Optimal Comprehensible Targeting*

Walter W. Zhang

University of Chicago Booth School of Business

September 21, 2023

Preliminary Draft
(Latest Version Here)

Abstract

Developments in machine learning and big data allow firms to fully personalize and target their marketing mix. However, data and privacy regulations, such as those in the European Union (GDPR), incorporate a "right to explanation", which is fulfilled when targeting policies are comprehensible to customers. This paper provides a framework for firms to navigate right-to-explanation laws. First, I introduce a new method called Policy DNN, which combines policy learning and deep neural networks, to form a profit-maximizing black box benchmark and provide theoretical guarantees on its performance. In contrast to prior approaches that use a two-step method of estimating treatment effects before assigning individuals their treatment group, Policy DNN directly estimates treatment assignment, which improves efficiency. Second, I construct a class of comprehensible targeting policies that is represented by a sentence. Third, I show how to optimize over this class of policies to find the profit-maximizing comprehensible policy. I demonstrate that it is optimal to estimate the comprehensible policy directly from the data, rather than projecting down the black box policy into a comprehensible policy. Finally, I apply my framework empirically in the context of price promotions for a durable goods retailer using data from a field experiment. I quantify the cost of explanation, which I define as the difference in expected profits between the optimal black box and comprehensible targeting policies. The comprehensible targeting policy reduces profits by 7% or 22 cents per customer when compared to the black box benchmark.

Keywords: Interpretable AI, Targeting, Policy Learning, Deep Learning, Data and Privacy Regulation

*Email: walterwzhang@chicagobooth.edu. Preliminary draft — results may change. I thank Sanjog Misra, Günter J. Hitsch, Pradeep K. Chintagunta, Tengyuan Liang, and Avner Strulov-Shlain for their encouragement and support. This paper greatly benefited from discussions with Max H. Farrell, Yuexi Wang, Karthik Srinivasan, James W. Kiselik, Olivia R. Natan, Reuben Bauer, Benedict Guttman-Kenney, and participants at various seminars.

1 Introduction

Black box algorithms dominate marketing decisions today (Katsov, 2017).[1] These algorithms are fast, personalized, and, most importantly, designed to be profit-maximizing for the firm.[2] Modern algorithms apply profit-maximizing marketing decisions at the individual level. At the extreme, customers can face their own marketing mix.

However, full personalization of the marketing mix may harm the firm. Consider the following example in promotions management: two customers go to check out at the register but only one of the two customers is given a 20%-off promotion. To justify the exclusion of the other customer, the firm's sales representative cannot simply say it was profit-maximizing to do so. The representative instead needs to provide the customer a comprehensible explanation for why she did not receive the promotion. If a suitable explanation cannot be found or understood, then the excluded customers may feel slighted by the firm (Dietvorst and Bartels, 2022).

More generally, customers may desire an explanation to either learn from the algorithmic decision (van Osselaer and Alba, 2000) (if they unexpectedly got 20% off, then they want to know how to get it again), or to understand the algorithm (in the case where they were not given the promotion when another customer was) (Dietvorst et al., 2015). Similarly, firms need an explanation of the algorithm because it can provide long-term brand equity by addressing customer needs, makes its algorithmic decisions easier to justify by its human representatives, and allows them to self-diagnose their own algorithmic decision making.

This example underscores a need for comprehensible algorithms by both customers and firms; such algorithms require more than just a simple explanation of the black box algorithm.[3] Comprehensible algorithms are both transparent and complete. They are transparent if the algorithm is explainable to customers. They are complete if that explanation is identical to the policy being implemented. To give a concrete example of a comprehensible targeting policy, the policy "target a customer with a promotion if she has not bought in the last thirty days and lives in Chicago" is both transparent and complete.

Recent regulation in General Data Protection Regulation (GDPR) includes a "right to explanation" (European Commission, 2016), which suggests that firms need to fully explain their algorithmic decisions to their consumers. Firms violating GDPR face a hefty fine of 4% of global revenues or 20 million Euros, whichever is higher. As the proposed AI Act (European Commission, 2021) gains adoption in Europe and regulatory measures like the California Consumer Privacy Act (CCPA) emerge in other jurisdictions, the scope of regulatory oversight around right-to-explanation legislation is set to expand.

[1] A black box algorithm is one that users can observe the outcomes, but the internal mechanisms remain opaque.

[2] Modern algorithms can automate all steps of the marketing decision process instead of just providing decision support. The marketing decision support system (MDSS) proposed in Little (1979) has now expanded to forming decisions directly instead of just supporting them.

[3] The literature is divided as to whether explanations of the black box algorithm are sufficient for regulators (Edwards and Veale, 2017; Gillis and Spiess, 2019).

This paper provides a framework for firms to design and analyze optimal comprehensible targeting policies. These policies are optimal in that they are profit-maximizing and are comprehensible to customers, regulators, and the firm's own representatives. My framework allows firms to quantify the profit differences from implementing optimal comprehensible policies as opposed to black box policies and to compare the two targeting policies. There are three components in my framework: (1) forming the black box algorithm benchmark, (2) constructing a class of comprehensible policies, and (3) finding the optimal comprehensible policy.

First, I propose a new black box algorithm for marketing decisions that directly learns the optimal decision from the data. This new algorithm provides a benchmark for what the firm could implement as the black box targeting policy. Standard approaches in the literature first run an RCT, then estimate the distribution of heterogeneous treatment effects, and finally form the optimal targeting policy from the estimated effects (Hitsch et al., 2023; Simester et al., 2020). The new approach obviates the second step and directly learns the optimal policy from the data. The key insight is that learning the optimal decision for discrete treatments is easier than learning the distribution of heterogeneous treatment effects. I use deep neural networks (DNN) to represent the policy function and call the new method Policy DNN. I provide inference around the expected profits generated from the black box algorithm. I then show the new approach learns the optimal policy better in Monte Carlo simulations and generates more profits in the empirical application than standard approaches.

Second, I construct a class of comprehensible targeting policies. These comprehensible policies are targeting policies that consist of a sentence formed from conditional clauses joined by logic operators (e.g., "Target a customer with a promotion if she has not bought in the last thirty days and lives in Chicago").[4] The structure of the these comprehensible policies is motivated from the explainable AI and philosophy literatures.[5] These comprehensible policies are subsets of decision trees as well as logic trees (Schwender and Ruczinski, 2010). I further show how I can feature engineer these clauses out from standard database marketing datasets.

Third, I find the optimal comprehensible policy that maximizes firm profits. The components of the comprehensible policy which I optimize over are the conditional clauses and the logic operators that combine the clauses. I show how to solve the optimization problem using brute force and greedy algorithms. Then, I demonstrate how to conduct inference around the expected profits generated from the optimal comprehensible policy.

[4]The prior example, "Target a customer with a promotion if she has not bought in the last thirty days and lives in Chicago" is a comprehensible targeting policy with a length of two conditional clauses. The conditional clauses directly formed from RFM data are "if she has not bought in the last thirty days" and "if she lives in Chicago." The logic operator joining the two conditional clauses is "and." The sentence is under five clauses, so it is conversational. This sentence embeds a contrastive causal framework as the counterfactual state if she bought in the last thirty days or does not live in Chicago will lead to her not being targeted.

[5]The explainable AI literature suggests that sentences with more than five constitutive clauses are not commonly understandable and are not conversational (Miller, 2019). The philosophy literature notes that understanding is based on contrastive causation and counterfactual inference (Lipton, 1990). These two concepts are built into the constructed class of comprehensible policies.

Using my proposed framework I perform a cost analysis as firms adopt comprehensible targeting policies. I first compare whom the optimal black box and the optimal comprehensible policies target. I then compute the profit difference between the two policies. Since the comprehensible policy is less expressive and personalized than the black box policy, the comprehensible policy should be less profitable than the black box policy. I call this difference in profits the *cost of explanation* that the firm faces if it adopts the comprehensible policy.

I implement this cost analysis with an empirical application in promotions management. I use the dataset from Ni et al. (2012) as a case study of the framework. The dataset includes a randomized control trial of a $10-off promotion randomly mailed to 176,961 households for a durable goods retailer. The outcome of interest is sales during the promotional period of December 2003. First, I document how Policy DNN produces a more profitable targeting policy than other standard black box approaches. Second, I show how to generate and engineer different comprehensible policies from this RFM dataset. Third, I find the optimal comprehensible policy and compare it to the black box targeting policy. These three steps let me document (1) how the comprehensible targeting policy differs in whom it targets compared to the black box targeting policy and (2) the firm's cost of explanation if it implemented the optimal comprehensible policy instead of the black box policy.

In the empirical application, the optimal comprehensible policy with three clauses targets those who spend a lot during the holiday period but not in the spring, or who spent less than average in prior holiday period.[6] Comparing the optimal comprehensible policy to the best-performing black box policy, I find that the optimal comprehensible policy does not systematically overtarget or undertarget; it appears to be limited in its ability to capture customer heterogeneity due to the comprehensibility constraint. I further find that the cost of explanation is 22 cents per person, which constitutes a 7% loss in profits compared to the optimal black box policy and a 38 cents per person (or 16%) gain in profits over a blanket targeting policy.

These results are of substantive interest. I provide an exercise where I benchmark the loss the firm faces if it moves away from the black box policy to comply with the right-to-explanation legislation with the GDPR penalty. In the empirical application, if I assume a basis of 10 million customers, the implied 22 cents cost of explanation per person leads to an expected loss of $2.2 million per month. If the GDPR penalty is $20 million and is enforced at a 10% rate, then the expected penalty is $2 million. The firm thus may decide to not comply depending on how much it believes the 10% enforcement rate and its willingness to break the law for monetary gain. From the regulator's perspective, it may consider raising the penalty or the enforcement rate to guarantee compliance.

Moving from the empirical application back to theory, I show that forming the optimal comprehensible policy directly from the data is more profitable than forming it by ex post projecting

[6] More specifically, the optimal comprehensible policy is "Target a customer if she spent in the top half of spenders who spend during Christmas over the last two years *and* did not spend among the top half of spenders in spring over the last two years *or* is among the bottom half of spenders during last year's holiday promotion."

down the black box policy to a comprehensible policy. The latter procedure is motivated by the explainable AI (XAI) literature. Methods in XAI often provide a locally approximative model of the black box in order to shed light upon the black box's decisions (Biran and Cotton, 2017; Miller, 2019; Mothilal et al., 2019; Rai, 2020; Senoner et al., 2022). I show that the direct approach generates more profits than the ex post approach theoretically and validate it using the empirical application. I also show how I can conduct inference on the ex post comprehensible policy by recentering the empirical process results from Kitagawa and Tetenov (2018). As a result, firms should directly form the comprehensible policy from the data instead of first finding the optimal black box and then projecting it down.

This paper builds on many different extant literatures. The implementation of algorithms in decision making and their effects are well documented (Kleinberg et al., 2015, 2017). In marketing, algorithmic decisions are often made by forming optimal targeting policies in various domains (Ascarza, 2018; Chintagunta et al., 2023; Ellickson et al., 2022; Hitsch et al., 2023; Karlinsky-Shichor and Netzer, 2019; Rossi et al., 1996; Simester et al., 2020; Smith et al., 2022; Yoganarasimhan et al., 2022; Zhang and Misra, 2022) and the literature is reviewed in Rafieian and Yoganarasimhan (2022). This paper's methodological contribution, Policy DNN, builds on results from the policy learning literature (Athey and Wager, 2021; Kallus and Zhou, 2018; Kitagawa and Tetenov, 2018; Kitagawa et al., 2021; Mbakop and Tabord-Meehan, 2021).[7] Specifically, I combine results from using statistical surrogates for policy learning (Zhao et al., 2012) with DNN (Farrell et al., 2021) and provide inference around the expected profits generated from the optimal targeting policy by utilizing the inference engine for DNN from (Farrell et al., 2020). Methodologically, this paper introduces and provides inference for a state-of-the-art methodology (DNN) used in policy learning with discrete treatment variables.

The comprehensible policy class builds on the interpretable AI literature where expert systems and logic rules are designed to be interpretable by human agents (Angelino et al., 2018; Cawsey, 1991, 1992, 1993; Rudin, 2019; Weiner, 1980). This paper deviates from that literature in that the comprehensible policy is much simpler and is conversational—a firm's representative is able to fully state the targeting policy in a sentence. Edwards and Veale (2017) and Kleinberg et al. (2018) document issues that regulators have in diagnosing and understanding algorithms. My framework for finding and evaluating optimal comprehensible policies provides a solution for firms to comply with regulatory demands while balancing profitability.

The rest of the paper is organized as follows: Section 2 provides an overview of the methodology and sets up the mathematical notation. Section 3 introduces Policy DNN. Section 4 constructs a class of comprehensible targeting policies and Section 5 shows how to find the optimal comprehensible policy. I show that generating a comprehensible policy from projecting down a black box is less profitable than learning the policy directly from the data in Section 6. I provide the

[7] For multiperiod settings, policy learning has been explored in the marketing literature for solving reinforcement learning problems (Ko et al., 2022; Liu, 2022). The use of policy learning to find optimal targeting policies directly from the data in static settings is not well explored in the marketing literature.

empirical application in Section 7: in Section 7.1, I document the differences in whom the black box and the optimal comprehensible targeting policy target and in Section 7.2, I quantify the cost of explanation. I then discuss what firms should consider when deciding to implement comprehensible policies in Section 8 and provide an example of how the firm in the empirical application will be impacted by GDPR's right-to-explanation regulation in Section 8.1. I conclude in Section 9.

2 Framework

In this section, I provide an overview of the general methodology of forming and evaluating comprehensible policies. Figure 1 summarizes the methodology, which I discuss in Section 2.1. I introduce the mathematical framework in Section 2.2. A key methodological theme for both the black box and optimal comprehensible policies is policy learning, where the optimal targeting policy is directly formed by maximizing profits.

2.1 Framework overview

Figure 1 illustrates the perceived tradeoff between comprehension and expected profits. Comprehension can be considered as 1/(Model Complexity) and can be made mathematically rigorous by using the targeting policy's Vapnik-Chervonenkis dimension or Rademacher complexity as a proxy for model complexity.

As targeting policies become more comprehensible, they become less personalized and cannot capture customer heterogeneity as well as incomprehensible black box methods. Each dot on the figure represents a specific targeting policy. Targeting policies that are conversational, or are simple enough to be explained in a conversation, are the right of the dashed line. To provide an example, a targeting policy that blanket targets everyone would be very comprehensible but not profitable, and it would be represented by a point on the bottom right side of the figure near the horizontal axis.

The proposed methodological framework evaluates and compares black box targeting policies to comprehensible targeting policies. There are three components for the framework: (1) forming the black box policy, (2) constructing a class of comprehensible policies, and (3) finding the optimal profit-maximizing comprehensible policy. I now describe the three pieces to form the framework while using Figure 1 as a guide.

I first focus on the class of black box policies. Specifically, I use the class of deep neural networks (DNN). I denote this class of functions $\mathcal{F}_{DNN}$, and it is represented on the left side of the figure with the blue curve. DNNs are used in many state-of-the-art machine learning applications and possess a uniform approximation property that allows them to approximate any function (Goodfellow et al., 2016). As the blue curve moves from right to left it trades off profitability for comprehensibility. For example, a DNN with only one hidden layer with a handful of nodes

would be a point on the right side of the blue curve while a DNN that is both deep and wide would be on the left side of the curve. The curve trends down after a certain point to represent overfitting when the DNN is too complex.

I show that I can find the optimal black box in Section 3 by combining policy learning with DNN and that I can conduct inference around the profits from the optimal targeting policy. The optimal black box policy function is represented as $d^*_{DNN}(x)$ in the figure and generates expected profits of $\Pi^*_{DNN}(x)$. The targeting policy itself is represented as the point on the blue curve.

Second, I construct a class of comprehensible policies ($\mathcal{F}_{\text{Comp}}$) in Section 4 and represent it with the red curve on the right side of the figure. I show how to generate the clauses for the comprehensible policies from marketing data and then form the policies. The number of clauses represents model complexity in this function class. To trace out the red curve, the example targeting policy "Target a customer if she has not bought in the last thirty days" would be a point on the very right side of the red curve as it is very comprehensible and is conversational but is limited in its profitability. The targeting policy "Target a customer if she has not bought in the last thirty days and lives in Chicago" will also be on the curve but is to the left of the former example. Since the sentence has two clauses, it will be less comprehensible but more profitable. The red curve also trends down after a point to represent overfitting; a targeting policy that contains twenty clauses can overfit and is not conversational.

Third, I show how to optimize over this class of comprehensible policies in Section 5 by using brute force and greedy algorithms. I find the point on the red curve that generates the highest profits and denote this targeting policy as $d^*_{\text{Comp}}(x)$ which generates expected profits Π^*_{Comp}. I also show that finding the optimal comprehensible policy directly from the data is more profitable than finding it by projecting down the black box in Section 6.

Using the empirical application, I then compare the two targeting policies from the optimal black box $d^*_{DNN}(x)$ and the optimal comprehensible policy to examine the differences in the targeting policies in Section 7.1. I then document the cost of explanation, or the difference in profits from the optimal black box to the optimal comprehensible policy ($\Delta\Pi = \Pi^*_{DNN} - \Pi^*_{\text{Comp}}$), in Section 7.2. These comparisons enable me to conduct a cost analysis documenting the lost expected profits from using a comprehensible policy instead of the black box policy as well as the targeting differences between the two policies.

This framework allows me to substantively evaluate the firm's economic losses in the empirical application faces when it follows right-to-explanation laws for its targeting policy. I compare the cost of explanation to the expected GDPR penalty. This measurement exercise allows me to evaluate the penalty's impact on the firm and is discussed in Section 8.1. I discuss how these losses play a role for managerial decision making in Section 8

2.2 Mathematical framework

I define (X, W, Y) as the data tuple of the covariates, the treatment, and the outcome variable respectively. I observe this tuple for each individual i and assume the data is $i.i.d.$ for n observations. I consider a binary treatment $W \in \{0, 1\}$ and the data X has dimension p. I further define $Y(1), Y(0)$ as the potential outcomes for the binary treatment with observed outcome variable $Y = WY(1) + (1 - W)Y(0)$.

To provide a concrete example, I can think of Y as sales, W as a promotional mailing, and X as consumer characteristics. The firm cares about maximizing expected profits from its promotions management campaign where profits are a function of the sales.

Consider the standard inverse propensity weighted profit estimator for a given policy function $d : \mathbb{R}^p \to \{0, W\}$ that maps the covariates to the targeting rule,

$$\hat{\Pi}(d) = \sum_{i=1}^{n} \frac{W_i}{e(x_i)} \pi_i(1) d(x_i) + \frac{1 - W_i}{1 - e(x_i)} \pi_i(0)(1 - d(x_i)). \tag{1}$$

This profit estimator is an unbiased estimator for the profits from the targeting policy

$$\Pi(d) = \sum_{i=1}^{n} \pi_i(1) d(x_i) + \pi_i(0)(1 - d(x_i)) \tag{2}$$

where $\pi_i(1) = mY_i(1) - c, \pi_i(0) = mY_i(0)$ are the counterfactual profit values when individual i is respectively targeted and not targeted and $e(x_i) = P(W_i \mid X = x_i)$ is the propensity score.[8]

From Equation (2), I see that the optimal policy function is

$$d^* = \mathbf{1}\{\pi_i(1) > \pi_i(0)\} = \mathbf{1}\{m(Y_i(1) - Y_i(0)) > c\} \tag{3}$$

in which individual i is targeted if and only if her counterfactual profits are higher under that treatment assignment, and where $\mathbf{1}\{\cdot\}$ represents the indicator function and m, c respectively represent the profit margins and the cost of issuing the treatment.

Since I do not observe $Y_i(1), Y_i(0)$ because of the fundamental problem of causal inference, I cannot form $Y_i(1) - Y_i(0)$. Instead, I want to find a representation for the difference in the potential outcomes. I make a few assumptions to do so. First, I have a structural assumption that customers have a utility function

$$u_i = \alpha(x_i) + \beta(x_i) W_i,$$

which maps to the outcome variable with function $Y_i = G(u_i) + \epsilon_i$. In the example with the

[8] The inverse propensity weighted profit estimator is also known as the Horvitz-Thompson profit estimator (Imbens and Rubin, 2015), and $E[\hat{\Pi}] = \Pi$ under the standard assumptions of unconfoundedness, overlap, and SUTVA (Hitsch et al., 2023).

promotional mailing, I assume a linear function for $G(\cdot)$, $G(u_i) = u_i$. The assumption simplifies the problem to

$$Y_i = \alpha(x_i) + \beta(x_i)W_i + \epsilon_i, \tag{4}$$

where I can interpret $\alpha(x_i)$ as the baseline sales if customer i was not given a treatment and $\beta(x_i)$ as the incremental effect of issuing the treatment on sales for customer i. I allow the intercept term $\alpha(x_i)$ and the coefficients $\beta(x_i)$ to depend on the individual's pretreatment covariates, which provide both heterogeneity in the coefficients and the ability to forecast $\alpha(x_i), \beta(x_i)$ for a customer with given x_i.

I make three additional assumptions to ensure I can recover $\beta(x_i)$ as a causal effect of the treatment W_i (Imbens and Rubin, 2015). These assumptions are typical in experimental settings. The first two are the unconfoundedness and the overlap assumptions. These two assumptions are provided under a properly run randomized control trial (RCT). I then assume the stable unit treatment value assumption (SUTVA) holds. This assumption effectively implies that there are no spillover effects from sharing the promotional mailing in the running example. I formally state the assumptions below.

