Separating functions incentivize the consumer to use the assigned profile while not affecting their expected payment if they do so. Thus, they preserve the convexity of the game and the budget balance of the core payments. We provide a rank-based condition for the existence of separating functions which shows that they exist except for a set of instances with measure 0.

Using data from Pecanstreet.org [Rhodes *et al.*, 2014], we learn household utility models which we use as the basis for our empirical analysis.

The major contributions of Chapter 4 are:

- We extend POU games to support multiple profiles.
- We show that the extension remains convex.
- We introduce a new incentive problem that emerges from the partial observability of the demand profiles and address that problem using separating functions.
- We experimentally test POU and MPOU games using utility models learned from electricity use data. We show that the social welfare of MPOU games is greater than that of the fixed-rate tariff, which is greater than that of POU games.

We find that coordination systems developed in Chapters 3 and 4 improve efficiency, but they place large attentional demands on consumers. To mitigate these demands, we propose that consumers should be represented by autonomous agents, who manage their electricity consumption and interact with the coordination systems on their behalf. In Chapter 5, we focus on the problem of how such agents are informed of the preferences of the consumers they represent. We develop and study experiential elicitation, a new approach to preference elicitation motivated by the electricity management setting. In experiential elicitation, the cost and accuracy of a query are dependent on the current state of the system at the time of the query. This presents a technical challenge by linking the problems of control and elicitation.

We develop an instance of the experiential elicitation framework for the electricity management setting. The decision problem is modeled with a Markov decision process. We introduce a new query type, the relative value query, motivated by our intuitions about what is natural and easy to respond

to in that setting. We show that the relative value query has a synergistic relationship with Gaussian processes, which can aggregate functional difference information. We test our framework on synthetic data and show that our approach to the problem outperforms several natural baselines.

The major contributions of Chapter 5 are:

- We introduce and motivate the experiential elicitation approach for a variety of AI systems that interact with users over time;
- We theoretically connect optimal experiential elicitation to other, well-known, problems in AI;
- We study the interplay (and synergies) of *relative value queries (RVQs)*, which are natural in HVAC settings, with a Gaussian-process (GP) model of preference prediction;
- We develop a system for experiential elicitation in a smart-home HVAC setting using RVQs and GPs, and analyze it empirically using a combination of real and synthetic data, showing that it accrues higher reward than natural baselines.

6.2 Future Work

We now discuss interesting directions for future research.

6.2.1 Integrating Chapters 3 and 4

A natural place to start is integrating the approaches of Chapters 3 and 4. These two chapters approach different aspects of the same problem of coordinating electricity consumption. The former focuses on issues of demand "shape" whereas the latter focuses on predictability. Ideally, a deployed system should solve both parts of the problem simultaneously. We now discuss how this could be done.

The organizing framework should come from Chapter 3. The objective in that chapter is to maximize social welfare, but the form of the objective does not impact the methods as long as it is computationally tractable. Thus, we could add a component that captures the predictability aspect we study in Chapter 4. The optimization will likely be more difficult, but because we solve the optimizations in both chapters as mixed-integer programs, they should be combinable in principle. It is possible that there is a problematic detail, such as an issue with the objective "pressure" required by the constraints.

Combining these two approaches loses some of the useful features of Chapter 4's model. In particular, the combined model won't be convex because the theoretical analysis that shows that the model of Chapter 3 has a high price of stability (Section 3.4) would apply. We could instead model stability in a way similar to Chapter 3, using payments based on smoothed envy-freeness such as SBEF or the Shapley value. If we took this route, we would be faced with the same multi-objective optimization that we encounter in Chapter 3. The sampling-based approach used there may or may not be as effective with different objective functions.

There are questions about how people interact with cooperative game theoretic stability notions in the real world. We discuss these in the next section.

The other approach to combining the two chapters is to expand out from the prediction-of-use games model. Some of this work has been done already. For example, Robu et al. [2017] consider POU games with multiple available tariffs, which we do not (note that they consider only a single profile). It is likely that a larger class of convex games exists. The problem is that, once we start adding in the key

features of the demand shape problem, we will lose convexity. Thus, its unclear how useful results that rely on convexity are. However, there are some aspects of the POU setting that we do not understand very well that are probably worth exploring in more detail, such as the independence of prediction errors assumption and the idea of joint users being able to make predictions contingent on public factors (such as the weather). We discuss these aspects in the future work section of Chapter 4.

The principal-agent problem that occurs in POU games will also arise if POU games are combined with other models. We should be able to use separating functions to address them in essentially the same way as in Chapter 4. Designing separating functions can be done after performing the matching optimization. It "solidifies" the result of the optimization, incentivizing agents to use their assigned profiles. The calculation of separating functions should not be affected by non-independent prediction errors.

Abstracted away from the electricity setting, the overall goal can be viewed as constructing a general pipeline for a cooperative game modeling a market. We want to be able to coordinate the use of scarce resources, eliciting utility functions truthfully from the agents, while taking stability into account and raising enough revenue to pay for the resources.[1]

In the remaining sections, we discuss key challenges on the path to this goal. The next section discusses stability from a practical perspective in cooperative games and the following incentive issues.

