2

Human Behavioral Models Using Utility Theory and Prospect Theory

Anuradha M. Annaswamy and Vineet Jagadeesan Nair

Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

2.1 Introduction

Analysis and synthesis of large-scale systems require the understanding of cyber physical human systems. Interactions between humans and cyber-components that interact with the physical system are varied and depend on a variety of factors and the goals of the large-scale systems. If the problem at hand concerns the behavior of the cyber–physical human systems (CPHS) under emergency conditions, the interactions between humans and automation need to focus on a shared control architecture (Erslan et al. 2023) with appropriate granularity of task allocation and timeline. Typically, the human in this context is an expert and when anomalies occur, either takes over control from automation or provides close supervision to the automation to ensure an overall safe CPHS. Under normal circumstances, the interactions may include other architectures. The role of the human is not necessarily that of an operator or an expert, but a user. The human may be a component in the loop, responding to outputs from the automation, and making decisions that serve in turn as inputs or reference signals r to the physical system (see Figure 2.1).

Typical examples of such interactions have begun to occur both in power grids and transportation and can be grouped under the rubric of transactive control. The transactive control concept (Chassin et al. 2004; Bejestani et al. 2014; Annaswamy and Nudell 2015) consists of a feedback loop resulting from incentives provided to consumers. Introduced in the context of smart grids, a typical transactive controller consists of an incentive signal sent to the consumer from the infrastructure and a feedback signal received from the consumer, and together the goal is to ensure that the underlying resources are optimally utilized. This introduces a feedback loop, where empowered consumers serve as actuators into an infrastructure, and transactive control represents a feedback control design that ensures that the goals of the infrastructure are realized, very similar to Figure 2.1. The use of transactive control in smart grids can be traced to homeostatic control proposed in Schweppe (1978) and Schweppe et al. (1980), which suggested that demand-side assets can be engaged using economic signals. Transactive control in this context has come to denote market-based control mechanisms that incentivize responsive loads and engage them in providing services to the grid where and when of great need (as described in Katipamula et al. (2006), Li et al. (2015), Hao et al. (2016), Somasundaram et al. (2014), Hammerstrom et al. (2008), Widergren et al. (2014), Melton (2015), Kok (2013), and Bernards et al. (2016)), and has the ability to close the loop by integrating customers via the right incentives to meet their local objectives such as lowering their electric bill as well as meeting global system objectives such as voltage and frequency regulation.

Cyber–Physical–Human Systems: Fundamentals and Applications, First Edition.
Edited by Anuradha M. Annaswamy, Pramod P. Khargonekar, Françoise Lamnabhi-Lagarrigue, and Sarah K. Spurgeon.

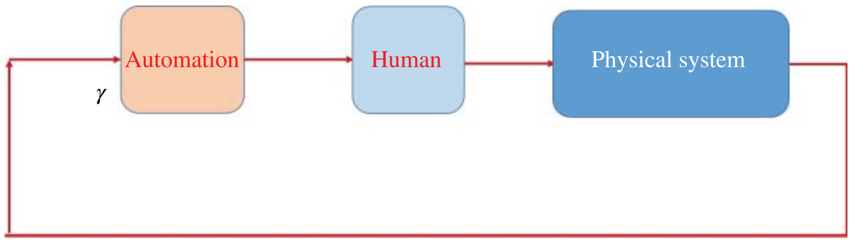

Figure 2.1 An Example of a CPHS with human-in-the–loop.

Another example of transactive control is in the context of congestion control in transportation (Phan et al. 2016; Annaswamy et al. 2018) – where dynamic toll pricing, determined by the controller, is used as an incentive signal to the drivers, who then decide whether or not to enter a tolled segment, thereby regulating traffic density and possibly alleviating congestion. In both examples, it is clear that an understanding of the human behavior is important, and the modeling of the overall socio-technical system that includes the human behavior and their interaction with the physical system is an important first step in designing the feedback controller.

This chapter focuses on two different tools that have been proposed for deriving behavioral models of human as a consumer. These tools include Utility Theory and Prospect Theory, and are described in Sections 2.2 and 2.3. Prospect Theory is a framework introduced by Nobel prize-winning behavioral economists and psychologists that has been extensively shown to better represent decision-making under uncertainty. It builds upon Utility Theory models by introducing additional nonlinear transformations on agents' objective utilities to model their irrational and subjective behaviors. Cumulative Prospect Theory (CPT) is an extension of Prospect Theory that also considers distortion of subjective probabilities through a weighting function. Examples are drawn from transportation to elucidate the impact of these tools, particularly in terms of predicting mode choice probabilities of passengers.

2.2 Utility Theory

The starting point for Utility Theory (Ben-Akiva et al. 1985) is the assignment of a value to an outcome in the form of a utility function. If in general there are N different possible choices, then U_i is utility of outcome $i, i = 1, \dots, N$. The benefit of the utility function is that it provides a substrate for modeling a human's decision when faced with these choices. In particular, the probability that the human will select choice ℓ is determined using the utility function as:

$$p^{\ell} = \frac{e^{U^{\ell}}}{\sum_{j=1}^{N} e^{U^{\ell}}} \quad \ell \in \{1, \dots, N\} \tag{2.1}$$

Equation (2.1) then serves as a simple behavioral model of the human which can be appropriately utilized in the overall problem of interest. For a simple problem with only two choices A and B, the probability p^A is given by

$$p^A = \frac{1}{1 + e^{-\Delta U}} \tag{2.2}$$

where $\Delta U = U^A - U^B$. Modeling the behavior of the human in the problem, whether as a consumer, an operator, a user, or an expert, the Utility Theory-based approach entails the characterization of all possible outcomes and the determination of the utility of each of the outcome i as U^i. A variety of factors contributes to the utility and hence determining the behavioral model in Equation (2.1) is nontrivial. Quantitative aspects related to economy, qualitative aspects such as comfort, and hard-to-identify aspects such as strategic behavior, negative externalities, global, and network-based expectations, are all factors that may need to be simultaneously accounted for. Nevertheless, Equation (2.1) serves as a cornerstone for many problems, where CPHS models have to be derived.

2.2.1 An Example

Suppose we consider a transportation example where a passenger intends to travel from points A to B and has two choices for transport: a shared ride service (SRS) and public transit. The utility function for this trip can be determined as:

$$u = \mathbf{a}^{\intercal}\,\mathbf{t} + b\gamma + c \tag{2.3}$$

where the components of $\mathbf{t} = [t_{walk}, t_{wait}, t_{ride}]^{\intercal}$ denote the walking, waiting, and riding times, respectively, γ denotes the ride tariff, $\mathbf{a} = [a_{walk}, a_{wait}, a_{ride}]^{\intercal}$ are suitable weights, and c denotes all other externalities that do not depend on either travel time or ride tariff. Both the weights $\mathbf{a}$ and tariff coefficient b are assumed to be negative since these represent disutilities to the rider arising from either longer travel times or higher prices, while c can be either positive or negative depending on the characteristics of the given travel option. All of the parameters $\mathbf{a}, b, c$ determine the behavior of a rider and could vary with time, the environment, or other factors. Equation (2.3) indicates that the choice of the rider of the SRS over other options like public transit is determined by whether $u(SRS) > u(\text{public transit})$ for a given ride.