Assumption 1. *(Unconfoundedness) The potential outcomes $Y_i(1), Y_i(0)$ are statistically independent of the treatment variable W_i which is represented formally as $\{Y_i(1), Y_i(0)\} \perp W_i$.*

Assumption 2. *(Overlap) The propensity score $e(x_i) = P(W_i \mid X = x_i)$ is bounded between zero and one which is represented formally as $0 < e(x_i) < 1$.*

Assumption 3. *(SUTVA) The potential outcomes for any individual do not vary with the treatment assignments for other individuals. For each individual, and there are no different forms of treatments that lead to different potential outcomes.*

These three assumptions allow me to map the identify the conditional expectation of potential outcomes and then map them to the observed data. I have that

$$\begin{aligned}
\alpha(x_i) &= E[Y_i(0) \mid X = x_i, W = 0] = E[Y_i \mid X = x_i, W = 0] \\
\alpha(x_i) + \beta(x_i) &= E[Y_i(1) \mid X = x_i, W = 1] = E[Y_i \mid X = x_i, W = 1] \\
\beta(x_i) &= E[Y_i(1) \mid X = x_i, W = 1] - E[Y_i(0) \mid X = x_i, W = 0] \\
&= E[Y_i \mid X = x_i, W = 1] - E[Y_i \mid X = x_i, W = 0]
\end{aligned}$$

and $\beta(x_i)$ is the heterogeneous treatment effect (HTE) or the conditional average treatment effect (CATE) for treatment W_i.

Circling back to the optimal policy in Equation (3), while I cannot learn the true optimal policy d^* because of the fundamental problem of causal inference, I can instead learn the optimal policy

function as a function of covariates x_i,

$$d^*(x_i) = d^*(\beta(x_i)) = \mathbf{1}\{m(E[Y_i(1) \mid X = x_i, W = 1] - E[Y_i(0) \mid X = x_i, W = 0]) > c\} \quad (5)$$
$$= \mathbf{1}\{\beta(x_i) > \frac{c}{m}\},$$

where I use the conditional expectation of the potential outcomes in place of the potential outcomes.

Recent approaches to form the optimal policy take a three-step or two-step approach.[9] I first outline the three-step approach: it estimates the conditional expectation functions $E[Y_i(1) \mid X, W = 1]$ and $E[Y_i(0) \mid X, W = 0]$ using regressions, then forms the treatments effects $\hat{\beta}(x) = \hat{E}[Y_i(1) \mid X, W = 1] - \hat{E}[Y_i(0) \mid X, W = 0]$, and lastly constructs the optimal policy $\hat{d}^*(x_i) = \mathbf{1}\{\hat{\beta}(x_i) > \frac{c}{m}\}$ following the plug-in rule. The two-step approach first directly estimates the heterogeneous treatment effect $\hat{\beta}(x)$ and then constructs the optimal policy $\hat{d}^*(x_i) = \mathbf{1}\{\hat{\beta}(x_i) > \frac{c}{m}\}$.

I now make three remarks on the standard three-step or two-step approach to motivate my proposed procedure of Policy DNN that combines policy learning and deep neural networks (DNN).

Remark 4. I outline some inefficiencies for both three-step and two-step approaches. I first note that to determine the optimal policy function $d(x_i)$ in Equation (5), I only need to know whether $\hat{\beta}(x)$ is greater than the scaled cost of treatment $\frac{c}{m}$. This implies to form the optimal policy only $\text{sign}\{\hat{\beta}(x) > \frac{c}{m}\}$ is required. Then $\alpha(x)$ and the exact value of $\beta(x)$ are then nuisance parameters. The three-step approach estimates $\hat{\alpha}(x_i)$ and $\hat{\beta}(x_i)$ and the two-step approach estimates $\hat{\beta}(x_i)$.

Instead, I focus on estimating the optimal policy function, $\mathbf{1}\{\beta(x) > \frac{c}{m}\}$, directly. The key idea is that just knowing $\mathbf{1}\{\beta(x) > \frac{c}{m}\}$ (or equivalently the sign of $\beta(x) - \frac{c}{m}$) is both necessary and sufficient for finding the optimal policy.

Remark 5. To build intuition around the prior remark, Figure 2 provides a visualization of different distributions of $\beta(x)$ that all produce the same targeting policy.[10] The solid black vertical line represents the cutoff $\frac{c}{m}$: all individuals to the right of the cutoff should be targeted and all individuals to the left of the line should not be targeted. The dashed vertical line represents an individual's $\beta(x_i)$. I emphasize that the targeting rule $d^*(x_i)$ is the same for individual i as long as her $\beta(x_i)$ is above the cutoff $\frac{c}{m}$.

Remark 6. The proposed approach allows for a more flexible representation function than the

[9] See Hitsch et al. (2023) for an overview of different methods to form optimal targeting policies from conditional average treatment effects and Künzel et al. (2019) for an overview of different three-step approaches for estimating conditional average treatment effects.

[10] The top panel provides a density plot of $\beta(x) \sim N(c/m, 5^2)$. The optimal targeting policy is to target those with $\beta(x) > \frac{c}{m}$ and to not target those with $\beta(x) < \frac{c}{m}$. In the next two panels I provide a monotonic transformation of the $\beta(x)$ for those individual i above and below the cutoff for targeting. The optimal targeting rule remains the same despite the transformation. Thus, many different densities of the CATEs, $\beta(x)$, can produce the same optimal targeting policy $d^*(x)$.

direct policy estimator espoused in Kitagawa and Tetenov (2018) and in the policy tree two-step approach introduced in Athey and Wager (2021). In both papers, the authors derive minimax rates for the welfare regret of policy functions to the true policy function. Both papers restrict the complexity of the learned policy function, ensuring the square root of the Vapnik-Chervonenkis (VC) dimension cannot grow faster than $\sqrt{n}$ rate following the fundamental theorem of statistical learning (Shalev-Shwartz and Ben-David, 2013; Vapnik, 2000), which in practice leads to the policy function represented by linear combinations of or shallow trees of covariates.

Instead of looking at minimax rates, I focus on semiparametric inference for policy learning, which provides a guide for practical use but is less conservative than minimax rates (Mou et al., 2022).[11] Minimax rates provide guidance for the decision making under the worst case scenario for the decision maker while semiparametric inference provides guidance for a learned decision rule. The latter will be less conservative because it does not assume the worst case scenario.

In the next section, I leverage the results from (Farrell et al., 2020, 2021) to learn a flexible policy function that can be represented by a DNN. The proposed approach provides a more pragmatic approach to learning the policy function. DNN inherently capture more complex policy functions than shallow trees or linear threshold functions and are better suited for algorithmic decision making in practice.

3 Policy learning with deep neural networks

In this section, I propose Policy DNN. The new black box technique leverages the idea that it is easier to directly learn the optimal targeting policies than the standard approach of learning the heterogeneous treatment effects first and then forming the optimal targeting policy. I leverage surrogate functions with a deep neural network framework for the new methodology in Section 3.1. I then show how to attain inference around the profits under the optimal targeting policy with Policy DNN in Section 3.2. Monte Carlo results in Appendix Section B suggest that Policy DNN learns the optimal policy function better than Causal DNN of Farrell et al. (2021).

3.1 Defining the Policy DNN estimator

I espouse a more direct approach to learn the policy function $d(x_i)$ from the profit estimator in Equation (1). However, the profit-maximizing estimator to the policy function $d(x_i)$ is not a smooth function so it cannot be optimized over computationally. To remedy this, I propose surrogate function and show that the policy function from the surrogate is consistent to the optimal policy function. These results build on Bartlett et al. (2006) and Zhao et al. (2012). Bartlett et al. (2006) proposed excess risk bounds for convex surrogate functions and Zhao et al. (2012)

[11] The literature has long discussed using maximin and minimax regret for statistical decision making (Savage, 1951) with the former deemed too conservative (Manski, 2004).

analyzed policy learning a surrogate loss function for support vector machines and proposed a outcome weighted classifier approach for policy learning.

My approach combines the results from Zhao et al. (2012) and Farrell et al. (2020) by using deep neural nets to directly learn $d(x_i)$ using a surrogate function. I leverage the inference engine in Farrell et al. (2020) to conduct inference for expected profits under the learned optimal policy function.

Specifically, I propose a surrogate for the optimal policy function $d^*(x_i) = \mathbf{1}\{\beta(x_i) > \frac{c}{m}\}$ to be

$$\tilde{d}(x_i) = \tilde{d}(m\tilde{\beta}(x_i) - c) = f(m\tilde{\beta}(x_i) - c), \tag{6}$$

where $f : \mathbb{R} \to [0, 1]$ is a monotonically increasing Lipschitz function that maps the representation of $\beta(x_i)$, denoted as $\tilde{\beta}(x_i)$, to a relaxed version of the decision rule $\tilde{d}(x_i)$. For example, I can consider $f(z) = \frac{\tanh(z)+1}{2}$ or a sigmoid function as possible functions for $\tilde{d}(x_i)$.

I emphasize that $\tilde{d}(x_i)$ is the surrogate or relaxed version of the optimal targeting $d^*(x) = \mathbf{1}\{\beta(x_i) > \frac{c}{m}\}$. To convert the surrogate to a targeting policy, I need to threshold it. To give an example of a thresholded targeting rule, $\mathbf{1}\{\tilde{d}(x_i) > 0.5\}$ is a targeting policy that targets customer i if $\tilde{d}(x_i)$ is greater than 0.5 and does not target otherwise.

I form the surrogate profit function by first plugging in $\tilde{d}(x_i)$ for $d(x_i)$ in Equation (1):

$$\hat{\Pi}(\tilde{d}) = \sum_{i=1}^{n} \frac{W_i}{e(x_i)} \pi_i(1)\tilde{d}(x_i) + \frac{1 - W_i}{1 - e(x_i)} \pi_i(0)(1 - \tilde{d}(x_i)). \tag{7}$$

Then I use a DNN to represent $\tilde{\beta}(x_i)$. I use the results from from Farrell et al. (2020) to estimate $\tilde{\beta}(x_i)$ as the parameter of interest while using the negative of Equation (7) as the loss function to be minimized.

Figure 3 provides a visualization of the architecture. In contrast, Causal DNN proposed in Farrell et al. (2020) has the two structural parameters, $\alpha(x)$ and $\beta(x)$, in an additional parameter layer. Here, I use the surrogate for the policy function $\tilde{d}(x)$, which combines with the treatment indicator W and the outcome variable Y to form the surrogate loss function (Equation 7). I establish my setup satisfies the general framework proposed by Farrell et al. (2020).

Proposition 7. *(Suitable loss) The negative profit loss*

$$\mathscr{L} = -\sum_{i=1}^{n} \frac{W_i}{e(x_i)} \pi_i(1)\tilde{d}(x_i) + \frac{1 - W_i}{1 - e(x_i)} \pi_i(0)(1 - \tilde{d}(x_i))$$

to estimate $\tilde{\beta}(x_i)$ satisfies Assumption 1 in Farrell et al. (2020).

I verify the proposition holds with the proposed loss function in Appendix Section A.1 which establish Lipschitz and curvature and conditions for the loss function. I further choose a DNN

architecture for $\tilde{\beta}(x)$ to satisfy Assumption 2 in Farrell et al. (2020).[12] This setup allows me to leverage the results from Theorem 1 in Farrell et al. (2020) to estimate $\tilde{\beta}(x)$ from the DNN. I then use the DNN estimates to produce a targeting policy and then conduct inference around expected profits from following the targeting policy.

I show that the sign of $\mathbf{1}\{m\tilde{\beta}(x) - c > 0\} = \mathbf{1}\{\tilde{d}(x) > 0.5\}$ from the surrogate approach will be consistent to the sign of $\mathbf{1}\{m\beta(x) - c > 0\} = d^*(x)$, which is the optimal policy function, in the population.[13] Because the final targeting policy depends only on the sign of $m\beta(x) - c$, individuals are targeted if and only if the sign is positive. The surrogate approach will provide a consistent targeting policy to that of the two-step approach, which first estimates $\hat{\beta}(x)$ and then forms the targeting policy. In essence, my procedure produces a targeting policy that is consistent for the optimal policy function in the population.

Proposition 8. *(Sign consistency of the surrogate) The surrogate policy function* $\mathbf{1}\{\tilde{d}(x) > 0.5\}$ *produces the same targeting rule as the optimal policy function* $d^*(x) = \mathbf{1}\{\beta(x) > \frac{c}{m}\}$. *In other words, the targeting policy from* $\tilde{d}(x)$ *is sign consistent to that of* $d(x)$.

I provide the proof in Appendix Section A.1 and where I adapt the results from Zhao et al. (2012) for my loss function. As a direct result of Proposition 8, the profits generated from the surrogate policy function will be equivalent to the profits generated from the optimal policy function. The proof is also provided in Appendix Section A.1.

Corollary 9. *Expected profits from the targeting policy generated from* $\tilde{d}(x)$ *are consistent for the expected profits generated from* $d^*(x)$ *or* $E[\Pi(d^*(x))] = E[\Pi(\mathbf{1}\{\tilde{d}(\tilde{\beta}(x)) > 0.5\})]$.

In summary, I have shown that I can use a surrogate for the policy function, use a DNN to estimate parameters of the surrogate policy function ($\tilde{\beta}(x)$), and then generated the optimal policy function from the surrogate function. The thresholded policy function from the surrogate ($\mathbf{1}\{\tilde{d}(x) > 0.5\}$) will be sign consistent to the optimal targeting policy ($d^*(x) = \mathbf{1}\{m\beta(x) > c\}$) and the profits generated from the two policies will also be consistent in the population. Appendix Section B contains a Monte Carlo simulation study comparing Policy DNN and the Causal DNN from Farrell et al. (2021).

3.2 Inference for Policy DNN

I now shift my focus to inference around the profits generated from the Policy DNN targeting policy. I provide an overview of the main theorem and leave the proof and the formal statement to Appendix Section C.

[12] Assumption 2 adds constraints on the smoothness of the approximating function class.

[13] *Sign consistency* in this setting implies the proposed policy function's targeting rule is consistent to the optimal targeting rule. In the binary treatment case for the surrogate $\tilde{\beta}(x)$, when $m\tilde{\beta}(x) - c > 0 \iff m\beta(x) - c > 0$, the individual should be targeted ($d^*(x) = 1$). Further when $m\tilde{\beta}(x) - c < 0 \iff m\beta(x) - c < 0$, the individual should not be targeted ($d^*(x) = 0$).

Theorem 10. *(Inference for Policy DNN). Under some mild conditions,*

$$\sqrt{n}\left(\hat{\Pi}(\tilde{d}) - \Pi(d^*)\right) \xrightarrow{d} N(0, V)$$

for finite V defined in Equation 18 in Appendix Section C.

The theorem is formally stated and proved in Appendix Theorem 15. I now provide a sketch of the proof.

Expanding the left hand side in Theorem 10, I obtain three terms

$$\sqrt{n}\left(\hat{\Pi}(\tilde{d}) - \Pi(d^*)\right) = \underbrace{\sqrt{n}\left(\hat{\Pi}(\tilde{d}) - \Pi(\tilde{d})\right)}_{(1)} + \underbrace{\sqrt{n}\left(\Pi(\tilde{d}) - \Pi(\mathbf{1}\{\tilde{d} > 0.5\})\right)}_{(2)} + \underbrace{\sqrt{n}\left(\Pi(\mathbf{1}\{\tilde{d} > 0.5\}) - \Pi(d^*)\right)}_{(3)}.$$

The first term represents the difference between the sample surrogate profits and the population surrogate profits. With my loss function (Equation 7) and Proposition 7, I use results from Farrell et al. (2020) to show

$$\sqrt{n}\left(\hat{\Pi}(\tilde{d}) - \Pi(\tilde{d})\right) \xrightarrow{d} N(0, V)$$

for finite variance V (Equation 18 in Appendix Section C), which is the variance of the influence function. Because I learn the optimal targeting policy directly from the data, I can conduct inference around the surrogate profits at the firm's optimal targeting strategy. Further, I use an envelope theorem argument to cancel out the correction term of the influence function. In practice, this provides an efficiency gain visualized in the Monte Carlo simulations (Appendix Section B); Policy DNN has smaller standard errors than Causal DNN around their targeting policies' accuracy rate to the ground truth.

The second term represents the difference between the surrogate profits and the profits from a targeting policy formed by thresholding the surrogate targeting policy. I impose a margin assumption (Assumption 14) dictating how apart $\tilde{d}(x_i)$ and $\mathbf{1}\{\tilde{d}(x_i) > 0.5\}$ can be and use it to show

$$\sqrt{n}\left(\Pi(\tilde{d}) - \Pi(\mathbf{1}\{\tilde{d} > 0.5\})\right) \to o_p(1).$$

I provide a concrete example of the margin assumption for the $\tilde{d}(k, x) = \frac{\tanh(k\beta(x)+1)}{2}$ function with scale parameter k. Here, $k \asymp \ln(n)$ is needed for the margin assumption to hold.

The third term represents the difference between the profits from the thresholded surrogate policy to the profits from the optimal targeting policy. I use Corollary 9 to show

$$\Pi(\mathbf{1}\{\tilde{d} > 0.5\}) - \Pi(d^*) = 0.$$

Combining these results, I have

$$\sqrt{n}\left(\hat{\Pi}(\tilde{d}) - \Pi(d^*)\right) \xrightarrow{d} N(0, V) + o_p(1) + 0 \simeq N(0, V),$$

which supplies the main theorem. A discussion of the implications of the margin assumption is provided before the formal statement of Assumption 14. Lastly, I provide two remarks around the key technical contributions at the end of Appendix Section C.

4 Forming a comprehensible class of policies

This section constructs a class of comprehensible policies. I focus on targeting policies that can be represented by a sentence such as

"Target customer if she lives in Chicago and she has not bought in the last thirty days."

This sentence has two conditional clauses "if she lives in Chicago" and "[if] she has not bought in the last thirty days" that are linked by the "and" logic operator. Thus the customer will be targeted if both clauses are true and will be not targeted otherwise. This class of comprehensible policies is denoted by $\mathcal{F}_{\text{Comp}}$.

I describe what makes a targeting policy comprehensible in Section 4.1, construct the comprehensible policies class in Section 4.2, and show how I can engineer the clauses from standard recency, frequency, and monetary (RFM) marketing data in Section 4.3.

4.1 What is comprehensibility?

The sentence, "Target customer if she lives in Chicago and she has not bought in the last thirty days" is comprehensible in the sense it is *transparent* and *complete*. For completeness, the sentence is the targeting policy being implemented by the firm. There is a one-to-one mapping between the explanation provided to the implemented policy. For transparency, the firm can state the sentence to the customer or regulator, and the sentence is easy to explain and parse. The number of clauses captures the complexity of the targeting policy and a sentence with more unique clauses will necessarily be more complex.

To ensure these comprehensible policies remain understandable, I limit the number of clauses that can be used in the sentence. Miller (2019) suggests that explanations with that are too long are not understandable. While a sentence with more than five clauses may be grammatically correct, it becomes difficult for a sales representative to explain it to the customer. To capture this restriction, I focus on comprehensible targeting policies with at most five clauses to ensure the comprehensible policy is *conversational*.

The comprehensible policies are constructed to embed *contrastive* explanations (Lipton, 1990).

The explainable AI literature uses the idea of contrastive explanations as a core component of the explanation itself (Biran and Cotton, 2017; Halpern and Pearl, 2005a,b; Miller, 2019; Mothilal et al., 2019).[14] In the running example, a contrastive explanation means that since customers are only targeted if they live in Chicago and have not bought in the last thirty days, the negation of either conditional clause—"if they do not live in Chicago" or "have bought in the last thirty days"—implies the customer would have not been targeted. By providing the customer with the targeting rule and counterfactual states in which the customer would not be targeted, the firm is able to explain the comprehensible targeting in a manner that the customer can understand.

Fundamentally, the choice of this class of comprehensible policies is motivated by the emphasis on *transparent*, *complete*, and *conversational* targeting policies that embed contrastive explanations. This class of policies deviates from those proposed in the interpretable AI literature, which suggests constructing long rule lists (Angelino et al., 2018; Rudin, 2019), expansive decision trees (Weiner, 1980), or large flow charts (Cawsey, 1992).[15] As a result, the comprehensible policy class will be simpler than those used in interpretable AI: Large rule lists and decisions trees may be transparent and complete, but they are not conversational.

4.2 Comprehensible policies

I consider the class of sentences that consistent of clauses linked by logic operators as the proposed class of policies. I call this class of models *comprehensible policies*. In essence, I want to capture targeting rules that can be represented as a coherent and grammatically correct sentences which say,

$$\begin{aligned} &\text{"Target if customer } i \text{ has } \textbf{this}\text{"} \\ &\text{"Target if customer } i \text{ has } \textbf{this} \textit{ and } \textbf{that}\text{"} \\ &\text{"Target if customer } i \text{ has } \textbf{this} \textit{ or } \textbf{that} \textit{ and } \textbf{not this}\text{"} \end{aligned} \tag{8}$$

where "**this**" and "**that**" are two conditional clauses described by covariates and are linked by the logic operators "*and*," *"or*," and "*xor*".[16] The clauses can be negated so "**this**" can be formed into "**not this**" by prefixing the clause with the *"not"* operator.[17] In this setup, the complexity of

[14] In western philosophy, Spinoza holds the strongest position about explanation, namely that everything is part of a causal chain of explanation. While some explanations are difficult to find, Spinoza equates denying these explanations with seeking refuge in "the sanctuary of ignorance" (Spinoza, 1985). This account refutes contrastive explanation: each explanation is already determined by others. As a contemporary Spinoza commentator writes, "our place in the world is simply the way in which we are explained by certain things and can serve to make intelligible—i.e., explain—certain other things" (Della Rocca, 2008).