6.2.2 Rationality of Players

Cooperative game theoretic stability concepts have received little testing in real environments. Standard Shapley payments have an obvious failure mode as any number of agents could have a potentially large incentive to defect, which could cause the entire mechanism to unravel. To an extent, Shapley payments rely on agents recognizing that although it may be in their short-term interest to leave the mechanism, their long term interest is to stay in it. Core-stable payments have the problem that, despite their theoretical strength, they can appear quite unfair, as we saw in Section 2.1.4. The high computational complexity of some cost-sharing schemes indicates that they can have quite unintuitive outcomes. It can be hard to understand why one is unable to raise one's payment by making a side deal.

We summarize two empirical studies of how cost sharing is performed in a cooperative game by Williams [1988] and Michener and Yuen [1982]. Williams studies multiple-purpose river developments constructed by the United States Army Corps of Engineers. The purposes include flood control, hydroelectric power, irrigation, etc. The cost and how it is divided across state, federal and local bodies is fixed given the purposes that are selected for the project. What makes William's analysis possible is that the Corps of Engineers calculates the cost and benefit of many of the possible combinations of purposes. The author finds that all observed cost divisions are core-stable in the case where a core allocation exists, even in the case when other solution concepts, such as the Shapley value, are not.

Michener and Yuen perform laboratory experiments, where participants are told the characteristic value function of a game and then rewarded monetarily proportional to the payment they receive in the game. The authors come to the opposite conclusion of Williams: that the core is less predictive than other solution concepts.

One explanation for the conflicting results is the level of rationality of the players. In Williams, the negotiating parties are city representatives acting in a professional capacity, whereas in Michener and Yuen, they are probably undergraduates being paid to participate in the study. The electricity use

[1] Note that standard VCG's weaknesses are in the latter two aspects.

setting is likely closer to the latter than the former. This may be a negative indicator of the applicability of the core in practice.

Another potential explanation of the difference in outcomes between the two papers is a difference in starting conditions. In Williams, the initial proposed division is based on the Corps of Engineers' cost-sharing algorithm, which heavily favors the core. In Michener and Yuen, there is no initial proposal given to the participants. If this is the key difference, it would indicate that players are persuadable. This may be the case because the players are not initially familiar with cost sharing in cooperative games, and could develop stronger preferences over time.

One way to form a general theory of real-world cost sharing is through an analogy to the way assumptions are handled in non-cooperative games. There, the area of *behavioral game theory* [Camerer, 2011] is devoted to the problem of predicting player actions. Behavioral game theory has developed a number of different theories to explain the results of laboratory tests. In practice, systems perform robust optimization (e.g., Tambe [2011]) against a range of different behavior models.

A direct analogue of behavioral game theory for cooperative games would predict a distribution over payments to players for a particular cooperative game. This could be seen as a quantitative extension of the empirical cooperative games work previously discussed. However, the philosophy of mechanism design indicates that the outcome is proposed by the system, and the players choice is whether to accept it or not.

We could follow this logic and develop a probabilistic model of whether a group of players would accept an outcome or not. Electricity coordination schemes motivate us to think in the large, in contrast to behavioral game theory which usually focuses on two-player games. For example, a model that predicts the chance to defect for each agent and treats those chances as independent has a lot of theoretical appeal. In a large market, however, there would almost always be at least one defector, no matter how small the individual probabilities were.

One key intuition is that egalitarian fairness concepts have a role to play, such as similarity-based envy-freeness or DHPRZ. Giving a guarantee that the outcome they receive is similar to that of agents with similar capabilities seems to address a key concern. Coming to an agreement on what constitutes similar (on both dimensions) is difficult. For example, consider the instance with odd and even players of Section 2.1.4, where core-stable allocations appear unfair. From a mathematical perspective, we view the two players as completely different—there are two types of capabilities in the game and they are polar opposites. Our intuition leads us to view them as similar because their roles are reflections of each other's.

6.2.3 Incentives to Misreport

Questions remain around the issue that players may be incentivized to misreport their utility functions. We discuss this briefly in Section 3.7, and extend that discussion here. Our idea is to follow Boutilier and Sandholm's [2006] approach to combinatorial auctions by prioritizing preference elicitation concerns and achieve truthfulness on the mechanism side. In this section, we focus on the idea of extending the VCG mechanism to achieve this goal. It is possible that an ad hoc approach where we bound the incentive of the agent to misreport with every query, or an empirical approach where we measure the incentive to misreport in data, are better options, but more can be said about the VCG approach.

We begin with an abstract model of Chapters 3 and 4. We have a set of agents N, each with a private utility function $U_i : O \to \mathbb{R}$ mapping from outcome space to the utility received. We have a

cost function $C : O \to \mathbb{R}$ that maps from outcome space to the amount of revenue we need to raise to execute the outcome. Each agent i has a set of demand profiles Π_i. An outcome consists of a demand profile assignment to each agent. We use the standard mechanism design notation for types. Let the set of types be Θ, agent i's true type be θ_i, agent i's reported type be $\hat{\theta}_i$ and the joint type report of all agents be $\hat{\theta}$.

The optimization problem faced by the mechanism is to select an outcome that maximizes an organizing function minus the cost of the outcome and determines payments to each agent such that the amount of revenue raised is greater than the cost of the outcome. In standard mechanism design, the organizing function is social welfare, i.e., the sum of the utilities of the individual agents. Our organizing function differs from the standard one in two main ways. First, we want to take the cost of the outcome into account. Second, we may want to add other factors to the objective, such as stability.

