

The Limits of Manipulation within Personalized Incentive Policies

Andrea Araldo¹, Alessa Mayer², Ravi Seshadri³, Amir Brudner^{4,5}

¹SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France

²Institut Polytechnique de Paris

³Technical University of Denmark

⁴Massachusetts Institute of Technology

⁵The Hebrew University Business School

Abstract

Demand management policies aim to shift individual behavior toward sustainable choices (e.g., goods, transport modes). Personalized incentive policies tailor subsidies to agents' needs and constraints, but personalization also exposes the regulator to agents' strategic manipulative behavior. The extent to which agents can profitably manipulate such policies remains poorly understood, which limits their applicability.

This paper establishes analytically the conditions under which manipulation benefits the agent. We formalize a Stackelberg game in which an agent may simulate a different profile to extract higher incentives from a regulator. Previous work typically assumes that the regulator would *instantaneously* classify agents into the simulated user profile. In contrast, we explicitly model the regulator the process of learning the agent profile, based on observed past behavior. This makes manipulation harder, since it forces the agent to consistently behave according to a simulated profile, thus renouncing to follow their true preferences. Manipulative agents thus incur a *Cost of Manipulation*, which has been typically neglected. By accounting for it, we get a counterintuitive finding: although classification uncertainty reduce personalization accuracy, it may benefit the regulator by deterring manipulation. We establish an upper bound on the agent's gain from manipulation, which becomes stricter as profiles become harder to distinguish for the regulator. Simulations on the London Passenger Mode Choice dataset illustrate this deterrence effect. The framework can therefore be used to stress-test the robustness of personalized incentive policies before deployment.

1 Introduction

Behavioral change remains a major challenge in achieving sustainability goals, especially as technological solutions alone are insufficient. Demand Management Policies (DMP) favor changes in the agents' behavior toward sustainable choices (IEA 2021). Potential greenhouse gas reduction achievable via DMP is estimated to exceed 40% by 2050, compared to baselines (IPCC 2023; Creutzig et al. 2022). We will consider DMP in which a regulator, e.g., a government or any type of authority, applies a signal.

Negative signals (e.g., prices, taxes, bans) generally lack public acceptability (to the point of violent protests (Dabla-Norris et al. 2023)) and are often inadvertently inequitable (Blasco, Guillaud, and Zemmour 2023). For instance, increasing taxes on fuel particularly impacts low-income rural

households, who are dependent on their cars to perform their everyday activities, and who spend, for transport, already up to 30% of their income (Le Bras 2019). On the other hand, positive signals (e.g., incentives, subsidies) are expensive for the regulator and may not be sufficiently strong to induce relevant behavioral shifts (Gneezy and Rustichini 2000).

Personalized DMP (PDMP) can help overcome the above issues (Javaudin, Araldo, and de Palma 2023; Ascarza and Israeli 2022; Xie et al. 2024) by adapting signals to the condition and the preferences of each agent. Personalized signals can target agents for whom a behavioral shift is less costly, thus mitigating inequitable effects. Thus, PDMP have the potential to be more effective and acceptable than non-personalized DMP.

However, PDMP are vulnerable to manipulative agents, who may make dishonest choices in order to conceal their true preferences and manipulate the regulator's signals in their favor. For instance, it has been shown that some individuals maliciously reduced their labor supply to get a subsidy dedicated to the under-employed (Firpo et al. 2014).

In this work, we focus on this critical vulnerability by modeling the interaction between a regulator and a potentially manipulative agent in a Stackelberg game. The regulator (the leader) estimates agent (follower) profiles based on observed choices, while the agent may adopt a simulated (i.e., fake) utility function to secure higher incentives. We investigate how the regulator's continuous estimation of agent utility impacts manipulation. Extensive work in artificial intelligence on strategic behavior (§2) models agents in ways that are not behaviorally-grounded, and thus not compliant with the way human agents make choices. They are thus not applicable to demand-side mitigation. Our contribution is the first behaviorally-grounded model of repeated interaction between a regulator and an agent within a demand-side mitigation scenario, consisting of a personalized incentive policy. We model agents' choices, based on Random Utility Theory (RUT) (Strzalecki 2025; McFadden 1981). In contrast with previous work, in which the agent simply reports their profile, we explicitly model the regulator's *preference elicitation* (Chapman and Fisher 2025), with which the regulator learns the agent's profile.

Previous results showed that the agent could greatly manipulate the regulator. Our analysis shows instead that such findings are over-pessimistic. Indeed, we theoretically show

that realistic preference elicitation deters manipulative behavior and, consequently, personalized incentive policies can effectively reduce social cost. Results are validated in Monte Carlo simulations. The proofs that are not in the main paper are in the Supplementary Material.

2 Related Work

Strategic interaction and manipulation are studied in artificial intelligence, game and learning theory. In Adversarial Interaction Attacks (Nguyen et al. 2019), a defender allocates resources to defend targets and an attacker chooses the target to attack. The defender seeks to learn the attacker’s profile to better allocate defense resources. The attacker can manipulate the defender, by implementing certain sequences of attacks that steer the defender’s learning in a way that increases the utility of the attacks. Our case, instead, cannot be simultaneous: first, the regulator must set the incentives for the different available alternatives; then, the agent chooses an alternative. Our setting is thus better described by a Stackelberg game, where the leader (the regulator) takes their action (setting incentives) and *then* the follower (the agent) takes their action (choosing an alternative).

In recent work, when the leader has limited information about the follower, the latter can manipulate the outcome of the interaction. (Gan et al. 2019a,b; Nguyen 2019; Birmpas et al. 2020; Dawkins, Han, and Xu 2021) show that the follower can implement an optimal manipulative strategy, which gives the follower higher payoff: the leader is thus greatly vulnerable to manipulation, which is simply performed by the follower via strategic reporting. Such previous findings would tend to discourage the application of personalized policies. We however observe that their setting is unrealistic and too pessimistic for the leader. Indeed, manipulation is “too easy”: “To manipulate the [leader], the [follower] reports a fake [profile]”, and the leader just considers the reported profile when deciding their action. The previous quote is from (Gan et al. 2019a), but this assumption is common to all the aforementioned work. Similarly, framing the setting within Bayesian Persuasion (Kamenica and Gentzkow 2011; Jain and Perchet 2024), an agent could manipulate the regulator by choosing the information to report.

We consider instead a more realistic setting, in which the regulator (leader) *learns* the agent (follower) profile, by interacting with them and observing their previous choices, performing *preference elicitation* (Chapman and Fisher 2025). In the following, we will thus refer to *instantaneous* and *realistic preference elicitation*, to contrast, respectively, the setting of the aforementioned work (where the leader immediately considers the profile reported by the agent when setting incentives) and the more realistic setting of our work (where the leader considers instead the profile they learn by observing the agent’s behavior). Related are screening contracts, with which a regulator could design a menu of options, specifically to distinguish among agent profiles (Einav et al. 2022). Our setting differs because the options (alternatives) are exogenous and agent choices are stochastic and thus do not perfectly reveal their profile.

In (Sessa et al. 2020) the agent can *adversarially* choose one profile at every timeslot. However, their setting is too

pessimistic (and thus inappropriate) for studying personalized policies, for different reasons. First, we assume that agents behave according to Random Utility Theory (RUT) (Strzalecki 2025; McFadden 1981), which is more realistic in the economic settings we are tackling. This assumption constrains the agent’s alternatives to be probabilistically chosen, at every timeslot, according to a single profile. The manipulation of the agent thus consists in committing to a “fake” profile, instead of adversarially selecting a new profile at every timeslot as in (Sessa et al. 2020). Moreover, we assume agents aim to maximize their utility, and not necessarily act adversarially (e.g., aim to maximize the loss of the regulator). Second, (Sessa et al. 2020) assume the leader adopts probabilistic policies, while in our case incentive policies are constrained to be deterministic. Indeed, it would not be acceptable that a regulator (e.g., a government) gives randomized incentives to individuals.

Ionescu et al. (2023) study strategic behavior in matching algorithms, while we consider demand-side management policies. Moreover, their manipulation is implemented by “returning agents”, which take the role of our regulator. We instead study user-side manipulation. A different, but related, problem is the personalized reserved price setting in a repeated auction, considered by (Nedelec et al. 2020). In their work, however, the set of leaders (corresponding to the buyers) are those who can potentially manipulate the follower (the seller). As a consequence, their findings do not hold in our case, in which agents (followers) can manipulate the regulator (leader).