[15] A separate strand of the interpretable AI literature used in the marketing literature incorporates interpretable structures in black box models (Rai, 2020; Fong et al., 2021; Wang et al., 2022).

[16] The operator "xor" represents the exclusive or which captures A or B but not A and B.

[17] The running example "Target customer if she lives in Chicago and she has not bought in the last thirty days" has two total conditional clauses and they are linked by the logic operator "and." The conditional clauses themselves

the explainable policy is defined as the number of conditional clauses used and is denoted as ℓ.

While comprehensible policies are simple to state, optimizing over the sentences can be combinatorially difficult. If there are k distinct possible clauses and three logic operators for a sentence of length ℓ (that uses ℓ total clauses), then the total possible combinations for that sentence is $3^{\ell-1}(2k)^{\ell}$, which is exponential in the number of possible clauses.[18] I provide an example that enumerates the different combinations for a sentence with two clauses below.

Example 11. (Comprehensible policies with two clauses) Consider two clauses A and B. For a sentence with $\ell = 2$ clauses, there are twelve combinations where A is in the first clause spot and B is in the second clause spot.

A and B	A and not B
A or B	A or not B
A xor B	A xor not B

not A and B	not A and not B
not A or B	not A or not B
not A xor B	not A xor not B

The comprehensible policies can be expressed in the most general form in logic trees (Schwender and Ruczinski, 2010).[19] Logic trees take in binary covariates and generate a sentence that uses the covariates as clauses and links them using logic operators. The elements of the tree are the binary covariates and the logic operators are *and*, *or*, and *not*. The logic trees can then be collapsed into a sentence of the form in Example 8. The comprehensible policies are simpler versions of decisions trees and a comprehensible policy with ℓ clauses can be represented by a decision tree of ℓ layers. The procedure to generate decision trees from the comprehensible policies is described in Appendix Section D.

4.3 Generating clauses

Comprehensible policies require clauses with binary values, and I show how I can generate these clauses from standard RFM marketing data in this section. The clauses depend on the data available to the firm, and I assume the data takes on the standard tabular form where each column of the data is understandable. In the RFM case, a data column that describes sales in the past 12 months is understandable to mangers, customers, and regulators.

are "she lives in Chicago" and "she has not bought in the last thirty days."

[18] For a two clause sentence, allowing for all permutations of the clauses (e.g., including sentences such as, "Target A or not A") leads to $3^1 4^2 = 48$ different combinations.

[19] Logic trees can be directly mapped to decision trees as is shown in Lemma 18.

Binary covariates in the data do not require further processing since they can be represented by clauses directly. For example, a binary covariate that is 1 if a customer is "a new user" and 0 otherwise is represented by the clause "is a new user."

Categorical covariates are expanded to binary covariates and then turned into the clauses. For example, the type of mobile phone that a customer has can be converted into binary variables ("has an iPhone," "has an Android," and "has a flip phone") and these are converted into clauses directly. The clauses from the categorical variables are more expressive, as a clause like "does not have a flip phone" would capture the customer having either an "iPhone" or an "Android".

For continuous covariates, I discretize them into three bins that are each represented by a binary indicator. I implement this discretization with RFM data in mind. To provide a concrete example, Figure 4 shows the unconditional distribution for the past November sales covariate from the empirical application in the upper panel and the conditional distribution for nonzero past November sales in the lower panel. I first note that 97% of the observations are zero so I first construct the clause "has *zero* past November sales."

I then look at the conditional distribution of past November sales for the customers that spend during November in the lower panel. The median value is \$129.99 so I classify all customers above this value to be "high" and the customer below this value and above zero to be "low." Their respective clauses will be "has *low* past November sales among spenders" and "has *high* past November sales among spenders." I follow this procedure in the empirical application for the RFM dataset. In other settings, researchers can construct three clauses of low, medium, or high values by cutting up the continuous variable at the empirical quantiles.[20]

Naturally, data scientists can transform the data into different representations or embeddings, but I consider the transformed variables largely to be not interpretable. At an extreme, a data scientist can estimate $\hat{\alpha}(x_i)$ and $\hat{\beta}(x_i)$ using a black box method.[21] Then the comprehensible policy would naturally recover the optimal policy function ($d^*(x_i)$) that says, "Target customer if she has $\hat{\beta}(x_i) > c/m$." Such a targeting policy would have the structure of the targeting sentence but is not transparent since the process of attaining $\hat{\beta}(x_i)$ is opaque.

5 Optimal comprehensible policy

I now find the optimal comprehensible policy among the class of comprehensible policies that I constructed in Section 4. I denote this class of comprehensible policies as $\mathcal{F}^{\ell}_{\text{comp}}$ and now make the dependence on the number of clauses ℓ explicit. Recall that ℓ represents the complexity of the comprehensible policy since a sentence with more clauses can provide a finer partition of the customer base. As an example, $\mathcal{F}^{2}_{\text{comp}}$ represents this class of comprehensible policies that are

[20] An even more general approach can also discretize the continuous covariates into smaller bins based on deciles or more precise quantiles but these will significantly expand the total number of clauses to search over in the optimization step.

[21] The terms $\hat{\alpha}(x_i)$ and $\hat{\beta}(x_i)$ follow the notation in Section 3.1.

covered by sentences with $\ell = 2$ clauses.

I now focus on optimizing over this class of policies for a given ℓ value using policy learning. Specifically, I use the direct empirical welfare maximization framework (Kitagawa and Tetenov, 2018) where I directly find the optimal comprehensible policy $d^*_{\text{comp}}(x) \in \mathcal{F}^{\ell}_{\text{comp}}$ that maximizes the sample profit estimator in Equation 1,

$$d^*_{\text{comp}}(x) = \underset{d' \in \mathcal{F}^{\ell}_{\text{comp}}}{\arg\max}\, \hat{\Pi}(d'). \tag{9}$$

I propose two procedures that directly maximize profits over this class of comprehensible policies. The first is brute force optimization, which guarantees the globally optimal solution but can be computationally intensive. The second is a greedy algorithm that is computationally tractable but may result in a locally optimal solution. I then provide inference around the optimal comprehensible policy by leveraging the results from Kitagawa and Tetenov (2018).

5.1 Brute force algorithm

I first propose the brute force optimization approach where I first enumerate over all possible comprehensible policies and then choose the one that yields the highest expected profits. The algorithm's step are detailed in Algorithm 1.

Algorithm 1 Brute force optimization

Setup: Number of clauses ℓ:

1. Discretize the p covariates into q pieces to get the clauses
2. Combinatorially iterate through all targeting policies combinations of the pq clauses and logic operators
3. Choose the policy with ℓ clauses that maximizes profits

To outline the algorithm, consider the problem of finding a $\ell = 3$ clause sentence that has the structure:

Target if {A} <and/or/xor> {B} <and/or/xor> {C}

where {A}, {B}, {C} are the clauses in the sentence and <and/or/xor> are the logic operators. To find the optimal comprehensible policy with three clauses, the brute force approach would first enumerate all possible clauses in {A}, all possible clauses in {B}, and all possible clauses in {C} (and those in {not A}, etc.) as well as the possible logic operators, <and/or/xor>, between the clauses. This approach is computationally intensive because the number of all possible combinations grows exponentially in the total number of clauses ℓ. For p covariates discretized into q candidate clauses this leads to $3^{\ell-1}(2pq)^{\ell}$ different combinations. To put that number into context, for thirty variables discretized into three candidate clauses there are approximately 5.8 million different combinations to search over for a three-clause sentence.

Practically, I can use the brute force algorithm to enumerate all possible targeting policies when $\ell \leq 2$ and the dataset itself is not too large. I do so in the empirical application to provide a comparison to the greedy algorithm's solution.

5.2 Greedy algorithm

Since the brute force algorithm is not computationally tractable in many scenarios, I propose a greedy version of the algorithm. The algorithm is computationally feasible but may only find a local optimal solution that generates lower profits than that of the globally optimal solution.

Greedy algorithms are commonly used in the marketing literature (Lilien et al., 1992) and also used in estimating decisions trees in the statistics literature (Breiman, 1984). Since the local optimal solution may be suboptimal to the global solution in generating profits, the results form the greedy algorithm can be viewed as a lower bound of the profits from the globally optimal comprehensible policy. Algorithm 2 details the algorithm's steps.

Algorithm 2 Greedy algorithm

Setup: Number of clauses ℓ:

1. Discretize the p covariates into q pieces to get the clauses
2. Find the single best clause that maximizes profits
3. For $l \in \{2, \dots \ell\}$:
   (a) Iterate all clause and logic operator combinations while holding the $l-1$ clauses and logic operators fixed
   (b) Choose the combination that maximizes the profits

To outline the greedy algorithm, consider the case of finding a $\ell = 3$ clause sentence that has the structure:

$$\text{Target if } \underbrace{\{\text{A}\}}_{(1)} \underbrace{\text{<and/or/xor> } \{\text{B}\}}_{(2)} \underbrace{\text{<and/or/xor> } \{\text{C}\}}_{(3)}$$

where {A}, {B}, {C} are the clauses in the sentence and <and/or/xor> are the logic operators. The greedy algorithm breaks up the combinatorially difficult problem by solving it piece by piece. In the example, the greedy algorithm would first find the best single clause sentence or optimize over the possible clauses {A} (the first piece). Then, it would hold the one clause targeting rule (the solution to {A}) fixed and then find the best logic operator and {B} combination (the second piece). Lastly, it would hold the two clause targeting rule (the solution to {A} <and/or/xor> {B}) fixed and find the best logic operator and {C} combination (the third piece).

The greedy algorithm evaluates a smaller set of combinations of comprehensible policies since it does not enumerate over all possible combinations of clauses and logic operators. As a result, it searches over $6(\ell-1)pq\ell$ combinations for a comprehensible policy with ℓ clauses and p covariates discretized in q candidate clauses. For thirty variables discretized into three candidate clauses, there are approximately 1.1 thousand combinations to search over for a three-clause

sentence, which is many orders of magnitude smaller than the brute force approach's 5.8 million combinations.

In the empirical application, I use the greedy algorithm to solve for the optimal comprehensible policy. Since the combinatorial space is dramatically reduced, searching for the optimal comprehensible policy with the greedy algorithm for $\ell = 10$ takes around a minute while using the brute force algorithm for only $\ell = 3$ has an estimated run time of over three weeks. Both algorithms were implemented using the R package for `torch` as the backend (Falbel and Luraschi, 2023).

5.3 Inference for optimal comprehensible policies

To conduct inference around the optimal comprehensible policy, whether it is found though the brute force algorithm or the greedy algorithm, I can adopt results from Kitagawa and Tetenov (2018).[22] I am using the empirical welfare maximization framework, albeit with a more specific function class $\mathcal{F}^{\ell}_{\text{comp}}$ with a fixed ℓ. I summarize how I implement their theoretical results in this framework and further theoretical details can be found in their paper.

Kitagawa and Tetenov (2018) provide a minimax optimal rates for policy learning via empirical welfare maximization. They study expected welfare regret of a candidate policy function's welfare to the optimal policy function's welfare. Their minimax rates around expected regret provide worst case guarantees for finding the optimal policy function and the rates scale at $K\sqrt{VC(d)/n}$ where $VC(d)$ represents the Vapnik-Chervonenkis (VC) dimension of the policy function, K is a constant, and n is the number of observations. The VC dimension of the class of policy functions inherently needs to be finite. To adapt the results, I define the class of policy functions to be the class of comprehensible policies $\mathcal{F}^{\ell}_{\text{comp}}$ with ℓ fixed.[23]

In this setting, I only need rates for statistical inference that are a different objective than minimax rates. Confidence intervals can be attained around the estimated policy function using the empirical process bootstrap outlined in Algorithm B.1 in Appendix B of Kitagawa and Tetenov (2018). I follow this procedure to attain inference around the optimal comprehensible policy in the empirical application.

6 Projecting down the black box

In this section, I show how comprehensible policies can be formed by projecting down a black box policy and demonstrate that doing so will yield less profitable comprehensible policies than finding them directly from the data. I call this projection down procedure the *ex post* approach

[22] To complete their setup, I would need to further assume the outcome variable (Y) is bounded in addition to the standard assumptions of Assumptions 1, 2, and 3.

[23] The number of clauses ℓ need to be fixed and finite in order for the VC dimension of the comprehensible policy to be finite (Lemma 19 in Appendix Section D).

and call the procedure for finding the policy directly from the data (Section 5) the *direct* approach.

I provide an analytical justification for why the *direct* approach will generate more profitable comprehensible policies than the *ex post* approach in Section 6.1. I then show how I can attain inference around the projected down optimal comprehensible policy by recentering the empirical process results from Kitagawa and Tetenov (2018) in Section 6.2.

The literature in explainable AI (XAI) studies a local approximation of a black box algorithm where they project down the black box to a simpler, more explainable model (Biran and Cotton, 2017; Miller, 2019). However, I will show that using the projection down procedure to form optimal comprehensible targeting policies leads to less profitable comprehensible policies than those formed directly from the data (as in Section 5).

To project down the black box, I first consider a profit loss function from two candidate targeting policies $d(x)$, $d'(x)$. I consider profits as the outcome of interest because they are a direct measure of producer surplus while other metrics such as AUC or classification accuracy to the black box method do not have a direct economic interpretation.

I choose the absolute profit difference as the loss between two policies $d(x)$ and $d'(x)$, expressed as

$$\begin{aligned}\mathcal{L}(d, d') &= |\hat{\Pi}(d) - \hat{\Pi}(d')| \\ &= \sum_{i=1}^{n} \underbrace{\mathbf{1}\{d(x_i) \neq d'(x_i)\}}_{\text{Classification loss}} \underbrace{\left| \frac{W_i}{e(x_i)}\pi_i(1) - \frac{1 - W_i}{1 - e(x_i)}\pi_i(0) \right|}_{\text{Weight}}. \end{aligned} \tag{10}$$

The first term, $\mathbf{1}\{d(x_i) \neq d'(x_i)\}$, can be interpreted as the classification loss.[24] The second term, $|\frac{W_i}{e(x_i)}\pi_i(1) - \frac{1-W_i}{1-e(x_i)}\pi_i(0)|$, can be interpreted as the classification weight for customer i. This loss function is similar to the outcome weighted learning setup from Zhao et al. (2012) and the weighted classifier setup in Zhang et al. (2012). Intuitively, the loss is nonzero for an individual if the two policies differ (the classification loss) and the difference is scaled by the absolute profit difference from disagreeing for that individual. Further, the expected weight for the observation is $E\left[\frac{W_i}{e}\pi_i(1) - \frac{1-W_i}{1-e}\pi_i(0)\right] = E[\pi_i(1) - \pi_i(0)]$ or the expected individual level profit difference.

Using the absolute profit difference loss function in the XAI framework, I aim to find the explainable policy $\xi(x)$ that solves the equation

$$\check{\xi}(x) = \arg\min_{d'} \mathcal{L}(d, d') + \Omega(d') \tag{11}$$

where $\mathcal{L}(d, d')$ is the difference in the profits of the two policy functions d, d' as in Equation 10 and $\Omega(d')$ represents the complexity of explainable policy d'. The generalized loss function in Equation 11 trades off between minimizing the loss between d and d' and minimizing the

[24] Appendix Section E provides the derivation of the loss function.

complexity of d'.

Equation 11 captures the general framework used the XAI literature where an explainable, transparent box model d' is used to approximate the black box model d (Biran and Cotton, 2017; Miller, 2019). Methods in this domain include LIME (Ribeiro et al., 2016) and SHAP (Lundberg and Lee, 2017), which both provide a local linear decomposition of the black box to its covariates.

I take a step further and directly embed explainability as a constraint in the problem. Instead of considering a trade-off between model complexity and the performance of the explainable policy, I choose a level of complexity and then find the best performing explainable policy. Consider an explainable class of policies $D(\Omega_l)$ that all have the same complexity Ω_l, which is indexed by some constant l. I now want to find the best performing explainable policy $d' \in D(\Omega_l)$ and solve the equation

$$\xi_{D'}(x) = \underset{d' \in D(\Omega_l)}{\arg\min}\, \mathcal{L}(d, d') \tag{12}$$

where I embed the level explainability as a direct constraint in the optimization problem.

I let d be the estimated Policy DNN policy function $d^*_{DNN}(x)$ from Section 3. For the explainable policy, I use the class of comprehensible policies constructed in Section 4 and the complexity parameter l can just be the number of clauses ℓ in the sentence. To unify the notation, I define $\mathcal{F}^{\ell}_{\text{comp}} = D(\Omega_l)$ as the class of comprehensible policies of length ℓ.

6.1 Projecting down the black box is not profit-maximizing

I now show that projecting down the black box policy to find the comprehensible policy following Equation 12 leads to a less profitable comprehensible policy than directly optimizing the comprehensible policy by maximizing sample profits (Section 5). I first compare the two objectives used in the two approaches,

$$\min_{d' \in \mathcal{F}_{\text{comp}}} \left| \max_{d \in \mathcal{F}_{DNN}} \hat{\Pi}(d) - \hat{\Pi}(d') \right| \quad \text{vs.} \quad \max_{d' \in \mathcal{F}_{\text{comp}}} \hat{\Pi}(d'). \tag{13}$$

The left-hand side represents the profit level from forming a projected down optimal comprehensible policy using Equation 12 and the right-hand side represents the profit level from forming the optimal comprehensible policy directly from the data. When finding the optimal comprehensible policy directly, $d^*_{\text{comp}}(x) = \arg\max_{d' \in \mathcal{F}_{\text{comp}}} \hat{\Pi}(d')$, the property of the maximization operator says the *direct* approach will find the maximum profits for the comprehensible policy class. Thus, the comprehensible policy from the *ex post* approach will generate weakly less profits than the direct approach.

The intuition behind the result is lies in the structure of the loss functions for the two approaches. Because the ex post approach minimizes the loss between the comprehensible policy and the black box, when the black box does not classify customers well, the ex post approach will also not do well. In contrast, the direct approach learns the policy straight from the data as

it does not depend on the black box policy's results. In fact, if the black box does better than the comprehensible policy in generating a more profitable policy for all individuals in the data, then the two approaches will be the same.

I formalize the last statement to show that the two objectives in Equation 13 will be equal if the black box policy outperforms the comprehensible policy for all customers. This is formalized in the following assumption and I denote the individual-level profits from targeting policy $d(x_i)$ as $\pi(d(x_i))$.

Assumption 12. *The profits generated from the policy DNN are weakly greater than that of the comprehensible policy, so* $\pi(d_{DNN}(x_i)) \geq \pi(d(x_i)), \forall d \in \mathcal{F}^{\ell}_{comp}$.

Under Assumption 12,

$$\begin{aligned}\min_{d' \in \mathcal{F}^{\ell}_{\text{comp}}} \left| \max_{d \in \mathcal{F}_{DNN}} \hat{\Pi}(d) - \hat{\Pi}(d') \right| &= \min_{d' \in \mathcal{F}^{\ell}_{\text{comp}}} \left(\max_{d \in \mathcal{F}_{DNN}} \hat{\Pi}(d) - \hat{\Pi}(d') \right) \\ &= \min_{d' \in \mathcal{F}^{\ell}_{\text{comp}}} -\hat{\Pi}(d') \\ &= \max_{d' \in \mathcal{F}^{\ell}_{\text{comp}}} \hat{\Pi}(d'),\end{aligned}$$

where I used the assumption in the first line to remove the absolute value as $\max_{d \in \mathcal{F}_{DNN}} \hat{\Pi}(d) = \hat{\Pi}(d^*_{DNN}(x)) \geq \hat{\Pi}(d')$ for $d' \in \mathcal{F}^{\ell}_{\text{comp}}$ and I used that d' does not show up in the first term to get to the second line.

I interpret these results as cautionary guidance for marketing managers on forming optimal comprehensible targeting strategies. The XAI approach of projecting down the black box provides an approximative model to the black box. However, for finding the optimal comprehensible policy to be implemented in practice, firms should use direct empirical welfare maximization, or maximizing expected profits in the data directly, to find the optimal comprehensible policy.

6.2 Inference for projected down optimal comprehensible policies

To conduct inference around the *ex post* approach, I recenter the empirical process results from Kitagawa and Tetenov (2018) for the loss function in Equation 12. By showing their results apply to my framework, I can conduct inference for the *ex post* comprehensible policy.

I first define the notation for the theorem and suppress the dependence of the comprehensible policy targeting rule d on the length of the sentence ℓ for notational simplicity. These results holds for a finite number of clauses ℓ, implying the comprehensible policy has finite VC dimension from Appendix Lemma 19. I first define $\pi(d)$ to be the individual-level profits from targeting policy d,

and then define

$$\check{d}^*(x) = \arg\max_{d \in \mathcal{F}_{\text{comp}}} E_P\left[\left|\pi(d^*_{DNN}) - \pi(d)\right|\right]$$
$$\hat{d}_{ex}(x) = \arg\min_{d \in \mathcal{F}_{\text{comp}}} E_n\left[\left|\pi(d^*_{DNN}) - \pi(d)\right|\right],$$

where the first line has the expectation taken over the population distribution P and the second line has the expectation taken over the sample analog. The second line produces the *ex post* optimal comprehensible policy $\hat{d}_{ex}(x)$ and is equivalent to finding the optimal comprehensible policy by solving the sample analog of Equation 12.[25]

I impose sample splitting for estimating the $d^*_{DNN}(x)$ for the optimal black box policy in one data sample, for finding the optimal comprehensible policy $\hat{d}_{ex}(x)$ in another data sample, and conducting inference in the last data sample. I further define $\Gamma(d) = E_P\left[\left|\pi(d^*_{DNN}) - \pi(d)\right|\right]$ to be the profit loss for the targeting rule d evaluated in the population and $\Gamma_n(d) = E_n\left[|\pi(d_{DNN}) - \pi(d)|\right]$ to be the profit loss for targeting rule d in the sample.