Both of these problems can be addressed by adding dummy agents to the problem. For the cost function, we create a dummy agent with utility equal to the negative cost of the outcome. We do the same for the stability. Now we sum up agent utilities in the usual way, but we do not collect a payment from the dummy agents. Because the dummy agents can be introduced within the standard mechanism design model, the useful properties of VCG with Clarke payments are maintained, such as truthfulness, individual rationality, and non-negative payments. Note that this assumes we can maximize the new objective exactly, which was not true for the social welfare objective with stability in Chapter 3.

Adding dummy agents representing cost and stability resolves another issue with Clarke payments. Clarke payments rely on the existence of an externality imposed by some agent that it can "tax." Consider the following example.

Example 6.1. Suppose we have 1000 agents and two outcomes o_1 and o_2. 498 of the agents have utility 1 for outcome o_1 and 0 for o_2 and 502 have utility 0 for o_1 and 1 for o_2. In this scenario, VCG selects o_2 because it has a higher social welfare of 502 vs. 498. However, because removing one agent from the problem does not change the outcome, the payment under the Clarke pivot rule is 0 for all agents.

The problem of having a zero payment always arises when there are enough agents that removing a single agent does not change the outcome. It would seem that this can easily occur in a large enough problem. In the electricity setting, adding a new agent will often not change the optimal profiles of the other agents or the optimal generation strategy, but it will increase the cost.

It is more difficult to ensure that the mechanism raises the right amount of revenue from the agents to pay the cost. In the absence of other externalities, whether the Clarke payments raise too little or too much revenue depends on whether the cost function is concave or convex. The following two examples illustrate.

Example 6.2. Let there be 100 identical agents. Each agent demands 1 unit of electricity and accrues a utility of 0.11 if they receive it and 0 otherwise. There is only a single agent type, so untruthful reporting is impossible. Let the cost function be equal to the square root of the demand. We use the VCG mechanism with Clarke payments. With a dummy agent for the cost function, the outcome that serves all of the demands is selected because each agent receives more utility than the cost of serving them. Each agent's payment is $\sqrt{100} - \sqrt{99} \approx 0.05$. Thus, the Clarke payments raise only about half of the revenue needed.

Example 6.3. Consider the previous example except the cost function is the square of the demand, and each agent accrues a utility of 100 if they receive their unit of electricity and 0 otherwise. Social

welfare is maximized when 50 agents are served. Each agent's payment is $50^2 - 49^2 = 99$. The Clarke payments raise twice the revenue required.

In each example, we would like to scale the payments in order to make the revenue raised match the revenue required, which is permitted by the mechanism as long as our choice of scaling constant for each agent is not affected by that agent's report. This is obviously impossible when there are a small number of agents: each agent's report has too much of an impact on the outcome. However, in a large problem, we may be able to approximate the scaling factor well.

Concretely, suppose we want to calculate a scaling factor c_i for agent i's payment, which is a VCG payment with the Clarke pivot rule, using a dummy cost agent. If we did not care about incentive compatibility, we could raise the requisite revenue by setting the scaling factor equal to the ratio of the cost and the total payment collected from all agents:

$$c_i = \frac{C(f(\hat{\theta}))}{-\sum_{j \in N} t(\hat{\theta})(j)} \tag{6.1}$$

This would violate incentive compatibility because both the numerator and the denominator depend on i's reported type $\hat{\theta}_i$. If i's impact on the problem is small enough, we could approximate $f(\hat{\theta})$ with $f(\hat{\theta}_{-i})$ and $t(\hat{\theta})(j)$ with $t(\hat{\theta}_{-i})(j)$. The former is free computationally (as long as we are computing Clarke pivot rule payments), while the latter is very expensive: it requires n more optimizations of size $n-2$ per agent (where n is the number of agents).

Approximating these quantities to save computation would be necessary in practice. The complexity of computing Clarke payments alone is likely too high for large problems as it requires a separate optimization for each agent. However, many properties of VCG with Clarke payments are robust to payment approximation, in contrast to approximation of the organizing function. Recall that it is important for the organizing function optimization to be very accurate—if it is approximated, each agent may have an incentive to misreport that is as large as the difference in objective function values between the approximate and the exact solutions (see Section 2.3.4 for the proof).

Since the pivot rule can be any function, as long as it does not depend on the agent's report, truthfulness is never affected by the inaccuracy in the pivot rule calculation. We just need to begin our approximation with the true sum of the other agents' utilities in the selected outcome (which is cheap to compute).

Our main incentive for the approximation to be accurate is individual rationality. It would be possible to check whether the payment respects individual rationality given the report, but it would violate truthfulness to do so.

Bibliography

[Abadi *et al.*, 2015] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[Abraham *et al.*, 2007] David J. Abraham, Avrim Blum, and Tuomas Sandholm. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In *Proceedings of the Eighth ACM Conference on Electronic Commerce (EC'07)*, pages 295–304. ACM, 2007.

[Akasiadis and Chalkiadakis, 2013] Charilaos Akasiadis and Georgios Chalkiadakis. Agent cooperatives for effective power consumption shifting. In *Proceedings of the Twenty-seventh AAAI Conference on Artificial Intelligence (AAAI-13)*, pages 1263–1269, Bellevue, WA, 2013.

[Altman, 1999] Eitan Altman. *Constrained Markov Decision Processes*. Chapman and Hall, London, 1999.

