# Towards Robust Multi-objective Optimization Under Model Uncertainty for Energy Conservation

Jun-young Kwak, Pradeep Varakantham*, Rajiv Maheswaran, Milind Tambe,
Timothy Hayes, Wendy Wood, Burcin Becerik-Gerber
University of Southern California, Los Angeles, CA, 90089
*Singapore Management University, Singapore, 178902
{junyounk,maheswar,tambe,hayest,wendy.wood,becerik}@usc.edu,
*pradeepv@smu.edu.sg

## ABSTRACT

Sustainable energy domains have become extremely important due to the significant growth in energy usage. Building multiagent systems for real-world energy applications raises several research challenges regarding scalability, optimizing multiple competing objectives, model uncertainty, and complexity in deploying the system. Motivated by these challenges, this paper proposes a new approach to effectively conserve building energy. This work contributes to a very new area that requires considering large-scale multi-objective optimization as well as uncertainty over occupant preferences when negotiating energy reduction. There are three major contributions. We (i) develop a new method called HRMM to compute robust solutions in practical situations; (ii) experimentally show that obtained strategies from HRMM converge to near-optimal solutions; and (iii) provide a systematic way to tightly incorporate the insights from human subject studies into our computational model and algorithms. The HRMM method is verified in a validated simulation testbed in terms of energy savings and comfort levels of occupants.

## Categories and Subject Descriptors

I.2.11 [**ARTIFICIAL INTELLIGENCE**]: Distributed Artificial Intelligence—*Multiagent systems*

## General Terms

Algorithms, Experimentation, Human Factors

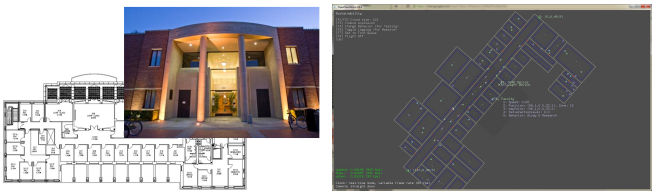
## Keywords

Energy, Sustainable Multiagent System, Multi-objective Optimization, Model Uncertainty, Human Subject Study

## 1. INTRODUCTION

The rapid growth in energy usage from commercial buildings in the U.S. has made the need for systems that aid in reducing energy consumption a top priority. Commercial buildings in the U.S. spent 18.5 QBtu in 2008, representing 46.2% of building energy consumption and 18.4% of U.S. energy consumption [1]. To that end, this work studies an innovative multiagent system to conserve building energy, specifically focusing on new algorithms to be deployed in commercial buildings (e.g., Ralph & Goldy Lewis Hall (RGL) at the University of Southern California (Figure 1(a))).

The purpose of a sustainable energy system is to efficiently conserve energy in the real-world, which raises three major technical research challenges. First, there are inherently multiple competing objectives like limited energy supplies, demands to satisfy occupants' comfort levels, and additional costs to maintain the system. This makes the problem harder as we need to explicitly consider multi-objective optimization techniques. Second, in such a complex domain, precisely knowing the world model is very challenging. Furthermore, as human occupants are directly involved in the optimization procedures, we must understand human behavior models and simultaneously reason about such model uncertainty in the domain. In addition, we should address novel scenarios that require agents to negotiate with groups of building occupants to conserve energy; previous work has typically focused on agents' negotiation with individual occupants in residential buildings. Third, this system should actually be deployed and verified in real testbed buildings, which adds another layer of complexity.



(a) Educational Building (RGL) (b) Simulation Testbed

**Figure 1: The actual research testbed at USC and our simulator**

Researchers have been developing multiagent systems to conserve energy for deployment in smart grids and buildings [8, 10, 12, 17, 18]. However, their work has been done with a particular focus on residential buildings and has not considered the combination of the above research challenges in sustainable energy domains. In addition, although human occupants in commercial buildings play a key role to effectively save energy, there has been little effort to conduct human subject studies and tightly incorporate fundamental understandings regarding human behaviors into the computational models. To overcome the weaknesses of prior work, we have proposed a new approach to efficiently compute robust policies for sustainable energy problems. In particular, we provide three major contributions in this paper. First, we develop a new method called HRMM (*Heuristics for Robust Multi-objective optimization under Model uncertainty*) to compute the MINIMAX and MINAVG strategies for *Bounded-parameter Multi-objective Markov Decision Problems* (BM-MDPs) that were proposed in [9] to optimize multiple competing objectives under uncertainty. Second, the MINIMAX and MINAVG strategies are experimentally shown to converge, which gives an insight into finding the solution bounds. Third, we provide a systematic way to tightly connect human subject studies from social psychology and our computation model and algorithm. Specifically, we suggest a new modeling method

to understand the underlying human behavior models for negotiations and construct more refined models for an input to BM-MDPs, which are used to compute robust strategies. We show that the generated solutions from HRMM substantially reduce the overall energy consumption compared to the existing control method while achieving comparable average comfort levels for occupants.

In Section 2, we describe our domain problems and testbeds. In Section 3, we describe a general problem formulation to achieve the desired goal in our domain, and detailed approaches are presented in Section 4. Section 5 provides evaluations and discussions.

## 2. MOTIVATING DOMAIN & TESTBEDS

Jointly performed with the university facility management team, this research is based on actual occupant preferences and schedules, actual energy consumption and loss data, real sensors and hand-held devices, etc. Figure 1(a) shows one of the real testbed buildings (Ralph & Goldy Lewis Hall) in which this work is to be deployed and the floor plan of the $3^{rd}$ floor. This campus building has three floors in total and is composed of classrooms, offices for faculty/staff, and conference rooms for meetings. Each floor has a large number of rooms and zones (a set of rooms that is controlled by specific equipment). The building includes components such as HVAC (Heating, Ventilating, and Air Conditioning) systems, lighting systems, office electronic devices like computers and AV equipment, and human occupants are classified either permanent (faculty, staff, researchers, etc.) or temporary (students or faculty attending classes/meetings, etc.).