Theorem 13. *(Uniform convergence rate of ex post comprehensible policies) Under Assumptions 1, 2, and 3 and the assumption that the outcome variable (Y) is bounded, for a hypothesis space of comprehensible policies with ℓ clauses ($\mathcal{H} = \mathcal{F}^{\ell}_{comp}$) that has bounded VC dimension ($VC(\mathcal{H}) < V < \infty$),*

$$\sup_P E_P\left[\Gamma(\check{d}^*) - \Gamma(\hat{d}_{ex})\right] \leq C\sqrt{\frac{V}{n}} = O_p\left(1/\sqrt{n}\right).$$

I provide the proof of Theorem 13 in Appendix Section A.2. The technical implication is that the rate for the difference in the profit loss scales at $\sqrt{VC(d_{ex})/n}$ for *ex post* comprehensible policy d_{ex}.[26] This is an upper bound and the lower bound results from Kitagawa and Tetenov (2018) can be similarly recentered to provide minimax rates for the profit differences. Inference can be attained around this *ex post* approach by using the empirical process bootstrap.

7 Empirical application

In this section, I provide an application to promotions management for a durable goods retailer as a proof of concept of the methodological framework. I first show that Policy DNN is the best performing black box to establish the optimal black box policy benchmark $d^*_{\text{DNN}}(x)$. Then, I find the optimal comprehensible policy $d^*_{\text{comp}}(x)$ and compare the targeting differences between the two targeting policies. I denote the profit difference of the Policy DNN to the comprehensible

[25] The formal definition of $\pi(d)$ is in Equation 22 in Appendix Section E.

[26] Since the rates are of $O_P(1/\sqrt{n})$, they are not fast enough to avoid sample splitting with the data used to estimate the Policy DNN; a $o_p(1/\sqrt{n})$ rate or faster is needed to avoid asymptotic bias when using the influence function for inference (Fisher and Kennedy, 2021; Kennedy, 2022). However, the same sample can be used if the bias correction term for $C\sqrt{V/n}$ from Theorem 13 is estimated fast enough and adjusted for in the confidence interval.

policy as the cost of explanation of implementing the comprehensible policy.[27] I then show that finding the optimal comprehensible policy directly produces a more profitable targeting policy than finding it by projecting down the black box targeting policy.

I use the second IMSI Durable Goods dataset from Ni et al. (2012) that contains a price promotion randomized control trial (RCT) for a durables goods store in 2003. The items available are mainly electronics and they encompass a range of products from small ticket to large ticket items.[28] The price promotion is a \$10-off coupon that is valid on the next purchase in the store.

The RCT contains 176,961 customers, and the treated customers were sent a promotion with probability 50%. The control group was not mailed any promotion. The dataset contains approximately 150 recency, frequency, and monetary (RFM) covariates that describe the customers' past behavior with the firm. The outcome of interest is sales during December 2003 (the promotional period) and the price promotion was mailed to the customers before December 2003.

To check for evidence of the covariate balance and the overlap assumption, I run a logistic regression to check if I can statistically predict the treatment variable of getting the promotion with the RFM data. I find that only the intercept value is statistically significant in the regression. Figure 5 shows the density of estimated propensity scores for the treated and the not treated groups from the logistic regression. The two densities essentially fully overlap which suggests that overlap and covariate balance hold in the data. Further checks for covariate balance can be found in Ni et al. (2012).

These results suggest the RCT was run correctly so the unconfoundedness and overlap assumptions should hold in the data. I further make the assumptions that customers used the coupon on their next possible purchase, there's no gaming of coupons, and there are no spillover effects from the mailed promotions to satisfy the stable unit treatment value assumption (SUTVA).

With the three standard assumptions satisfied, I can estimate the average treatment effect (ATE) of the price promotion on December 2003 and find the ATE to be 2.68 with a 95% confidence interval of $(1.63, 3.73)$. This result suggests that there is a statistically significant effect of the price promotion on December sales.

From the sales data, I also see that only 3.6% of all customers purchase during December. Further, conditional on purchase, the median spend size is \$149.99. There are a handful of individuals in the data who spend over five thousand in the store. Since I did not collect the data, I am not sure if these are outliers or errors in the data. In my subsequent analysis, I drop those that spend more than \$800 at the store (the top 0.29% of spenders) in the data.[29]

I impose profit margins $m = 45\%$ and cost of mailing $c = 37¢$ to complete the setup. The latter

[27] This framework is more general than shown in this application. A similar analysis can be performed for any black box policy class and any comprehensible policy class.

[28] To provide a concrete example, a sample small ticket item is something like a drip coffee maker and a sample large item is something like a refrigerator.

[29] I further motivate this data cleaning procedure by assuming that if a customer decides to spend a few thousand at the store than they are less influenced by the \$10-dollar off promotion to make the purchase because it is effectively a smaller percentage off the base price.

is the price of mailing a letter at USPS during the time period. Alternative numbers for the profit margin can be used in the framework and they can even be set to vary by customer covariates.

I first randomly split the data 80/20, estimate the models with 80% of the data, and evaluate different black box methods' targeting policies out of sample using 20% of the data.[30] I use the Policy DNN, Causal Forest, and Lasso black box methods. The Policy DNN follows the setup from Section 3. I use a DNN architecture of three hidden layers each with 12 rectified linear units (ReLU) nodes, a ADAM optimizer with a learning rate of 0.005, an early stopping criterion (following the training procedure described in Appendix Section B) after 2000 epochs, and the surrogate policy function $f(z) = \frac{\tanh(z+1)}{2}$. The Policy DNN learns the optimal targeting policy directly from the data and I denote the Policy DNN targeting rule as $d^*_{DNN}(x)$.

I use the Causal Forest and Lasso methods as proxies for the standard approach used in the literature. The Causal Forest is commonly used as a state of the art procedure in applied economics and marketing literatures (Wager and Athey, 2018). The Causal Forest bootstrap aggregates the Causal Trees from Athey and Imbens (2016). The Lasso is a popular machine learning method when linearity of the baseline model and the heterogeneous treatment model is assumed (Hastie et al., 2015; Taddy, 2019). Both the Causal Forest and Lasso first estimate $\hat{\beta}(x)$ which is then plugged in optimal targeting policy function. These two methods provide benchmark black box models that are commonly used in the literature.

With the black box targeting policies I can evaluate the expected profits under the targeting policy using Equation 1.[31] Table 1 provides the out of sample individual expected profits from these black box models as well as from the blanket targeting policy where everyone is sent the promotion. I see that the Policy DNN ($3.02) does better than both the Causal Forest ($2.74) and Lasso ($2.70) procedures in generating profits. All black box methods perform better than the blanket mailing policy ($2.42).

I interpret the profit gap between the Policy DNN and the Causal Forest as the difference from the policy learning approach where the optimal targeting policy is learned directly from the data to the standard approach that first learns the heterogeneous treatment effects and then plugs them into the optimal treatment rule. Less information is needed to learn the optimal policy function directly than to learn the heterogeneous treatment effects from the data. By focusing on only learning what I need for targeting policy, the procedure seems to perform better in the dataset.

I then interpret the small profit gap between the Causal Forest and the Lasso as reflecting that the heterogeneous treatment effects can be well approximated by a sparse linear functional. Lastly, the gap between the Lasso and blanket mailing represents the difference between doing any personalization using a black box algorithm and doing no personalization. This profit gap is

[30] I denote evaluating models in the 80% training data as in sample evaluation and in the 20% validation data as out of sample evaluation.

[31] The standard errors for Lasso and Causal Forest represent implementation uncertainty around the profits generated from Lasso and Causal Forest targeting policies. The implementation uncertainty does not account for the first-stage estimation uncertainty in the $\beta(x_i)$ parameters.

as large as the gap between the Policy DNN and the Causal Forest methods.

I form the optimal comprehensible targeting policy following Section 4 and Section 5. I denote the optimal comprehensible policy as $d^*_{\text{comp}}(x)$, and the three-clause optimal comprehensible policy using the greedy optimization algorithm is:

Target customer if she:

(

1. has bought *high* amount of items during Christmas over the last two years

and

2. did **not** have *high* spending during Spring over the last two years

)

or

3. has *low* spending during last years holiday mailer promotional period.

The sentence is read left to right and the parenthesis highlight the effect of the "and" logic operator.[32] The customers that are not described by the targeting policy are not targeted. The construction of the clauses with *low* and *high* descriptors from the RFM dataset follow the discussion in Section 4.3.

I interpret this targeting sentence as largely targeting two major segments of customers. The first group is described by the first two clauses. These are customers who buy a lot during Christmas but not a lot in the Spring, or people who focus their spending during the holiday period. Since the outcome of interest is December sales, these are individuals who spend a lot during the target period and may be more price sensitive.

The second group are those who are on the RFM customer list but have not spent a lot during the promotional December period the prior year. I consider this group to be customers who have spent at the store in years before but then either forgot about the store or went to another store the previous year. Then, the optimal comprehensible policy is suggesting retargeting these customers to incentivize them to come back to the store. Customers in either of these two groups will be targeted by the optimal three-clause comprehensible policy.

[32]The greedy algorithm finds the clauses left to right so the optimal one-clause comprehensible policy is just the first clause in the three-clause optimal policy and the optimal two-clause policy is the first two clauses in the three-clause optimal policy. The brute force algorithm is computationally tractable for the optimal one-clause and two-clause policies and matches the greedy algorithm's optimal policies. I use the greedy optimization algorithm to form optimal comprehensible policies for the remainder of the paper. Since the greedy algorithm can lead to a suboptimal, local solution, I can treat the profits from the optimal comprehensible policy as a lower bound in profits to that of the globally optimal comprehensible policy.

7.1 Do the targeting policies differ?

I first compare the targeting differences between Policy DNN $d^*_{DNN}(x)$ and the optimal comprehensible policy $d^*_{\text{comp}}(x)$, which I call the *direct* method in the figures and tables.[33] Figure 6 shows the targeting percentage of the customers for $\ell \in \{1, \dots 10\}$ number of clauses in sample.[34] Policy DNN targets 18.7% of customers. I see that with one clause, the optimal comprehensible policy targets more than Policy DNN. However, with three clauses, it gets to the closest in targeting percentage to Policy DNN. Adding more clauses in this setting appears to reduce the overall targeting percentage of the comprehensible policy.

I provide a visual demonstration of the targeting policy differences with a two-clause optimal policy in Figure 7. The axes represent two RFM covariates in the dataset: *spring sales over the last 24 months* and *items bought during Christmas over the last 24 months*. The jittered grey circular points represent the customers in the raw dataset. The jittered green triangular points are those customers that Policy DNN targets. Since Policy DNN learns a higher order representation of the data, it is not clear what its targeting rule is when visualized on these two dimensions. In contrast, the two-clause optimal comprehensible policy is represented by the pink rectangle, and everyone covered by the rectangle will targeted by the comprehensible policy. The two-clause comprehensible policy is the just the first two clauses of three clause targeting policy, "Target if customer has high amount of items during Christmas over the last two years and did not have high Spring spending over the last two years."

I now consider the three-clause optimal comprehensible policy and provide the confusion matrix in Table 4 in sample. I see that the three-clause policy targets 18.1% of the customer base and there is a 78% overlap between Policy DNN and the three-clause targeting policy. More specifically, the two policies agree in targeting 6.5% of the customer and not targeting 69.7% of the customers. The three-clause comprehensible policy targets 11.6% of customers who are not targeted by Policy DNN and Policy DNN targets 12.2% of the customers not targeted by the three-clause comprehensible policy. Since the overtargeting and undertargeting differences are relatively balanced when comparing the comprehensible policy to Policy DNN, it seems that the comprehensible policy is capturing similar variation as Policy DNN but cannot personalize as finely due to its comprehensibility constraint.

7.2 Cost of explanation

I now quantify the profit differences from Policy DNN and the optimal comprehensible policy and denote this gap as the *cost of explanation.* In Figure 8, I visualize the expected individual profits for the optimal comprehensible policy by the number of clauses $\ell \in \{1, \dots, 10\}$. I focus on the *direct*

[33] I make the distinction between the *direct* and *ex post* method of finding the optimal comprehensible policy in Section 6.

[34] These percentages are similar out of sample since the data is independent to splitting rule since it is a random 80/20 data split.

method for now and see that in sample as the number of clauses increases, the comprehensible policy does better. This result captures the fact that with more clauses the comprehensible policy can make more partitions of customers and better personalize the targeting policy.

Out of sample, the profits increase with more clauses but decrease slightly after eight clauses. These results suggests that with nine or ten clauses, the direct method can overfit in the training data. I also see that out of sample, the gap between the optimal comprehensible policy and Policy DNN is smaller because Policy DNN is likely to be overfitting in sample. After three clauses, the comprehensible policy's expected profits is not statistically significantly different from that of Policy DNN.

I evaluate the cost of explanation, or the profit difference of the two policies, for the three-clause optimal comprehensible targeting policy in Table 3. Per person, the optimal comprehensible policy generates \$2.80 in expected profits and the out of sample cost of explanation is 22 cents. This implies that implementing the three-clause optimal comprehensible policy policy instead of the Policy DNN black box policy will lead to a 22 cents loss in expected profits per person. This is a 7% loss compared to the Policy DNN profits and the comprehensible policy provides a 16% gain in profits compared to a blanket mailing policy.

Further, the difference between Policy DNN and the blanket mailing is 60 cents so using the optimal comprehensible policy provides a 38 cents gain over blanket targeting, or recouping $(60-22)/(60) = 63\%$ of the expected losses from implementing the blanket mailing policy. Thus, the firm does notably better by implementing the three-clause comprehensible policy than a blanket mailing policy.

Lastly, it is worth noting that the optimal comprehensible targeting policy (\$2.80) performs better than the standard Causal Forest (\$2.74) and Lasso (\$2.70) procedures (Table 1) in out of sample expected profits. While the profit differences are not statistically significant, they reflect the advantage of policy learning over the standard two-step approach of first estimating the heterogeneous treatment effects and then forming the optimal targeting policy. In general, I anticipate the optimal comprehensible policy to do well if there is not a lot of heterogeneity in the data and the true optimal policy is relatively simple. However, if there is a lot of heterogeneity in customers' reaction to the treatment or enough data to learn from, then a more complicated black box model (i.e., a neural net or forest) using the standard two-step approach should outperform the optimal comprehensible policy.

7.3 Projecting down the black box

I apply the *ex post* approach from Section 6 to project the black box model down to a comprehensible policy. I use the same 80/20 data split to evaluate the methods but split the 80% training sample again: I use the DNN trained on 60% of the data, project down the DNN and form the *ex post* comprehensible policy in 20% of the data, and conduct inference in the last 20% of the data.

Figure 8 plots the individual expected profits for the *direct* and *ex post* methods. I see that

the *direct* method outperforms the *ex post* method both out of sample. In expectation, the direct method generates \$2.80 per person while the ex post procedure only generates \$2.70 per person. The *ex post* procedure's targeting policy is,

Target customer if she:

1. has *high* Christmas spending over the last two years

   and

2. did **not** have *high* spending during Spring over the last two years

   and

3. has *high* total spending over the last three years.

Comparing this targeting policy that of the *direct* approach, I see that their first clauses are similar, but their second and third clauses as well as their logic operators are different. The differences imply that the two approaches are capturing different partitions of the customer base to target.

I further interpret the targeting policy from the *ex post* approach as targeting a different group of customers than the targeting policy from the *direct* approach. The *ex post* approach mainly targets one segment of customers. It targets customers who buy a lot during Christmas, curb their spending during the Spring, and are heavy spenders at the store. These are the customers who consistently exhibit significant spending patterns at the store, particularly during the holiday season.

Table 3 shows the cost of explanation for the two methods with three clauses. I see that the cost of explanation for the *ex post* procedure (32 cents) is higher than that of the the *direct* procedure out of sample (22 cents). These results match the profit differences visualized in Figure 8 and suggest the *direct* approach does better in generating profitable comprehensible policies.

Overall, this application empirically verifies the analytical results from Section 6.1. Optimal comprehensible policies should be found directly from the data (Section 5) rather than found by projecting down from the black box policy (Section 6).

8 Discussion

In this section, I study how firm managers can use the the proposed framework to analyze their decision to stay with a black box algorithm or to move to a comprehensible policy when forming targeting policies for their marketing mix. Circling back to the framework overview in Figure 1, I revisit the trade-off between profits and comprehensibility. Managers compare the complete producer surplus generated by the two policies, or

$$\underbrace{\Pi_{DNN} - R_{DNN} + B_{DNN}}_{\text{Black Box}} \text{ vs. } \underbrace{\Pi_{comp} - R_{comp} + B_{comp}}_{\text{Comprehensible Policy}},$$

to make the decision. I first highlight three components of the complete producer surplus to study the manager's problem of whether to stay with the black box or move to a comprehensible policy.

The first term (Π_{DNN}, Π_{comp}) represents the short term profit loss from moving away from the black box to the optimal comprehensible policy. The proposed framework constructs the black box and optimal comprehensible policy, and the analysis in Sections 7 quantifies the short term profit loss, or the cost of explanation, in the empirical example.

The second term (R_{DNN}, R_{comp}) represents the regulatory penalty that the firm faces while implementing its chosen targeting algorithm. If enforced, right-to-explanation laws will penalize black boxes and but not comprehensible targeting policies. Thus, the expected regulatory penalty will be higher for the black box policy ($R_{DNN} > R_{comp}$).

The third term (B_{DNN}, B_{comp}) represents the long-term effects of offering a comprehensible policy. Consumers can firms can benefit from comprehension as it can build better brand equity for the firm and makes implementing the targeting policy easier by its representatives. If comprehension leads to long-term benefits or costs, then it should be considered by firm managers when making the decision.

In Section 8.1, I focus on the first two terms that balance profitability and with the expected regulation penalty ($\Pi_{DNN} - R_{DNN}$ vs. $\Pi_{comp} - R_{comp}$). I leverage the proposed framework to study the effect of right-to-explanation law on firm's profits as the firm moves to an optimal comprehensible policy. This calibration exercise quantifies the economics damages that right-to-explanation legislation imposes on firms if enforced.

In Section 8.2, I then discuss the possible long-term effects of comprehensible targeting policies. Even though I do not have the data to study the benefits of offering a comprehensible policy in my empirical application, I outline potential factors that firms should consider when forming their decision.

8.1 Effect of GDPR's "right to explanation" on firms

GDPR suggests customers have a "right to explanation," which would require firms operating in the European Union to provide "an explanation of the decision involved in full" and be "given access to meaningful information about the logic involved" by the firm's human representative (European Commission, 2016).[35] If the right-to-explanation clause is enforced, the penalties for violating GDPR are the larger of 4% of global revenues or 20 million Euros. Although how exactly the right-to-explanation laws apply to firms is a subject of ongoing legal debate (Wachter et al., 2016), it is prudent for forward-looking firms to consider their potential implications. With my

[35] Specifically, the GDPR legislation states:

"The data subject to have the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached in full"

"[and] given access to meaningful information about the logic involved"

(European Commission, 2016)

framework, I can evaluate their impact on profits as a firm transitions from a black box targeting policy to an optimal comprehensible targeting policy to comply with the right-to-explanation legislation.

From the empirical application, I showed the cost of explanation ($\Pi_{DNN} - \Pi_{comp}$) was 22 cents per person for the three-clause optimal comprehensible policy. For a customer basis of 10 million, this implies 2.2 million dollars of lost profits due to moving away from the black box targeting policy.

GDPR litigation and enforcement has been publicly focused on multinational technology firms, but all firms under EU jurisdiction must abide by the law. To put the lost profits in perspective, I provide the following stylized calibration exercise. I assume the firm in the empirical application is small enough for the 20 million euros to be the penalty, a one-to-one exchange rate of euros to dollars, and a perceived enforcement rate of 10%. With these simplified assumptions, the expected penalty of noncompliance is 2 million dollars. In the framework, I set R_{DNN} as 2 million dollars and R_{comp} to be zero. The firm now compares the expected profit loss ($\Pi_{DNN} - \Pi_{comp} = 2.2$ million dollars) to the expected regulatory penalty ($R_{DNN} - R_{comp} = 2$ million dollars).

From a regulatory perspective, the firm may or may not abide with the right-to-explanation clause since the firm loses 2.2 million dollars from moving away from the black box targeting policy but faces an expected penalty of 2 million dollars from GDPR. As such, the regulator may consider increasing the lump sum penalty or increasing the perceived enforcement rate to ensure full compliance by firms.

On a flip side, ensuring compliance with the right-to-explanation clause has a nontrivial impact on the firm's bottom line. The expected profit loss of 2.2 million dollars is for one month of sales; scaled up annually, that is 26.4 million dollars in lost profits if the firm ran a promotional strategy every month. As a result, the impact of right-to-explanation laws can be quite substantial and regulators should consider these downstream impacts as they seek to implement data and privacy laws in other jurisdictions.

This calibration exercise can be readily extended to capture more complex settings. The regulatory penalty can depend on the complexity of the targeting policy as more complex targeting rules can face higher enforcement rates.[36] Then, R_{DNN} can increase in the complexity of the black box. For comprehensible policies that are have more than five clauses and are not conversational, R_{comp} can be set to increase with the number of clauses in the comprehensible targeting policy.[37] Other extensions to this exercise can be added to tailor it to different scenarios.