[Anand and Aron, 2003] Krishnan S. Anand and Ravi Aron. Group buying on the web: A comparison of price-discovery mechanisms. *Management Science*, 49(11):1546–1562, 2003.

[Ausubel *et al.*, 2006] Lawrence M. Ausubel, Peter Cramton, and Paul Milgrom. The clock-proxy auction: A practical combinatorial auction design. In Martin Bichler and Jacob K. Goeree, editors, *Handbook of Spectrum Auction Design*, chapter 6, pages 120–140. MIT Press, Cambridge, MA, 2006.

[Ausubel, 2004] Lawrence M. Ausubel. An efficient ascending-bid auction for multiple objects. *American Economic Review*, 94(5):1452–1475, 2004.

[Beigman and Vohra, 2006] Eyal Beigman and Rakesh Vohra. Learning from revealed preference. In *Proceedings of the Seventh ACM Conference on Electronic Commerce (EC'06)*, pages 36–42, Ann Arbor, 2006.

[Bichler *et al.*, 2013] Martin Bichler, Pasha Shabalin, and Jürgen Wolf. Do core-selecting combinatorial clock auctions always lead to high efficiency? An experimental analysis of spectrum auction designs. *Experimental Economics*, 16(4):511–545, 2013.

[Biró *et al.*, 2014] Péter Biró, David F. Manlove, and Iain McBride. The hospitals/residents problem with couples: Complexity and integer programming models. In *Experimental Algorithms*, pages 10–21. Springer, 2014.

[Blumrosen and Nisan, 2010] Liad Blumrosen and Noam Nisan. On the computational power of demand queries. *SIAM Journal on Computing*, 39(4):1372–1391, 2010.

[Bonilla *et al.*, 2010] Edwin V. Bonilla, Shengbo Guo, and Scott Sanner. Gaussian process preference elicitation. In *Advances in Neural Information Processing Systems 23 (NIPS-10)*, Vancouver, 2010.

[Boutilier and Lu, 2016] Craig Boutilier and Tyler Lu. Budget allocation using weakly coupled, constrained Markov decision processes. In *Proceedings of the Thirty-second Conference on Uncertainty in Artificial Intelligence (UAI-16)*, New York, 2016.

[Boutilier *et al.*, 2006] Craig Boutilier, Relu Patrascu, Pascal Poupart, and Dale Schuurmans. Constraint-based optimization and utility elicitation using the minimax decision criterion. *Artifical Intelligence*, 170(8–9):686–713, 2006.

[Boutilier, 2002] Craig Boutilier. A POMDP formulation of preference elicitation problems. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-02)*, pages 239–246, Edmonton, 2002.

[Braithwait *et al.*, 2007] Steven Braithwait, Dan Hansen, and Michael O'Sheasy. Retail electricity pricing and rate design in evolving markets. *Edison Electric Institute*, pages 1–57, 2007.

[Braziunas and Boutilier, 2010] Darius Braziunas and Craig Boutilier. Assessing regret-based preference elicitation with the UTPREF recommendation system. In *Proceedings of the Eleventh ACM Conference on Electronic Commerce (EC'10)*, pages 219–228, Cambridge, MA, 2010.

[Breiman *et al.*, 1984] Leo Breiman, Jerome Friedman, Charles J. Stone, and Richard A. Olshen. *Classification and Regression Trees*. Chapman and Hall/CRC, 1984.

[Breiman, 2001] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[Brero *et al.*, 2017] Gianluca Brero, Benjamin Lubin, and Sven Seuken. Probably approximately efficient combinatorial auctions via machine learning. In *Proceedings of the Thirty-first AAAI Conference on Artificial Intelligence (AAAI-17)*, pages 397–405, 2017.

[Camerer, 2011] Colin F. Camerer. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, Princeton, New Jersey, 2011.

[Castro *et al.*, 2009] Javier Castro, Daniel Gómez, and Juan Tejada. Polynomial calculation of the Shapley value based on sampling. *Computers & Operations Research*, 36(5):1726–1730, 2009.

[Chajewska *et al.*, 2000] Urszula Chajewska, Daphne Koller, and Ronald Parr. Making rational decisions using adaptive utility elicitation. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-00)*, pages 363–369, Austin, TX, 2000.

[Charlin *et al.*, 2012] Laurent Charlin, Richard Zemel, and Craig Boutilier. Active learning for matching problems. In *Proceedings of the Twenty-ninth International Conference on Machine Learning (ICML-12)*, pages 337–344, Edinburgh, 2012.

[Chen and Roma, 2010] Rachel R. Chen and Paolo Roma. Group buying of competing retailers. *Production and Operations Management*, 20(2):181–197, 2010.

[Chen *et al.*, 2007] Jian Chen, Xilong Chen, and Xiping Song. Comparison of the group-buying auction and the fixed pricing mechanism. *Decision Support Systems*, 43(2):445–459, 2007.

[Chu and Ghahramani, 2005] Wei Chu and Zoubin Ghahramani. Preference learning with Gaussian processes. In *Proceedings of the Twenty-second International Conference on Machine Learning (ICML-05)*, pages 137–144, Bonn, 2005.

[Clarke, 1971] Edward H. Clarke. Multipart pricing of public goods. *Public Choice*, 11(1):17–33, 1971.

[Conitzer and Sandholm, 2006] Vincent Conitzer and Tuomas Sandholm. Complexity of constructing solutions in the core based on synergies among coalitions. *Artificial Intelligence*, 170(6-7):607–619, 2006.