In this domain, there are two types of energy-related occupant behaviors that this work can influence to conserve energy use: individual and group behaviors. Individual behaviors only affect an environment where the individual is located, and group behaviors lead to changes in shared spaces and require negotiation with a group of occupants.

As an important first step in deploying this work in the actual building, we have constructed a realistic simulation testbed (Figure 1(b)) based on the open-source project OpenSteer [1] and validated the simulation testbed using real building energy and occupancy data. This validated simulation environment has been used to evaluate novel models and algorithms in terms of energy savings and occupants' comfort levels. Readers are referred to [9] for a more detailed description regarding the testbed.

## 3. BACKGROUND & PROBLEM STATEMENT

The desired goal in this work is to optimize multiple criteria, i.e., achieve maximum energy savings without sacrificing the comfort level of occupants in commercial buildings. This objective is achieved by two types of agents: room agents and proxy agents (Figure 2). There is a dedicated room agent per office and conference room, in charge of reducing energy consumption in that room, focusing on group negotiations with occupants. A proxy agent [19, 20] is on an individual occupant's hand-held device and it has the corresponding occupant's models. Proxy agents communicate on behalf of an occupant to the room agent based on their adjustable autonomy — when to interrupt a user and when to act autonomously.

The room agent is responsible for planning simple and complex tasks. These tasks include negotiating with groups of individuals to relocate meetings to smaller rooms to save energy, negotiating with multiple occupants of a shared office to reduce energy usage

[1]http://opensteer.sourceforge.net/

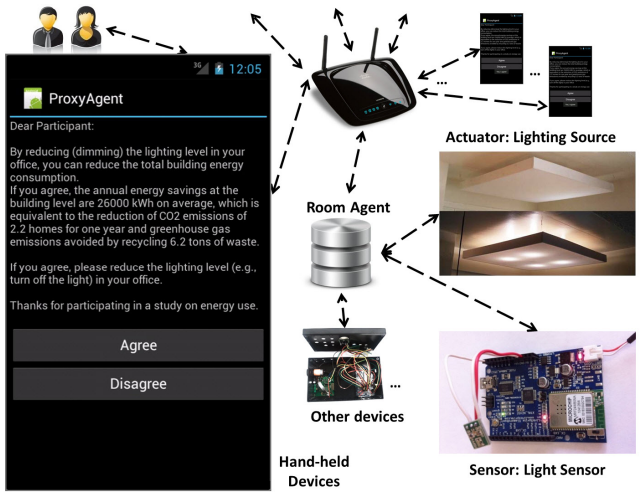



**Figure 2: Agents & Communication Equipment. An agent sends feedback including energy use to occupants.**

in the form of lights or HVACs, and others. We model this planning problem using BM-MDPs [9], which are a hybrid of *Multi Objective MDPs* (MO-MDPs) [5, 13] and *Bounded MDPs* (BMDPs) [7]. To achieve the given goal, BM-MDPs must reason with multiple objectives, but simultaneously must reason with given model uncertainty, as precisely knowing the model is very challenging in a complex domain. Before explaining BM-MDPs, we first briefly explain MO-MDPs on which BM-MDPs are built.

The negotiation scenarios require us to consider multiple objectives simultaneously: energy consumption and the comfort level of multiple individuals. To handle multiple objectives, MDPs have been extended to take into account multiple criteria assuming no *model uncertainty*. MO-MDPs are defined as an MDP where the reward function has been replaced by a vector of rewards. Specifically, MO-MDPs are described by a tuple $\langle S, A, T, \{R_i\}, p\rangle$, where $S$ is a finite set of world states, $A$ is the finite set of actions of an agent, $T$ is the transition function, $R_i$ is the reward function for objective $i$ and $p$ denotes the starting state distribution ($p(s) \geq 0$). In particular, the multiple reward functions, $\{R_i\}$, include energy consumption (e.g., the reduction in energy usage in moving from a conference room to a smaller office), and comfort level defined separately for each individual based on her/his preferences. The key principle we rely on, given the current domain of non-residential buildings is one of fairness; we wish to reduce energy usage, but we cannot sacrifice any one individual's comfort entirely in service of this goal. To meet this requirement, we focus on minimizing the *maximum regret* instead of maximizing the reward value based on a min-max optimization technique [14] to get a well-balanced solution across objectives under given model uncertainty.

BM-MDPs [9] now extend MO-MDPs to add capability to explicitly handle model uncertainty, and they are described by a tuple $\hat{M} = \langle S, A, \hat{T}, \{\hat{R}_i\}, p\rangle$, where $\hat{R}_i$ represents the reward function for objective $i$. The transition ($\hat{T}$) and reward functions ($\{\hat{R}_i\}$) in BM-MDPs have closed real intervals similar to BMDPs, which are limited to optimizing a single objective case (i.e., the BMDP model requires one unified reward function).

In BM-MDPs, to minimize the maximum regret, we first need to compute the optimal value for each objective $i$ using the MDP framework relying on the following formulation:

$$\min V_i^*(s) \tag{1}$$

$$\text{s.t.} \quad V_i^*(s) \geq \hat{R}_i(s,a) + \gamma \sum_{s' \in S} \hat{T}(s,a,s') \cdot V_i^*(s'), \tag{2}$$

$$0 \leq \gamma < 1, \tag{3}$$

where $V_i^*$ is an optimal value for objective $i$, and $\gamma$ is a discount factor.

We define the regret in BM-MDPs as follows:

**Definition** Let $H_i^\pi(s)$ be the *regret* with respect to a policy $\pi$ for objective $i$ and state $s$. Formally,

$$H_i^\pi(s) = V_i^{\pi_i^*}(s) - V_i^\pi(s), \tag{4}$$

where $V_i^{\pi_i^*}(s)$ is the value of the *optimal* policy, $\pi_i^*$, pre-calculated by (1)–(3) of the MDP formulation, and $V_i^\pi(s)$ is the value of the policy $\pi$ for objective $i$ and state $s$.