Lastly, I showed the cost of explanation the calibration study for one specific class of comprehensible policies that I proposed in Section 4. This class of targeting sentences conservatively complies with the right-to-explanation clause. Naturally, other classes of comprehensible tar-

[36] Lambin and Raizonville (2023) study an incomplete information game between firms and regulators where firms can offer explainable algorithms or black box algorithms and regulators can choose to audit the firm for compliance with right-to-explanation laws.

[37] Non-conversational comprehensible policies can make it difficult for the firm's human representative to explain the policy to customers in order to comply with the "right to explanation" clause.

geting policies can be considered in the general framework and the calibration exercise can be repeated with different classes of comprehensible targeting policies and black box targeting policies.

8.2 Long-term effects of comprehensibility

In many settings, offering a comprehensible policy can lead to downstream benefits for the firm. Customers can learn from a comprehensible policy but cannot learn from a black box. If the policy benefits customers, they can learn how to get the treatment again which can build further brand equity with the firm. On the flip side, customers who were excluded from the treatment can understand why they were left out and what they need to do in order to get the treatment.

For the firm's perspective, implementing a comprehensible policy is simpler than implementing a black box policy. If the firm representatives need to implement the targeting policy, then it is easier for them to train and follow a comprehensible policy. For example, training salespeople to follow a comprehensible policy will be simpler than doing so for a black box policy. It is also easier for these salespeople to explain these to customers.[38] Comprehensible policies are also easier to diagnose by the firm, and the firm can audit the targeting policy to ensure it does not use information from protected classes.

However, for some settings, comprehensibility of the targeting policy may not be beneficial for customers and firms. Customers may even dislike a transparent explanation of the algorithmic decision policy in certain settings. For example, in online dating, an explanation of the matchmaking algorithm may draw ire from customers. For firms operating in a competitive environment, offering comprehensible targeting policies can give competing firms insight about the firm's profitable customer base and its decision making. In equilibrium, it may not be beneficial to the firm to reveal such information to its competitors.

As a result, managers need to consider the long-term benefits and costs of comprehensibility for the black box policy (B_{DNN}) to that of the comprehensible policy (B_{comp}). In many cases, it seems that the benefits for comprehensibility are positive and are long-term ($B_{comp} > B_{DNN}$). Future research can explore the long-term effect of comprehensibility for firms and customers.

9 Conclusion

Data and privacy regulations like GDPR and CCPA are swiftly gaining traction worldwide. With GDPR, regulators in Europe now ask for a "right to explanation" where a black box algorithm's decisions need to be explainable to customers by the firm's human representatives. They have increasingly cracked down on firms violating data and privacy laws, and proposals to expand the

[38] The human-computer interaction literature studies how human agents can interact with explanations from black box policies and is overviewed in Chen et al. (2022).

regulation have only increased.[39]

This paper provides a framework for firms to navigate right-to-explanation laws. This framework is composed of both a methodology for firms to optimally form comprehensible marketing policies that comply with right-to-explanation regulation and a cost analysis to quantify the cost of doing so. In service to the framework, I first propose a new black box algorithm, Policy DNN, that combines policy learning and deep neural networks as a new profit-maximizing black box benchmark. This methodological contribution leverages the fact that learning the optimal policy for discrete treatments directly is more efficient than first learning heterogeneous treatment effects and then plugging them into the optimal policy function as done in the current literature.

I then propose a class of comprehensible policies that would satisfy the new regulatory constraints and that takes on the form of sentences. These sentences are conditional clauses linked by logic operators. I further show how to find the optimal, profit-maximizing, comprehensible policy for a sentence of given clause length.

With the established framework, I document how the two targeting policies differ and then quantify the cost of explanation, or the profit loss from implementing the optimal comprehensible policy to the black box policy. I provide an application for sending $10-off promotions for a durable goods retailer. I find that the proposed Policy DNN does better than other standard black box methods and find the cost of explanation to be 22 cents per person for a three-clause optimal comprehensible targeting policy, which is a 7% loss in profits from the optimal black box policy.

I quantify the profit loss that the firm will face from complying with right-to-explanation regulation. In the application, for a basis of 10 million customers, the cost of explanation leads to a 2.2 million profit loss from the firm's promotional strategy. These losses represent the economic damages for the firm from abiding by data and privacy regulation. While GDPR fines have been mainly levied on large multinational technology companies, my framework provides a localized way for any company under its jurisdiction to quantify and evaluate the regulation's impact on its bottom line.

This framework can be extended to capture benefits along with costs. While I only provide a cost analysis of how comprehension in marketing policies acts as a constraint on the firm's personalization and targeting strategies, there may be benefits from comprehension for the firm's customers. Customers appear to have a disdain for algorithmic decisions in certain scenarios (Dietvorst et al., 2015; Dietvorst and Bartels, 2022; Yalcin et al., 2023), and providing them a comprehensible explanation may lead them to foster future goodwill toward the firm. I leave exploring the benefits of comprehension to future research.

More generally, my framework enables firms to assess the impact of practical marketing constraints on their objectives of interest. In my setting, I use profits, or producer surplus, as the

[39]Meta Platforms was fined 1.2 billion euros for not abiding by GDPR rules on May 22, 2023 (European Data Protection Board, 2023). Although the GDPR violation did not specifically pertain to the "right to explanation" clause, it highlights increasing enforcement of GDPR regulations. Consequently, forward-looking firms should factor these regulations into their decision making process.

objective and the constraint is the comprehensibility of the targeting policy. The cost of explanation reflects how this constraint affects the firm's profits. This framework has the flexibility to explore alternative objectives like consumer surplus or total surplus and can accommodate different constraints, such as privacy or fairness considerations.

My paper links the theoretical targeting and personalization literature to what is done in practice by accounting for regulatory constraints. As black box algorithms gain more regulatory scrutiny with increasingly widespread use of generative artificial intelligence (AI) models and with the proposed AI Act (European Commission, 2021), firms need to navigate the regulatory environment if they decide to continue to leverage modern advances in AI for their day-to-day operations. This paper assesses the cost of right-to-explanation legislation for a firm's targeting policies. Future research in evaluating the effects of rapidly expanding data and privacy regulation on firms and customers is encouraged to further bridge theory and practice.

References

Angelino, E., N. Larus-Stone, D. Alabi, M. Seltzer, and C. Rudin (2018): "Learning certifiably optimal rule lists for categorical data," *Journal of Machine Learning Research*, 18, 1–78.

Ascarza, E. (2018): "Retention Futility: Targeting High-Risk Customers Might be Ineffective," *Journal of Marketing Research*, 55, 80–98.

Athey, S. and G. Imbens (2016): "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences of the United States of America*, 113, 7353–7360.

Athey, S. and S. Wager (2021): "Policy Learning With Observational Data," *Econometrica*, 89, 133–161.

Bartlett, P. L., M. I. Jordan, and J. D. McAuliffe (2006): "Convexity, Classification, and Risk Bounds," *Journal of the American Statistical Association*, 101, 138–156.

Biran, O. and C. Cotton (2017): "Explanation and Justification in Machine Learning: A Survey," *IJCAI 2017 Workshop on Explainable Artificial Intelligence (XAI)*, 8–13.

Breiman, L. (1984): *Classification And Regression Trees*, Routledge, 1 ed.

Cawsey, A. (1991): "Generating Interactive Explanations," *Aaai.Org*, 86–91.

——— (1992): *Explanation and Interaction: The Computer Generation of Explanatory Dialogues*, MIT Press.

——— (1993): "Planning interactive explanations," *International Journal of Man-Machine Studies*, 38, 169–199.

Chen, C., S. Feng, A. Sharma, and C. Tan (2022): "Machine Explanations and Human Understanding," .

Chintagunta, P. K., L. Huang, W. Miao, and W. Zhang (2023): "Measuring Seller Response to Buyer-initiated Disintermediation: Evidence from a Field Experiment on a Service Platform," *SSRN Electronic Journal*.

Della Rocca, M. (2008): *Spinoza*, New York, NY, 1 ed.

Dietvorst, B. J. and D. M. Bartels (2022): "Consumers Object to Algorithms Making Morally Relevant Tradeoffs Because of Algorithms' Consequentialist Decision Strategies," *Journal of Consumer Psychology*, 32, 406–424.

Dietvorst, B. J., J. P. Simmons, and C. Massey (2015): "Algorithm aversion: People erroneously avoid algorithms after seeing them err." *Journal of Experimental Psychology: General*, 144, 114–126.

Edwards, L. and M. Veale (2017): "Slave to the Algorithm? Why a Right to Explanation is Probably Not the Remedy You are Looking for," *SSRN Electronic Journal*.

Ellickson, P. B., W. Kar, and J. C. Reeder (2022): "Estimating Marketing Component Effects: Double Machine Learning from Targeted Digital Promotions," *Marketing Science*.

European Commission (2016): "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC," .

——— (2021): "Regulation of the European parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts," .

European Data Protection Board (2023): "1.2 billion euro fine for Facebook as a result of EDPB binding decision," .

Falbel, D. and J. Luraschi (2023): "torch: Tensors and Neural Networks with 'GPU' Acceleration," .

Farrell, M. H., T. Liang, and S. Misra (2020): "Deep Learning for Individual Heterogeneity: An Automatic Inference Framework," .

——— (2021): "Deep Neural Networks for Estimation and Inference," *Econometrica*, 89, 181–213.

Fisher, A. and E. H. Kennedy (2021): "Visually Communicating and Teaching Intuition for Influence Functions," *American Statistician*, 75, 162–172.

Fong, H., V. Kumar, and K. Sudhir (2021): "A Theory-Based Interpretable Deep Learning Architecture for Music Emotion," *SSRN Electronic Journal*.

Gillis, T. B. and J. L. Spiess (2019): "Big Data and Discrimination," *University of Chicago Law Review*, 86, 459–487.

Goodfellow, I., Y. Bengio, and A. Courville (2016): *Deep Learning*, MIT Press.

Halpern, J. Y. and J. Pearl (2005a): "Causes and Explanations: A Structural-Model Approach. Part I: Causes," *The British Journal for the Philosophy of Science*, 56, 843–887.

——— (2005b): "Causes and Explanations: A Structural-Model Approach. Part II: Explanations," *The British Journal for the Philosophy of Science*, 56, 889–911.

Hastie, T., R. Tibshirani, and M. Wainwright (2015): *Statistical learning with sparsity: The lasso and generalizations*, CRC Press, 1 ed.

Hitsch, G. J., S. Misra, and W. W. Zhang (2023): "Heterogeneous Treatment Effects and Optimal Targeting Policy Evaluation," *SSRN Electronic Journal*.

Imbens, G. W. and D. B. Rubin (2015): *Causal Inference for Statistics, Social, and Biomedical Sciences*, Cambridge University Press.

Kallus, N. and A. Zhou (2018): "Policy Evaluation and Optimization with Continuous Treatments," in *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, ed. by A. Storkey and F. Perez-Cruz, PMLR, vol. 84 of *Proceedings of Machine Learning Research*, 1243–1251.

Karlinsky-Shichor, Y. and O. Netzer (2019): "Automating the B2B Salesperson Pricing Decisions: Can Machines Replace Humans and When?" *SSRN Electronic Journal*.

Katsov, I. (2017): *Introduction to Algorithmic Marketing*.

Kennedy, E. H. (2022): "Semiparametric doubly robust targeted double machine learning: a review," .

Kitagawa, T., S. Sakaguchi, and A. Tetenov (2021): "Constrained Classification and Policy Learning," .

Kitagawa, T. and A. Tetenov (2018): "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice," *Econometrica*, 86, 591–616.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan (2017): "Human Decisions and Machine Predictions*," *The Quarterly Journal of Economics*.

Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer (2015): "Prediction Policy Problems," *American Economic Review*, 105, 491–495.

Kleinberg, J., J. Ludwig, S. Mullainathan, and C. R. Sunstein (2018): "Discrimination in the Age of Algorithms," *Journal of Legal Analysis*, 10, 113–174.

Ko, R., K. Uetake, K. Yata, and R. Okada (2022): "When to Target Customers? Retention Management using Dynamic Off-Policy Policy Learning," *SSRN Electronic Journal*.

Künzel, S. R., J. S. Sekhon, P. J. Bickel, and B. Yu (2019): "Metalearners for estimating heterogeneous treatment effects using machine learning," *Proceedings of the National Academy of Sciences*, 116, 4156–4165.

Lambin, X. and A. Raizonville (2023): "From Black Box to Glass Box: Algorithmic Explainability as a Strategic Decision," *SSRN Electronic Journal*.

Lilien, G. L., P. Kotler, and K. S. Moorthy (1992): *Marketing Models*, Prentice-Hall.

Lipton, P. (1990): "Contrastive Explanation," *Royal Institute of Philosophy Supplement*, 27, 247–266.

Little, J. D. C. (1979): "Decision Support Systems for Marketing Managers," *Journal of Marketing*, 43, 9.

Liu, X. (2022): "Dynamic Coupon Targeting Using Batch Deep Reinforcement Learning: An Application to Livestream Shopping," *Marketing Science*.

Luedtke, A. and A. Chambaz (2020): "Performance guarantees for policy learning," *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56.

Lundberg, S. M. and S.-I. Lee (2017): "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Curran Associates, Inc., vol. 30.

Manski, C. F. (2004): "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 72, 1221–1246.

Mbakop, E. and M. Tabord-Meehan (2021): "Model Selection for Treatment Choice: Penalized Welfare Maximization," *Econometrica*, 89, 825–848.

Milgrom, P. and I. Segal (2002): "Envelope theorems for arbitrary choice sets," *Econometrica*, 70, 583–601.

Miller, T. (2019): "Explanation in artificial intelligence: Insights from the social sciences," *Artificial Intelligence*, 267, 1–38.

Mothilal, R. K., A. Sharma, and C. Tan (2019): "Explaining Machine Learning Classifiers through Diverse Counterfactual Explanations," .

Mou, W., M. J. Wainwright, and P. L. Bartlett (2022): "Off-policy estimation of linear functionals: Non-asymptotic theory for semi-parametric efficiency," .

Ni, J., S. A. Neslin, and B. Sun (2012): "Database Submission The ISMS Durable Goods Data Sets," *Marketing Science*, 31, 1008–1013.

Rafieian, O. and H. Yoganarasimhan (2022): "AI and Personalization," *SSRN Electronic Journal*.

Rai, A. (2020): "Explainable AI: from black box to glass box," *Journal of the Academy of Marketing Science*, 48, 137–141.

Ribeiro, M. T., S. Singh, and C. Guestrin (2016): ""Why Should I Trust You?"," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, 1135–1144.

Rossi, P. E., R. E. McCulloch, and G. M. Allenby (1996): "The value of purchase history data in target marketing," *Marketing Science*, 15, 321–340.

Rudin, C. (2019): "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, 1, 206–215.

Savage, L. J. (1951): "The Theory of Statistical Decision," *Journal of the American Statistical Association*, 46, 55.

Schwender, H. and I. Ruczinski (2010): "Logic Regression and Its Extensions," 25–45.

Senoner, J., T. Netland, and S. Feuerriegel (2022): "Using Explainable Artificial Intelligence to Improve Process Quality: Evidence from Semiconductor Manufacturing," *Management Science*, 68, 5704–5723.

Shalev-Shwartz, S. and S. Ben-David (2013): *Understanding machine learning: From theory to algorithms*, vol. 9781107057.

Simester, D., A. Timoshenko, and S. I. Zoumpoulis (2020): "Efficiently Evaluating Targeting Policies: Improving on Champion vs. Challenger Experiments," *Management Science*, 66, 3412–3424.

Smith, A. N., S. Seiler, and I. Aggarwal (2022): "Optimal Price Targeting," *Marketing Science*.

Spinoza, B. (1985): *The Collected Works of Spinoza*, vol. 1, Princeton: Princeton University Press.

Taddy, M. (2019): *Business data science : combining machine learning and economics to optimize, automate, and accelerate business decisions*, McGraw Hill, 1 ed.

van Osselaer, S. M. J. and J. W. Alba (2000): "Consumer Learning and Brand Equity," *Journal of Consumer Research*, 27, 1–16.

Vapnik, V. N. (2000): *The Nature of Statistical Learning Theory*, New York, NY: Springer New York.

Wachter, S., B. Mittelstadt, and L. Floridi (2016): "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation," *SSRN Electronic Journal*.

Wager, S. and S. Athey (2018): "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, 113, 1228–1242.

Wang, T., C. He, F. Jin, and Y. J. Hu (2022): "Evaluating the Effectiveness of Marketing Campaigns for Malls Using a Novel Interpretable Machine Learning Model," *Information Systems Research*, 33, 659–677.

Weiner, J. (1980): "BLAH, a system which explains its reasoning," *Artificial Intelligence*, 15, 19–48.

Yalcin, G., E. Themeli, E. Stamhuis, S. Philipsen, and S. Puntoni (2023): "Perceptions of Justice By Algorithms," *Artificial Intelligence and Law*, 31, 269–292.

Yoganarasimhan, H., E. Barzegary, and A. Pani (2022): "Design and Evaluation of Optimal Free Trials," *Management Science*.

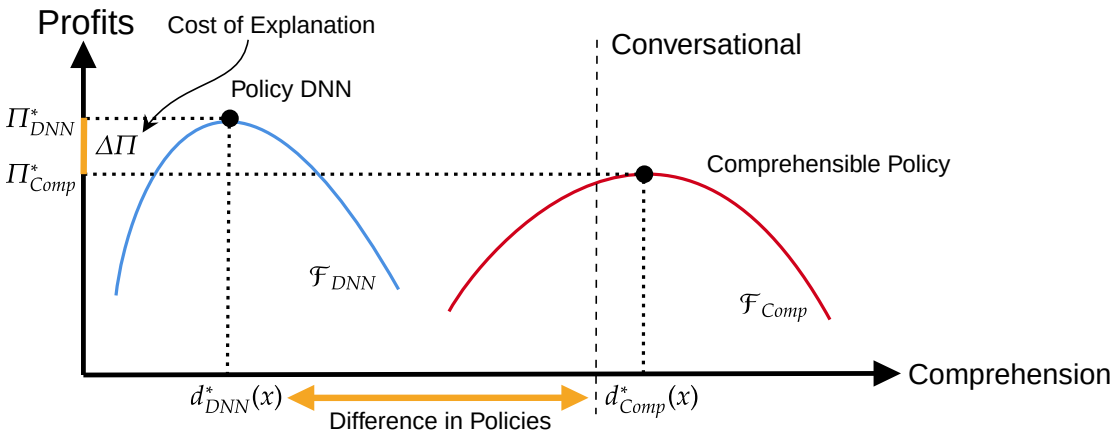
Zhang, B., A. A. Tsiatis, M. Davidian, M. Zhang, and E. Laber (2012): "Estimating optimal treatment regimes from a classification perspective," *Stat*, 1, 103–114.

Zhang, W. W. and S. Misra (2022): "Coarse Personalization," .

Zhao, Y., D. Zeng, A. J. Rush, and M. R. Kosorok (2012): "Estimating Individualized Treatment Rules Using Outcome Weighted Learning," *Journal of the American Statistical Association*, 107, 1106–1118.