[Cramton *et al.*, 2005] Peter Cramton, Yoav Shoham, and Richard Steinberg, editors. *Combinatorial Auctions*. MIT Press, Cambridge, Massachussetts, 2005.

[Cremer and McLean, 1988] Jacques Cremer and Richard P McLean. Full extraction of the surplus in Bayesian and dominant strategy auctions. *Econometrica*, pages 1247–1257, 1988.

[Demange *et al.*, 1986] Gabrielle Demange, David Gale, and Marilda Sotomayer. Multi-item auctions. *Journal of Political Economy*, 94:863–872, 1986.

[Denholm *et al.*, 2015] Paul Denholm, Matthew O'Connell, Gregory Brinkman, and Jennie Jorgenson. Overgeneration from solar energy in California. a field guide to the duck chart. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2015.

[d'Epenoux, 1963] F d'Epenoux. A probabilistic production and inventory problem. *Management Science*, 10(1):98–108, 1963.

[Dickerson *et al.*, 2013] John P. Dickerson, Ariel D. Procaccia, and Tuomas Sandholm. Failure-aware kidney exchange. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce (EC'13)*, pages 323–340. ACM, 2013.

[D&R International, Ltd., 2012] D&R International, Ltd. *2011 Buildings Energy Data Book*. Department of Energy, Office of Energy Efficiency and Renewable Energy, 2012.

[Drummond and Boutilier, 2014] Joanna Drummond and Craig Boutilier. Preference elicitation and interview minimization in stable matchings. In *Proceedings of the Twenty-eighth AAAI Conference on Artificial Intelligence (AAAI-14)*, pages 645–653, Québec City, 2014.

[Drummond *et al.*, 2015] Joanna Drummond, Andrew Perrault, and Fahiem Bacchus. SAT is an effective and complete method for solving stable matching problems with couples. In *Proceedings of the Twenty-fourth International Joint Conference on Artificial Intelligence (IJCAI-15)*, pages 518–525, 2015.

[Dwork *et al.*, 2012] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pages 214–226, Cambridge, Massachusetts, 2012. ACM.

[Farhangi, 2010] Hassan Farhangi. The path of the smart grid. *Power and Energy Magazine, IEEE*, 8(1):18–28, 2010.

[Friedman *et al.*, 2009] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*. Springer-Verlag New York, 2009.

[Gale and Shapley, 1962] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, 69(1):9–15, 1962.

[Gent and Prosser, 2002] Ian P. Gent and Patrick Prosser. SAT encodings of the stable marriage problem with ties and incomplete lists. In *Proceedings of Theory and Applications of Satisfiability Testing (SAT)*, pages 133–140, 2002.

[Gent *et al.*, 2001] Ian P. Gent, Robert W. Irving, David F. Manlove, Patrick Prosser, and Barbara M. Smith. A constraint programming approach to the stable marriage problem. In *Principles and Practice of Constraint Programming (CP)*, pages 225–239, 2001.

[Gescheider, 2013] George A. Gescheider. *Psychophysics: the fundamentals*. Psychology Press, 2013.

[Gillies, 1959] Donald B. Gillies. Solutions to general non-zero-sum games. *Contributions to the Theory of Games*, 4(40):47–85, 1959.

[Goto *et al.*, 2016] Masahiro Goto, Atsushi Iwasaki, Yujiro Kawasaki, Ryoji Kurata, Yosuke Yasuda, and Makoto Yokoo. Strategyproof matching with regional minimum and maximum quotas. *Artificial Intelligence*, 235:40–57, 2016.

[Groves, 1973] Theodore Groves. Incentives in teams. *Econometrica*, 41:617–631, 1973.

[Holloway and White, 2003] Hillary A. Holloway and Chelsea C. White, III. Question selection for multiattribute decision-aiding. *European Journal of Operational Research*, 148:525–543, 2003.

[Huang *et al.*, 1999] Joe Huang, James Hanford, and Fuqiang Yang. Residential heating and cooling loads component analysis. Technical report, Building Technologies Department, Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory, University of California, 1999.

[Hudson and Sandholm, 2004] Benoit Hudson and Tuomas Sandholm. Effectiveness of query types and policies for preference elicitation in combinatorial auctions. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-04)*, pages 386–393, Washington, DC, USA, 2004. IEEE Computer Society.

[Hui and Boutilier, 2008] Bowen Hui and Craig Boutilier. Toward experiential utility elicitation for interface customization. In *Proceedings of the Twenty-fourth Conference on Uncertainty in Artificial Intelligence (UAI-08)*, pages 298–305, Helsinki, 2008.

[Judah *et al.*, 2014] Kshitij Judah, Alan Paul Fern, Prasad Tadepalli, and Robby Goetschalckx. Imitation learning with demonstrations and shaping rewards. In *Proceedings of the Twenty-eighth AAAI Conference on Artificial Intelligence (AAAI-14)*, pages 1890–1896, Quebec City, 2014.

[Kahneman, 2011] Daniel Kahneman. *Thinking, fast and slow*. Macmillan, 2011.

[Kiekintveld *et al.*, 2009] Christopher Kiekintveld, Manish Jain, Jason Tsai, James Pita, Fernando Ordóñez, and Milind Tambe. Computing optimal randomized resource allocations for massive security games. In *Proceedings of the Eighth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-09)*, pages 689–696, Budapest, 2009.