A behavioral model of a driver as above was applied to the congestion control problem on a highway segment to determine a dynamic tolling price strategy (Ben-Akiva et al. 1985; Phan et al. 2016). It was shown that with the alternative u_0 corresponding to travel on a no-toll road, the toll price can be determined using a nonlinear proportional-integral (PI) controller using a socio-technical model that was a cascaded system with the behavioral model as in Equations (2.2) and (2.3) and an accumulator model of the traffic flow. Using actual data from a highway segment in the US city of Minneapolis, it was shown that such a model-based dynamic toll price leads to a much more efficient congestion control (Annaswamy et al. 2018). More recently, this approach has been extended to a more complex highway section with multiple merges and splits, and applied to data obtained from a highway section near Lisbon, Portugal (Lombardi et al. 2022). Here too, a behavioral model of the driver similar to Equation (2.3) was employed. The results obtained displayed a significant improvement compared to existing traffic flow conditions, with minimal changes to the toll price.

2.3 Prospect Theory

The behavioral model based on Utility Theory has two deficiencies. The first is that the utility function is embedded in a stochastic environment, causing the utility function model to be more complex than that considered in Equation (2.3). The second is that the model as considered in Equation (2.2) may not be adequate in capturing all aspects of decision-making of a human. Strategic decision-making, adjustments based upon the framing effect, loss aversion, and probability distortion are several key features related to subjective decision-making of individuals when facing uncertainty. It is in this context that *Prospect Theory* (Kahneman and Tversky 2012) in general, and *Cumulative Prospect Theory* (CPT) (Tversky and Kahneman 1992), in particular, provide an alternate tool that may be more appropriate. CPT builds upon Prospect Theory by using a probability weighting function to represent the agent's distortion of perceived probabilities of outcomes and uses these probability weights to compute the subjective utilities from subjective values. We briefly describe this tool below.

We first introduce a stochastic component into the problem and utilize the transportation example as the starting point. As travel times are subject to stochasticity, u becomes a random process. For simplicity, suppose we assume that there are only two possible travel time outcomes, $\overline{\mathbf{t}}$ and $\underline{\mathbf{t}}$ ($\underline{\mathbf{t}} \leq \overline{\mathbf{t}}$) having corresponding utilities $\underline{u}$ and $\overline{u}$ ($\underline{u} \leq \overline{u}$), occurring with probabilities of $p \in [0, 1]$ and $1 - p$, respectively. It follows that the utility function for the SRS is given by

$$\begin{aligned} \underline{u} &= \mathbf{a}_{sm}^{\intercal}\ \overline{\mathbf{t}} + b_{sm}\gamma_{sm} + c_{sm} \\ \overline{u} &= \mathbf{a}_{sm}^{\intercal}\ \underline{\mathbf{t}} + b_{sm}\gamma_{sm} + c_{sm} \end{aligned} \tag{2.4}$$

If these outcomes follow a Bernoulli distribution, its cumulative distribution function (CDF) is defined on the support $[\underline{u}, \overline{u}]$:

$$F_U(u) = \begin{cases} 0 & \text{if } u < \underline{u} \\ p & \text{if } \underline{u} \leq u < \overline{u} \\ 1 & \text{if } u \geq \overline{u} \end{cases} \tag{2.5}$$

Suppose the alternative choice has a utility function u_o. If $u_o \leq \underline{u}$, the customer would always choose the SRS since it offers strictly better outcomes and conversely if $u_o \geq \overline{u}$. For all other cases, the underlying model becomes a combination of Equations (2.2) and (2.5).

We now address the second deficiency in Utility Theory. Conventional Utility Theory postulates that consumers choose among travel options based on their respective expected utilities (Fishburn 1988; Von Neumann and Morgenstern 2007). Alternatively, random utility models are another framework within Utility Theory that predict choice probabilities based on the utilities of different alternatives (computed using logit models) without accounting for risk, by assuming certain distributions for unobserved factors and error terms. However, both of these are inadequate when there is significant uncertainty involved. Prospect Theory (PT) is an alternative to Utility Theory that better describes subjective human decision-making in the presence of uncertainty and risk (Tversky and Kahneman 1992; Kahneman and Tversky 2012), and CPT is a variant of PT that weighs different outcomes using distorted subjective probabilities as perceived by passengers. This is needed since individuals have been shown to consistently underestimate the likelihood of high probability outcomes while overestimating the likelihood of less likely events (Tversky and Kahneman 1992). To describe CPT, we introduce a value function $V(\cdot)$ and a probability distortion $\pi(\cdot)$ given by Guan et al. (2019a) and Prelec (1998), with $\pi(0) = 0$ and $\pi(1) = 1$ by definition. These nonlinearities map the objective utilities (u) and probabilities (p) of each possible outcome to subjective values, as perceived by the passengers. Note here that the probability weighting function $\pi(\cdot)$ as described in Equation (2.7) is unique to CPT. The graphs in Figure 2.2 show examples of how the value and probability weighting functions may vary according to the objective utility u and actual probability p, respectively.

$$V(u) = \begin{cases} (u - R)^{\beta^+} & \text{if } u \geq R \\ -\lambda(R - u)^{\beta^-} & \text{if } u < R \end{cases} \tag{2.6}$$

$$\pi(p) = e^{-(-ln(p))^{\alpha}} \tag{2.7}$$

The CPT parameters here describe loss aversion (λ), diminishing sensitivity in gains (β^+) and losses (β–), and probability distortion (α). The reference R is the baseline against which users compare uncertain prospects. These can vary across individuals and also depending on the particular

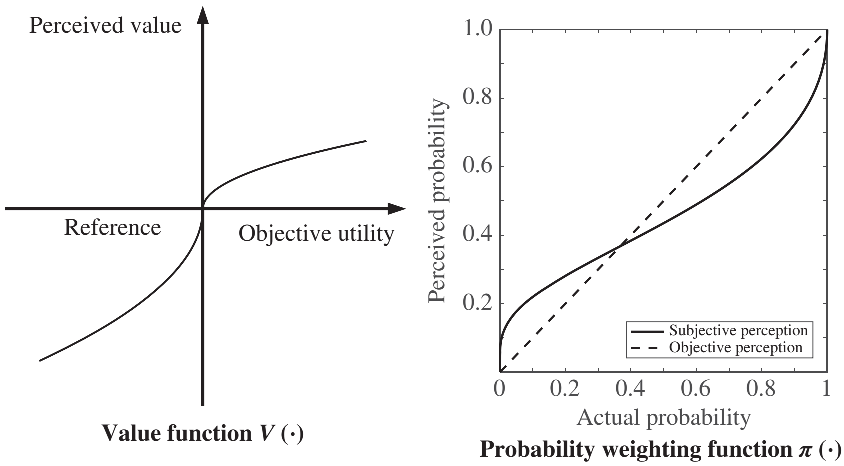

Figure 2.2 Illustrations of the CPT value function and probability weighting functions. Source: Adapted from (Guan et al. 2019a).

set of alternatives the customer is facing. Further details on the different possible types of references are provided in Section 2.3.2.