The objective is to find an optimal $\pi$, minimizing the maximum regret over all objectives given a noisy model $\hat{M}$, $U_{\hat{M}}^*(\pi)$.

$$\min_\pi U_{\hat{M}}^*(\pi), \tag{5}$$

$$\text{s.t.} \quad U_{\hat{M}}^*(\pi) \geq \sum_{s \in S} p(s) \cdot H_i^\pi(s), \forall i \in I, \tag{6}$$

$$V_i^\pi(s) = \sum_{a \in A} \pi(s,a) \left[ \hat{R}_i(s,a) + \gamma \sum_{s' \in S} \hat{T}(s,a,s') \cdot V_i(s') \right], \tag{7}$$

$$\sum_{a \in A} \pi(s,a) = 1, \forall s \in S, \tag{8}$$

$$0 \leq \gamma < 1, \tag{9}$$

where $I$ is a set of objectives, and $\pi$ is a randomized policy.

# 4. APPROACH

This work is driven by challenges of multi-objective optimization as well as model uncertainty, leading to two main ideas. First, we describe a novel approach to solve BM-MDPs to compute a robust and near-optimal policy. Second, we propose a new modeling method to build realistic human behavior models, which reduces the degree of model uncertainty in BM-MDPs so that we can obtain practical, applicable strategies to real-world situations.

## 4.1 Algorithm

Two previous heuristics have been proposed to compute optimistic and pessimistic BM-MDP policies in [9], however, those methods are limited to provide only two extreme solutions and they do not provide any solution bounds and/or show the convergence of computed solutions. To overcome these two limitations, we propose a new algorithm called HRMM based on the formulation described in Section 3, which provides a generic solution framework for BM-MDPs and empirical solution bounds. The HRMM method relies on two major features: i) sampling and ii) policy selection using cross-validation.

Algorithm 1 describes the overall flow of HRMM to solve BM-MDPs. The HRMM method is general and universal to solve BM-MDPs where each function in the algorithm can be independently replaced without affecting other parts.

We first generate BM-MDPs as an input considering refined human behavior models, which will be discussed in the next section

**Algorithm 1** HRMM

1: **BM-MDP** ← GETREFINEDMODEL()
2: {Considering model uncertainty and multiple objectives; This function is defined based on the math model considering more realistic human behavior functions described in Section 2}
3:
4: **for** $m = 1 \ldots N \in M$ **do**
5: {$M$ is a set of sampled models.}
6: $\mathbf{M}_m$ ← GETRANDOMMOMDPSAMPLE(**BM-MDP**)
7: $\pi_m$ ← SOLVEMOMDP($\mathbf{M}_m$)
8: {$\pi_m$ is an optimal policy computed from $\mathbf{M}_m$, a sampled MO-MDP model, based on the min-max formulation.}
9:
10: $\pi$ ← POLICYSELECTION($\{\pi\}, \{\mathbf{M}\}$)
11: {Given a set of sampled models and their corresponding policies, choose the final BM-MDP policy using various heuristics.}
12: **return** $\pi$

**Table 1: Cross Validation Matrix**

| | $M_1$ | $M_2$ | ... | $M_N$ |
|---|---|---|---|---|
| $\pi_1$ | $U_{M_1}^*(\pi_1)$ | $U_{M_2}^*(\pi_1)$ | ... | $U_{M_N}^*(\pi_1)$ |
| $\pi_2$ | $U_{M_1}^*(\pi_2)$ | $U_{M_2}^*(\pi_2)$ | ... | $U_{M_N}^*(\pi_1)$ |
| ... | | ... | | |
| $\pi_N$ | $U_{M_1}^*(\pi_N)$ | $U_{M_2}^*(\pi_N)$ | ... | $U_{M_N}^*(\pi_N)$ |

(line 1). We then randomly generate $N$ MO-MDP samples from the given BM-MDPs using a probability distribution over model uncertainty and solve each sampled model to compute policies based on the multi-objective optimization formulation presented in Section 3 (lines 4–8). For sampling, we assume a uniform distribution as a default option. In line 10, we compute an optimal BM-MDP policy $\pi$ using cross-validation, details will be presented below.

Now we focus on two main ideas in HRMM, i) sampling and ii) policy selection, and explain them in detail. To handle model uncertainty, we first sample $N$ MO-MDP models from the given BM-MDPs built upon such model uncertainty. In particular, for the reward functions, the value is randomly drawn from a given range based on the probability distribution over model uncertainty for each objective. The transition function is selected by a similar way based on the notion of *Order-Maximizing MDPs* [7], which selects the transition probabilities from the given intervals. More specifically, we randomly generate an order of states based on the probability distribution and take this order as an input to construct the transition function. To give some intuition behind this operation, we provide the following simple example to show how transition values are assigned from their intervals using the given order.

**Sampling Example** Consider a BM-MDP with three states: $s_1$, $s_2$, $s_3$. The transition ranges are $\hat{T}(s_1,a,s_1)$ = [0.2, 0.9], $\hat{T}(s_1,a,s_2)$ = [0.1, 0.3], $\hat{T}(s_1,a,s_3)$ = [0.2, 0.8]. Let us assume that the given order of states is $s_2$–$s_3$–$s_1$. The high-level idea is that we require movement to $s_2$ as much as possible within the given range of transition probability, and $s_3$ next, and so forth. Therefore, the transition probabilities would be $T(s_1,a,s_2)$ = 0.3 and $T(s_1,a,s_3)$ = 0.5 because $T(s_1,a,s_1)$ should be at least 0.2, and $T(s_1,a,s_1)$ = (1 - 0.3 - 0.5).

For each sampled model $M_m$, we solve equations (5)–(9) with $M_m$ and compute an optimal policy $\pi_m$ for the given model. For each model, we iterate this procedure and construct the matrix as shown in Table 1. In Table 1, $U_{M_n}^*(\pi_m)$ is the maximum regret value when $\pi_m$ is evaluated against a model $M_n$.