[Kingma and Ba, 2015] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *Proceedings of the 3rd ACM SIGKDD International Conference on Learning Representations (ICLR-15)*, San Diego, 2015.

[Kirschen and Strbac, 2004] Daniel S. Kirschen and Goran Strbac. *Fundamentals of Power System Economics*. John Wiley & Sons, Chichester, UK, 2004.

[Kota *et al.*, 2012] Ramachandra Kota, Georgios Chalkiadakis, Valentin Robu, Alex Rogers, and Nicholas R Jennings. Cooperatives for demand side management. In *Proceedings of the Twenty-First European Conference on Artificial Intelligence (ECAI-12)*, pages 969–974, Montpellier, France, 2012.

[Koutsoupias and Papadimitriou, 1999] Elias Koutsoupias and Christos Papadimitriou. Worst-case equilibria. In *Proceedings of the Sixteenth Symposium on Theoretical Aspects of Computer Science (STACS-99)*, pages 404–413, Trier, Germany, 1999.

[Leyton-Brown *et al.*, 2017] Kevin Leyton-Brown, Paul Milgrom, and Ilya Segal. Economics and computer science of a radio spectrum reallocation. *Proceedings of the National Academy of Sciences*, 114(28):7202–7209, 2017.

[Lu and Boutilier, 2012] Tyler Lu and Craig Boutilier. Matching models for preference-sensitive group purchasing. In *Proceedings of the Thirteenth ACM Conference on Electronic Commerce (EC'12)*, pages 723–740, Valencia, Spain, 2012.

[Ma *et al.*, 2016] Hongyao Ma, Valentin Robu, Na Li, and David C. Parkes. Incentivizing reliability in demand-side response. In *Proceedings of the Twenty-five International Joint Conference on Artificial Intelligence (IJCAI-16)*, pages 352–358, New York, 2016.

[Ma *et al.*, 2017] Hongyao Ma, David C. Parkes, and Valentin Robu. Generalizing demand response through reward bidding. In *Proceedings of the Sixteenth International Conference on Autonomous Agents and Multiagent Systems (AAMAS-17)*, pages 60–68, São Paolo, 2017.

[Manisterski *et al.*, 2008] Efrat Manisterski, David Sarne, and Sarit Kraus. Enhancing cooperative search with concurrent interactions. *Journal of Artificial Intelligence Research*, 32(1):1–36, 2008.

[Manlove *et al.*, 2007] David Manlove, Gregg O'Malley, Patrick Prosser, and Chris Unsworth. A constraint programming approach to the hospitals/residents problem. In *Integration of AI and OR Techniques in Constraint Programming (CPAIOR)*, pages 155–170, 2007.

[Maschler *et al.*, 1979] Michael Maschler, Bezalel Peleg, and Lloyd S. Shapley. Geometric properties of the kernel, nucleolus, and related solution concepts. *Mathematics of Operations Research*, 4(4):303–338, 1979.

[Meir *et al.*, 2014] Reshef Meir, Tyler Lu, Moshe Tennenholtz, and Craig Boutilier. On the value of using group discounts under price competition. *Artificial Intelligence*, 216:163–178, 2014.

[Meir *et al.*, 2017] Reshef Meir, Hongyao Ma, and Valentin Robu. Contract design for energy demand response. In *Proceedings of the Twenty-sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, pages 1202–1208, Melbourne, 2017.

[Michener and Yuen, 1982] H. Andrew Michener and Kenneth Yuen. A competitive test of the core solution in sidepayment games. *Behavioral Science*, 27(1):57–68, 1982.

[Müller *et al.*, 2002] Alfred Müller, Marco Scarsini, and Moshe Shaked. The newsvendor game has a nonempty core. *Games and Economic Behavior*, 38(1):118–126, 2002.

[Ng and Russell, 2000] Andrew Ng and Stuart Russell. Algorithms for inverse reinforcement learning. In *Proceedings of the Seventeenth International Conference on Machine Learning (ICML-00)*, pages 663–670, Stanford, CA, 2000.

[Nguyen *et al.*, 2014] Thanh Hong Nguyen, Amulya Yadav, Bo An, Milind Tambe, and Craig Boutilier. Regret-based optimization and preference elicitation for stackelberg security games with uncertainty. In *Proceedings of the Twenty-eighth AAAI Conference on Artificial Intelligence (AAAI-14)*, pages 756–762, Québec City, 2014.

[Odom and Natarajan, 2016] Phillip Odom and Sriraam Natarajan. Active advice seeking for inverse reinforcement learning. In *Proceedings of the Fifteenth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-16)*, pages 512–520, Singapore, 2016.

[Osborne and Rubinstein, 1994] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. MIT Press, Cambridge, 1994.

[Parkes, 2005] David C. Parkes. Auction design with costly preference elicitation. *Annals of Mathematics and Artificial Intelligence*, 44(3):269–302, 2005.

[Pedregosa *et al.*, 2011] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[Peranson and Roth, 1999] Elliott Peranson and Alvin E. Roth. The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American Economic Review*, 89(4):748–780, 1999.

[Perrault and Boutilier, 2015] Andrew Perrault and Craig Boutilier. Approximately stable pricing for coordinated purchasing of electricity. In *Proceedings of the Twenty-fourth International Joint Conference on Artificial Intelligence (IJCAI-15)*, Buenos Aires, 2015.