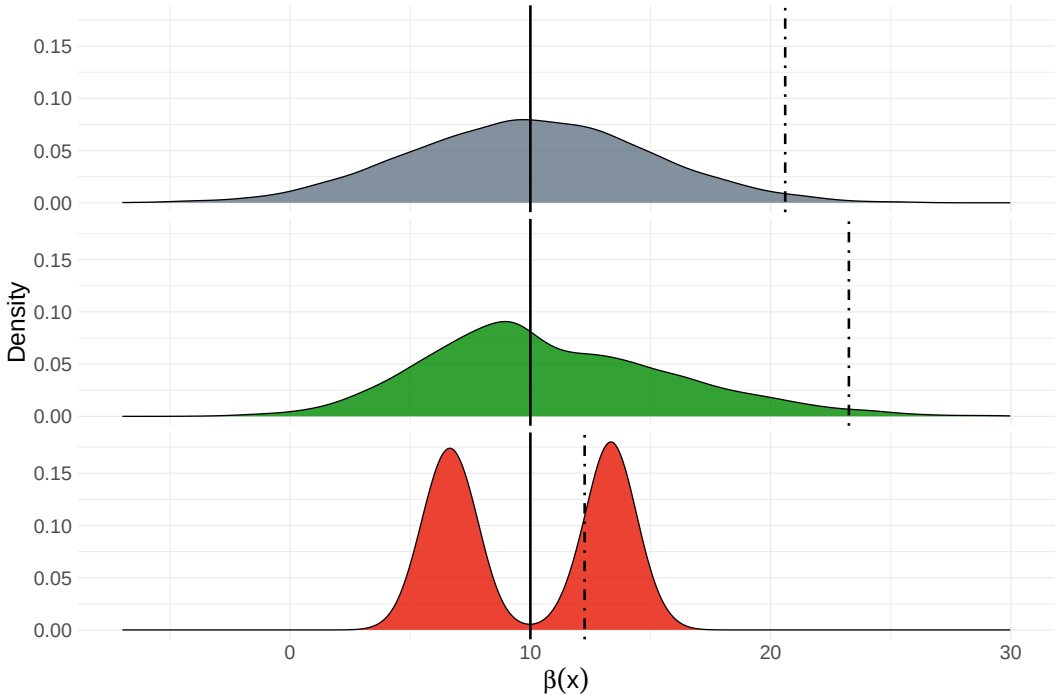
Figures

Figure 1: Methodological overview



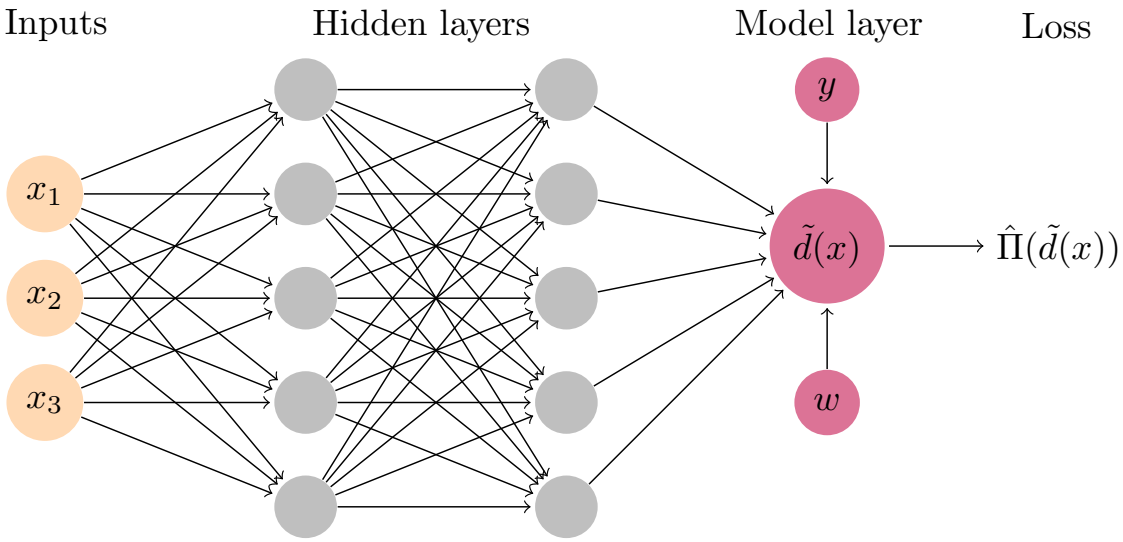
Note: This figure provides an overview of this paper's methodology and it is described in Section 2. The axes demonstrate the tradeoff between profitability and comprehensibility. Comprehensibility can be thought of 1/(Model Complexity). Section 3 proposes a new class of black box algorithms $\mathcal{F}_{DNN}$ for Policy DNN (the blue curve on the left) and finds the optimal targeting policy $d^*_{DNN}(x)$ and its corresponding profits. Section 4 traces out a class of comprehensible policies (the red curve on the right). Section 5 shows how to find the optimal comprehensible targeting policy $d^*_{\text{Comp}}(x)$ among the set of comprehensible policies. In the empirical application, Section 7.1 documents the differences in $d^*_{DNN}(x)$ and $d^*_{\text{Comp}}(x)$ which is denoted here as the differences in policies. Section 7.2 describes the profits differences of $\Delta\Pi = \Pi^*_{DNN} - \Pi^*_{\text{Comp}}$ which is denoted here as the cost of explanation.

Figure 2: Different $\beta(x)$ distributions with the same optimal targeting policy $d^*(x) = \mathbf{1}\{\beta(x) > c/m\}$



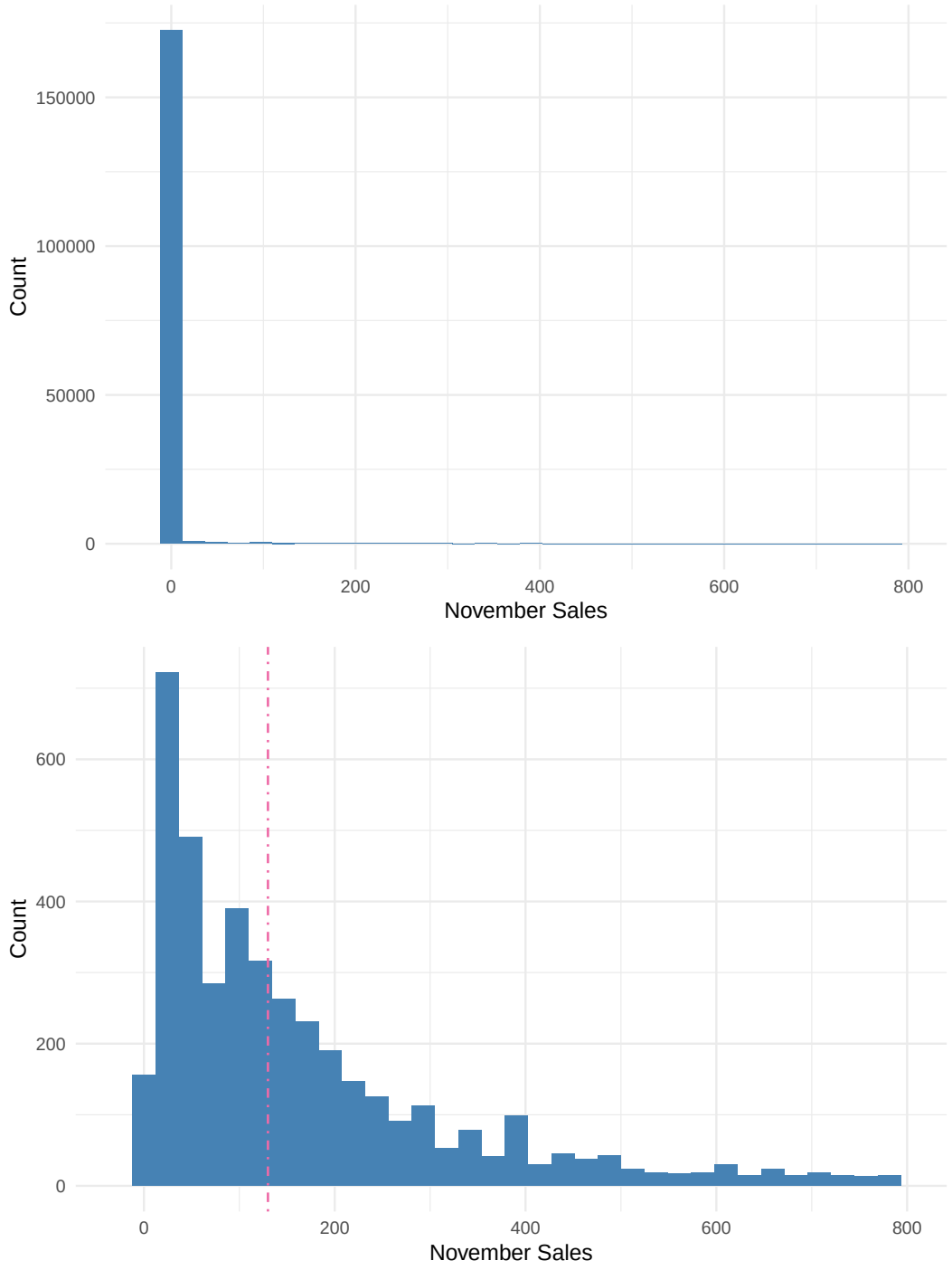
Note: These three CATE or $\beta(x)$ distributions produce the same optimal targeting rule $d^*(x)$. The dashed line represents the $\beta(x)$ for one individual. As long as the line does not cross the targeting threshold (the solid vertical line), the decision rule for that individual does not change and that individual should be targeted.

Figure 3: Policy Deep Neural Net Architecture



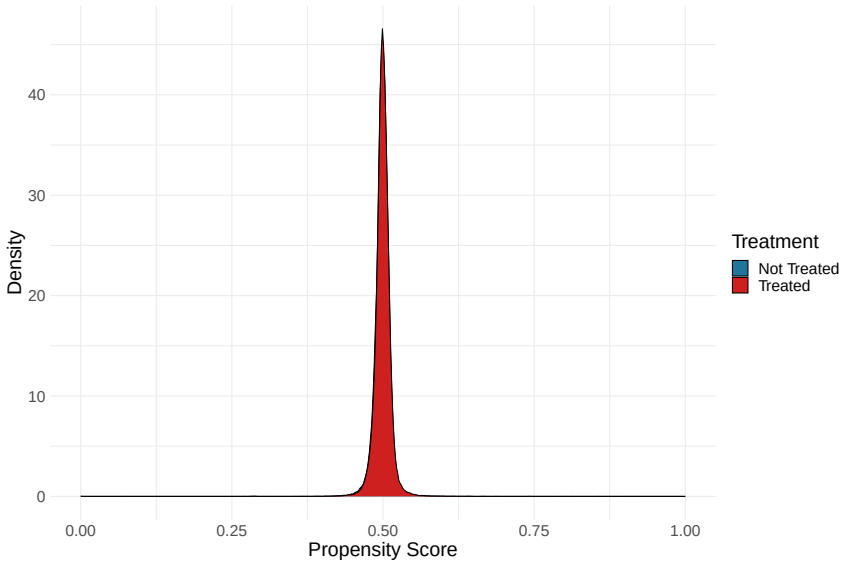
Note: This figure describes the architecture for the Policy DNN black box procedure.

Figure 4: Distribution of past November sales



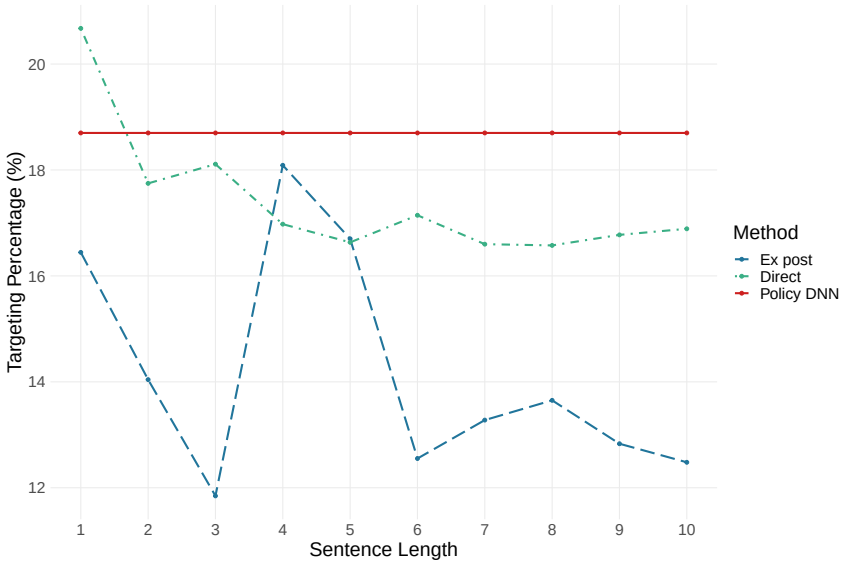
Note: These two figures plot the distribution of past November sales variable from the empirical application's RFM dataset. The upper panel shows the unconditional distribution in which 97% of the customers do not buy anything during November. The lower panel shows the conditional distribution of customers with non-zero spend during November. The median spend for the conditional distribution is $129.99 and is denoted by the vertical dashed line.

Figure 5: Propensity score density



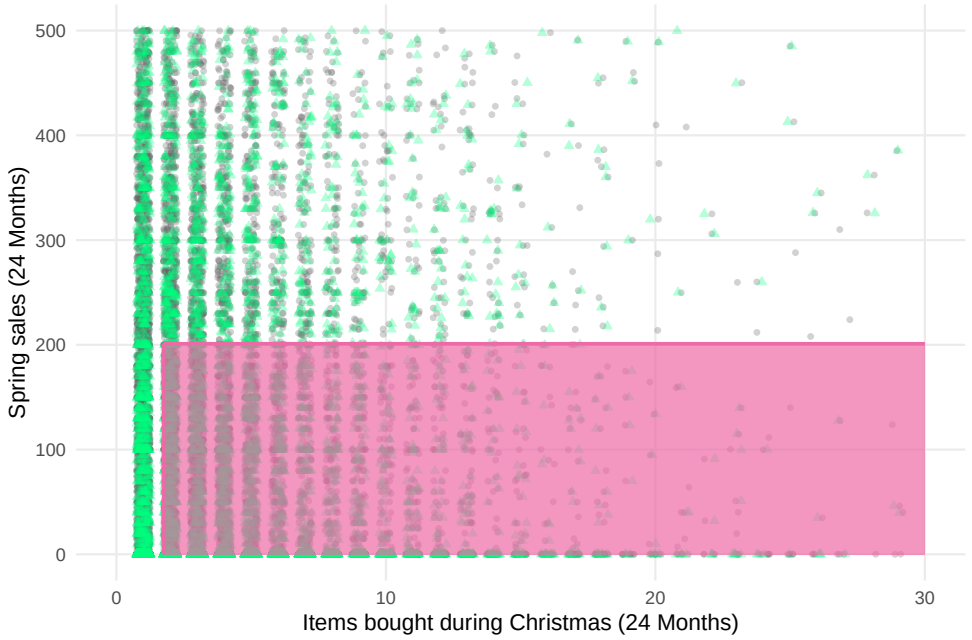
Note: This figure plots the density of the predicted propensity scores for the treated and not treated groups. The propensity score model was estimated using a logistic regression of the treatment indicator on the RFM covariates in the data set. The two densities essentially fully overlap with one another. These results suggest the overlap assumption holds in the data.

Figure 6: Comprehensible policy targeting percentages by clauses



Note: This figure plots the targeting percentage of the optimal comprehensible policies by number of clauses in sample. The *direct* method learns the comprehensible policy from the data and the *ex post* method learns the comprehensible policy from projecting down the Policy DNN to a comprehensible policy. The Policy DNN is represented by the dashed horizontal line and targets 18.7% of customers.

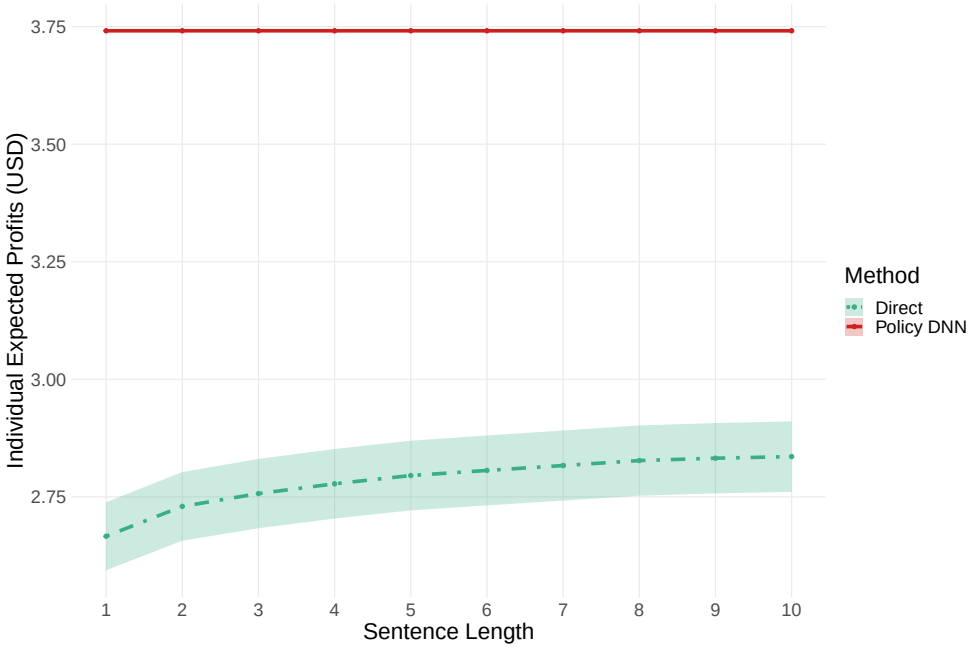
Figure 7: Two clause comprehensible policy vs. Policy DNN policy

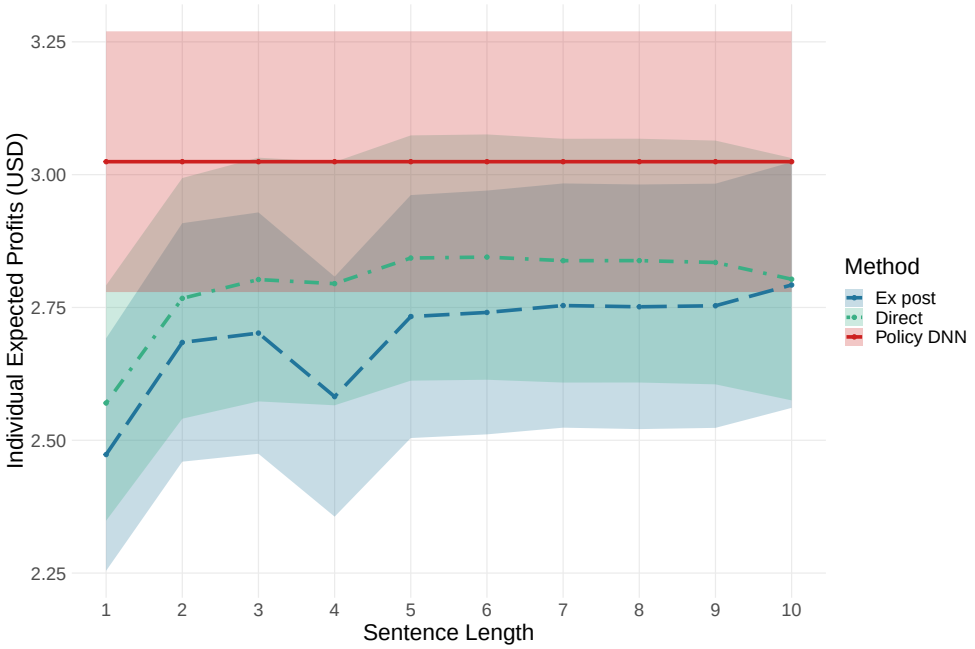


Note: This figure contrasts the two-clause comprehensible targeting policy to the Policy DNN targeting policy. The jittered grey circular points represents the data observations in the raw data for the RFM covariates spring sales (24 months) and items bought during Christmas (24 months). The jittered green triangular points represents which of customers the Policy DNN targets. The points within the pink rectangle will be targeted by the two-clause comprehensible policy, which is, "Target if customer has high amount of items during Christmas over the last two years and did not have high Spring spending over the last two years". High amount of Christmas items over two years is buying two or more items. High spending in spring over the last two years is spending at least \$200. The optimal comprehensible policy is formed using the *direct* method that learns the comprehensible policy from the data.

Figure 8: Individual expected profits by method

(a) In sample expected profits





(b) Out of sample expected profits

Note: These figures plots the expected profits by the number of clauses. The upper panel is the in sample expected profits and the bottom panel is the out of sample expected profits. In both panels, the *direct* method learns the comprehensible policy from the data and the *ex post* method learns the comprehensible policy from projecting down the Policy DNN to a comprehensible policy. The Policy DNN is represented by the dashed horizontal line and there are no standard errors for the Policy DNN in sample. The bands represent one standard error.

Tables

Table 1: Individual expected profits (out of sample)

Method	Individual Profits	SE
Policy DNN	3.026	0.245
Causal Forest	2.743	0.227
Lasso	2.705	0.231
Blanket	2.419	-

Note: These are the out of sample, expected profits for an individual when using the optimal targeting policy $d^*(x)$ from the Policy DNN, Causal Forest, and Lasso black box policies. The blanket targeting policy represents the expected profits if everyone was mailed the promotion. The standard errors for the Policy DNN are computed following Section 3.2. The standard errors for Causal Forest and Lasso represent the implementation uncertainty for the optimal targeting policy.

Table 2: Targeting policy differences

	Policy DNN		
Comprehensible Policy	Targeted	Not targeted	
Targeted	11,539	17,220	28,759
Not targeted	18,156	111,891	130,047
	29,695	129,111	158,806

Note: This a confusion table for the targeting policy differences for the optimal three-clause comprehensible policy and for the Policy DNN policy for the in-sample data. The Policy DNN targets 18.7% of customers while the comprehensible policy targets 18.1% of the customers. The two targeting policies overlap on 78% of customers. The optimal comprehensible policy is formed using the *direct* method that learns the comprehensible policy from the data.

Table 3: Cost of explanation for three-clause comprehensible policies

	In sample		**Out of sample**		**Comprehensible Targeting Policy**
	Profits	CoE	Profits	CoE	Target customer if:
Direct	\$2.76	98¢	\$2.80	22¢	high XMAS items (2Y) **and** *not* high Spring sales (2Y) **or** low spend last year holiday promo
Ex post	-	-	\$2.70	32¢	high XMAS sales (2Y) **and** *not* high Spring sales (2Y) **and** high total sales (3Y)
Policy DNN	\$3.74	–	\$3.02	–	

Note: This table provides the cost of explanation (CoE) of the three-clause optimal comprehensible policies and states their targeting policy. The *direct* method learns the comprehensible policy from the data and the *ex post* method learns the comprehensible policy from projecting down the Policy DNN to a comprehensible policy.

Appendix

A Supporting proofs

A.1 Proofs for Section 3.1

I verify Propositions 7 and 8 and Corollary 9 from Section 3.1 in this section. Their statements are restated for ease of exposition. For ease of notation, the variables without the subscript to i represent their vector quantity (e.g., $\pi(1) = (\pi_1(1), \ldots, \pi_n(1))'$).

Proposition. *(Suitable loss) The negative profit loss*

$$\mathscr{L} = -\sum_{i=1}^{n} \frac{W_i}{e(x_i)}\pi_i(1)\tilde{d}(x_i) + \frac{1-W_i}{1-e(x_i)}\pi_i(0)(1-\tilde{d}(x_i))$$

to estimate $\tilde{\beta}(x_i)$ satisfies Assumption 1 in Farrell et al. (2020).