The final step to compute a BM-MDP policy is policy-selection using Table 1. We provide a method to choose the MINIMAX and

MINAVG policies.

**MINIMAX:** MINIMAX finds an optimal policy, $\pi$, minimizing the worst-case maximum regret over models. Formally, we compute the policy as following:

$$\pi \leftarrow \arg\min_{m} \max_{n} U^{*}_{M_n}(\pi_m), \forall m, n \in M, \tag{10}$$

where $M$ is a set of sampled models.

**MINAVG:** MINAVG finds a policy, $\pi$, minimizing the average maximum regret over models. Thus, we consider the average performance of each policy and choose the best one among them.

$$\pi \leftarrow \arg\min_{m} \text{avg}_n U^{*}_{M_n}(\pi_m), \forall m, n \in M \tag{11}$$

## 4.2 Refined Human Behavior Modeling

One of the central challenges in a complex domain such as ours is to obtain/construct a correct model of the problem we try to address. More specifically, as human occupants actively engage in reasoning (e.g., negotiations to conserve energy in commercial buildings), understanding the underlying human behavior models becomes crucial. Thus, in this work, jointly performed with social psychologists, we propose a sophisticated modeling procedure to construct more realistic models of human behavior. We then leverage insights and expertise from these models as input to BM-MDPs.[2]

There is significant related literature in social psychology on understanding human behavior models with respect to negotiation processes [3, 4, 11, 16]. Among them, we specifically focus on *the effects of repeated exposure* (i.e., irritation factor), which fundamentally explains the phenomenon that agreement with advocacy would increase, then decrease as exposure frequency increases. There are multiple properties we need to consider regarding this effect including the complexity of stimuli, message exposure frequency, heterogeneity of the message, and the degree of learning (or recall), etc. Among those properties, the primary factor is the message complexity. In general, a complex message contains more heterogeneous and stronger arguments than a simple one, but is not necessarily lengthy.

The objective of this study is i) to understand what types of messages are most effective to affect occupants' energy-related decisions, and ii) to figure out the most effective means to convey those messages considering different occupants' preferences. This study can be investigated via a survey to collect responses of actual occupants. The analyzed results are eventually used to construct a refined BM-MDP model, which will be used as an input to compute an optimal policy in Algorithm 1.

We designed a survey to assess individual differences and preferences.[3] In this study, we measured participants' compliance rates to given energy suggestions within either a simple or complex message and corresponding comfort levels while varying lighting, temperature and meeting relocation/reschedule conditions in a commercial building. In particular, the survey is composed of four separate sections to assess individual characteristics, the lighting preference, the temperature preference, and the meeting relocation/reschedule preference. For each section, to measure preferences, the message type (either simple or complex) shown to subjects is randomly selected. For instance, while measuring the lighting preference, some participants are repeatedly exposed to a simple message (and others are exposed to a complex message) for a fixed number of times (e.g., if you reduce your lighting for the hours you work, the annual energy savings at the building level is equivalent to the reduction of $CO_2$ emissions for 1.6 homes for one year. Will you dim the lighting level in your office?) and the compliance rates and irritation levels caused by the provided messages are measured as message frequency increases. Changes in their comfort level are also measured as the lighting level changes. Likewise, for the meeting relocation/reschedule preferences, some participants are repeatedly exposed to a complex message, including heterogeneous environmental motives and tailored individual energy rates (and again, others are exposed to a simple message). Their comfort level changes are also measured with various meeting relocation options, specifically focused on the degree of displacement (i.e., how much difference between the original meeting and new meeting in terms of location and time) and occupant density in the room (i.e., $\frac{\#\text{ of meeting participants}}{\text{Room size}}$). At the end of the survey, each participant is asked how much of the conveyed messages they are able to recall. Figure 3 shows an actual screenshot of the online survey we have conducted, and the preliminary results will be provided in the evaluation section.


**Figure 3: Survey Screenshot**

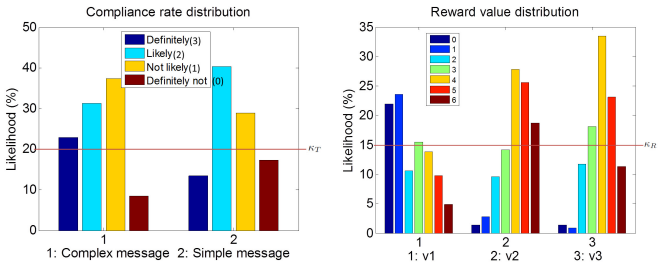
## 4.3 Constructing BM-MDPs

We now elaborate how to incorporate the survey outcomes into our computational model (i.e., BM-MDPs) for a more robust policy computation. In short, the measured compliance rate is used to construct the transition function ($\hat{T}$), and various comfort levels are incorporated into the reward function ($\{\hat{R}_i\}$). To give more concrete ideas on how to construct the BM-MDP model, we illustrate the following simple example.

Consider a meeting that has been scheduled in a medium meeting room (II in Figure 8(c)) that has more light sources and appliances than smaller offices. Assuming the meeting has few attendees, the

[2]Note that we only focus on modeling in this paper and leave the validation as future work.

[3]http://www-scf.usc.edu/ junyounk/survey2/index.php




(a) Compliance distribution (b) Reward value distribution

**Figure 4: Value distribution from the survey of the example scenario described in Section 4.3**

room agent will persuade an attendee ($P_1$) to relocate the meeting to nearby small, sunlit offices (III in Figure 8(c)), which can lead to significant energy savings. The room agent specifically provides the following simple suggestion repeatedly: "If you change your current meeting room to a smaller room, the average annual energy savings are equivalent to $CO_2$ emissions from 6.2 barrels of oil consumed. Will you change the meeting room (II) to a smaller room (III)?" Then, $P_1$ responds to the given energy suggestion to decide whether or not (s)he will agree to relocate the meeting. We assume that the above simple message was provided to $P_1$ two times on that day, the maximum capacity of a small office is 5, the current number of meeting attendees is 2 (including $P_1$), and the distance between a medium meeting room (II) and a small office (III) is less than 10 minutes by foot. The reward scale is [0, 100].