[Perrault and Boutilier, 2017] Andrew Perrault and Craig Boutilier. Multiple-profile prediction-of-use games. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (ICJAI-17)*, pages 366–373, Melbourne, 2017.

[Perrault and Boutilier, 2018] Andrew Perrault and Craig Boutilier. Experiential preference elicitation for autonomous HVAC systems. Unpublished, 2018.

[Perrault *et al.*, 2016] Andrew Perrault, Joanna Drummond, and Fahiem Bacchus. Strategy-proofness in the stable matching problem with couples. In *Proceedings of the Fifteenth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-16)*, pages 132–140, Singapore, 2016.

[Poupart *et al.*, 2006] Pascal Poupart, Nikos Vlassis, Jesse Hoey, and Kevin Regan. An analytic solution to discrete Bayesian reinforcement learning. In *Proceedings of the Twenty-third International Conference on Machine Learning (ICML-06)*, pages 697–704, Pittsburgh, 2006.

[Program, 2013] National Resident Matching Program. National resident matching program, results and data: 2013 main residency match, 2013.

[Prosser, 2014] Patrick Prosser. Stable roommates and constraint programming. In *Integration of AI and OR Techniques in Constraint Programming (CPAIOR)*, pages 15–28, 2014.

[Puterman, 1994] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, New York, 1994.

[Ramchurn *et al.*, 2012] Sarvapali D. Ramchurn, Perukrishnen Vytelingum, Alex Rogers, and Nicholas R. Jennings. Putting the "smarts" into the smart grid: A grand challenge for artificial intelligence. *Communications of the ACM*, 55(4):86–97, 2012.

[Rasmussen and Williams, 2006] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[Rassenti *et al.*, 2003] Stephen J. Rassenti, Vernon L. Smith, and Bart J. Wilson. Controlling market power and price spikes in electricity networks: Demand-side bidding. *Proceedings of the National Academy of Sciences*, 100(5):2998–3003, 2003.

[Regan and Boutilier, 2009] Kevin Regan and Craig Boutilier. Regret-based reward elicitation for Markov decision processes. In *Proceedings of the Twenty-fifth Conference on Uncertainty in Artificial Intelligence (UAI-09)*, pages 454–451, Montreal, 2009.

[Regan and Boutilier, 2011] Kevin Regan and Craig Boutilier. Eliciting additive reward functions for Markov decision processes. In *Proceedings of the Twenty-second International Joint Conference on Artificial Intelligence (IJCAI-11)*, pages 2159–2164, Barcelona, 2011.

[Reinhold, 2016] Arnold Reinhold. California hourly electric load vs. load less solar and wind (duck curve). `https://commons.wikimedia.org/wiki/File:Duck_Curve_CA-ISO_2016-10-22.agr.png`, 2016. Accessed May 25, 2018.

[Rhodes *et al.*, 2014] Joshua D. Rhodes, Charles R. Upshaw, Chioke B. Harris, Colin M. Meehan, David A. Walling, Paul A. Navrátil, Ariane L. Beck, Kazunori Nagasawa, Robert L. Fares, Wesley J. Cole, et al. Experimental and data collection methods for a large-scale smart grid deployment: Methods and first results. *Energy*, 65:462–471, 2014.

[Robu *et al.*, 2017] Valentin Robu, Meritxell Vinyals, Alex Rogers, and Nicholas Jennings. Efficient buyer groups with prediction-of-use electricity tariffs. *IEEE Transactions on Smart Grid*, 2017.

[Ronn, 1990] Eytan Ronn. NP-complete stable matching problems. *Journal of Algorithms*, 11(2):285–304, 1990.

[Rose *et al.*, 2012] Harry Rose, Alex Rogers, and Enrico H Gerding. A scoring rule-based mechanism for aggregate demand prediction in the smart grid. In *Proceedings of the Eleventh International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-12)*, pages 661–668, Valencia, Spain, 2012.

[Roth, 2002] Alvin E. Roth. The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica*, 70(4):1341–1378, 2002.

[Rothkopf and Dimitrakakis, 2011] Constantin A. Rothkopf and Christos Dimitrakakis. Preference elicitation and inverse reinforcement learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 34–48, Athens, 2011.

[Rothkopf *et al.*, 1998] Michael H. Rothkopf, Aleksander Pekeč, and Ronald M. Harstad. Computationally manageable combinatorial auctions. *Management Science*, 44(8):1131–1147, 1998.

[Samuelson, 1948] Paul A. Samuelson. Consumption theory in terms of revealed preference. *Economica*, 15(60):243–253, 1948.

[Sandholm and Boutilier, 2006] Tuomas Sandholm and Craig Boutilier. Preference elicitation in combinatorial auctions. In P. Crampton, Y. Shoham, and R. Steinberg, editors, *Combinatorial Auctions*, pages 233–264. MIT Press, Cambridge, MA, 2006.

[Sandholm *et al.*, 1999] Tuomas Sandholm, Kate Larson, Martin Andersson, Onn Shehory, and Fernando Tohme. Coalition structure generation with worst case guarantees. *Artificial Intelligence*, 111(1–2):209–238, 1999.

[Sarne and Kraus, 2005] David Sarne and Sarit Kraus. Cooperative exploration in the electronic marketplace. In *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05)*, pages 158–163, Pittsburgh, 2005.