Proof. Assumption 1 Farrell et al. (2020) in requires both a Lipschitz and the curvature condition for the loss function. I show two below and both derivations rely on using properties of the Lipschitz function $f(\cdot)$. I assume the outcome variable Y is bounded and that constants C_l, c_1, c_2 are both bounded and bounded away from zero. Recall that $\tilde{d} = f$ from the definition of $\tilde{d}(\cdot)$.

I first verify the Lipschitz condition, $|\mathscr{L}(y,t,\beta_0(x)) - \mathscr{L}(y,t,\beta_1(x))| \leq C_l||\beta_0(x) - \beta_1(x)||_2$, holds for the loss function.

$$\begin{aligned}
&\left|\left(\frac{W}{e(x)}\pi(1)f(\beta_1(x)) + \frac{1-W}{1-e(x)}\pi(0)(1-f(\beta_1(x)))\right) - \left(\frac{W}{e(x)}\pi(1)f(\beta_0(x)) + \frac{1-W}{1-e(x)}\pi(0)(1-f(\beta_0(x)))\right)\right| \\
&\qquad = \left|\left(\frac{W}{e(x)}\pi(1) - \frac{1-W}{1-e(x)}\pi(0)\right)(f(\beta_1(x)) - f(\beta_0(x)))\right| \\
&\qquad \leq C_l' \,|(f(\beta_1(x)) - f(\beta_0(x)))| \\
&\qquad \leq C_l \,|\beta_1(x) - \beta_0(x)|
\end{aligned}$$

where I chose $C_l' = max_i(\frac{W_i}{e(x_i)}\pi_i(1) - \frac{1-W_i}{1-e_i(x)}\pi_i(0))$ and the last line follows because f is a Lipschitz function.

I then verify the curvature condition $c_1||\beta_0(x) - \beta_1(x)||_2^2 \leq E[\mathscr{L}(y,t,\beta_0(x)) - \mathscr{L}(y,t,\beta_1(x))] \leq c_2||\beta_0(x) - \beta_1(x)||_2^2$ holds for the loss function.

$$\begin{aligned}
E[\mathscr{L}(y,t,\beta_0(x)) - \mathscr{L}(y,t,\beta_1(x))] &= E[\pi(1)\big(f(\beta_1(x)) - f(\beta_0(x))\big) - \pi(0)\big(f(\beta_1(x)) - f(\beta_0(x))\big)] \\
&= E[\big(\pi(1) - \pi(0)\big)\big(f(\beta_1(x)) - f(\beta_0(x))\big)] \\
&\leq c_2' E[f(\beta_1(x)) - f(\beta_0(x))] \\
&\leq c_2 E[\beta_1(x) - \beta_0(x)]
\end{aligned}$$

where I chose $c_2' = \max(\pi_i(1) - \pi_i(0))$ and the last line follows because f is a Lipschitz function. The lower bound for the curvature condition is shown similarly. □

Proposition. *(Sign consistency of the surrogate) The surrogate policy function* $\mathbf{1}\{\tilde{d}(x) > 0.5\}$ *produces the same targeting rule as the optimal policy function* $d^*(x) = \mathbf{1}\{\beta(x) > \frac{c}{m}\}$*. In other words, the targeting policy from* $\tilde{d}(x)$ *is sign consistent to that of* $d(x)$*.*

Proof. I prove this statement by extending the proof of Proposition 3.1 from Zhao et al. (2012). I first note that for conditional expected profits $E[\Pi(d)|X = x]$, the optimal policy $d^*(x) = \mathbf{1}\{\beta(x) > \frac{c}{m}\}$. Now consider a decision function $g(x)$ and its profits $\Pi(g)$. Taking the conditional expectation to x, I have

$$\begin{aligned}
E[\Pi(g) \mid X = x] &= E[\pi(1) \mid X = x]g(x) + E[\pi(0) \mid X = x](1 - g(x)) \\
&= E[\pi(1) - \pi(0) \mid X = x]g(x) + E[\pi(0) \mid X = x] \\
&= (mE[Y(1) \mid X = x] - c - mE[Y(0)])g(x) + mE[Y(0) \mid X = x] \\
&= (m(E[Y \mid X = x, W = 1] - E[Y \mid X = x, W = 0]) - c)g(x) + mE[Y \mid X = x, W = 0] \\
&= (m\beta(x) - c)g(x) + mE[Y \mid X = x, W = 0]
\end{aligned}$$

where I used the definitions of the counterfactual profits $\pi(1), \pi(0)$, $\beta(x)$ from Equation 4, and the unconfoundedness assumption.

To maximize the expected profits, I see that $g(x)$ must be positive when $m\beta(x) - c > 0$ and $g(x)$ is negative when $m\beta(x) - c < 0$. I then choose $g(x) = \tilde{d}(x) - 0.5$. Then, when $\tilde{d}(x) > 0.5$, I have $g(x) > 0$ which corresponds to $m\beta(x) - c > 0$ and when $\tilde{d}(x) < 0.5$, I have $g(x) < 0$ which corresponds to $m\beta(x) - c < 0$. Thus, that targeting policy from $\tilde{d}(x)$, $\mathbf{1}\{\tilde{d}(x) > 0.5\}$, is sign consistent to the optimal policy function $d^*(x) = \mathbf{1}\{\beta(x) > \frac{c}{m}\}$ in the population. □

Corollary. *Expected profits from the targeting policy generated from* $\tilde{d}(x)$ *are consistent for the expected profits generated from* $d^*(x)$ *or* $E[\Pi(d^*(x))] = E[\Pi(\mathbf{1}\{\tilde{d}(\tilde{\beta}(x)) > 0.5\})]$.

Proof. I need to show $E[\Pi(d^*(x))] - E[\Pi(\mathbf{1}\{\tilde{d}(\tilde{\beta}(x)) > 0.5\})] = 0$ for the two targeting policies $d^*(x)$ and $\mathbf{1}\{\tilde{d}(\tilde{\beta}(x)) > 0.5\})$. Taking the expectation of the profit function (Equation 2), I attain for targeting policy d,

$$E[\Pi(d)] = \sum_{i=1}^{n} E[\pi_i(1)d] + E[\pi_i(0)(1 - d)] = E[\pi(1)d] + E[\pi(0)(1 - d)].$$

Taking the differences of the two targeting policies and suppressing the dependence on x_i for notational simplicity, I attain

$$E[\Pi(d^*)] - E[\Pi(\mathbf{1}\{\tilde{d} > 0.5\})] = E[\pi(1)(d^* - \mathbf{1}\{\tilde{d} > 0.5\})] + E[\pi(0)(\mathbf{1}\{\tilde{d} > 0.5\} - d^*)].$$

From Proposition 8, I have $\mathbf{1}\{\tilde{d} > 0.5\} - d^* = 0$, so the right-hand side of the term is zero and $E[\Pi(d^*)] - E[\Pi(\mathbf{1}\{\tilde{d} > 0.5\})] = 0$. □

A.2 Proofs for Section 6

I provide the proof for Theorem 13 in Section 6. I restate the theorem for ease of exposition.

Theorem. *(Uniform convergence rate of ex post comprehensible policies) Under Assumptions 1, 2, and 3 and the assumption that the outcome variable (Y) is bounded, for a hypothesis space of comprehensible policies with ℓ clauses ($\mathcal{H} = \mathcal{F}^{\ell}_{comp}$) that has bounded VC dimension ($VC(\mathcal{H}) < V < \infty$),*

$$\sup_{P} E_P\left[\Gamma(\check{d}^*) - \Gamma(\hat{d}_{ex})\right] \leq C\sqrt{\frac{V}{n}} = O_p\left(1/\sqrt{n}\right).$$

Proof. I extend the analogous proof of Theorem 2.1 in Kitagawa and Tetenov (2018) to provide the upper bound. First, I see that

$$\begin{aligned}
\Gamma(\check{d}) - \Gamma(\hat{d}_{PWL}) &= \Gamma(\check{d}) - \Gamma_n(\hat{d}_{ex}) + \Gamma_n(\hat{d}_{ex}) - \Gamma(\hat{d}_{PWL}) \\
&\leq \Gamma(\check{d}) - \Gamma_n(\hat{d}_{ex}) + \sup_{d\in\mathcal{F}^{\ell}_{\text{comp}}} |\Gamma_n(d) - \Gamma(d)| \\
&\leq \Gamma(\check{d}) - \Gamma_n(\check{d}) + \sup_{d\in\mathcal{F}^{\ell}_{\text{comp}}} |\Gamma_n(d) - \Gamma(d)| \\
&\leq 2 \sup_{d\in\mathcal{F}^{\ell}_{\text{comp}}} |\Gamma_n(d) - \Gamma(d)|
\end{aligned} \tag{14}$$

where used that $\Gamma_n(\hat{d}_{ex}) \geq \Gamma_n(\check{d})$ from the optimality of $\hat{d}_{ex}$ to get to the third line. Since the bound holds for all $\Gamma(\check{d})$, it also holds for $\Gamma(\check{d}^*)$ so

$$\Gamma(\check{d}^*) - \Gamma(\hat{d}_{ex}) \leq 2 \sup_{d\in\mathcal{F}^{\ell}_{\text{comp}}} |\Gamma_n(d) - \Gamma(d)|.$$

I now need to bound the empirical process term $\sup_{d\in\mathcal{F}^{\ell}_{\text{comp}}} |\Gamma_n(d) - \Gamma(d)|$ to complete the proof. From the Assumptions 1, 2, and 3 and the bounded outcome variable (Y) assumption, the $\Gamma(d)$ function is bounded with $||\Gamma(d)||_\infty < \bar{\Gamma}$. Assumption 2 provides strict overlap (for propensity score $e(x)$, I have $\epsilon < e(x) < 1 - \epsilon, \forall x$ and for some ϵ). Lastly, the hypothesis space of comprehensible policies with a fixed length l has a bounded VC dimension $VC(\mathcal{H}) < V < \infty$ from Lemma 19.

I then use Lemma A.4 in the supplement of Kitagawa and Tetenov (2018) to bound the the empirical process term. The lemma adapted to my notation provides

$$E_P\left[\sup_{d\in\mathcal{F}^{\ell}_{\text{comp}}} |\Gamma_n(d) - \Gamma(d)|\right] \leq C_1\bar{\Gamma}\sqrt{\frac{V}{n}}$$

for some constant C_1. Applying this result to Equation 14, I have that

$$E_p\left[\Gamma(\check{d}) - \Gamma(\hat{d}_{ex})\right] \leq C\sqrt{\frac{V}{n}}.$$

Then taking the supremum over the possible probability distributions P that satisfy the assumptions yields,

$$\sup_P E_p\left[\Gamma(\check{d}) - \Gamma(\hat{d}_{ex})\right] \leq C\sqrt{\frac{V}{n}} = O_p(1/\sqrt{n})$$

as V and C do not depend on n. □

B Simulation Study for Policy DNN

I provide a Monte Carlo simulation study for Policy DNN below and compare it to Causal DNN (Farrell et al., 2021, 2020). This section aims to demonstrate the differences between Causal DNN to Policy DNN in learning the optimal treatment policy.

B.1 Monte Carlo Simulations

I consider the following data generating process (DGP). I have $P = 10$ total covariates that are individually denoted as $x_p, \forall p \in P$, the treatment indicator is W, the outcome variable is Y, and $\alpha(x), \beta(x)$ respectively are the heterogeneous intercept and treatment effect terms.

$$x_p \sim \begin{cases} Unif(-1,1) & p \text{ is odd} \\ Bernoulli(0.5) & p \text{ is even} \end{cases} \tag{15}$$
$$W \sim Bernoulli(0.5)$$

$$\begin{aligned} \alpha(x) &= x_2 + \min\{0, x_3\} \\ \beta(x) &= -0.55 + \max\{0, x_1\} + \max\{x_3, x_4, x_5\} \\ y &= \alpha(x) + \beta(x)W + \epsilon \end{aligned}$$

where $\epsilon \sim N(0,1)$ and I let the number of observations n be $20,000$ for both the training set and the test set. Following the running example from the introduction, I can consider W to be a mailed catalog or marketing treatment and Y to be sales for the retailer in the weeks following the treatment.

I first make two comments on the structure of the DGP. First, the covariate space is mixture of continuous and discrete variables. These reflect practical settings where there are binary indicators for variables such as race or gender and continuous variables to represent variables such

as age or purchase history. Second, the $\alpha(x), \beta(x)$ functions are highly non-linear and use information from only the first five covariates. These two components of the DGP make the optimal treatment policy a highly non-linear and non-smooth function.

I choose the profit margin to be $m = 0.9$ and the cost of treatment to be $c = 0.3$. Then the profits attained from treatment are $\pi(1) = mY(1) - c$ and the profits from no treatment are $\pi(0) = mY(1)$. The optimal treatment regime if I observe both potential outcomes would be $d^* = \mathbf{1}\{\pi(1) > \pi(0)\} = \mathbf{1}\{m\beta(x) > c\}$.

I use the same architecture and the training process for the two DNN make the results more comparable. Both DNN have three hidden layers with $(3, 8, 3)$ ReLu nodes each. The Adam optimizer is implemented with a weight decay value of 10^{-5} and a learning rate of 0.001. The learning rate has a scheduler and it goes down by a exponential factor of 0.9 after 500 training epochs. The training data is split and randomly shuffled before every epoch into a training and validation set. The model trains for a minimum of $2,000$ epochs and training stops when the 100 epoch moving average for the validation set's loss is higher than the 100 epoch prior moving average validation loss. For Policy DNN, I use $\tilde{d}(x_i) = \frac{\tanh(m\tilde{\beta}(x_i)-c)+1}{2}$ as the surrogate function.

The outcome of interest is the learned optimal treatment policy $\hat{d}(x_i)$ from Causal DNN and Policy DNN. Because I know the true DGP, I can compare the classification accuracy of $\hat{d}(x_i)$ to the true optimal treatment policy d^*. I find that across 1,000 Monte Carlo iterations, the average classification accuracy of Policy DNN is 72.2% with standard deviation 4.0%. In contrast, the average classification accuracy of Causal DNN is 65.8% with standard deviation 6.6%. A random targeting rule would have classification accuracy of 50%.

B.2 Sensitivity Analysis

I now consider the case where additional noise variables are injected into the data. This procedure is done to study how robust Policy DNN and Causal DNN are to extraneous noise variables. This setting reflects many big data settings where the number of covariates is large but the true DGP is relatively sparse.

I keep the same setup, architecture, and training procedure from Section B and only increase the number of covariates from the original $P = 10$. I follow the same covariate generation process in Equation 15 to generate the additional covariates and run the Monte Carlo simulation $1,000$ times.

Table 4 demonstrates the changes in classification accuracy as I increase the number of extraneous covariates. The case with no extraneous covariates is equivalent to the scenario in Section B. As I add extraneous covariates, the performance of both methods degrades. Causal DNN appears to perform worse than Policy DNN. At 500 extraneous covariates ($P = 510$), the two methods perform similarly. Further Policy DNN seems to have a smaller standard deviation of its performance across simulations than Causal DNN.

These results suggest that Policy DNN is more robust to noise injections than Causal DNN

and may be more efficient at learning the true policy. Thus for many settings with big data that contain a large set of continuous and discrete covariates along with a sparse DGP, I expect Policy DNN to do better than Causal DNN in finding the optimal policy. However, I caution that these results are from only one Monte Carlo example and additional theoretical and empirical analysis should be done to delineate the performance of the two.

C Inference for Policy DNN

In this section, I provide an exposition around the proof to build intuition and then state the general theorem. I then provide two remarks around the key technical implications.

To conduct inference around the optimal targeting policies, I want to show that the difference in estimated profits from the surrogate approach ($\hat{\Pi}(\tilde{d})$) to profits from the optimal targeting policy $\Pi(d^*)$ converges to a normal distribution. Mathematically, I want to show

$$\sqrt{n}\left(\hat{\Pi}(\tilde{d}) - \Pi(d^*)\right) \xrightarrow{d} N(0, V)$$

for some finite variance V. Expanding the term on the left hand side, I attain three terms,

$$\begin{aligned}\sqrt{n}\left(\hat{\Pi}(\tilde{d}) - \Pi(d^*)\right) = &\underbrace{\sqrt{n}\left(\hat{\Pi}(\tilde{d}) - \Pi(\tilde{d})\right)}_{(1)} + \underbrace{\sqrt{n}\left(\Pi(\tilde{d}) - \Pi(\mathbf{1}\{\tilde{d} > 0.5\})\right)}_{(2)} \\ &+ \underbrace{\sqrt{n}\left(\Pi(\mathbf{1}\{\tilde{d} > 0.5\}) - \Pi(d^*)\right)}_{(3)}.\end{aligned}$$

I discuss how I control for each of the three terms below and suppress the dependence of the functions on x for notational simplicity.

The first term, $\hat{\Pi}(\tilde{d}) - \Pi(\tilde{d})$, represents the difference between the sample surrogate profits to the population surrogate profits. As discussed in Section 3.1, I verify the surrogate profit loss falls into the framework in Farrell et al. (2020) using Proposition 7 and by assuming Assumption 2 in Farrell et al. (2020) holds. This yields

$$\sqrt{n}\left(\hat{\Pi}(\tilde{d}) - \Pi(\tilde{d})\right) \xrightarrow{d} N(0, V) \tag{16}$$

for a finite variance V. Specifically, I use their general framework and choose the parameter of interest as surrogate profits, $H(x, \tilde{\beta}(x)) = \Pi(\tilde{d}) = \Pi(\tilde{d}(\tilde{\beta}(x)))$. Here, the structural parameter, $\tilde{\beta}(x)$, is directly represented by the DNN. Further, from the policy learning framework, the parameter of interest is itself the loss function when training the DNN in Section 3.1. In the terminology of Farrell et al. (2020), I have $\mu_0 = E[\Pi(\tilde{d})]$ and $\hat{\mu} = \hat{\Pi}(\tilde{d})$ which represent the population

expectation and the sample analog of the parameter of interest respectively. I use sample splitting to split the data into training and validation datasets: I estimate the DNN on the training dataset and conduct inference in the validation dataset.

I form the uncentered influence function,

$$\psi(\omega, \tilde{\beta}, \Lambda) = H(x, \tilde{\beta}(x)) - H_{\tilde{\beta}}\Lambda(x)^{-1}\mathcal{L}_{\tilde{\beta}} \tag{17}$$

where $H_{\tilde{\beta}}$ represents the gradient of the parameter of interest to the structural parameter $\tilde{\beta}$, $\mathcal{L}_{\tilde{\beta}}$ is the gradient of the loss function, and $\Lambda(x) = E[\mathcal{L}_{\tilde{\beta}\tilde{\beta}}|X = x]$ is the conditional expectation of the Hessian of the loss function, and ω represents the data tuple (X, W, Y).

Crucially, I want to conduct inference around the optimal targeting policy implemented by the firm. I leverage the fact that the loss in estimating the structural parameter $\tilde{\beta}(x)$ is the same as the parameter of interest, so for each individual with covariates x_i, the general envelope theorem (Milgrom and Segal, 2002) tells us

$$H_{\tilde{\beta}}\Big|_{\tilde{\beta}=\tilde{\beta}^*} = \frac{\partial \Pi}{\partial \tilde{d}}\frac{\partial \tilde{d}}{\partial \tilde{\beta}}\Big|_{\tilde{\beta}=\tilde{\beta}^*} = 0$$

as $\frac{\partial \tilde{d}}{\partial \tilde{\beta}}\Big|_{\tilde{\beta}=\tilde{\beta}^*} = 0$ from the envelope theorem. Thus $H_{\tilde{\beta}} = 0$ will hold pointwise for each individual with covariates x_i.

The uncentered influence function then becomes $\psi(\omega, \tilde{\beta}, \Lambda) = H(x, \tilde{\beta}(x)) = \Pi(\tilde{d})$ and the plug-in's standard deviation will be the standard error around the profits,

$$\sqrt{n}\left(\hat{\Pi}(\tilde{d}) - \Pi(\tilde{d})\right) = \sqrt{n}\left(\hat{\mu} - \mu_0\right) \xrightarrow{d} N(0, V)$$

for variance term

$$V = \frac{1}{n_{\mathrm{v}}}\sum_{i=1}^{n_{\mathrm{v}}}(\psi_i - \hat{\mu}_i)^2 \tag{18}$$

in which n_{v} represents the data in the validation set.

The second term, $\Pi(\tilde{d}) - \Pi(\mathbf{1}\{\tilde{d} > 0.5\})$, represents how close the surrogate profits from the surrogate targeting policy $\tilde{d}$ are to the thresholded surrogate targeting policy $\mathbf{1}\{\tilde{d} > 0.5\}$. I want to show that

$$\sqrt{n}\left(\Pi(\tilde{d}) - \Pi(\mathbf{1}\{\tilde{d} > 0.5\})\right) \to o_p(1). \tag{19}$$

Ensuring the difference in profits decays at a rate fast enough will require an extra assumption that is parallel to the margin assumption used in the literature (Zhao et al., 2012; Kitagawa and Tetenov, 2018). Essentially, this says if the distribution of heterogeneous treatment effects is well separated enough around the optimal targeting policy's cutoff value, then surrogate profits will

match that of the profits using the thresholded targeting rule from the surrogate. I state the assumption more formally in Assumption 14.

Assumption 14. *(Margin) There exists a scale parameter $k \in \mathbb{R}$ such that $|\tilde{d}(k,x) - \mathbf{1}\{\tilde{d}(x) > 0.5\}| \to o_p(1/\sqrt{n})$ as $k \to \infty$.*

I now make this assumption more concrete. Recall f is the Lipschitz function that maps $\tilde{\beta}(x)$ to the relaxed decision rule $\tilde{d}(x)$. I further parameterize f by with scale parameter k, such that $f(k,x) \to \mathbf{1}\{f(k,x) > 0.5\}$ as $k \to \infty$. To give a concrete example, let $f(k,x) = \frac{\tanh(kx)+1}{2}$ and as $k \to \infty$, I will derive the rate for k such that that $f(k,x) \to \mathbf{1}\{f(k,x) > 0.5\}$ is satisfied.