In Performative Prediction (PP) (Perdomo et al. 2020), predictions produced by a supervised learning model alter the data-generating process itself, thereby changing the distribution on which future predictions are evaluated. However, in PP the action of the leader is to change classification parameters, whereas the action of our leader (the regulator) is to propose incentives. Moreover, different from PP, the goal of our leader is not to predict, but to propose incentives. In PP, the leader aims to reduce classification loss, whereas our leader aims to reduce generalized social cost. In PP, the agent influences the classifier parameters, which is not our case. Also, PP is not applicable to demand-side policies: in (Perdomo et al. 2020, §5.1), the regulator is allowed to change the classifier over time. However, for transparency constraints, a regulator, such as a public authority, must fix and announce the rules in advance; they would otherwise be unacceptable and perceived as black-box or arbitrary by citizens. This is why we fix the agent’s profile classification is performed via the Maximum Likelihood Estimator, according to real preference elicitation cases (Chapman and Fisher 2025).

Our setting has some similarity with the Vickrey–Clarke–Groves (VCG) mechanism. The latter computes prices so as to maximize social welfare, while ensuring truthful behavior of the agents. One could be tempted to use a VCG-like mechanism in our setting. This is however not possible, because in VCG, the social welfare is the sum of the agent’s utilities, while in our case the social cost that the regulator aims to minimize is not such a sum, but is rather related to sustainability goals (e.g., emissions).

Despite the rich literature in the analysis of manipulative

behavior, the settings considered are either too simplistic - the agent can manipulate by simply misreporting - or too pessimistic - the manipulative behavior is not appropriately constrained to be economically and behaviorally grounded. This paper aims to fill this gap.

3 Model

Notation is entirely defined in the paper. For ease of reference, it is summarized in Table 1 of the the Suppl. Material. We formulate a repeated Stackelberg game (as (Birmpas et al. 2020)) with a leader (regulator) and a follower (agent). At each iteration, the agent chooses an alternative $i \in \mathcal{I}$, which generates externality $c_i \geq 0$. Let alternatives $\mathcal{I} = \{1, \dots, I\}$ be sorted in the increasing order of externalities, i.e., from the most to the less sustainable. To encourage the choice of the most sustainable alternative, at the beginning of each iteration the regulator proposes incentive $y \geq 0$.

To give our model some context, we briefly mention some real-world applications. For instance, consumers (agents) can choose to buy a short-lasting, unethically produced clothing item, which would correspond to alternative 2, with c_2 representing the environmental and social externality caused by producing this good. On the other hand, alternative 1 would be to buy this clothing item second-hand, which has low externality c_1 . Incentive y would be a subsidy to encourage buying second-hand. Another application consists in the choice of a traveler (agent) of public transport (alternative 1) or taking the car (alternative 2) for a trip. Here, externality c_2 would represent the car-related externality, e.g. pollution, induced congestion. The government could provide an incentive y to encourage taking public transport, which produces lesser externality c_1 . A last example is the decision of a firm (agent) to prefer electricity generated from renewable energies (alternative 1) or traditional, fossil-fuel-based electricity (alternative 2), which imposes a higher externality c_2 in the form of CO₂ emissions. The regulator can set a subsidy y to encourage the firm to use renewable sources.

3.1 Random Utility Modeling

We model agent behavior according to random utility theory (RUT) (Strzalecki 2025; McFadden 1981). The agent has a *utility function* $v(i, y)$, associating a certain *systematic utility* (or, briefly, utility (Cascetta 2009, §3.2)) to alternative i when the regulator proposes incentive y . When the agent chooses alternative $i \in \mathcal{I}$, they get random utility: systematic utility $v(i, y)$ plus a random variable ϵ_i , called *random term*. Following RUT, we assume ϵ_i are identically and independently drawn from a Gumbel distribution with 0-mean and scale parameter $\theta \geq 0$. The random term accounts for some inherent randomness in agent's behavior (Cascetta 2009, p100), which makes the agent's choice not easily predictable for the regulator. This randomness may be due to the inability of the agent to estimate the exact alternatives' utility or to an additional utility disturbance, which is completely observable only after the choice is made. The agent adopts in general a mixed *strategy*, i.e., they make choices according to a probability distribution over the available alternatives. We focus here on a class of mixed strategies, well-known in RUT, namely the *logit model* (McKelvey and Palfrey 1995, §3): if

the agent has a systematic utility function $v(\cdot, \cdot)$ and the regulator proposes incentive y , they will choose alternative i with probability:

$$\mathbb{P}_i^v(y) = \frac{e^{v(i, y)/\theta}}{\sum_{i' \in \mathcal{I}} e^{v(i', y)/\theta}}. \quad (1)$$

The probabilistic nature of the choice accounts for the bounded rationality of the agent (Wheeler 2018), who does not always make the theoretically optimal decisions, due to limited information. In our case, the agent does not know in advance the actual outcome $v(i, y) + \epsilon_i$ of their choice, which is revealed only *after* the choice is made. This is because the actual realization of ϵ_i is not known before alternative i is chosen. Note that, the larger the scale parameter, the greater the randomness in the agent's choice.

The behavior of the agent is determined by their *profile*, capturing inherent agent's preferences or needs impacting their choices. A profile is mapped to a certain systematic utility function. The set of profiles is $\mathcal{K} = \{0, 1, \dots, K - 1\}$ and their respective utility functions are $u_k(\cdot, \cdot)$.

We allow any finite number K of agent profiles. According to standard RUT ((Cascetta 2009, §3.2.5a)), each alternative i is described with a (possibly) agent-dependent attribute vector $\mathbf{x}_{i,k}$. Considering the case in which alternatives represent transport modes, the components of $\mathbf{x}_{i,k}$ may represent monetary cost, walking travel time, in-vehicle time, comfort, etc. If alternatives represent different products, $\mathbf{x}_{i,k}$ may instead represent characteristics such as the price, the reputation of the brand, the weight, etc. Some components are agent-dependent: e.g., the in-vehicle time to reach a certain destination depends on the starting location, e.g., the residence, which varies from one agent to another. This is why $\mathbf{x}_{i,k}$ is indexed by k . Each profile is associated with a coefficient vector α_k , which captures how individuals with profile k value the various attributes. For instance, older individuals may assign greater importance to walking time compared to younger individuals. The systematic utility is thus $\alpha_k \cdot \mathbf{x}_{i,k}$ in the absence of any incentive policy.

Lemma 1. *Assuming utility is quasi-linear with respect to income¹ utility function $u_k(\cdot, \cdot)$ can be written as*

$$u_k(i, y) = \begin{cases} a_k + y & i = 1 \\ b_k & i \neq 1 \end{cases}; \quad \begin{aligned} a_k &:= \alpha_k \cdot \mathbf{x}_{1,k}; \\ b_k &:= \theta \ln \sum_{i \geq 0} e^{\frac{\alpha_k \cdot \mathbf{x}_{i,k}}{\theta}} \end{aligned}$$

Also, $p_k(y) := \mathbb{P}_1^{u_k}(y) = \sigma\left(\frac{a_k - b_k + y}{\theta}\right)$ (2)

where $\sigma(\cdot)$ is the sigmoid

This equation summarizes the fact that, depending on their profiles, some agents might have higher disutility than other when choosing alternative 1, and we compensate them for this loss, adapting incentive y to their case. This personalized compensation introduces vertical equity (Black et al. 2022, §3.2), since it distributes incentives to those who need it the

¹This is a standard assumption in econometrics ((Javaudin, Araldo, and de Palma 2023)), which means that both $\alpha_k \cdot \mathbf{x}_{i,k}$ and the incentive y are expressed in the same unit as income and that y has an additive effect on utility.

most to change their behavior. Quantity $a_k - b_k$ interestingly captures how much profile k is *inherently responsible*, i.e., how much they prefer the more sustainable alternative $i = 1$ over the others, independent of the proposed incentive. We call $a_k - b_k$ the *responsibility intensity* of that profile. With no lack of generality, we order all profiles accordingly:

$$a_{k-1} - b_{k-1} \geq a_k - b_k, \text{ for } k = 1, \dots, K - 1 \quad (3)$$

Lemma 2. *Let us consider an agent following a certain profile $k \in \mathcal{K}$. Then, the expectation of the externalities caused by them and of the incentive collected are, respectively:*

$$\begin{aligned} \text{ext}^k(y) &= p_k(y) \cdot c_1 + (1 - p_k(y)) \cdot c_{-1}; \\ \text{inc}^k(y) &= p_k(y) \cdot y \\ \text{with } c_{-1} &:= \sum_{i>1} \frac{e^{\frac{u_0(i,y)}{\theta}}}{\sum_{i'>1} e^{\frac{u_0(i',y)}{\theta}}} c_i > c_1 \end{aligned} \quad (4)$$

The generalized social cost when the agent chooses i consists of social cost c_i and the cost of incentive y_i provided by the regulator, the latter term weighted by a parameter $\beta > 0$, which expresses the relative importance the regulator gives to incentive-related expenditures versus externalities and also normalizes the units of measure. The expected social cost is

$$\tilde{f}^k(y) = \text{ext}^k(y) + \beta \cdot \text{inc}^k \quad (5)$$

Note that computing the expected value above requires knowing the utility $\hat{v}(\cdot, \cdot)$ that the agent is adopting. This information is not available to the regulator, which can only thus only estimate this expectation, based on estimate $\hat{v}(\cdot, \cdot)$, inferred via preference elicitation.