With the above distortions in the value function, the overall utility function for a given stochastic outcome gets modified. Note that the utility function U is now a random variable of the form

$$U = X + b\gamma = \mathbf{a}^{\intercal}\ \mathbf{T} + b\gamma \tag{2.8}$$

where the $b\gamma$ term is deterministic based on the SRS ride offer tariff and the random variable for the disutility (due to travel times) $X = \mathbf{a}^{\intercal}\ \mathbf{T}$ is stochastic and represents the uncertainty in travel times (walking, waiting, riding) $\mathbf{T} = [\underline{\mathbf{T}}, \overline{\mathbf{T}}]$. Similar to Equation (2.4), we can derive $X = [\underline{x}, \overline{x}]$ with CDF $F_X(x) = F_U(x + b\gamma)$, where $\underline{x}$ and $\overline{x}$ correspond to the shortest and longest travel times. These correspond to the best- and worst-case utilities $\overline{u}$ and $\underline{u}$, respectively, with $U = [\underline{u}, \overline{u}]$.

The Utility Theory-based derivation of the objective utility is as follows: if U takes on discrete values $u_i \in \mathbb{R}, \forall i \in \{1, \ \ldots, \ n\}$ and the outcomes are in ascending order, i.e. $u_1 < \cdots < u_n$, where $n \in \mathbb{Z}_{>0}$ is the number of possible outcomes, one can determine the objective utility U^o as the expectation of U according to Utility Theory as (Von Neumann and Morgenstern 2007), i.e.

$$U^o = \sum_{i=1}^{n} p_i u_i \tag{2.9}$$

where $p_i \in (0, 1)$ is the probability of outcome u_i, and $\sum_{i=1}^{n} p_i = 1$. In contrast, a Prospect Theory-based derivation of utility function includes the value function V rather than u_i and the probability distortion $\pi(.)$ is applied to the probabilities p_i. Suppose we define the corresponding utility function as U_R^s, which is the subjective utility perceived by the passenger according to Cumulative Prospect Theory, then

$$U_R^s = \sum_{i=1}^{n} w_i V(u_i) \tag{2.10}$$

where R denotes the reference corresponding to the framing effect mentioned above, and w_i denotes the weighting that represents the subjective perception of p_i. Suppose that k out of the n outcomes are losses, $0 \leq k \leq n, k \in \mathbb{Z}_{\geq 0}$, and the rest are nonlosses, i.e. $u_i < R$ if $1 \leq i \leq k$ and $u_i \geq R$

if $k < i \leq n$. We can then derive the subjective probability weights assigned by the decision-maker to each of these discrete outcomes from the cumulative distribution function of U given by $F_U(u)$, as follows:

$$w_i = \begin{cases} \pi\left[F_U(u_i)\right] - \pi\left[F_U(u_{i-1})\right], & \text{if } i \in [1, k] \text{ (losses)} \\ \pi\left[1 - F_U(u_{i-1})\right] - \pi\left[1 - F_U(u_i)\right], & \text{otherwise (non-losses)} \end{cases} \tag{2.11}$$

where we have assumed $F_U(u_0) = 0$ for ease of notation.

It is clear that in contrast to U^o, U_R^s is centered on R, loss aversion is captured by choosing $\lambda > 1$, and diminishing sensitivity by choosing $0 < \beta^+, \beta^- < 1$. The probability distortion is quantified by choosing $0 < \alpha < 1$. The extension from Equation (2.10) to the continuous case of U_R^s is

$$U_R^s = \int_{-\infty}^{R} V(u)\frac{d}{du}\left\{\pi\left[F_U(u)\right]\right\} du + \int_{R}^{\infty} V(u)\frac{d}{du}\left\{-\pi\left[1 - F_U(u)\right]\right\} du \tag{2.12}$$

With the above objective evaluation of a utility function as in Equation (2.9) and subjective evaluation as in Equation (2.10), one can now determine the probability of acceptance of an outcome as follows. As has been shown in Equation (2.1), the evaluation of the probability of acceptance of an outcome ℓ requires the utility of all alternate outcomes. Without loss of generality, suppose there are only two alternatives, with the objective and subjective utility of option $i \in \{1, 2\}$, given by U_i^0 and $U_{i_R}^s$, respectively. Then the objective probability of acceptance of choice 1 using Utility Theory is given by

$$p_1^0 = \frac{e^{U_1^0}}{e^{U_1^0} + e^{U_2^0}} \tag{2.13}$$

while the subjective probability of acceptance of option 1 using CPT is given by

$$p_{1_R}^s = \frac{e^{U_{1_R}^s}}{e^{U_{1_R}^s} + e^{U_{2_R}^s}} \tag{2.14}$$

2.3.1 An Example: CPT Modeling for SRS

We illustrate the Prospect Theory model using the transportation example considered above, extended to the case where a passenger now has three choices $i \in \{1, 2, 3\}$, (i) public transit like buses or the subway, (ii) using a shared ride pooling service (SRS), and (iii) another which may be an exclusive ride hailing service (such as UberX). The SRS has greater uncertainty in pick up, drop off, and travel times when compared to the UberX and transit alternatives due to the possibility of more passengers being added en route. Thus, both UberX and transit can be treated as certain prospects when compared to the SRS option, which has uncertain outcomes. CPT can then be used to model the passenger's risk preferences to predict their decision-making under such uncertainty. The discussions above show that a number of parameters related to the CPT framework have to be determined. These include $\alpha, \beta^+, \beta^-, \lambda$ defined in $V(\cdot)$ and $\pi(\cdot)$, which are in addition to the parameters $\mathbf{a}, b, c$ defined in (Equation (2.3)) associated with the travel times $t_{\text{walk}}, t_{\text{wait}}, t_{\text{ride}}$ and tariff coefficients, for all three travel modes. In order to estimate these parameters, we designed and conducted a comprehensive survey eliciting travel and passenger risk preferences from $N = 955$ respondents in the greater Boston metropolitan area (Jagadeesan Nair 2021). Note that the constant terms in the utility function, c_{UberX} and c_{SRS} were measured relative to public transit as a baseline, i.e. $c_{transit} = 0$.

Table 2.1 summarizes the mean values and standard deviations of the parameters that we estimated for the discrete mode choice model using maximum simulated likelihood estimation, along with their standard errors. We can also use the estimated travel time and price coefficients

Table 2.1 Parameters describing the discrete choice logit models for SRS and UberX.