This situation is captured as follows: The current state ($s_1$) indicates a medium meeting room (II in Figure 8(c)), the message frequency is 2, the message complexity is simple and the suggestion type is the meeting relocation. The only change in the target state ($s_1'$) from $s_1$ is the meeting room change to a small office room (III in Figure 8(c)). The basic idea to choose a range for the transition probability and reward value is to select the minimum and maximum values that have a higher relative frequency than a given threshold ($\kappa_T$ for $T$ & $\kappa_R$ for $R$ [4]). In this example, we set $\kappa_T$ to 20% and $\kappa_R$ to 15%. Then, $P_1$'s transition probability of complying to the given suggestion ($\hat{T}(s_1, \text{Relocate}, s_1')$) is [0.33, 0.67] since the minimum and maximum values exceeding $\kappa_T$ are "Not likely" (raw value is 1 out of 3: 0.33) and "Likely" (raw value is 2 out of 3: 0.67) as shown in Figure 4(a). $P_1$'s reward function considers multiple factors: i) irritation level ($v_1$) by the provided message (Figure 9(c)), ii) comfort level change ($v_2$) according to occupant density of the target room (Figure 9(e)), and iii) comfort level change ($v_3$) by the degree of displacement (Figure 9(f)). Then, $v_1$ is [0.0, 50.0] since the minimum and maximum values exceeding $\kappa_R$ are 0 (out of 6: 0.0) and 3 (out of 6: 0.5) (see Figure 4(b)). $v_2$ is [66.67, 100.0] as the minimum value is 4 (out of 6: 0.67) and the maximum value is 6 (out of 6: 1.0) as shown in Figure 4(b). Similarly, $v_3$ is [50.0, 83.33] since the distance is less than 10 minutes by foot, the minimum and maximum values are 3 and 5 (out of 6), respectively as shown in Figure 4(b). Assuming that $P_1$'s weight values over these factors are 20%, 50%, and 30%, $\hat{R}(s_1, \text{Relocate}, s_1') = -0.2 * v_1 + 0.5 * v_2 + 0.3 * v_3 = [38.34, 74.99]$. In this way, we complete the baseline BM-MDP model considering all possible states and actions. For future work, we will consider machine learning techniques to refine the constructed model further based on real-world signals.

[4]Since the transition and reward functions often have different scales, we define a separate threshold parameter for each function. However, we can set the same value for both parameters for the convenience.

# 5. EVALUATION

In this section, we evaluate the performance of the HRMM method and experimentally show that its policies converge to near-optimal solutions. At the end of this section, we provide preliminary results in a validated simulation testbed (Figure 1(b)) and survey results that we have conducted. The experiments were run on Intel Core2 Duo 2.53GHz CPU with 4GB main memory. All techniques were evaluated for 100 independent trials throughout this section and we report the average values. For the sampling, we assumed a uniform distribution.

## 5.1 Evaluation of HRMM

We first compared the performance of the MINIMAX and MINAVG policies with other competitors' policies. Figure 5 shows the average maximum regret on the y-axis of different strategies. As we plot the regret, the lower value on the y-axis indicates better performance. Each policy was evaluated against 10000 samples. In the figure, BEST-AVG and WORST-AVG are the lowest and highest average maximum regret when all sampled policies are evaluated against 10000 samples, respectively. Assuming we do not explicitly consider model uncertainty in the world (i.e., MO-MDPs), we get one single policy computed from the given MO-MDP model that is essentially identical to one of sampled models from BM-MDPs when the number of samples is large enough. Thus, the BEST-AVG and WORST-AVG policies indicate the best and worst possible MO-MDP policies when being evaluated in the real-world under such model uncertainty. We also compare with the standard MDP with a unified reward based on the weighted sum method [24]. The uniform weight distribution was applied to the weighted sum method.

As shown in Figure 5, both MINIMAX and MINAVG showed very good performance, which were close to the average ideal case that we can assume in the world under given model uncertainty. This means the strategies from HRMM outperform almost all MO-MDPs and the standard MDP. In addition, MINAVG showed better performance than MINIMAX *on average*, but MINIMAX had lower variance when being evaluated, which means it reliably reacts to the worst possible case.

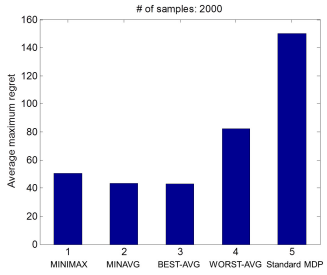
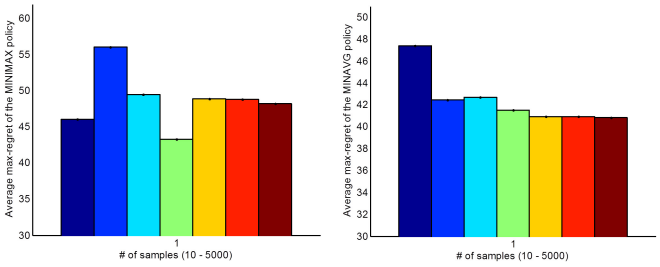



**Figure 5: Performance Comparison**

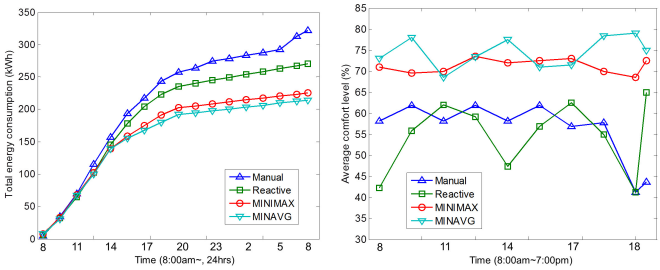
We then experimentally show that two strategies computed by HRMM converge as the number of samples increases. The y-axis in Figure 6 shows the average maximum regret of each MINIMAX and MINAVG policy, and the x-axis indicates the number of samples from 10 to 5000. Figure 6 shows that both policies converged when the number of samples increased (particularly from around 1000 samples). This result also confirms that MINAVG shows generally better performance than MINIMAX on average.