3.2 Interaction between Regulator and Agent

The interaction between the regulator and the agent is a sequential game. In each timeslot t :

1. The regulator infers profile $\hat{k}^t = \psi(\mathbf{h}^{t-1})$ of the agent, via a deterministic inference rule ψ applied on previous observations $\mathbf{h}^{t-1} = y^1, i^1, \dots, y^{t-1}, i^{t-1}$, where y^s, i^s are the incentives proposed by the regulator and the alternative selected by the agent at iteration $s = 1, \dots, t-1$. In the numerical results, inference rule $\psi(\cdot)$ is the maximum likelihood estimator, details in (66) of the supplemental material.
2. The regulator takes its *action*. An action is the incentive y^t that the regulator proposes to the agent at timeslot t for choosing alternative 1. The regulator proposes incentive $y^t = \phi(\hat{k}^t)$, which is associated with inferred profile $\hat{k}^t$ via deterministic function $\phi(\cdot)$. We assume that this function returns the incentive so as to minimize the estimated expectation of the generalized social cost (5).
3. The agent chooses alternative $i^t \in \mathcal{I}$, according to $\mathbb{P}_i^v(y^t)$ formulated in (1).
4. The regulator transfers incentive y^t to the agent, if $i^t = 1$, or no incentive if $i^t > 1$.
5. The agent gets utility $v(i^t, y^t) + e_{i^t}$, where e_{i^t} is a realization of random noise ϵ_{i^t} .

The regulator may interact with multiple agents. We focus on one regulator-agent interaction, assuming cross-agent effects do not materially alter individual choices. This standard assumption (Javaudin, Araldo, and de Palma 2023) is appropriate because realistic incentive policies, even when producing non-negligible effects, rarely induce shifts massive enough to substantially change the surrounding choice environment, e.g., affecting congestion. Since in this case aggregate social cost is additive: the system then decomposes into separate regulator-agent interactions, and our analysis applies to each agent. Regulator budget constraints could be added in step 2, which we do not do as it would not add conceptual added value. Moreover, the regulator could solve complex optimization problems when calculating y^t , e.g., to maximize some long-term cumulative objective. However, such complex optimizations would not be implementable by real decision makers: for political acceptance and transparency to citizens, policies must be based on simple, understandable rules, to prevent citizens from wrongly perceiving that incentives are calculated by an arbitrary black box. Our personalized incentives are instead easy to understand by the agents, since they are based on the simple concept that an agent with profile k deserves a certain incentive y_k^* , whose value can be transparently announced by the regulator a priori. Therefore, $y^t = y_{k^t}^*$. Prop. 6 will detail the way in which such a value is calculated.

4 Analysis

4.1 Optimal Incentives

There are unique incentives that minimize the estimated expectation of the generalized social cost $\tilde{f}^{\hat{v}^t}(y)$:

Proposition 3. *There is a unique optimal incentive for inferred profile $u_k(\cdot, \cdot)$:*

$$y_{u_k}^* = \max \left(\frac{c_{-1} - c_1}{\beta} - \theta \left(1 + W \left(e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1} - c_1}{\theta\beta} - 1} \right) \right), 0 \right)$$

where $W(\cdot)$ is the Lambert's W-function. Moreover,

Corollary 4. *Let y_k^* be the optimal incentives as defined in Proposition 3. Then y_k^* is non-increasing with responsibility intensity $a_k - b_k$.*

This proposition tells us which profiles expect to have higher incentives. Thanks to ordering (3) and Cor. 4 we can thus determine the order of proposed incentives.

The following proposition prevents optimal incentives from being infinitely large, which is crucial when evaluating the economic feasibility of the regulator's policy.

Proposition 5. *The optimal incentive is upper-bounded, i.e.,*

$$0 \leq y_{\hat{v}^t}^* < \frac{c_{-1} - c_1}{\beta}.$$

As expected, the previous bound grows with the externality c_{-1} generated when not choosing alternative 1 and decreases with the social cost c_1 of alternative 1 since the regulator can prevent more social cost by encouraging alternative 1 if there is larger difference between the costs caused by alternatives 1 and the others.

4.2 Agent's Strategic Behavior

We focus here on an agent that is inherently responsible, i.e., their true profile is the *responsible profile* $v(\cdot, \cdot) = u_0(\cdot, \cdot)$ (see ordering (3)) and who has a natural preference toward sustainable alternatives, i.e., $a_0 > b_0$. We study in which case such an agent may be better off manipulating, by adopting (i.e., simulating) another profile.

We assume that the agent has *bounded rationality*, i.e., they seek to maximize their utility subject to cognitive and computational limitations, which prevents them from adopting arbitrarily complex optimization. This is compliant with human behavior (Wheeler 2018) and rules out sophisticated Markovian or RL planning strategies that require accounting for long-term rewards. We thus assume that, if the agent decides to be manipulative, they will imitate another profile $r \in \mathcal{K}$, via adopting utility $u_r(\cdot, \cdot)$, i.e., consistently choosing alternative $i^t = 1$ or $i^t > 1$ according to probability $p^r(\cdot)$ (see Lemma 1) in all the timeslots. In the following analysis, among profiles $\mathcal{K}$, three deserve particular attention:

- $k = 0$ is the most responsible profile. We focus on a user whose true profile is 0 and study under which conditions they are better off manipulating.
- r is the profile adopted by the agent, which determines probability $p^r(\cdot)$ to choose the sustainable alternative. If the agent behaves honestly, $r = 0$, otherwise, $r > 0$.
- $\hat{k}^t$ is the regulator's estimate at time t of the agent's profile.

The true utility collected by the agent when the regulator proposes incentive y is

$$u_0(y) = \begin{cases} a_0 + y & \text{if choosing } i = 1 \\ b_0 & \text{if choosing } i > 1, \end{cases} \quad (6)$$

no matter whether the agent is honest ($r = 0$) or not ($r > 0$).

When a user is manipulative, it incurs a loss of true utility, due to the fact that they do not choose alternatives in accordance with their true profile. On the other end, they have a gain, obtained via shifting the probability distribution of the proposed incentive Y_t toward a more favorable one. We capture this loss and gain in the definitions of Gain from Manipulation and Cost of Manipulation, which will be given in Lemma 7. We need first to introduce some notation.

Definition 6. [Notation of the random variables] Let $\hat{K}^t = \psi(\mathbf{H}^{t-1})$ be the random variable (r.v.) denoting the profile inferred by the regulator at time t . It is the result of applying a deterministic function $\psi(\cdot)$ (see (66)) to $\mathbf{H}^{t-1}$, which is the r.v. representing the history. $\mathbf{H}^{t-1}$ is stochastic since the agent's choices are stochastic (Lemma 1). Let $Y^t = \phi(\hat{K}^t)$ denote the incentive proposed at time t . It is a r.v., whose possible values are y_k^* , $k \in \mathcal{K}$.

Let $U_0(y)$ denote the random variable representing such a utility. It takes values $a_0 + y$ and b_0 with probability $p_k(y)$ and $1 - p_k(y)$, respectively (Lemma 1). The true utility collected by the agent at time t is thus random variable $U_0(Y^t)$. Its stochasticity comes from the stochasticity of Y^t , as well as the stochasticity of $U_0(y)$ for any value of y .

For notational compactness, we introduce the following conditional expectation and probability operators

$$\mathbb{E}_k[\cdot] = \mathbb{E}[\cdot | \text{agent adopts profile } k], \quad (7)$$

$$\mathbb{P}_k[\cdot] = \mathbb{P}[\cdot | \text{agent adopts profile } k]. \quad (8)$$

Let r.v. Y_k^t and $\hat{K}_k^t$ represent the incentive proposed at time t and the profile inferred by the regulator, when the agent adopts profile k . Observe that all r.v. $\hat{K}_k^t$, $\forall k \in \mathcal{K}$, $\forall t$ have the same domain $\mathcal{K}$. They only differ in how the probability mass, i.e., $P_k[A]$, $\forall A \subseteq \mathcal{K}$ (8), is distributed in the domain. Similarly, all r.v. Y_k^t , $\forall k \in \mathcal{K}$, $\forall t$ have the same domain $\{y_k^*\}_{k \in \mathcal{K}}$. This notation makes it clear that, by adopting a certain profile $k \in \mathcal{K}$, the agent induces a certain probability distribution over the inferred profile, and thus over the incentive proposed.