Parameter	Mean	SE	SD	SE
a_{walk} [min^{-1}]	−0.0586	0.0053	0.1412	0.0079
a_{wait} [min^{-1}]	−0.0113	0.0182	0.1491	0.0356
$a_{ride,\ transit}$ [min^{-1}]	−0.0105	0.0013	0.0284	0.0017
$a_{ride,\ UberX}$ [min^{-1}]	−0.0086	0.0014	0.0058	0.0010
$a_{ride,\ SRS}$ [min^{-1}]	−0.0186	0.0013	0.0095	0.0007
b [$\$^{-1}$]	−0.0518	0.0050	0.0597	0.0042
c_{UberX}	−2.5926	0.1800	2.3034	0.1558
c_{SRS}	−2.2230	0.1497	1.8175	0.1530

Table 2.2 Value of time spent on different modes, obtained from the random parameters logit model.

Trip leg or mode	VOT (in $/h)
Walking	67.8702
Waiting	13.1480
Transit ride	12.1703
Exclusive ride hailing	9.9466
Pooled ride sharing	21.5549

to determine the passengers' value of time (VOT) spent on different modes. The value of time is defined as the extra tariff that a person would be willing to pay or cost incurred to save an additional unit of time, i.e. it measures the willingness to pay (WTP) for extra time savings. In absolute terms, the VOT spent on mode i can be calculated as the ratio between the marginal utilities of travel time and trip cost:

$$VOT_i = \frac{\frac{\partial U_i}{\partial t}}{\frac{\partial U_i}{\partial \gamma}} = \frac{a_i}{b_i} \tag{2.15}$$

The VOT parameters estimated from our survey are shown in Table 2.2. For the CPT model, estimated, we slightly modified the model to allow for different probability distortion effects in the gain and loss regimes. Thus, the modified weighting function is given by

$$\pi_{\pm}(p) = e^{-(-ln(p))^{\alpha_{\pm}}} \tag{2.16}$$

The CPT risk parameters were estimated using the method of certainty equivalents (Rieger et al. 2017; Wang and Zhao 2019), by presenting surveyed passengers with a series of chance scenarios asking them to choose a travel mode by comparing the uncertain or risky SRS versus the certain UberX and transit options. Nonlinear least squares curve fitting was then used to estimate the CPT parameter values. A schematic for the overall estimation process is shown in Figure 2.3. More details on the survey design and estimation procedures can be found in Jagadeesan Nair (2021).

Having estimated these parameters, we can compute the subjective or perceived value of the SRS outcomes, and the passenger's subjective probability of accepting the shared ride service offer using

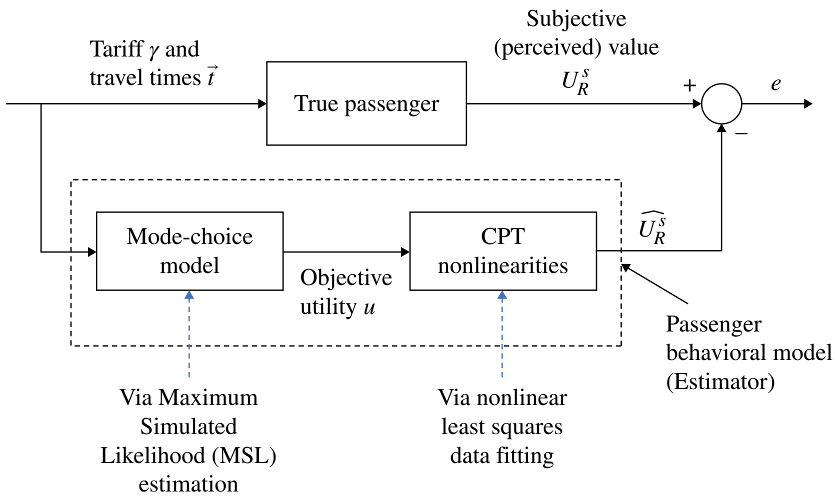

Figure 2.3 Estimation of mode choice models and CPT parameters from passenger survey data.

Equations (2.14)–(2.6). Since the CPT model better accounts for passengers' irrational preferences when faced with uncertainty and risk, this subjective acceptance probability is more accurate than what would be predicted using conventional Utility Theory alone.

2.3.1.1 Detection of CPT Effects via Lotteries

In addition to estimating the numerical values parametrizing the subjective value and probability weighting functions, we also used the survey responses to detect the key CPT effects, which are:

- **Framing effect**: Individuals value prospects with respect to a reference point instead of an absolute value, and perceive gains and losses differently.
- **Diminishing sensitivity**: In both gain and loss regimes, sensitivity diminishes when the prospect gets farther from the reference. Therefore, the perceived value is concave in the gain regime and convex for losses, implying that people are generally risk-averse in gains and risk-seeking in the loss regime.
- **Probability distortion**: Individuals overweight small probability events and underweight large probability events.
- **Loss aversion**: Individuals are affected much more by losses than gains.

In order to estimate these effects, we considered simplified choice scenarios involving monetary lotteries. Survey respondents were also asked a series of hypothetical lottery questions after completing the SRS travel choice scenarios. We can then test for the existence of CPT-like behaviors based on their lottery responses, as described in Rieger et al. (2017). More details on the survey and lottery scenarios can be found in Jagadeesan Nair (2021).

From Table 2.4, we see that the valid responses clearly display CPT effects. The reflection or framing effect is shown by nearly all the valid respondents, indicating that our proposed value function is likely an accurate descriptor of how the passengers perceive their gains and losses. The probability weighting effect is not as dominant, but it is still quite significant. We find that majority of them (>72%) show at least some overweighting of probabilities, and it is also most common in the lower probability ranges (between 10 and 60%). This agrees with CPT theory since it postulates that people tend to overestimate the likelihood of rare events. The relatively large value of the mean gain/loss ratio (>1) in the mixed lotteries indicates a significant degree of loss aversion

Table 2.3 Summary of CPT parameter estimates for SRS travel preferences.

	α^+	α^-	β^+	β^-	λ
Mean	0.4456	0.1315	0.2166	0.3550	20.0494
Median	0.4124	0.1320	0.2188	0.3649	11.8715
SD	0.1828	0.0448	0.0985	0.1906	25.8554

Table 2.4 Summary of key CPT effects observed from lotteries.

CPT effect tested	**% of valid responses**
Reflection effect (framing & diminishing sensitivity)	95.03
Probability overweighting between	
10% and 60% probability	62.56 %
60% and 90% probability	40.51 %
10% and 90% probability	51.05 %
Any probability weighting	72.44 %
Loss aversion	
Mean gain/loss ratio for mixed outcome lotteries	3.7254
Median gain/loss ratio for mixed outcome lotteries	1.0250

among the surveyed passengers. However, the median value is quite close to 1 indicating that loss aversion may not be as prevalent for a sizeable portion of the passengers sampled in this study. These results from Tables 2.3 and 2.4 demonstrate that we can use survey responses and data on passenger choices to (i) validate CPT behavioral effects and also (ii) estimate mathematical models and parameters to describe their risk preferences. Using real-time data from sources like ridesharing apps, these models can be continuously updated and improved in order to more accurately predict passenger choice and behavior.