Lastly, we tested the MINIMAX and MINAVG policies in terms of energy consumption (kWh) and average comfort level of occupants (%). Figure 7(a) shows that the cumulative total energy consumption on the y-axis in kWh measured during 24 hours for all control strategies and time on the x-axis. We report the average total energy consumption measured over 30 sample weekdays throughout different seasons (3 weekdays in 2011 Spring, 10 weekdays in




(a) Convergence: MINIMAX (b) Convergence: MINAVG

**Figure 6: Convergence Results**




(a) Total Energy Consumption (b) Average Comfort Level

**Figure 7: Performance Evaluation**

2011 Summer, 17 weekdays in 2011 Fall). In the figure, the manual strategy represents the current strategy operated by the facility management team in RGL. The reactive strategy has an additional automatic operation feature based on the presence of occupants, which can be easily implemented using cheap sensors in the real building.

The MINIMAX achieved energy savings of 30.03% with an actually measured compliance rate (68.18%) from [9] and up to 41.35% with the ideal compliance rate (i.e., occupants always accept the suggestions provided by the room agents) when compared to the manual control strategy, and 16.64% compared to the reactive control strategy. On the other hand, the MINAVG achieved energy savings of 33.44% and 20.71% compared to the manual and reactive control strategies, respectively.

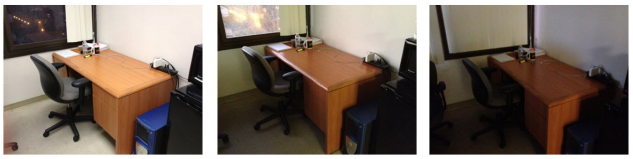
In addition to energy savings, we compared the average comfort level of human occupants under different control strategies in the simulation testbed. Figure 7(b) shows the average comfort level in percentage on the y-axis and time on the x-axis. As shown in the figure, both MINIMAX and MINAVG reliably showed higher average comfort level (about 70% or higher) than other control strategies as it plans ahead of the schedules considering uncertainty using BM-MDP policies.

## 5.2 Preliminary Survey Results

In this section, we provide preliminary results from the survey designed as described in Section 4.2. While conducting the survey, we provided concrete contextual information as shown in Figure 8. From this experiment, we answer the following questions by comparing change in energy behavior patterns and corresponding comfort levels.

HYPOTHESIS 1. *As exposure frequency increases, more complex feedback will shift "the effects of repeated exposure" and simultaneously lead to higher compliance rates to energy conservation suggestions than simpler feedback.*

HYPOTHESIS 2. *As exposure frequency increases, more complex feedback will make the message less irritating and thus mitigate the overall irritation level compared to simpler feedback.*



I. Maximum lighting level II. Medium lighting level III. Minimum lighting level

(a) Lighting preference

I. 60-70F II. 70-80F III. 80-90F

(b) Temperature preference

I. Large conference room II. Medium meeting room III. Small office

(c) Meeting relocation preference

**Figure 8: Survey Test Conditions**

We tested the hypotheses above as follows: we first recruited 220 participants on Amazon Mechanical Turk who work or have worked in an office (either at a company or school) regularly. We conducted this study for one week in the spring of 2012 and collected data from human subjects.

Table 2 shows that the average compliance rates to energy suggestions regarding the lighting, temperature, and meeting relocation/reschedule conditions. In the table, for the lighting condition, I→II represents the message suggesting that occupants dim the lighting level from the maximum level to the medium level and II→III indicates from the medium level to the minimum level. For the temperature condition, we provided the message to assess the likelihood of adjusting the desired temperature to a higher-level. Similarly, we measured how willing occupants are to relocate a meeting from a large conference room to a smaller office for the meeting preference condition. The values in the parentheses indicate the ratio of people who are exposed to specific type of message (simple vs complex).

The results shown in the table did not strongly support Hypotheses 1 as the overall compliance rate according to different types of messages did not show a significant difference except for some cases in lighting level and meeting relocation. However, in general, with the complex message, participants indicate stronger intentions when they agree to the given suggestions. In addition, for the meeting relocation, subjects agreed to relocate the meeting to a smaller space with a high compliance rate (about 78%) without suffering from a high degree of irritation, which means that there is a huge potential to effectively save energy use in buildings by leveraging this type of energy suggestion. As we described before, these com-