Lemma 7. Let $CM_r(y)$ be the Cost of Manipulation incurred by an agent simulating profile r when the proposed incentive is $y \in \{y_k^*\}_{k \in \mathcal{K}}$:

$$CM_r(y) := \mathbb{E}_0[U_0(y)] - \mathbb{E}_r[U_0(y)]. \quad (9)$$

Let CM_r^t be its mean:

$$CM_r^t := \mathbb{E}_0[CM_r(Y^t)] = \sum_{k \in \mathcal{K}} \mathbb{P}_0(Y^t = y_k^*) \cdot CM_r(y_k^*) \quad (10)$$

Let GM_r^t be the Gain from Manipulation when simulating profile r , defined as

$$GM_r^t := \sum_{k \in \mathcal{K}} \left(\begin{array}{c} \mathbb{P}_r[Y^t = y_k^*] \\ -\mathbb{P}_0[Y^t = y_k^*] \end{array} \right) \mathbb{E}_r[U_0(y_k^*)]. \quad (11)$$

Let ΔU_r^t be the manipulation advantage, defined as the additional expected true utility collected by the agent at time t when simulating profile r over acting honestly.

Such an advantage can be decomposed in

$$\Delta U_r^t := \mathbb{E}_r[U_0(Y^t)] - \mathbb{E}_0[U_0(Y^t)] = GM_r^t - CM_r^t. \quad (12)$$

Proof. By the law of total expectation, for each $j \in \{0, r\}$, $\mathbb{E}_j[U_0(Y^t)] = \sum_{k \in \mathcal{K}} \mathbb{P}_j(Y^t = y_k^*) \mathbb{E}_j[U_0(y_k^*)]$. Replacing these expressions in (12) and adding and subtracting $\sum_{k \in \mathcal{K}} \mathbb{P}_0(Y^t = y_k^*) \mathbb{E}_r[U_0(y_k^*)]$ gives the Lemma. $\square$

We now study how the agent can manipulate the probability distribution of the incentives. Let IE_r^t denote the *intensity of elicitation* of profile r at time t , i.e., how much the agent can change the probability distribution of the profile inferred by the regulator, by simulating profile $r > 0$:

$$IE_r^t := D_{\text{KL}}(\hat{K}_0^t \| \hat{K}_r^t), \quad (13)$$

which is the Kullback-Leibler (KL) divergence between the profile inferred by the regulator when the agent adopts true profile 0 or simulates profile r , respectively.

Similarly, let IP_r^t denote the *intensity of personalization* of profile r at time t , i.e., how much the agent can change the probability distribution of the incentives proposed by the regulator, by simulating profile $r > 0$:

$$IP_r^t := D_{\text{KL}}(Y_0^t \| Y_r^t) \leq IE_r^t, \quad (14)$$

where the inequality is obtained thanks to the contraction of divergence induced by function $\hat{Y}_k^t = \phi(\hat{K}_k^t)$ (Polyanskiy and Wu 2025, Theor. 7.4).

Theorem 8. *The gain from manipulation is bounded by*

$$0 \leq GM_r^t \leq R_r \sqrt{IP_r^t/2} \leq R_r \sqrt{IE_r^t/2}, \quad (15)$$

$$\text{where } R_r := \mathbb{E}_r[U_0(y_{K-1}^*)] - \mathbb{E}_r[U_0(y_0^*)] \quad (16)$$

is the range of expected true utility.

Proof. Thanks to Prop. 3, Y_r^t First-Order-Stochastically Dominates (FOSD) Y_0^t , which intuitively means that by simulating $r > 0$ the agent shifts the probability mass toward higher values of incentives. Observe that $y \mapsto \mathbb{E}[U_0(y)]$ is increasing. Looking at the sum defining GM_r^t (11), the terms with high values of $\mathbb{E}[U_0(y)]$ are multiplied by positive numbers ($\mathbb{P}_r[\cdot] > \mathbb{P}_0[\cdot]$, due to FOSD) and only low values of $\mathbb{E}[U_0(y)]$ are multiplied by negative numbers. Simple calculation thus yields a positive sum and thus $GM_r^t \geq 0$.

As for the upper bound, we will first bound GM_r^t in terms of total variation $\delta(Y_0^t, Y_r^t)$ between the probability distributions of Y_0^t and Y_r^t (def. in (Polyanskiy and Wu 2025, (7.3))).

Since $\sum_{k \in \mathcal{K}} (\mathbb{P}_r[Y^t = y_k^*] - \mathbb{P}_0[Y^t = y_k^*]) = 0$, we can introduce any constant A in the expression of GM_r^t (Lemma 7) without changing its value, as follows:

$$\begin{aligned} GM_r^t &= \sum_{k \in \mathcal{K}} \left(\frac{\mathbb{P}_r[Y^t = y_k^*]}{-\mathbb{P}_0[Y^t = y_k^*]} \right) \cdot (\mathbb{E}_r[U_0(y_k^*)] - A) \\ &= \frac{R_r}{2} \sum_{k \in \mathcal{K}} \left(\frac{\mathbb{P}_r[Y^t = y_k^*]}{-\mathbb{P}_0[Y^t = y_k^*]} \right) \cdot \frac{\mathbb{E}_r[U_0(y_k^*)] - A}{R_r/2} \\ &= \frac{R_r}{2} \sum_{k \in \mathcal{K}} \left(\frac{\mathbb{P}_r[Y^t = y_k^*]}{-\mathbb{P}_0[Y^t = y_k^*]} \right) \cdot f(y_k^*), \end{aligned}$$

where we define function $f(y_k^*) = \frac{\mathbb{E}_r[U_0(y_k^*)] - A}{R_r/2}$. The previous expression is equivalent to

$$GM_r^t = \frac{R_r}{2} (\mathbb{E}_r[f(Y^t)] - \mathbb{E}_0[f(Y^t)]) \quad (17)$$

Let us choose in particular

$$A = (\mathbb{E}_r[U_0(y_{K-1}^*)] + \mathbb{E}_r[U_0(y_0^*)]) / 2$$

Since $y \mapsto \mathbb{E}_r[U_0(y)]$ is an increasing function. Also, thanks to Prop. 3, y_k^* is non decreasing with k . Since A is the midpoint between the maximum and minimum expected value,

$$|A - \mathbb{E}_r[U_0(y_k^*)]| \leq R_r/2, \forall k \in \mathcal{K}. \quad (18)$$

and thus $|f(y_k^*)| \leq 1, \forall k \in \mathcal{K}$. We can thus apply the sup-representation of total variation from Polyanskiy and Wu (2025, (7.18)) and obtain

$$GM_r^t \leq R_r \cdot \delta(Y_0^t, Y_r^t). \quad (19)$$

Thanks to the Pinsker inequality first (Polyanskiy and Wu 2025, (7.27))

$$\delta(Y_0^t, Y_r^t) \leq \sqrt{\frac{1}{2} D_{\text{KL}}(Y_0^t \| Y_r^t)}, \quad (20)$$

Combining (19), (20) and (14) concludes the proof. $\square$

Theor. 9 formalizes that the more the intensity of personalization, the more the agent can, at each timeslot, gain from manipulation. Theor. 9 result adds that, complying with intuition, the more the interactions between the regulator and the agent, the more the latter can manipulate the former. However, Theor. 9 limits the rate of growth of the potential for manipulation: such a growth can be at most linear.