2.3.2 Theoretical Implications of CPT

Before discussing the theoretical implications of CPT for SRS, we briefly outline the reference R which is one of the most important parameters for CPT modeling There are mainly two different types of references Wang and Zhao (2018):

1. ***Static reference***: These are fixed values independent of the SRS offer. For example, this could be the objective utility of the alternative option to the SRS, i.e. $R = U_A^0 = A^0$.
2. ***Dynamic reference***: These depend directly on the uncertain prospect (SRS offer) itself. For example, the reference could be $R = \tilde{x} + b\gamma$, where the travel time disutility $\tilde{x}$ may be set as the best case ($\underline{x}$), worst case ($\overline{x}$), or some value in between $\tilde{x} \sim F_X(x)$.

In Section 2.3.1, we determined mode choice and CPT behavioral models for passengers for a large population and over many different possible travel scenarios. In the following, we consider the case of a single trip (ride request and offer), in order to draw a few key insights about passenger risk preferences and travel behavior using computational experiments. Here, the passenger makes a choice between the SRS option (uncertain prospect) and an exclusive ride-hailing service like UberX (certain prospect). This section is adapted from our earlier work applying CPT-based dynamic pricing for SRS (see Guan et al. (2019a) for more details).

A dynamic routing problem of 16 passengers using real SRS request data from San Francisco was considered (Annaswamy et al. 2018), and the request from one of these passengers was utilized to run the computational experiments shown here. An alternating minimization-based algorithm developed in Guan et al. (2019b) was applied to determine the optimal routes and corresponding travel times of the SRS. The possible delays in travel time were constrained to be at most four minutes of extra wait and ride time, respectively. For each ride request, travel times and prices of the SRS option (UberX) were retrieved from Uber.[1]

Using the travel times and prices, along with utility coefficients from Table 2.1, the objective utility A^o of the alternative (UberX) and $\underline{x}$, $\overline{x}$ of the SRS are calculated, using Equation (2.4). Note that $A^o, \underline{x}, \overline{x}$ are negative as they represent disutilities incurred by the passenger due to travel times and tariffs. Using this numerical setup, we explore the three implications via simulations: (i) fourfold pattern of risk attitudes, (ii) strong aversion of mixed prospects, and (iii) self-reference. We use the following key properties of CPT-based behavioral models for our analysis; more details and derivations of these can be found in Guan et al. (2019a):

Property 1 is related to static and dynamic reference points as described above. These help decide the dynamic tariff γ that drives the subjective acceptance probability p_R^s to approach the desired probability of acceptance p^*. Defining the expected values as $\bar{U} = \mathbb{E}_{f_U}(u)$ and $\overline{X} = \mathbb{E}_{f_X}(x)$, Properties 2 and 3 are related to the subjective utility and subjective acceptance probability when using $R = \bar{U}$, i.e. $U_{\bar{U}}^s$ and $p_{\bar{U}}^s$, respectively, Guan et al. (2019a).

1. Given any static reference point $R \in \mathbb{R}$ or dynamic reference point of the form of $R = \tilde{x} + b\gamma,\ \tilde{x} \in \mathbb{R}$, p_R^s strictly decreases with the tariff γ.
2. Given any uncertain prospect, there exists a λ^*, such that $\forall \lambda > \lambda^*$, $U_{\bar{U}}^s < 0$.
3. For any uncertain prospect, given that λ is sufficiently large such that $U_{\bar{U}}^s < 0$, and within the price range $\gamma \in [\underline{\gamma}, \overline{\gamma})$, where $\underline{\gamma}$ satisfies $\overline{X} + b\underline{\gamma} = A^o$, and $\overline{\gamma}$ satisfies $\left[A^o - (\overline{X} + b\overline{\gamma})\right]^{\beta^+} - U_{\bar{U}}^s = A^o - (\overline{X} + b\overline{\gamma})$, $p_{\bar{U}}^s < p^o$. In the SRS context, this implies that under the stated conditions, the uncertain SRS is relatively less attractive to the passenger than the alternative A in terms of their probability of accepting the ride offer.

All of the above properties do agree with our intuition, but they allow us to formally and precisely express these ideas to provide a rigorous mathematical framework. In Sections 2.3.2.1–2.3.2.3, we use these along with the previously defined CPT-based behavioral model to quantify the theoretical consequences of CPT for SRS applications.

2.3.2.1 Implication I: Fourfold Pattern of Risk Attitudes

The fourfold pattern of risk attitudes is regarded as "the most distinctive implication of Prospect Theory" (Tversky and Kahneman 1992), which states that while facing an uncertain prospect, individual risk attitudes can be classified into four categories:

1. Risk averse over high probability gains.
2. Risk seeking over high probability losses.
3. Risk seeking over low probability gains.
4. Risk averse over low probability losses.

1 https://www.uber.com/.

These risk attitudes are often used to justify the subjective decision-making of individuals for application such as SRS as well as other problems such as playing lotteries (as described in Section 2.3.1.1) and getting insurance coverage, among others.

We now illustrate the fourfold pattern in the SRS context using the following scenario, which uses the classic setup for the analysis of the fourfold pattern (Tversky and Kahneman 1992): individuals decide between two options, a certain prospect and an uncertain prospect with two outcomes. The uncertain prospect is the SRS (i.e. UberX), which we assume obeys a truncated Poisson distribution with $K = 1$, i.e. the passenger is subjected to at most one delay. Therefore, the two possible SRS outcomes are $(\underline{x} + b\gamma)$ and $(\overline{x} + b\gamma)$, with the following corresponding probabilities derived from the Poisson probability mass function (PMF) $f_X^P(x)$:

$$f_X^P(x) = \begin{cases} \frac{1}{Z^P}\frac{(\lambda^P)^k e^{-\lambda^P}}{k!}, & \text{if } x = \overline{x} - k\frac{\overline{x}-\underline{x}}{K} \\ 0, & \text{otherwise} \end{cases} \tag{2.17}$$

$$f_X^P(\underline{x}) = \frac{\lambda^P}{\lambda^P + 1}, \quad f_X^P(\overline{x}) = \frac{1}{\lambda^P + 1} \tag{2.18}$$

The four scenarios above are realized through suitable choices of R and λ^P as follows: depending on the choice of the dynamic reference point, the SRS is a gain if $R = \underline{x} + b\gamma$ and a loss if $R = \overline{x} + b\gamma$. The SRS is considered high or low probability when the outcome that is not the reference can be realized with a probability of p_{NR} or $(1 - p_{\text{NR}})$, respectively, where p_{NR} is close to 1. In the computational experiments presented in Figure 2.4, $p_{\text{NR}} = 0.95$. Moreover, the range of the tariff is chosen as follows:

$$\begin{cases} \underline{x} + b\gamma < A^o \text{ if } R = \underline{x} + b\gamma \\ \overline{x} + b\gamma > A^o \text{ if } R = \overline{x} + b\gamma \end{cases} \tag{2.19}$$

such that the objective utility of the certain prospect, A^o, lies in the same gain or loss regime as the SRS and therefore represents a reasonable alternative to the SRS.