**Table 2: Compliance Rates (%)**

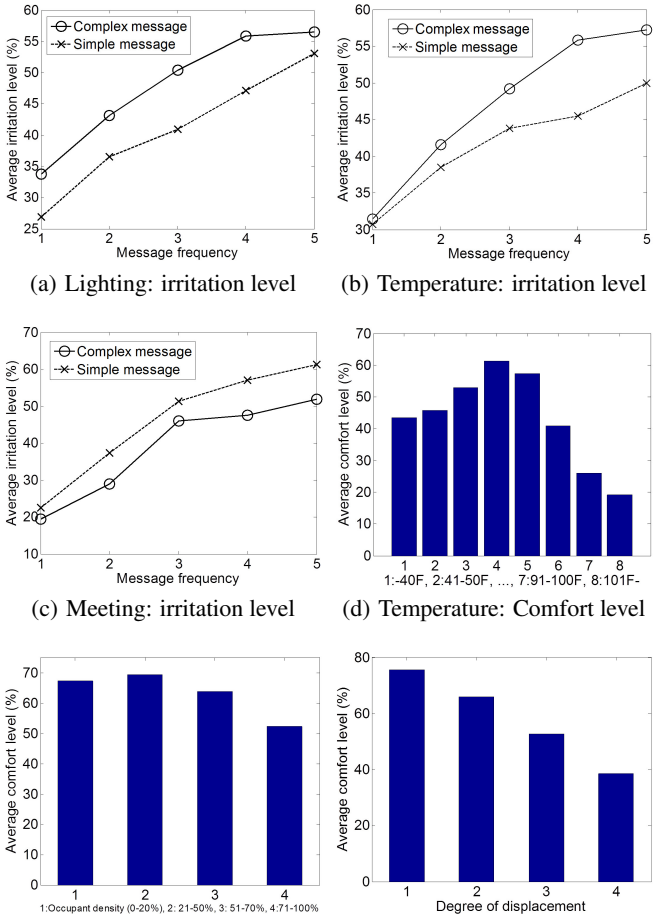
| | Lighting | | Temperature | | Meeting | |
|---|---|---|---|---|---|---|
| | C (49.0) | S (51.0) | C (52.8) | S (47.2) | C (43.6) | S (56.4) |
| I→II | 60.22 | 61.83 | 59.62 | 61.21 | 77.89 | 78.32 |
| II→III | 47.62 | 40.51 | 51.44 | 52.7 | 56.22 | 50.0 |

(C: complex message, S: simple message, I, II, and III are test conditions as shown in Figure 8.)

pliance values are directly used to construct the transition function.

Figure 9 shows comfort levels measured under various conditions, which includes irritation level change while repeating the energy suggestion message to human subjects. Figures 9(a)–9(c) show the average irritation level on the y-axis while varying the message frequency (x-axis). In the graph, the lower irritation level indicates the better result. As shown in these irritation results, as claimed in Hypothesis 2, the complex message either led to a lower irritation level or converged faster than a simple message, which means that a simple message might reach much higher irritation levels as we repeat the message further. On the other hand, Figures 9(d)–9(f) show the average comfort level (y-axis) while varying current conditions on the x-axis.[5] These results give more concrete ideas about how to model human occupants' comfort level change while handling various types of energy suggestions. Specifically, these outcomes are incorporated into the reward function.

Although we only provided a subset of results from the survey in this paper, there are many potential hypotheses we can verify to explain more interesting human behavior models. We use this data to construct a refined baseline human behavior model for computing robust strategies in BM-MDPs.




(a) Lighting: irritation level (b) Temperature: irritation level

(c) Meeting: irritation level (d) Temperature: Comfort level

(e) Meeting: Comfort lv. vs. Occupant density (f) Meeting: Comfort lv. vs. Degree of displacement

**Figure 9: Survey Outcomes**

## 6. RELATED WORK

In discussing related work, a key point we wish to emphasize is the uniqueness of our work in combining research on multiagent systems. Specifically, our BM-MDP algorithm handles uncertainty, and negotiations with human subjects, in an innovative application for energy savings. It is this specific combination of attributes that sets this work apart from previous research.

**Multiagent Energy Systems:** Multiagent systems have been considered to provide sustainable energy for buildings and smart grid management. Miller *et al.* [12] investigated how the optimal dispatch problem in the smart grid can be framed as a decentralized agent-based coordination problem and presented a novel decentralized message passing algorithm. Their work was empirically evaluated in large networks using real distribution network data. In addition, [8] addressed research challenges to integrate plug-in Electric Vehicles (EVs) into the smart grid.

To model and optimize building energy consumption, Mamidi *et al.* [10] developed smart sensing and adaptive energy management agents to decrease energy consumptions by HVACs in buildings. They showed that in the educational building, these sensor agents can be used to accurately estimate the number of occupants in each room and predict future occupancy relying on machine learning to intelligently control HVAC systems. Ramchurn *et al.* [17] considered more complex deferrable loads and managing comfort in the residential buildings. Rogers *et al.* [18] addressed the challenge of adaptively controlling a home heating system in order to minimize cost and carbon emissions within a smart grid using Gaussian processes to predict the environmental parameters. Our domain is different in focusing on energy savings in commercial buildings, and the representation and approaches are also different from previous work by allowing consumers (i.e., occupants) to play a part in optimizing the operation in the building instead of managing the optimal demand on buildings.

**Negotiation and Social Influence in Human Subject Studies:** Wood and Neal [23] have studied the potential of interventions to reduce energy consumption and they have shown that it is not only to change workplace energy consumption but also to establish energy use habits that maintain over time. Abrahmase *et al.* [2] also reviewed 38 interventions aimed to reduce household energy consumption, and they concluded that normative feedback about energy use is the most promising strategy for reducing and maintaining low consumption.

In social psychology, there has been a significant deal of work to figure out the correlation between irritation/distraction factors and persuasion. McCullough and Ostrom [11] and Cacioppo and Petty [4] discussed that message repetition would increase positive attitudes in a situation where highly similar communications are used and showed that there is a positive relationship between the number of presentations and attitude from general social psychology perspectives. Focusing on a commercial advertisement,

[5]The x-axis in Figure 9(f) indicates the following conditions:
1: Less than 5 mins walking distance or less than 30 mins shift
2: Less than 10 mins walking distance or less than 1 hr shift
3: Less than 15 mins walking distance or less than 2 hrs shift
4: Less than 30 mins walking distance or less than 4 hrs shift

Pechmann and Stewart [16] predicted the effectiveness of different strategies on advertising and examined the effects of message repetition on attitude changes. In addition, Baron *et al.* [3] discussed that distractions affect behavior decisions, but they are more or less effective in increasing persuasion depending upon whether people can easily ignore the distraction.

We leverage lessons and insights from social psychology in understanding and designing reliable and accurate human behavior models to compute robust strategies in the real-world.