Theorem 9. *Manipulation potential IE_r^t is bounded by*

$$IE_r^t \leq (t-1)IE_r \quad (21)$$

$$\text{where } IE_r := \max_{k \in \mathcal{K}} D_{\text{KL}}(\text{Ber}(p_0(y_k^*)) \| \text{Ber}(p_r(y_k^*))), \quad (22)$$

and $\text{Ber}(p)$ is a Bernoulli distribution with probability p . Moreover, the gain from manipulation is bounded by

$$0 \leq GM_r^t \leq R_r \sqrt{\frac{1}{2}(t-1)IE_r}. \quad (23)$$

Proof. The history of proposed incentives and chosen alternatives is $\mathbf{h}^{t-1} = y^1, i^1, \dots, y^{t-1}, i^{t-1}$. Therefore, the realization of $\hat{K}^t$ is obtained as a deterministic function $\psi(\mathbf{h}^{t-1})$. Let $\mathbf{H}_k^{t-1}$ be the random variables representing the possible histories when the agent adopts profile k . Then

$$IE_r^t = D_{\text{KL}}(\hat{K}_0^t \| \hat{K}_r^t) \leq D_{\text{KL}}(\mathbf{H}_0^{t-1} \| \mathbf{H}_r^{t-1}) \quad (24)$$

$$\leq \left[\sum_{s=1}^{t-1} \sum_{\mathbf{h} \in \mathcal{H}^{s-1}} \mathbb{P}_0[\mathbf{H}^{s-1} = \mathbf{h}] \cdot D_{\text{KL}} \left(\frac{\mathbb{P}_0(I^s | \mathbf{H}^{s-1} = \mathbf{h})}{\mathbb{P}_r(I^s | \mathbf{H}^{s-1} = \mathbf{h})} \right) \right] \quad (25)$$

$$= \left[\sum_{s=1}^{t-1} \sum_{\mathbf{h} \in \mathcal{H}^{s-1}} \mathbb{P}_0[\mathbf{H}^{s-1} = \mathbf{h}] \cdot D_{\text{KL}} \left(\frac{\text{Ber}(p_0(y_{\psi(\mathbf{h}^{s-1})}^*))}{\text{Ber}(p_r(y_{\psi(\mathbf{h}^{s-1})}^*))} \right) \right] \quad (26)$$

$$\stackrel{(22)}{\leq} (t-1)IE_r, \quad (27)$$

Inequality (24) simply exploits contraction of divergence induced by applying a function (Polyanskiy and Wu 2025, Theor. 7.4). Eq. (25) is the chain rule obtained as in (Okoroafor et al. 2023, (9)) and $\mathcal{H}^{s-1}$ is the set of all possible histories up to time $s-1$. Finally, Eq (23) is the combination of Theor. 8 and (21). $\square$

Is it thus possible to understand up to which timeslot manipulation is deterred, as formalized in the following corollary.

Corollary 10. *The manipulation of the regulator via imitation profile r is deterred in any timeslot t s.t.*

$$t < 1 + \frac{2}{IE_r} \left(\frac{CM_r^{\min}}{R_r} \right)^2 \quad (28)$$

$$\text{where } CM_r^{\min} = \min_{k \in \mathcal{K}} CM_r(y_k^*). \quad (29)$$

Proof. Lemma 7 implies that manipulation is beneficial only if $GM_r^t - CM_r^t > 0$:

$$0 < GM_r^t - CM_r^t \leq R_r \sqrt{\frac{1}{2}(t-1)IE_r} - CM_r^{\min} \quad (30)$$

The last inequality is obtained by applying (23) and (29). Therefore, (28) ensures manipulation is deterred. $\square$

5 Impact of Preference Elicitation

We now examine how preference elicitation limits the agent’s ability to manipulate the regulator. It not only makes it harder for the agent to convince the regulator that their adopted profile is $r > 0$, but also substantially reduces the expected utility gained from manipulation.

We compare two preference-elicitation models. The first, *Instantaneous Preference Elicitation (InstPE)*, is implicitly assumed in previous artificial intelligence works (Dawkins, Han, and Xu 2021; Birmpas et al. 2020; Gan et al. 2019a,b; Nguyen 2019). Under *InstPE*, the regulator immediately accepts the agent’s simulated utility profile, so that $\hat{k}^t = r$ for all t . This corresponds to strategic misreporting and allows an agent behaving according to profile r to immediately maximize their expected utility.

We instead introduce *Realistic Preference Elicitation (RealPE)*. Here, the regulator observes the agent’s choices over time and updates its belief ψ about the agent’s utility function at each timeslot t . The agent must therefore establish the simulated profile through repeated interaction and may meanwhile be classified under another profile. This repeated simulation is costly because it may require choosing alternatives that the agent does not prefer. To distinguish the two settings, we denote expectations by $\mathbb{E}^{InstPE}$ and $\mathbb{E}^{RealPE}$. Under *InstPE*, the agent can directly compare the expected utilities of honest and manipulative behavior.

Observation 11. *In InstPE (i) the regulator proposes incentive y_k^* if the agent declares profile $k \in \mathcal{K}$ and (ii) the agent always behaves according to their true profile 0, no matter the declared profile k . Therefore, the operators (8)-(7) are such that $\mathbb{P}_k[\cdot] = \mathbb{P}_0[\cdot]$ and $\mathbb{E}_k[\cdot] = \mathbb{E}_0[\cdot]$.*

The following Lemma shows that the cost of manipulation is null under *InstPE* and the gain from manipulation is directly the difference between the expected true utilities, under incentive y_r^* of simulated profile r and under incentive y_0^* associated to the true profile 0.

Lemma 12.

$$GM_r^{t, InstPE} = \mathbb{E}_0[U_0(y_r^*)] - \mathbb{E}_0[U_0(y_0^*)], \quad CM_r^t = 0, \quad \forall t$$

6 Case Study: Travel-Mode Incentives

We here validate the theoretical findings and show the deterrent impact of Realistic Preference Elicitation (*RealPE*) on manipulative users, compared to Instantaneous Preference Elicitation (*InstPE*). We run a comprehensive Monte Carlo simulation campaign calibrated on transport mode choices in London (open dataset of Hillel, Elshafie, and Jin (2018)), which reports choices of travelers (agents), with $i = 1$ (bike), $i = 2$ (public transport (PT)), $i = 3$ (car). We calibrated the parameters related to 3 profiles $k = 0, 1, 2$ and the respective behavioral parameters, and ordered them according to the responsibility intensity. Without imposing it, the calibration returned profiles with preference toward biking, PT, and car. We then run Monte Carlo simulation over multiple seeds, to implement the stochastic choice of the agent, the profile inference (preference elicitation) of the regulator and the calculation of the proposed incentive. The important parameters assigned based on previous experimented incentive policies

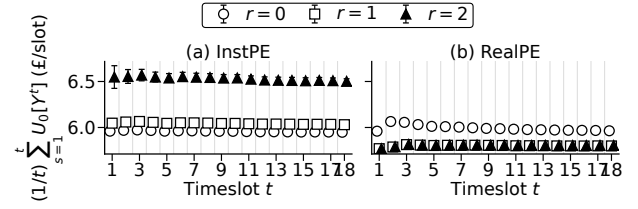


Figure 1: Time-averaged true agent utility. Averages across all the random realizations (corresponding each to a random seed) and 95% confidence intervals are represented

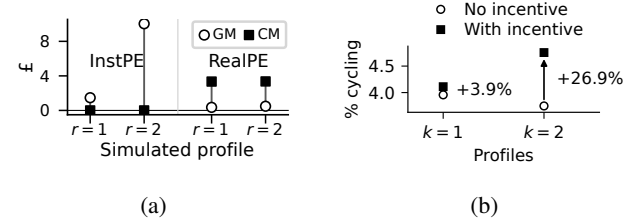


Figure 2: (a) Cost of and gain from manipulation (*CM*, *GM*); (b) change in the percentage of cycling trips

and CO₂ emission evaluations, as well as values obtained via calibration and (6) are $\theta = 3.7$, $\{a_k\}_k = \{6.1, -6.8, -6.6\}$, $\{b_k\}_k = \{5.8, -11.7, -11.9\}$, $\{c_{neg1,k}\}_k = \{2.8, 2.2, 2.6\}$, $\beta = 0.6$, $\{y_k^*\}_k = \{0, 0.1, 0.9\}(\mathcal{E})$.

A detailed explanation of the dataset, the hyperparameters, the calibration method and the parameters resulting from calibration §D1 of the of the Suppl.Material. A Python notebook that reproduces dataset downloading and preprocessing, calibration and simulation is provided in the Suppl.Material. It runs entirely in less than 2 minutes in Google Colab.

Fig. 2a shows that Cost of Manipulation is higher than the Gain under *RealPE*. Due to (12), time-averaged true utility $(1/t) \sum_{s=0}^t U_0(Y^s)$ is higher when acting honestly under *RealPE* (Fig. 1). Manipulation is thus deterred. Fig. 1-2a also shows that the situation would be inverted under the over-pessimistic *InstPE* setting. Therefore, under realistic conditions, personalized incentive policies can be effective in fostering sustainable behavior, as shown in Fig. 2b where we computed the induced shift in bike usage.

7 Conclusion

By explicitly modeling the regulator’s learning process in personalized incentive policies, we establish an upper bound on the gain achievable from manipulative agents. Interestingly, such a bound tightens as the regulator’s ability to distinguish profiles decreases. We characterize the resulting deterrence effect and show, through simulations calibrated on observed travel choices, how it improves policy effectiveness. Our analysis can serve as a pre-deployment test, helping regulators assess the policy’s robustness to manipulation.