With the uncertain and the certain prospect defined in the SRS context above, we illustrate the fourfold pattern in Figure 2.4 using four quadrants. According to the fourfold pattern (a)–(d) in

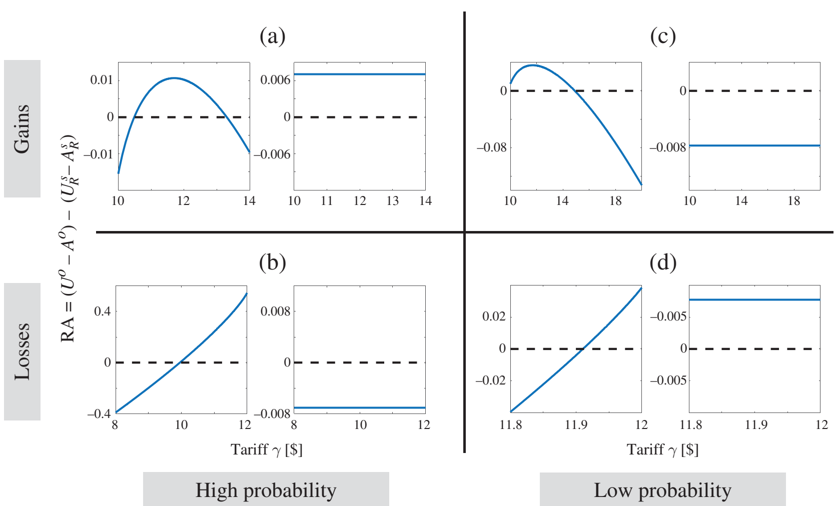

Figure 2.4 Illustration of the fourfold pattern of risk attitudes in the SRS context. (a) $R = \underline{x} + b\gamma$ and $f_X^P(\underline{x}) = 0.05$, (b) $R = \underline{x} + b\gamma$ and $f_X^P(\underline{x}) = 0.95$, (c) $R = \overline{x} + b\gamma$ and $f_X^P(\underline{x}) = 0.95$, and (d) $R = \overline{x} + b\gamma$ and $f_X^P(\underline{x}) = 0.05$.

Figure 2.4, the diagonal quadrants should correspond to risk-averse behavior while the off-diagonal ones are risk-seeking. In each quadrant, we plot a metric defined as $\text{RA} = (U^o - A^o) - (U_R^s - A_R^s)$ with respect to the tariff γ. This metric captures the Relative Attractiveness (RA) that the uncertain prospect has over the certain prospect for rational individuals (who can be modeled by conventional Utility Theory) versus individuals modeled with CPT. This follows from Equations (2.13) and (2.14), since $\text{RA} > 0 \implies p^o > p_R^s$. In Figure 2.4, we note that $\text{RA} > 0$ corresponds to all regions where the light gray curve is above zero and indicates risk-averse attitudes, as rational individuals have higher probability to accept the uncertain prospect than irrational ones. Similarly, $\text{RA} < 0$ corresponds to the light gray line being below zero and denotes risk-seeking attitudes. In each quadrant, two subplots are provided. The subplot on the right corresponds to a specific set of parameters $\beta^+ = \beta^- = \lambda = 1$ which completely removes the role of $V(\cdot)$ and hence removes the effects of framing, diminishing sensitivity and loss aversion (as specified in Section 2.3.1.1). The subplot on the left corresponds to standard CPT parameters chosen in the range $0 < \beta^+, \beta^- < 1,\ \lambda > 1$, thus producing a general CPT model. As explained before, each quadrant corresponds to a specific choice of R and λ^P, which together determine if an outcome is a gain or loss, and whether with high or low probability.

The most important observation from Figure 2.4 comes from the differences between the left and right subplots in each of the four quadrants. For example, from Figure 2.4a which is the case of high probability gains, all risk attitudes in the right subplot correspond to $\text{RA} > 0$ and therefore risk averse, while those on the left are only risk averse for a certain price range. That is, the fourfold pattern is violated in the left subplot. The same trend is exhibited in all four quadrants. This is because the fourfold pattern is due to the interplay between $\pi(\cdot)$ and $V(\cdot)$ and is valid only when the magnitude of $\pi(\cdot)$ is sufficiently large relative to that of $V(\cdot)$, such that probability distortion dominates (Harbaugh et al. 2009). This corresponds to the right subplots[2] as well as the left subplots within certain price ranges.

The insights drawn from this analysis of the fourfold risk attitudes pattern is that these four categories can suitably inform the dynamic pricing strategy for the SRS tariff design, through the left subplots in each quadrant of Figure 2.4. That is, it helps quantify two key qualitative statements: (i) presence of risk-seeking passengers allows for flexibility in charging higher tariffs, and (ii) presence of risk-averse passengers highlights the need for some additional constraints or limits on reasonable tariffs that can be imposed.

2.3.2.2 Implication II: Strong Risk Aversion Over Mixed Prospects

A mixed prospect is defined as an uncertain prospect whose portfolio of possible outcomes involves both gains and losses (Kahneman and Tversky 2012; Abdellaoui et al. 2008). Clearly, an uncertain prospect is always mixed when the reference point R is chosen to be its own expectation (i.e. expected value of its outcomes). The strong risk aversion of mixed prospects in the CPT framework stems from mainly loss aversion, as the impact of the loss often dominates its gain counterpart. This implication will be illustrated below in our SRS context using two different interpretations.

The first interpretation follows from property 2, which states that when $R = \bar{U}$, the subjective utility is strictly negative for a sufficiently large λ. Therefore, with $R = \bar{U}$ and such a large enough λ, the uncertain prospect is always subjectively perceived as a strict loss. This has been verified numerically with $\lambda > 1$. Since the objective utility determined by Utility Theory relative to the expectation is neutral (since $\bar{U} = U^0 = \mathbb{E}_{f_U}(u)$), strong aversion is exhibited.

The second interpretation follows from Property 3, which implies that when Property 2 holds, within the tariff range $[\underline{\lambda}, \overline{\lambda})$, the uncertain prospect is less likely to be accepted by the CPT-inclined passengers compared to the rational ones, as $p_{\bar{U}}^s < p^o$.

2 The right subplot in each quadrant corresponds to the case where individuals are risk neutral in the gain or loss regimes separately, and loss neutral, then $\pi(\cdot)$ alone is sufficient to generate the fourfold pattern.