**Multi-objective Optimization Techniques:** There has been a significant amount of work done on multi-objective optimization. The most common approaches to multi-objective optimization are to find Pareto optimal solutions [15], use the weighted sum method to aggregate multiple objectives using a prior preference [24], or consider the weighted min-max (or *Tchebycheff*) formulation that provides a nice theoretical property in terms of sufficient/necessary conditions for Pareto optimality [14].

Chatterjee *et al.* [5] considered MDPs with multiple discounted reward objectives. They theoretically analyzed the complexity of the proposed approach and showed that the Pareto curve can be approximated in polynomial time. Wiering and Jong [22] described a novel algorithm to compute Pareto optimal policies for deterministic multi-objective sequential decision problems. Authors proved that the algorithm converges to the Pareto optimal set of value functions and policies for deterministic infinite horizon discounted multi-objective Markov decision processes. Ogryczak *et al.* [13] focused on finding a compromise solution in multi-objective MDPs for a well-balanced solution. They compared their approach relying on the Tchebycheff scalarizing function to the weighted sum method. On the other hand, there has been some significant advances to handle model uncertainty on standard MDPs including [6, 7]. Recently, Soh and Demiris [21] extended the previous work and considered the multiple-reward POMDPs. They presented two hybrid multi-objective evolutionary algorithms that generate non-dominated sets of policies. Our work is different from them as we assume model uncertainty while simultaneously optimizing multiple criteria in MDPs.

# 7. CONCLUSION

In this work, we presented a new approach to conserve energy in commercial buildings via providing a robust strategy. There are several key contributions. We (i) developed a new method called HRMM to compute robust solutions in practical situations; (ii) experimentally showed that obtained strategies from HRMM converge to near-optimal solutions; and (iii) provided a systematic way to tightly incorporate the insights from human subject studies into our computational model and algorithms. We showed that the generated solutions from HRMM substantially reduce the overall energy consumption compared to the existing control method while achieving comparable average comfort levels for occupants.

# 8. REFERENCES


[1] *Buildings Energy Data Book*. U.S. Dept. of Energy, 2010.

[2] W. Abrahmase, L. Steg, C. Vlek, and T. Rothengatter. A review of intervention studies aimed at household energy conservation. *J Environ. Psychol.*, 25:273–291, 2005.

[3] R. Baron, P. Baron, and N. Miller. The relation between distraction and persuasion. *Psychological Bulletin*, 80(4):310, 1973.

[4] J. Cacioppo and R. Petty. Effects of message repetition on argument processing, recall, and persuasion. *Basic and Applied Social Psychology*, 10(1):3–12, 1989.

[5] K. Chatterjee, R. Majumdar, and T. A. Henzinger. Markov decision processes with multiple objectives. In *STACS*, 2006.

[6] K. V. Delgado, S. Sanner, L. N. de Barros, and F. G. Cozman. Efficient solutions to factored MDPs with imprecise transition probabilities. In *AAAI*, 2009.

[7] R. Givan, S. Leach, and T. Dean. Bounded-parameter Markov decision processes. *Artificial Intelligence*, 2000.

[8] S. Kamboj, W. Kempton, and K. S. Decker. Deploying power grid-integrated electric vehicles as a multi-agent system. In *AAMAS*, 2011.

[9] J. Kwak, P. Varakantham, R. Maheswaran, M. Tambe, F. Jazizadeh, G. Kavulya, L. Klein, B. Becerik-Gerber, T. Hayes, and W. Wood. SAVES: A sustainable multiagent application to conserve building energy considering occupants. In *AAMAS*, 2012.

[10] S. Mamidi, Y.-H. Chang, and R. Maheswaran. Improving building energy efficiency with a network of sensing, learning and prediction agents. In *AAMAS*, 2012.

[11] J. McCullough and T. Ostrom. Repetition of highly similar messages and attitude change. *Journal of Applied Psychology*, 59(3):395, 1974.

[12] S. Miller, S. D. Ramchurn, and A. Rogers. Optimal decentralised dispatch of embedded generation in the smart grid. In *AAMAS*, 2012.

[13] W. Ogryczak, P. Perny, and P. Weng. A compromise programming approach to multiobjective Markov decision processes. In *MCDM*, 2011.

[14] A. Osyczka. An approach to multicriterion optimization problems for engineering design. *Comput. Methods Appl. Mech. Eng.*, 15:309–333, 1978.

[15] V. Pareto. *Manuale di Economica Politica*. Societa Editrice Libraria, 1906.

[16] C. Pechmann and D. Stewart. Advertising repetition: A critical review of wearin and wearout. *Current issues and research in advertising*, 1988.

[17] S. D. Ramchurn, P. Vytelingum, A. Rogers, and N. R. Jennings. Agent-based control for decentralised demand side management in the smart grid. In *AAMAS*, 2011.

[18] A. Rogers, S. Maleki, S. Ghosh, and N. Jennings. Adaptive home heating control through Gaussian process prediction and mathematical programming. In *International Workshop on Agent Technology for Energy Systems (ATES)*, 2011.

[19] P. Scerri, D. Pynadath, L. Johnson, P. Rosenbloom, M. Si, N. Schurr, and M. Tambe. A prototype infrastructure for distributed robot-agent-person teams. In *AAMAS*, 2003.

[20] N. Schurr, J. Marecki, and M. Tambe. Improving adjustable autonomy strategies for time-critical domains. In *AAMAS*, 2009.

[21] H. Soh and Y. Demiris. Evolving policies for multi-reward partially observable markov decision processes (mr-pomdps). In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, pages 713–720. ACM, 2011.

[22] M. Wiering and E. De Jong. Computing optimal stationary policies for multi-objective markov decision processes. In *Approximate Dynamic Programming and Reinforcement Learning, 2007. ADPRL 2007. IEEE International Symposium on*, pages 158–165. IEEE, 2007.

[23] W. Wood and D. Neal. A new look at habits and the habit? goal interface. *Psychological Review*, 114:843–863, 2007.

[24] K. Yoon and C.-L. Hwang. *Multiple Attribute Decision Making, An Introduction*. Sage Publications, 1995.