[Scheffel *et al.*, 2012] Tobias Scheffel, Georg Ziegler, and Martin Bichler. On the impact of package selection in combinatorial auctions: an experimental study in the context of spectrum auction design. *Experimental Economics*, 15(4):667–692, 2012.

[Schlimmer and Granger, 1986] Jeffrey C. Schlimmer and Richard H. Granger. Incremental learning from noisy data. *Machine Learning*, 1(3):317–354, 1986.

[Schulz and Moses, 2003] Andreas S. Schulz and Nicolás E. Stier Moses. On the performance of user equilibria in traffic networks. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA-03)*, pages 86–87, San Francisco, 2003.

[Shann and Seuken, 2013] Mike Shann and Sven Seuken. An active learning approach to home heating in the smart grid. In *Proceedings of the Twenty-third International Joint Conference on Artificial Intelligence (IJCAI-13)*, pages 2892–2899, Beijing, 2013.

[Shann and Seuken, 2014] Mike Shann and Sven Seuken. Adaptive home heating under weather and price uncertainty using GPs and MDPs. In *Proceedings of the Thirteenth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-14)*, pages 821–828, Paris, 2014.

[Shapley and Shubik, 1966] Lloyd S. Shapley and Martin Shubik. Quasi-cores in a monetary economy with nonconvex preferences. *Econometrica: Journal of the Econometric Society*, pages 805–827, 1966.

[Shapley and Shubik, 1969] Lloyd S. Shapley and Martin Shubik. On market games. *Journal of Economic Theory*, 1(1):9–25, 1969.

[Shapley and Shubik, 1971] L. S. Shapley and M. Shubik. The assignment game I: The core. *International Journal of Game Theory*, 1(1):111–130, 1971.

[Shapley, 1953] Lloyd S. Shapley. A value for n-person games. In Harold W. Kuhn and Albert W. Tucker, editors, *Contributions to the Theory of Games II*, pages 307–317. Princeton University Press, Princeton, 1953.

[Shapley, 1971] Lloyd S. Shapley. Cores of convex games. *International Journal of Game Theory*, 1:11–26, 1971.

[Shoham and Leyton-Brown, 2008] Yoav Shoham and Kevin Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.

[Simon and Blume, 1994] Carl P Simon and Lawrence Blume. *Mathematics for economists*, volume 7. Norton New York, 1994.

[Smallwood and Sondik, 1973] Richard D. Smallwood and Edward J. Sondik. The optimal control of partially observable Markov processes over a finite horizon. *Operations Research*, 21:1071–1088, 1973.

[Snoek *et al.*, 2012] Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems 25 (NIPS-12)*, pages 2951–2959, Harrahs and Harveys, Lake Tahoe, 2012.

[Snoek *et al.*, 2015] Jasper Snoek, Oren Rippel, Kevin Swersky, Ryan Kiros, Nadathur Satish, Narayanan Sundaram, Mostofa Patwary, Prabhat, and Ryan Adams. Scalable Bayesian optimization using deep neural networks. In *Proceedings of the Thirty-second International Conference on Machine Learning (ICML-15)*, pages 2171–2180, Lille, 2015.

[Srivastava *et al.*, 2014] Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[Tambe, 2011] Milind Tambe. *Security and game theory: algorithms, deployed systems, lessons learned*. Cambridge University Press, 2011.

[Team, 2011] G. M. Team. Electricity and gas supply market report. Technical Report 176/11, The Office of Gas and Electricity Markets (Ofgem), December 2011.

[Thrall and Lucas, 1963] Robert M. Thrall and William F. Lucas. N-person games in partition function form. *Naval Research Logistics (NRL)*, 10(1):281–298, 1963.

[Unsworth and Prosser, 2005] Chris Unsworth and Patrick Prosser. A specialised binary constraint for the stable marriage problem. In *Abstraction, Reformulation and Approximation (SARA)*, pages 218–233, 2005.

[Vickrey, 1961] William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance*, 16(1):8–37, 1961.

[von Neumann, 1928] J von Neumann. Zur theorie der gesellschaftsspiele. *Mathematische annalen*, 100(1):295–320, 1928.

[Waddams *et al.*, 2014] Catherine Waddams, David Deller, Graham Loomes, Monica Giulietti, Ana Moniche, and Joo Young Jeon. *Who Switched at the Big Switch and Why?* Centre for Competition Policy, 2014.

[Ward Jr, 1963] Joe H. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

[Widmer and Kubat, 1996] Gerhard Widmer and Miroslav Kubat. Learning in the presence of concept drift and hidden contexts. *Machine Learning*, 23(1):69–101, 1996.

[Williams *et al.*, 2012] James H. Williams, Andrew DeBenedictis, Rebecca Ghanadan, Amber Mahone, Jack Moore, William R. Morrow, Snuller Price, and Margaret S. Torn. The technology path to deep greenhouse gas emissions cuts by 2050: the pivotal role of electricity. *science*, 335(6064):53–59, 2012.

[Williams, 1988] Michael A. Williams. An empirical test of cooperative game solution concepts. *Systems Research and Behavioral Science*, 33(3):224–237, 1988.

[Xu and Mannor, 2009] Huan Xu and Shie Mannor. Parametric regret in uncertain Markov decision processes. In *48th IEEE Conference on Decision and Control*, pages 3606–3613, Shanghai, 2009.