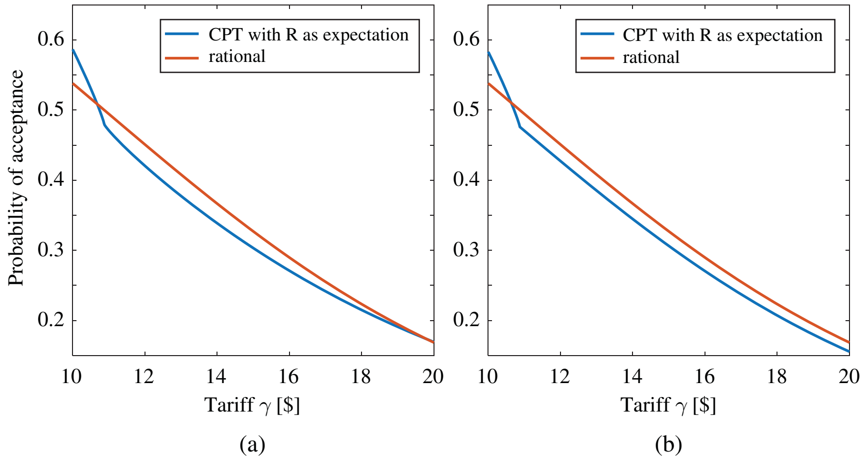

Figure 2.5 Comparison of $p^s_{\bar{U}}$ and p^o. For fair comparison, the tariff range of $\gamma \geq \frac{A^o - \overline{X}}{b}$ is plotted, where the alternative is nonloss. (a) CPT with $\beta^+ < 1$ and (b) CPT with $\beta^+ = 1$.

Figure 2.5 illustrates Property 3 with $f_X(x)$ obeying a Normal distribution, with the tariff range $\underline{\gamma}$ and $\overline{\gamma}$ approximated using the numerical setup. It is clear from the left subplot that within this price range, passengers exhibit strong risk aversion over the SRS, as the light gray curve is strictly above the dark gray one. It is interesting to note that when $\beta^+ = 1$, which corresponds to the scenario where passengers are risk neutral in the gain regime, the maximum possible tariff $\overline{\gamma} \to \infty$ (see Figure 2.5b).

The implication regarding strong risk aversion over mixed prospects is as follows: as the SRS has significant uncertainty, for passengers who regard the expected service quality as the reference, and when the alternative is relatively a nonloss prospect, strong risk aversion is exhibited. Hence, the SRS is strictly less attractive to these passengers when compared to rational ones. Therefore, the dynamic tariffs may need to be suitably designed by the SRS operator or server so as to compensate for these perceived losses. These could also take the form of special rebates, subsidies, or deals for SRS passengers.

2.3.2.3 Implication III: Effects of Self-Reference

In this section, we compare $p^s_{\bar{U}}$ with $p^s_{A^o}$, i.e. the subjective acceptance probability of the uncertain SRS while using (i) the expected SRS outcome as the dynamic reference versus, (ii) the objective utility of the certain alternative (e.g. UberX) as a static reference, respectively. Four different probability distributions are considered. In each case, how these two probabilities vary with the tariff γ was evaluated. The results are shown in Figure 2.6.

Figure 2.6 illustrates that for all four distributions, $p^s_{\bar{U}} \geq p^s_{A^o}, \forall\, \gamma$, which implies that the SRS is always more attractive when the reference is the expectation of itself, rather than the alternative. Note that $p^s_{\bar{U}} = p^s_{A^o}$ when $\gamma = \frac{A^o - \overline{X}}{b} \implies \bar{U} = A^o$ and hence the two reference points coincide.

The following summarizes the third implication inferred from Figure 2.6: $\bar{U}$ is essentially the rational counterpart of the uncertain prospect. Therefore, it could be argued that, when deciding between two prospects, the chance to accept one prospect is always higher if this prospect itself is regarded as the reference, compared with the case where the alternative is considered as the reference. This arises from loss aversion, i.e. $\lambda > 1$, and can be explained as follows: when one prospect is regarded as the reference, by definition, it would never be perceived as a loss and therefore not experience the magnified perception out of losses, whereas the alternative may be subject

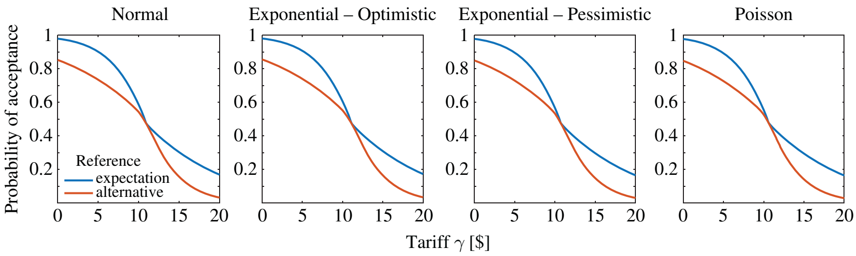

Figure 2.6 Comparison of $p^s_{\bar{U}}$ with $p^s_{A^o}$ using four different PDFs $f_X(x)$. In the truncated Poisson distribution, the parameters are set as $\lambda^P = 4$ and $K = 5$.

to being regarded as a loss and therefore can experience this skewed perception. In contrast, if the alternative is chosen as the reference, the roles are reversed.[3] Moreover, the statement is in fact intuitive as those passengers who regard the expectation as the reference have in some sense already subscribed to the SRS, and hence are naturally inclined to exhibit higher acceptance probabilities and therefore have higher willingness to pay. This partially explains the reason why converting customers to more uncertain SRS options from competitors and more traditional, predictable transportation modes (like driving, public transit, and exclusive ride-hailing) is typically more difficult than maintaining the current customer base. The last observation from Figure 2.6 is the invariance of the comparison with respect to the underlying probability distributions, which implies that the above results on self-reference are fairly general.

2.4 Summary and Conclusions

Several examples of CPHS include real-time decisions from humans as a necessary building block for the successful performance of the overall system. Many of these problems require behavioral models of humans that lead to these decisions. In this chapter, we describe two different tools that may be suitable for determining these behavioral models, which include Utility Theory and Prospect Theory. Tools from Utility Theory have been used successfully in several problems in transportation such as resource allocation and balance of supply and demand. This theory is described in Section 2.2 and illustrated using a transportation example that consists of a shared mobility problem where human riders are presented with the choice of different travel modes. We then show how these models can be used to address and mitigate traffic congestion. Cumulative Prospect Theory, an extension of Prospect Theory, is a modeling tool widely used in behavioral economics and cognitive psychology that captures subjective decision-making of individuals under risk or uncertainty. CPT is described in Section 2.3, with the same transportation example used to illustrate its application potential. Results from a survey conducted with about 1000 respondents are used to derive a CPT model and estimate its parameters. Lottery questions were also included in the survey to illustrate CPT effects, the results of which are described in this section as well. Finally, a few theoretical implications of CPT are presented that provide an overall quantitative structure to the qualitative behavior of humans in overall decision-making scenarios.

Human-in-the-loop behavioral models can be applied to several other applications beyond ridesharing, both within the transportation and mobility sector, and other domains that involve

3 Other effects of CPT due to $\alpha, \beta^+, \beta^- < 1$ may result in complicated nonlinearities which might alleviate loss aversion. Therefore, this statement is valid when λ is sufficiently larger than 1, such that loss aversion dominates.

humans interacting with cyber–physical systems. One such example includes demand response in power grids. Future research involves the development of more accurate models of human operators for which more quantitative data are required with human decision-makers as integral components of the overall infrastructure. For instance, these could include more sophisticated utility functions for different modes that also take into account factors other than price and travel time which could affect passengers' choices. In addition, more realistic data sets on ridesharing, mode choice, and dynamic pricing could be obtained either by conducting larger, representative surveys or through pilot studies and field trials performed in conjunction with ridesharing companies or transit authorities.