References

- Ascarza, E.; and Israeli, A. 2022. Eliminating unintended bias in personalized policies using bias-eliminating adapted trees (BEAT). *PNAS*, 119(11): e2115293119.
- Birmpas, G.; Gan, J.; Hollender, A.; Marmolejo, F.; Rajgopal, N.; and Voudouris, A. 2020. Optimally deceiving a learning leader in stackelberg games. *NeurIPS*, 33: 20624–20635.
- Black, E.; Elzayn, H.; Chouldechova, A.; Goldin, J.; and Ho, D. 2022. Algorithmic fairness and vertical equity: Income fairness with IRS tax audit models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 1479–1503.
- Blasco, J.; Guillaud, E.; and Zemmour, M. 2023. The inequality impact of consumption taxes: An international comparison. *J. Public Economics*, 222: 104897.
- Cascetta, E. 2009. *Transportation systems engineering*. Springer.
- Chapman, J.; and Fisher, G. 2025. Preference Elicitation: Common Methods and Potential Pitfalls. In *Handbook of Experimental Methodology, Volume 1*. Chapter 2.
- Creutzig, F.; Roy, J.; Devine-Wright, P.; Díaz-José, J.; Geels, F.; Grubler, A.; Maïzi, N.; Masanet, E.; Mulugetta, Y.; Onyige-Ebeniro, C.; et al. 2022. Demand, services and social aspects of mitigation.
- Dabla-Norris, E.; Helbling, T.; Khalid, S.; Khan, H.; Magistretti, G.; Sollaci, A.; and Srinivasan, K. 2023. Public Perceptions of Climate Mitigation Policies: Evidence from Cross-Country Surveys. *Staff Discussion Notes*, 2023(002).
- Dawkins, Q.; Han, M.; and Xu, H. 2021. The Limits of Optimal Pricing in the Dark. *NeurIPS*, 34: 26649–26660.
- Firpo, S.; Pieri, R.; Jr., E. P.; and Souza, A. P. 2014. Evidence of eligibility manipulation for conditional cash transfer programs. *Economia*, 15(3): 243–260.
- Gan, J.; Guo, Q.; Tran-Thanh, L.; An, B.; and Wooldridge, M. 2019a. Manipulating a Learning Defender and Ways to Counteract. In *NeurIPS*.
- Gan, J.; Xu, H.; Guo, Q.; Tran-Thanh, L.; Rabinovich, Z.; and Wooldridge, M. 2019b. Imitative follower deception in stackelberg games. In *EC*.
- Gneezy, U.; and Rustichini, A. 2000. Pay enough or don't pay at all. *Quarterly J. Econ.*, 115(3): 791–810.
- Hillel, T.; Elshafie, M. Z.; and Jin, Y. 2018. Recreating passenger mode choice-sets for transport simulation: A case study of London, UK. *Proceedings of the Institution of Civil Engineers-Smart Infrastructure and Construction*, 171(1): 29–42.
- IEA. 2021. Do we need to change our behaviour to reach net zero by 2050?
- Ionescu, S.; Du, Y.; Joseph, K.; and Hannak, A. 2023. Strategic Behavior in Two-sided Matching Markets with Prediction-enhanced Preference-formation. *NeurIPS*, 36: 36041–36052.
- IPCC. 2023. Climate Change 2023: Synthesis Report, Summary for Policymakers.
- Jain, A.; and Perchet, V. 2024. Calibrated Forecasting and Persuasion. In *EC*.
- Javaudin, L.; Araldo, A.; and de Palma, A. 2023. Personalised incentives with constrained regulator's budget. *Transportmetrica A: Transport Science*, 21(3): 2284353.
- Kamenica, E.; and Gentzkow, M. 2011. Bayesian persuasion. *American Economic Review*, 101(6): 2590–2615.
- Le Bras, H. 2019. Voiture, Gilets jaunes et Rassemblement national. *Études*, (4): 31–44.
- McFadden, D. 1981. Econometric models of probabilistic choice. 198272.
- McKelvey, R. D.; and Palfrey, T. R. 1995. Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1): 6–38.
- Nedelec, T.; Calauzenes, C.; Perchet, V.; and Karoui, N. E. 2020. Robust Stackelberg buyers in repeated auctions. *AIS-TATS*, 108.
- Nguyen. 2019. Imitative Attacker Deception in Stackelberg Security Games. *IJCAI*.
- Nguyen et al. 2019. Deception in finitely repeated security games. *AAAI*.
- Okoroafor, P.; Gupta, V.; Kleinberg, R.; and Goh, E. 2023. Non-stochastic CDF estimation using threshold queries. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 3551–3572. SIAM.
- Perdomo, J.; Zrnic, T.; Mendler-Dünner, C.; and Hardt, M. 2020. Performative Prediction. *ICML*, 119: 7599–7609.
- Polyanskiy, Y.; and Wu, Y. 2025. *Information Theory: From Coding to Learning*. Cambridge, UK: Cambridge University Press. ISBN 978-1-108-83290-8.
- Sessa, P. G.; Bogunovic, I.; Kamgarpour, M.; and Krause, A. 2020. Learning to Play Sequential Games versus Unknown Opponents. In *NeurIPS*.
- Strzalecki. 2025. *Ch.1 of Stochastic Choice Theory*. Cambridge Un. Press.
- Wheeler, G. 2018. Bounded rationality. *Stanford Encyclopedia of Philosophy*.
- Xie, Y.; Seshadri, R.; Zhang, Y.; and Akinepally, A. 2024. Real-time personalized tolling for managed lanes. *Tr.Res.P.* C, 163: 104629.

SUPPLEMENTAL MATERIAL:APPENDIX

A Table of notation (2 pages)

Table 1: Notation used throughout the paper and the place where each symbol is first introduced.

Symbol	Meaning	First introduced
<i>Alternatives, profiles, and utilities</i>		
$\mathcal{I} = \{1, \dots, I\}, i$	Set of alternatives, its cardinality, and an alternative index; alternative 1 is the most sustainable one.	Sect. 3, first paragraph
c_i	Externality generated by choosing alternative i .	Sect. 3, first paragraph
$y \geq 0$	Incentive offered for choosing alternative 1.	Sect. 3, first paragraph
$v(i, y)$	Systematic utility from alternative i under incentive y .	Sect. 3, before Eq. (1)
ϵ_i, θ	Random utility term for alternative i and its Gumbel scale parameter.	Sect. 3, before Eq. (1)
$\mathbb{P}_i^v(y)$	Probability that an agent with systematic utility v chooses alternative i under incentive y .	Eq. (1)
$\mathcal{K} = \{0, \dots, K - 1\}, k$	Set of profiles, its cardinality, and a profile index.	Sect. 3, before Lemma 1
$u_k(i, y)$	Systematic utility associated with profile k .	Sect. 3, before Lemma 1
$\mathbf{x}_{i,k}, \boldsymbol{\alpha}_k$	Attribute vector of alternative i for profile k , and the corresponding coefficient vector.	Sect. 3, before Lemma 1
a_k, b_k	Systematic utility of alternative 1 without incentives, and aggregate utility of the remaining alternatives, for profile k .	Lemma 1
$p_k(y), \sigma(\cdot)$	Probability that profile k chooses alternative 1, and the sigmoid function.	Eq. (2)
$a_k - b_k$	Responsibility intensity of profile k .	Immediately before Eq. (3)
$ext^k(y), inc^k(y)$	Expected externality and expected collected incentive when the agent adopts profile k .	Lemma 2
c_{-1}	Expected externality conditional on not choosing alternative 1.	Lemma 2
β	Weight assigned to incentive expenditures relative to externalities.	Eq. (5)
$\tilde{f}^k(y)$	Expected generalized social cost under profile k and incentive y .	Eq. (5)
$\hat{v}(\cdot, \cdot), \hat{v}^t(\cdot, \cdot)$	Regulator's estimate of the utility profile adopted by the agent.	After Eq. (5)
<i>Sequential interaction and optimal incentives</i>		
t, s	Current timeslot and a generic preceding-timeslot index.	Step 1 of Sect. 3.2
$\mathbf{h}^{t-1}$	Realized history of incentives and choices observed before timeslot t .	Step 1 of Sect. 3.2
$\hat{k}^t$	Profile inferred by the regulator at timeslot t .	Step 1 of Sect. 3.2
$\psi(\cdot)$	Deterministic profile-inference rule.	Step 1 of Sect. 3.2
y^t, i^t	Incentive proposed and alternative chosen at timeslot t .	Steps 2–3 of Sect. 3.2
$\phi(\cdot)$	Deterministic mapping from an inferred profile to an incentive.	Step 2 of Sect. 3.2
y_k^*	Optimal incentive associated with inferred profile k .	Prop. 3

Table 1: Notation used throughout the paper (continued).