The assumption requires $|\tilde{d}(k,x) - \mathbf{1}\{\tilde{d}(k,x) > 0.5\}| \leq 1/\sqrt{n}$ to ensure Equation 19 holds since the difference in profits $\Pi(\tilde{d}) - \Pi(\mathbf{1}\{\tilde{d} > 0.5\}) = C(\tilde{d}(k,x) - \mathbf{1}\{\tilde{d}(k,x) > 0.5\})$ from Equation 23 in Appendix Section E. The constant C will be bounded if the outcome variable (Y) is bounded and the overlap assumption holds (Assumption 2). Since k is a scale parameter, it will not affect the threshold rule so $\mathbf{1}\{\tilde{d}(k,x) > 0.5\} = \mathbf{1}\{\tilde{d}(x) > 0.5\}$. Then, I have that $|\tilde{d}(k,x) - \mathbf{1}\{\tilde{d}(x) > 0.5\}| \leq 1/\sqrt{n}$.

Consider the case where $\tilde{d}(k,x) = \frac{\tanh(k\tilde{\beta}(x))+1}{2}$. Without loss of generality, I assume $\tilde{\beta}(x) > 0$. I first see that $\frac{\tanh(k\tilde{\beta}(x))+1}{2} > 0.5$ under this regime so I have $\mathbf{1}\{\tilde{d}(x) > 0.5\} = 1$. Then, I need to show

$$
\begin{aligned}
|\tilde{d}(k,x) - 1| = \left| \frac{\tanh(k\tilde{\beta}(x)) + 1}{2} - 1 \right| &\leq \frac{1}{\sqrt{n}} \\
\left| \tanh(k\tilde{\beta}(x)) - 1 \right| &\leq \frac{2}{\sqrt{n}} \\
-\tanh(k\tilde{\beta}(x)) + 1 &\leq \frac{2}{\sqrt{n}} \\
\tanh(k\tilde{\beta}(x)) &\geq 1 - \frac{2}{\sqrt{n}}
\end{aligned}
$$

where I used that $\tanh(k\tilde{\beta}(x))) \leq 1$ to get from the second line to the third line.

I assume that $\tilde{\beta}(x)$ has a small probability to reside in $[0, \epsilon_n]$ where $\epsilon_n = \frac{1}{M_n}$ and M_n is some arbitrary slowly increasing large constant sequence. I then want to equivalently show that

$$
P\left(\tanh(k\tilde{\beta}(x)) \geq 1 - \frac{2}{\sqrt{n}} \right) \xrightarrow{a.s.} 1. \tag{20}
$$

To do so, I use the law of total probability to get

$$P\left(\tanh(k\tilde{\beta}(x)) \geq 1 - \frac{2}{\sqrt{n}}\right) \geq P\left(\tilde{\beta}(x) \in [\epsilon_n, \infty) \cap \tanh(k\tilde{\beta}(x)) \geq 1 - \frac{2}{\sqrt{n}}\right)$$
$$= \left(1 - P(\tilde{\beta}(x) \in [0, \epsilon_n))\right) P\left(\tanh(k\tilde{\beta}(x)) \geq 1 - \frac{2}{\sqrt{n}} \middle| \tilde{\beta}(x) \geq \epsilon_n\right).$$

In the last line, the first term will tend to 1 by the assumption. For the second term, I use that fact that since $\tanh(k\tilde{\beta}(x))$ is an increasing monotonic function in $\tilde{\beta}(x)$, I have

$$P\left(\tanh(k\tilde{\beta}(x)) \geq 1 - \frac{2}{\sqrt{n}}) \middle| \tilde{\beta}(x) \geq \epsilon_n\right) \geq P\left(\tanh(k\epsilon_n) \geq 1 - \frac{2}{\sqrt{n}}\right)$$

since $\tilde{\beta}(x) \geq \epsilon_n$.

Then to find the rate of k that satisfies Equation 20, I need

$$P\left(\tanh(k\epsilon_n) \geq 1 - \frac{2}{\sqrt{n}}\right) \xrightarrow{a.s.} 1$$
$$\Leftrightarrow k\epsilon_n \geq \tanh^{-1}\left(1 - \frac{2}{\sqrt{n}})\right)$$
$$= \frac{1}{2}\left(\ln(2 - \frac{2}{\sqrt{n}}) - \ln(\frac{2}{\sqrt{n}})\right)$$
$$\asymp \ln(\sqrt{n}) \asymp \ln(n).$$

Thus, $k \asymp \frac{1}{\epsilon_n}\ln(n)$ is required for Equation 19 to hold. To recap, Assumption 14 is satisfied for $\tilde{d}(k, x) = \frac{\tanh(k\tilde{\beta}(x))+1}{2}$ when $k \asymp \ln(n)$.

The third term, $\Pi(\mathbf{1}\{\tilde{d} > 0.5\}) - \Pi(d^*)$, is relatively easy to control for. From Corollary 9, I have that $\Pi(\mathbf{1}\{\tilde{d} > 0.5\})$ is consistent to $\Pi(d^*)$ in the population so

$$\Pi(\mathbf{1}\{\tilde{d} > 0.5\}) - \Pi(d^*) = 0. \tag{21}$$

Thus, combining my results for the three terms in Equations 16, 19, and 21, I show that $\sqrt{n}\left(\hat{\Pi}(\tilde{d}) - \Pi(d^*)\right) \xrightarrow{d} N(0, V)$ for variance term V defined in Equation 18. I now state the general theorem.

Theorem 15. *(Inference for Policy DNN). Under Assumptions 1, 2, 3, and 14 and the assumption that the outcome variable Y is bounded,*

$$\sqrt{n}\left(\hat{\Pi}(\tilde{d}) - \Pi(d^*)\right) \xrightarrow{d} N(0, V)$$

for finite V defined in Equation 18.

To recap, I use the results from Farrell et al. (2020) to attain inference around the surrogate profits and show the difference of the surrogate profits to the optimal profits is small enough. I provide two remarks around this procedure.

Remark 16. In the analysis of the first term, the envelope theorem eliminates the correction term in the influence function which provides a gain in efficiency. The envelope theorem argument holds because the loss function in the first stage and the parameter of interest are the same, which are both the surrogate profits. In standard approaches (e.g., Causal DNN) the loss function is the mean squared error of the model's fit and is not profits directly. As a result, the correction term will exist for the standard approach. This argument with the envelope demonstrates the additional efficiency of using policy learning over the standard approach in the Farrell et al. (2020) framework.

Remark 17. The use of surrogates to get rates of convergence is different from the approach used in Zhao et al. (2012) which builds on the work by Bartlett et al. (2006) for convex surrogates. In Zhao et al. (2012), they can get almost n-rate convergence with strong margin assumptions for using a convex surrogate loss function for support vector machines.[40] In my setting, the bottleneck on the rates come from the use of Farrell et al. (2020) inference framework and this allows my results to be more general. Any machine learning method that satisfies the assumptions of Farrell et al. (2020) can be used for policy learning with my framework, and this enables researchers to use a more general class of machine learning procedures.

D Comprehensible policies and decision trees

In this section, I demonstrate that comprehensible policies are subsets of decision trees. Recall that in Section 4, I constructed the comprehensible policy class to be targeting policies that can be represented by a sentence. A decision tree of ℓ layers can be more complex than a comprehensible policy of ℓ clauses. To show this, I visualize how to construct the decisions trees from comprehensible policies for one to three clauses and the provide an algorithm to generally do so. I then leverage the link from comprehensible policies to decisions trees to control the Vapnik–Chervonenkis (VC) dimension of comprehensible policies.

Lemma 18. *A comprehensible policy of length ℓ can be represented by a full decision tree of depth l.*

Proof. I provide a proof by construction. The $\ell = 1$ case is straightforward and I show the representation explicitly for a comprehensible policy of $\ell = 2$ clauses and $\ell = 3$ clauses. Then, I present algorithms to map a comprehensible policy of length ℓ to a decision tree of depth ℓ.

[40] Faster than $\sqrt{n}$-rate convergence under strong margin assumption for certain settings are also discussed in Luedtke and Chambaz (2020).

A one clause comprehensible policy can be represented as a decision tree of depth one which has one split. Figure 9 and Figure 10 provide the mapping between a comprehensible policy of $\ell = 2$ clauses and $\ell = 3$ clauses to their respective decision trees. A key result in these two examples is that a two clauses comprehensible policy can be represented by a decision tree of depth two and a three clauses comprehensible policy can be represented by a decision tree of depth three.

To complete the construction, Algorithm 3, 4, and 5 show how to grow the decision tree when "and", "or", and "xor" operators and a clause are added a comprehensible policy. Adding one additional clause the the comprehensible policy increases the decision tree by an extra level of depth. Thus, a comprehensible policy of length ℓ can be represented by a decision tree of depth l. □

Lemma 19. *A comprehensible policy of finite length $\ell \in \mathbb{N}$ has a finite Vapnik–Chervonenkis (VC) dimension for finite number of covariates $p \in \mathbb{N}$.*

Proof. Athey and Wager (2021) supply the asymptotic VC dimension for a decision tree of l layers, p covariates, and observations n as $VC(d_{DT}) = \tilde{O}(2^{l_n} \log_2(n))$ where $l_n = \lfloor \kappa \log_2(p) \rfloor, \kappa < 1/2$. Here, $f(n) = \tilde{O}(g(n))$ implies there is a function that scales polylogarithmically in its arguments for $f(n) < h(g(n))g(n)$.

Lemma 18 shows that a comprehensible policy of length ℓ can be represented by a decision tree of depth ℓ. This result implies that the VC dimension of the full decision tree of depth ℓ will be an upper bound for that of the comprehensible policy,

$$VC(d_{\text{comp}}) \leq VC(d_{DT}) = \tilde{O}(2^{l_n} \log_2(p)).$$

However since the number of clauses (and depth of the decision tree) ℓ is finite in my setting, I choose $l'_n = \min\{l_n, \ell\} \leq \ell$ for the upper bound of the decision tree's depth. This choice of the model structure breaks the dependence of the VC dimension of the tree to n as $2^l \log_2(p) = C$ for finite p. Then, I see that $VC(d_{\text{comp}}) \leq VC(d_{DT}) \leq C'$ where d_{comp} has ℓ clauses and d_{DT} has depth ℓ. Thus, the VC dimension for a comprehensible sentence of finite length ℓ has a finite VC dimension. □

E Profit loss function derivation

This section derives the profit loss function used in Section 6. The individual-level inverse propensity weighted (IPWE) profit loss estimator for policy function $d(x_i)$ is

$$\pi(d(x_i)) = \frac{1 - W_i}{1 - e(x_i)} \pi_i(0)(1 - d(x_i)) + \frac{W_i}{e(x_i)} \pi_i(1) d(x_i). \tag{22}$$

In this notation, the sample profits are then $\hat{\Pi}(d) = \sum_{i=1}^{n} \pi(d(x_i))$ with is an estimate for average population profits $\Pi(d) = E[\pi(d(x_i))]$.

The individual-level IPWE of profit difference of two policies $d(x_i), d'(x_i)$ is

$$\begin{aligned}\pi(d(x_i)) - \pi(d'(x_i)) &= \frac{1-W_i}{1-e(x_i)}\pi_i(0)(1-d(x_i)) + \frac{W_i}{e(x_i)}\pi_i(1)d(x_i) \\ &\quad - \frac{1-W_i}{1-e(x_i)}\pi_i(0)(1-d'(x_i)) - \frac{W_i}{e(x_i)}\pi_i(1)d'(x_i) \\ &= \frac{1-W_i}{1-e(x_i)}\pi_i(0)(1-d(x_i)-(1-d'(x_i))) + \frac{W_i}{e(x_i)}\pi_i(1)(d(x_i)-d'(x_i)) \\ &= \frac{1-W_i}{1-e(x_i)}\pi_i(0)(d'(x_i)-d(x_i)) + \frac{W_i}{e(x_i)}\pi_i(1)(d(x_i)-d'(x_i)) \\ &= (d(x_i)-d'(x_i))(\frac{W_i}{e}\pi_i(1) - \frac{1-W_i}{1-e(x_i)}\pi_i(0))\end{aligned} \tag{23}$$

To construct a loss function, I consider the absolute difference of the individual-level IPWE profit differences,

$$\begin{aligned}|\pi(d(x_i)) - \pi(d'(x_i))| &= \left|(d(x_i)-d'(x_i))(\frac{W_i}{e(x_i)}\pi_i(1) - \frac{1-W_i}{1-e(x_i)}\pi_i(0))\right| \\ &= \mathbf{1}\{d(x_i) \neq d'(x_i)\}\left|\frac{W_i}{e(x_i)}\pi_i(1) - \frac{1-W_i}{1-e(x_i)}\pi_i(0)\right|\end{aligned}$$

where I used the fact that $d(x_i)$ and $d'(x_i)$ are indicator functions to get to the second line. I treat the sample-level difference as the loss function of interest for two policies d, d'

$$\begin{aligned}\mathcal{L}(d,d') &= |\hat{\Pi}(d) - \hat{\Pi}(d')| \\ &= \sum_{i=1}^{n}|\pi(d(x_i)) - \pi(d'(x_i))| \\ &= \sum_{i=1}^{n}\mathbf{1}\{d(x_i) \neq d'(x_i)\}\left|\frac{W_i}{e(x_i)}\pi_i(1) - \frac{1-W_i}{1-e(x_i)}\pi_i(0)\right|.\end{aligned}$$

Algorithms

Algorithm 3 Adding an “and” operator and a clause to a decision tree

Objective: Add “and D” to the comprehensible policy.
Setup: “and” is the logic operator and “D” is the new clause being added. Take the decision tree representation of the comprehensible policy without the addition.

1. For all terminal nodes in the decision tree that have value 1, grow a split on clause D and that have terminal node value 1 if clause D is true and 0 otherwise

Algorithm 4 Adding an ”or” operator and a clause to a decision tree

Objective: Add “or D” to the comprehensible policy.
Setup: “or” is the logic operator and “D” is the new clause being added. Take the decision tree representation of the comprehensible policy without the addition.

1. For all terminal nodes in the decision tree that have value 0, grow a split on clause D and that have terminal node value 1 if clause D is true and 0 otherwise

Algorithm 5 Adding a ”xor” operator and a clause to a decision tree

Objective: Add “xor D” to the comprehensible policy.
Setup: “xor” is the logic operator and “D” is the new clause being added. Take the decision tree representation of the comprehensible policy without the addition.

1. For all terminal nodes in the decision tree that have value 0, grow a split on clause D and that have terminal node value 1 if clause D is true and 0 otherwise
2. For all terminal nodes in the decision tree that have value 1, grow a split on clause D and that have terminal node value 0 if clause D is true and 0 otherwise

Tables

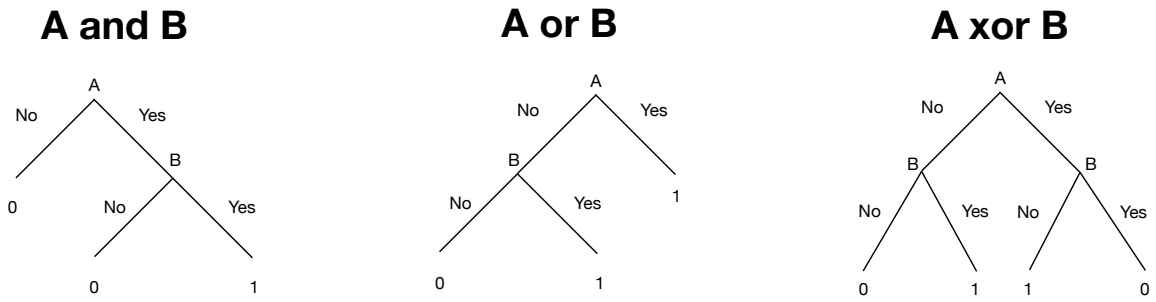
Table 4: Policy DNN vs. Causal DNN Classification Accuracy

		Policy DNN		**Causal DNN**	
P	Signal / Noise	Mean	SE	Mean	SE
10	50%	71.9%	3.9%	67.1%	6.3%
15	33%	70.9%	4.1%	66.9%	6.2%
20	25%	70.5%	4.1%	66.1%	6.4%
30	17%	69.8%	4.2%	64.5%	6.8%
60	8%	68.2%	4.4%	62.6%	7.6%
110	5%	65.3%	4.5%	59.7%	8.4%
260	2%	59.5%	4.5%	56.7%	8.0%
510	1%	56.2%	4.4%	57.0%	7.4%

Note: This table demonstrates the accuracy of Policy DNN's and Causal DNN's targeting policies to the oracle targeting policy with 1,000 Monte Carlo iterations. The true data generating process uses five covariates and P represents the number of covariates used in the simulation. As P increases, the signal to noise ratio in the data gets smaller. I see that Policy DNN is more accurate than Causal DNN and its standard error around the classification accuracy is smaller.

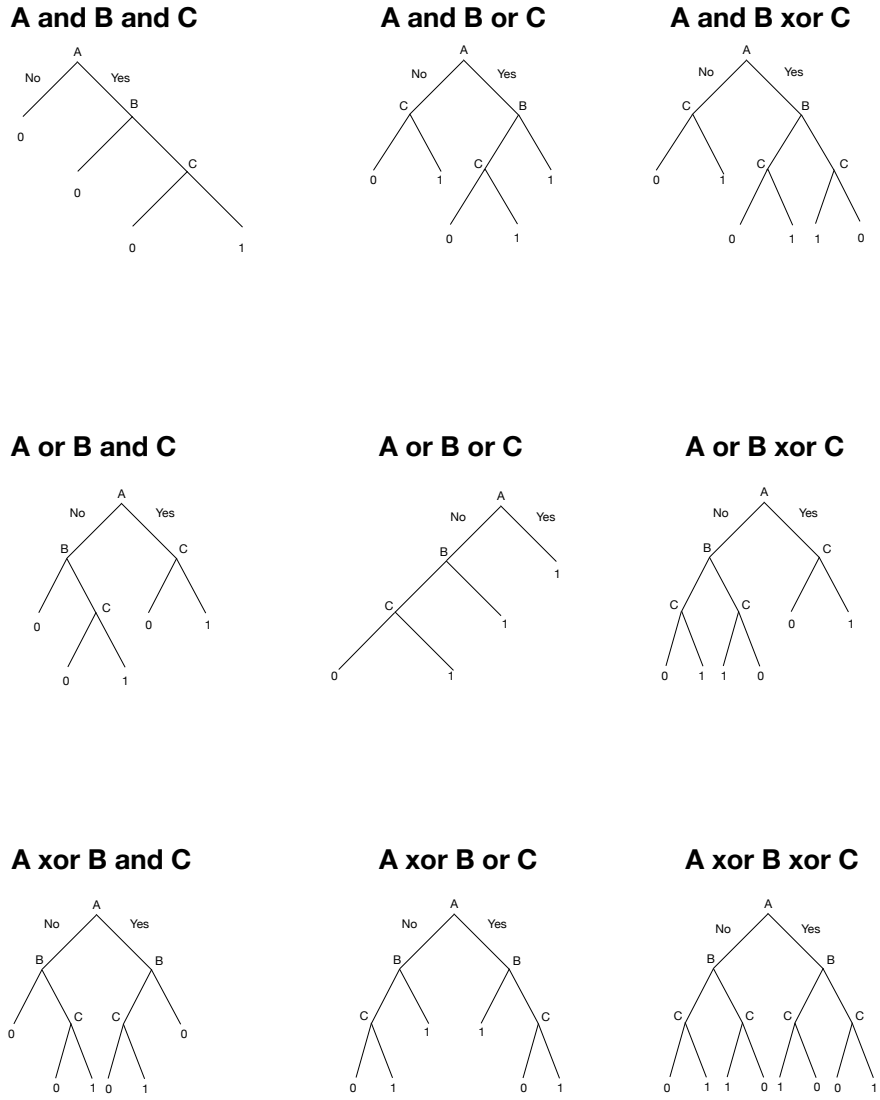
Figures

Figure 9: Two clause comprehensible policies and decision trees ($\ell = 2$)



Note: This figure demonstrates the mapping between a comprehensible policy with two clauses (A, B) to a decision tree of depth two. The comprehensible policy on the left states target if "A and B" in which a customer is targeted if the clause A and clause B are both true and not targeted otherwise. The corresponding decision tree first has a split whether clause A is true (yes/no) and then conditional on clause A being true it has another split on whether clause B is true. The terminal nodes of 1 indicate targeted and 0 indicated not targeted. The comprehensible policy and its respective decision tree will target and not target the same sets of customers. The other comprehensible policies and decision trees in the figure are mapped similarly.

Figure 10: Three clause comprehensible policies and decision trees ($\ell = 3$)



Note: This figure shows the mapping between a comprehensible policy with three clauses (A, B, C) to a decision tree of depth three. The comprehensible policy on the top left states target if "A and B and C" in which a customer is targeted if the clause A, clause B, and clause C are all true and not targeted otherwise. The corresponding decision tree first has a split whether clause A is true (yes/no), then conditional on clause A being true it has another split on whether clause B is true, and lastly conditional on A and B being true it has a split on whether clause C is true. The terminal nodes of 1 indicate targeted and 0 indicated not targeted. The comprehensible policy and its respective decision tree will target and not target the same sets of customers. The other comprehensible policies and decision trees in the figure are mapped similarly.