Acknowledgments

This work was supported by the Ford-MIT Alliance.

References

Abdellaoui, M., Bleichrodt, H., and l'Haridon, O. (2008). A tractable method to measure utility and loss aversion under prospect theory. *Journal of Risk and Uncertainty* 36 (3): 245.

Annaswamy, A. and Nudell, T. (2015). Transactive control–what's in a name. *IEEE Smart Grid Newsletter*

Annaswamy, A.M., Guan, Y., Tseng, H.E. et al. (2018). Transactive control in smart cities. *Proceedings of the IEEE* 106 (4): 518–537.

Bejestani, A.K., Annaswamy, A., and Samad, T. (2014). A hierarchical transactive control architecture for renewables integration in smart grids: analytical modeling and stability. *IEEE Transactions on Smart Grid* 5 (4): 2054–2065.

Ben-Akiva, M.E., Lerman, S.R., Lerman, S.R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, vol. 9. MIT Press.

Bernards, R., Reinders, J., Klaassen, E. et al. (2016). Meta-analysis of the results of European smart grid projects to quantify residential flexibility. *CIRED Workshop 2016*, 1–4. IET.

Chassin, D.P., Malard, J.M., Posse, C. et al. (2004). Modeling Power Systems as Complex Adaptive Systems. *Technical Report No. PNNL-14987*. Richland, WA (United States): Pacific Northwest National Lab. (PNNL).

Erslan, E., Yildiz, Y., and Annaswamy, A.M. (2023). *Safe Shared Control in the Face of Anomalies*. IEEE-Press.

Fishburn, P.C. (1988). *Nonlinear Preference and Utility Theory*, No. 5. Baltimore, MD: Johns Hopkins University Press.

Guan, Y., Annaswamy, A.M., and Tseng, H.E. (2019a). Cumulative prospect theory based dynamic pricing for shared mobility on demand services. *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2239–2244. IEEE.

Guan, Y., Annaswamy, A.M., and Tseng, H.E. (2019b). A dynamic routing framework for shared mobility services. *ACM Transactions on Cyber-Physical Systems* 4 (1): 6:1–6:28. https://doi.org/10.1145/3300181.

Hammerstrom, D.J., Ambrosio, R., Carlon, T.A. et al. (2008). Pacific Northwest GridWise™ testbed Demonstration Projects; Part I. Olympic Peninsula Project. *Technical Report No. PNNL-17167*. Richland, WA (United States): Pacific Northwest National Lab. (PNNL).

Hao, H., Corbin, C.D., Kalsi, K., and Pratt, R.G. (2016). Transactive control of commercial buildings for demand response. *IEEE Transactions on Power Systems* 32 (1): 774–783.

Harbaugh, W.T., Krause, K., and Vesterlund, L. (2009). The fourfold pattern of risk attitudes in choice and pricing tasks. *The Economic Journal* 120 (545): 595–611.

Jagadeesan Nair, V. (2021). Estimation of cumulative prospect theory-based passenger behavioral models for dynamic pricing & transactive control of shared mobility on demand. Master's thesis. Massachusetts Institute of Technology.

Kahneman, D. and Tversky, A. (2012). Prospect theory: an analysis of decision under risk. In: (ed. L.C. MacLean and W.T. Ziemba) *Handbook of the Fundamentals of Financial Decision Making*, *World Scientific Handbook in Financial Economics Series*, vol. 4, 99–127. World Scientific. ISBN 978-981-4417-34-1. https://doi.org/10.1142/9789814417358_0006. https://www.worldscientific.com/doi/abs/10.1142/9789814417358_0006.

Katipamula, S., Chassin, D.P., Hatley, D.D. et al. (2006). Transactive Controls: A Market-Based GridWise™ Controls for Building Systems. *Technical Report No. PNNL-15921*. Richland, WA (United States: Pacific Northwest National Lab. (PNNL).

Kok, K. (2013). *The PowerMatcher: Smart Coordination for the Smart Electricity Grid*, 241–250. The Netherlands: TNO.

Li, S., Zhang, W., Lian, J., and Kalsi, K. (2015). Market-based coordination of thermostatically controlled loads–Part I: a mechanism design formulation. *IEEE Transactions on Power Systems* 31 (2): 1170–1178.

Lombardi, C., Annaswamy, A., and Santos, L.P. (2022). Model-based dynamic toll pricing scheme for a congested suburban freeway with multiple access locations. *Journal of Intelligent Transportation Systems (provisionally accepted)*, 1–28.

Melton, R. (2015). Pacific Northwest Smart Grid Demonstration Project Technology Performance Report Volume 1: Technology Performance. *Technical Report PNW-SGDP-TPR-Vol.1-Rev.1.0; PNWD-4438, Volume 1*. Richland, WA (United States): Pacific Northwest National Lab. (PNNL).

Phan, T., Annaswamy, A.M., Yanakiev, D., and Tseng, E. (2016). A model-based dynamic toll pricing strategy for controlling highway traffic. *2016 American Control Conference (ACC)*, 6245–6252. IEEE.

Prelec, D. (1998). The probability weighting function. *Econometrica* 66 (3): 497–527. https://doi.org/10.2307/2998573.

Rieger, M.O., Wang, M., and Hens, T. (2017). Estimating cumulative prospect theory parameters from an international survey. *Theory and Decision* 82 (4): 567–596. https://doi.org/10.1007/s11238-016-9582-8.

Schweppe, F.C. (1978). Power systems2000': hierarchical control strategies. *IEEE Spectrum* 15 (7): 42–47.

Schweppe, F.C., Tabors, R.D., Kirtley, J.L. et al. (1980). Homeostatic utility control. *IEEE Transactions on Power Apparatus and Systems* (3): 1151–1163.

Somasundaram, S., Pratt, R.G., Akyol, B. et al. (2014). Reference Guide for a Transaction-Based Building Controls Framework. Pacific Northwest National Laboratory.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5 (4): 297–323. https://doi.org/10.1007/BF00122574.

Von Neumann, J. and Morgenstern, O. (2007). *Theory of Games and Economic Behavior (Commemorative Edition)*. Princeton University Press.

Wang, S. and Zhao, J. (2018). How Risk Preferences Influence the Usage of Autonomous Vehicles. *Technical Report number: 18-00785*. Washington DC, United States: Transportation Research Board 97th Annual Meeting.

Wang, S. and Zhao, J. (2019). Risk preference and adoption of autonomous vehicles. *Transportation Research Part A: Policy and Practice* 126: 215–229.

Widergren, S.E., Subbarao, K., Fuller, J.C. et al. (2014). AEP Ohio gridSMART Demonstration Project Real-Time Pricing Demonstration Analysis. *Technical Report No. PNNL-23192*. Richland, WA (United States): Pacific Northwest National Lab. (PNNL).