Symbol	Meaning	First introduced
<i>Manipulation analysis</i>		
$0, r$	Agent’s true profile and the profile imitated by a manipulative agent.	Sect. 4.2
$u_0(y)$	True realized utility under incentive y , expressed according to the selected alternative.	(6)
$\mathbf{H}^{t-1}, \hat{K}^t, Y^t$	Random history, inferred profile, and proposed incentive at timeslot t .	Def. 6
$U_0(y), U_0(Y^t)$	Random true utility under a fixed incentive y , and under the random incentive proposed at timeslot t .	Def. 6
$\mathbb{E}_k[\cdot], \mathbb{P}_k[\cdot]$	Expectation and probability conditional on the agent adopting profile k .	Eqs. (7)–(8)
$Y_k^t, \hat{K}_k^t$	Random incentive and inferred profile conditional on the agent adopting profile k .	After Eqs.(7)-(8)
$CM_r(y), CM_r^t$	Cost of Manipulation under incentive y , and its mean at timeslot t .	Eqs. (9)–(10)
GM_r^t	Gain from Manipulation at timeslot t when imitating profile r .	Eq. (11)
ΔU_r^t	Manipulation advantage: additional expected true utility obtained by imitating profile r .	Eq. (12)
$D_{\text{KL}}(\cdot \parallel \cdot)$	Kullback–Leibler divergence.	Eq. (13)
IE_r^t	Intensity of elicitation: distinguishability between the inferred-profile distributions induced by profiles 0 and r .	Eq. (13)
IP_r^t	Intensity of personalization: distinguishability between the corresponding incentive distributions.	Eq. (14)
R_r	Range of expected true utility over the incentives the regulator may propose.	Theor. 8, Eq. (16)
$\delta(\cdot, \cdot)$	Total-variation distance between probability distributions.	Proof of Theor. 8
IE_r	Maximum one-step KL divergence between the choices induced by profiles 0 and r .	Eq. (22)
$\text{Ber}(p)$	Bernoulli distribution with success probability p .	Theor. 9, Eq. (22)
$\mathbf{H}_k^{t-1}, \mathcal{H}^{s-1}, I^s$	Random history conditional on profile k , set of possible histories, and random alternative at timeslot s .	Proof of Theor. 9
$CM_r^{\min}$	Minimum Cost of Manipulation over all incentives that may be proposed.	Eq. (29)
<i>Preference-elicitation models</i>		
<i>InstPE, RealPE</i>	Instantaneous and Realistic Preference Elicitation.	Sect. 5
$\mathbb{E}^{\text{InstPE}}[\cdot], \mathbb{E}^{\text{RealPE}}[\cdot]$	Expectation under Instantaneous and Realistic Preference Elicitation.	Sect. 5
$Z^t := \mathbf{1}\{I^t = 1\}, z^t$	Binary random variable and its realization indicating whether the sustainable alternative 1 is chosen.	Appendix C
$L_k^t(\mathbf{h}^{t-1})$	Likelihood of observing history $\mathbf{h}^{t-1}$ under profile k .	Eq. (63)
$\ln L_k^t(\mathbf{h}^{t-1})$	Log-likelihood of observing history $\mathbf{h}^{t-1}$ under profile k .	Eq. (64)

B Proofs

Proof of Lemma 1

Lemma. Assuming utility is quasi-linear with respect to income,² utility function $u_k(\cdot, \cdot)$ can be written as

$$u_k(i, y) = \begin{cases} a_k + y & i = 1 \\ b_k & i \neq 1 \end{cases}; \quad a_k := \alpha_k \cdot \mathbf{x}_{1,k}; \quad b_k := \theta \ln \sum_{i>1} e^{\frac{\alpha_k \cdot \mathbf{x}_i}{\theta}} \quad (31)$$

$$\text{Also, } p_k(y) := \mathbb{P}_1^{u_k}(y) = \sigma\left(\frac{a_k - b_k + y}{\theta}\right) \quad (32)$$

where $\sigma(\cdot)$ is the sigmoid

Proof. The quasi-linearity of the utility with respect to income implies that the systematic utility associated with alternative 1 is $u_k(1, y) = \alpha_k \cdot \mathbf{x}_1 + y$, and thus $u_k(1, y) = a_k + y$. If the agent does not choose alternative 1, they will choose the best among the remaining alternatives $i > 1$. As explained in §3, the random utility of the agent associated with alternative i is $u_k(i, y) + \epsilon_i$, where ϵ_i is a Gumbel distribution with scale parameter³ θ . As suggested by (Strzalecki 2025, §1.8), it is reasonable to assume that, when choosing among alternatives $i > 1$, the agent chooses the best among them. The expected utility in this case is

$$b_k := \mathbb{E}[u_k(i, y) | i > 1] = \mathbb{E}\left[\max_{i>1} [u_k(i, y) + \epsilon_i]\right] \quad (33)$$

$$= \theta \ln \sum_{i>1} e^{\frac{v(i, y)}{\theta}} \quad (34)$$

where the last equality is obtained via (Strzalecki 2025, Prop. 3.6). Observe that we do not need to carry the specification of y in symbol b_k , because no incentive is given to any alternative $i > 1$.

By simple calculus $e^{\frac{b_k}{\theta}} = \sum_{i>1} e^{\frac{v(i, y)}{\theta}}$. Therefore, probability (1) can be written as:

$$\mathbb{P}_1^{u_k} = \frac{e^{\frac{a_k + y}{\theta}}}{e^{\frac{a_k + y}{\theta}} + \sum_{i>1} e^{\frac{v(i, y)}{\theta}}} = \frac{1}{1 + e^{\frac{b_k}{\theta} / e^{\frac{a_k + y}{\theta}}}} \quad (35)$$

$$= \frac{1}{1 + e^{-\frac{a_k + y - b_k}{\theta}}} \quad (36)$$

which is the sigmoid calculated in $\frac{a_k + y - b_k}{\theta}$. $\square$

Proof of Lemma 2

Lemma. Let us consider an agent following a certain profile $k \in \mathcal{K}$. Then, the expectation of the externalities caused by them and of the incentive collected are, respectively:

$$\text{ext}^k(y) = p_k(y) \cdot c_1 + (1 - p_k(y)) \cdot c_{-1,k};$$

$$\text{inc}^k(y) = p_k(y) \cdot y$$

$$\text{with } c_{-1,k} := \sum_{i>1} \frac{e^{\frac{u_k(i, y)}{\theta}}}{\sum_{i'>1} e^{\frac{u_k(i', y)}{\theta}}} c_i > c_1 \quad (37)$$

²This is a standard assumption in econometrics ((Javaudin, Araldo, and de Palma 2023)), which means that both $\alpha_k \cdot \mathbf{x}_i$ and the incentive y are expressed in the same unit as income and that y has an additive effect on utility.

³The scale parameter is also called “noise parameter”.

Proof. The result is obtained by applying the law of total expectation:

$$\begin{aligned} \text{ext}^k(y) &= \mathbb{P}_1^{u_k}(y) \mathbb{E}(\text{ext}(y) | \text{agent adopts } k, i = 1) \\ &\quad + (1 - \mathbb{P}_1^{u_k}(y)) \mathbb{E}(\text{ext}(y) | \text{agent adopts } k, i \neq 1) \\ &= p_k(y) \cdot c_1 \\ &\quad + (1 - p_k(y)) \cdot \sum_{i>1} \frac{e^{\frac{u_k(i, y)}{\theta}}}{\sum_{i'>1} e^{\frac{u_k(i', y)}{\theta}}} c_i \end{aligned} \quad (38)$$

The conditional distribution among the alternatives $i > 1$ does not depend on y , because the incentive modifies only the utility of alternative 1.

The expected incentive collected is computed similarly. $\square$

Lemma. The estimated expectation of the generalized social cost $\tilde{f}^k(y) : [0, +\infty[\rightarrow \mathbb{R}$ has a unique minimum. Let $\hat{y}_k$ be the root of

$$g^k(y) := y + \theta + \theta e^{(y + a_k - b_k)/\theta} - \frac{c_{-1,k} - c_1}{\beta} \quad (39)$$

Then,

$$\text{if } \frac{c_{-1,k} - c_1}{\beta} > \theta(1 + e^{(a_k - b_k)/\theta}) \quad (40)$$

$$\text{the min of } \tilde{f}^k : [0, +\infty[\rightarrow \mathbb{R} \text{ is at } \hat{y}_k > 0 \quad (41)$$

$$\text{else such a minimum is at } y = 0 \quad (42)$$

Proof. The estimated expectation of the generalized social cost, which the regulator aims to minimize, is given by (5). By Lemmas 2 and 1 this is

$$\tilde{f}^k(y) = \frac{e^{(a_k + y)/\theta}}{e^{(a_k + y)/\theta} + e^{b_k/\theta}} (c_1 + \beta y) \quad (43)$$

$$+ \frac{e^{b_k/\theta}}{e^{(a_k + y)/\theta} + e^{b_k/\theta}} c_{-1,k}. \quad (44)$$

By taking its derivative, we get

$$\begin{aligned} (\tilde{f}^k(y))' &= \frac{\beta y e^{(y + a_k + b_k)/\theta} + \beta \theta e^{(y + a_k + b_k)/\theta} + \beta \theta e^{2(y + a_k)/\theta}}{\theta(e^{(a_k + y)/\theta} + e^{b_k/\theta})^2} \\ &\quad + \frac{c_1 e^{(y + a_k + b_k)/\theta} - c_{-1,k} e^{(y + a_k + b_k)/\theta}}{\theta(e^{(a_k + y)/\theta} + e^{b_k/\theta})^2} \end{aligned} \quad (45)$$

To find the extrema we take $(\tilde{f}^k(y))' = 0$ and, thus, we obtain

$$\begin{aligned} \beta y e^{(y + a_k + b_k)/\theta} + \beta \theta e^{(y + a_k + b_k)/\theta} + \beta \theta e^{2(y + a_k - b_k)/\theta} \\ - (c_{-1,k} - c_1) e^{(y + a_k - b_k)/\theta} = 0 \end{aligned}$$

Dividing by $\beta e^{(y + a_k + b_k)/\theta}$, we obtain

$$\iff g^k(y) := y + \theta + \theta e^{(y + a_k - b_k)/\theta} - \frac{c_{-1,k} - c_1}{\beta} = 0 \quad (46)$$

So the extremum of $\tilde{f}^k(y) : \mathbb{R} \rightarrow \mathbb{R}$ is at the root of $g^k(y)$, which we will refer to as $\hat{y}_k$.

Note that $g^k(y)$ is an increasing function, which implies that its root $\hat{y}_k$ is unique.

Moreover, observe that $(\tilde{f}^k)'$ is equal to g^k multiplied by a positive quantity, namely $(\beta e^{(y+a_k+b_k)/\theta})/(\theta(e^{(a_k+y)/\theta} + e^{b_k/\theta})^2)$. Therefore, $g^k(y) > 0$ implies $(\tilde{f}^k(y))' > 0$, i.e., $\tilde{f}^k(y)$ increasing, and $g^k(y) < 0$ implies $(\tilde{f}^k(y))' < 0$, i.e., $\tilde{f}^k(y)$ decreasing.

For $y < \hat{y}_k$, function $\tilde{f}^k(y)$ is decreasing, since $y < \hat{y}_k$ implies $g^k(y) < g^k(\hat{y}_k)$, i.e. $g^k(y) < 0$. For $y > \hat{y}_k$, function $\tilde{f}^k(y)$ is increasing, since $y > \hat{y}_k$ implies $g^k(y) > g^k(\hat{y}_k)$, i.e. $g^k(y) > 0$. The root of g^k is, thus, the minimum of $\tilde{f}^k(y)$ on $\mathbb{R} \rightarrow \mathbb{R}$.

Suppose that the root $\hat{y}_k$ of g^k is non-positive. In this case, in the interval $[0, +\infty[$, function $\tilde{f}^k(y)$ is always increasing. Its minimum is thus at 0. Otherwise, the minimum of $\tilde{f}^k(y) : [0, +\infty[\rightarrow \mathbb{R}$ is the root $\hat{y}_k$ of g^k .

Condition (40) determines the sign of the root $\hat{y}_k$. Indeed

$$\text{If } \frac{c_{-1,k} - c_1}{\beta} > \theta(1 + e^{(a_k - b_k)/\theta}), \quad (47)$$

$$\text{then } 0 > \theta(1 + e^{(a_k - b_k)/\theta}) - \frac{c_{-1,k} - c_1}{\beta} \quad (48)$$

$$\Rightarrow g^k(\hat{y}_k) > g^k(0) \quad (49)$$

To obtain the last implication, we simply use the definition of $g^k(y)$ (46) (replacing $y = 0$) and the fact that it is zero at its root, i.e., $g^k(\hat{y}_k) = 0$.

Since $g^k(y)$ is increasing, equation (49) implies that $\hat{y}_k \geq 0$, and so the minimum of $\tilde{f}^k : [0, +\infty[\rightarrow \mathbb{R}$ corresponds to $\hat{y}_k$. On the contrary, if condition (40) does not hold, the root $\hat{y}_k < 0$ and the minimum of $\tilde{f}^k : [0, +\infty[\rightarrow \mathbb{R}$ is 0. $\square$

Proofs of Proposition 3 and Corollary 4

Proposition. 3. *There is a unique optimal incentive for profile $k \in \mathcal{K}$:*

$$y_k^* = \max \left(\frac{c_{-1,k} - c_1}{\beta} - \theta \left(1 + W \left(e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta}} - 1 \right) \right), 0 \right) \quad (50)$$

where $W(\cdot)$ is the Lambert's W-function.

Proof. The estimated expectation of the generalized social cost, that the regulator aims to minimize, is given by (5). By (2) this is

$$\begin{aligned} \tilde{f}^k(y) &= \frac{e^{(a_k+y)/\theta}}{e^{(a_k+y)/\theta} + e^{b_k/\theta}} (c_1 + \beta y) \\ &\quad + \frac{e^{b_k/\theta}}{e^{(a_k+y)/\theta} + e^{b_k/\theta}} c_{-1,k} \end{aligned}$$

Let $y_k^* = 0$ be point at which $\tilde{f}^k(y) : [0, +\infty[\rightarrow \mathbb{R}$ has its minimum (which is unique - see (39)).

Let us start by assuming that (40) holds. In this case, y_k^* is

the the root of $g^k(y)$ and is positive (see (39)). Therefore

$$\begin{aligned} y_k^* + \theta + \theta e^{(a_k - b_k + y_k^*)/\theta} - \frac{c_{-1,k} - c_1}{\beta} &= 0 \\ \Rightarrow \frac{y_k^*}{\theta} + 1 + e^{(a_k - b_k + y_k^*)/\theta} - \frac{c_{-1,k} - c_1}{\theta\beta} &= 0 \\ \Rightarrow e^{(a_k - b_k + y_k^*)/\theta} &= \frac{c_{-1,k} - c_1}{\theta\beta} - \frac{y_k^*}{\theta} - 1 \end{aligned}$$

We substitute $u = \frac{c_{-1,k} - c_1}{\theta\beta} - \frac{y_k^*}{\theta} - 1$ and, by inserting u and $\frac{y_k^*}{\theta} = \frac{c_{-1,k} - c_1}{\theta\beta} - 1 - u$ into the last equation, we get

$$\begin{aligned} e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1 - u} &= u \\ \Rightarrow e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1} &= u e^u \end{aligned}$$

Now we can apply the Lambert W-function on both sides

$$W(e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1}) = W(u e^u) = u$$

By reinserting $u = \frac{c_{-1,k} - c_1}{\theta\beta} - \frac{y_k^*}{\theta} - 1$, we obtain

$$W(e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1}) = \frac{c_{-1,k} - c_1}{\theta\beta} - \frac{y_k^*}{\theta} - 1$$

$$\Rightarrow y_k^* = \frac{c_{-1,k} - c_1}{\beta} - \theta - \theta \cdot W(e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1})$$

Summarizing, we obtain that, if (40) holds, then

$$y_k^* = \frac{c_{-1,k} - c_1}{\beta} - \theta \cdot \left(1 + W(e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1}) \right) > 0 \quad (51)$$

If, instead, (40) does not hold, then $y_k^* = 0$ (see (42)). Therefore,

$$y_k^* = \max \left(\frac{c_{-1,k} - c_1}{\beta} - \theta \cdot \left(1 + W(e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1}) \right), 0 \right) \quad (52)$$

Corollary. 4. *Let y_k^* be the optimal incentives as defined in Proposition 3. Then y_k^* is non-increasing in $a_k - b_k$.*

Proof. By Proposition 3,

$$y_k^* = \max \left(\frac{c_{-1,k} - c_1}{\beta} - \theta \left(1 + W(e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1}) \right), 0 \right)$$

First, note that $e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1}$ is increasing with $a_k - b_k$, because of the term $\frac{a_k - b_k}{\theta}$ inside the exponential. Following this,

$$W(e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1})$$

is increasing with $a_k - b_k$, since if the argument of the Lambert W function is positive and increasing, the Lambert W function is increasing as well. It follows that

$$-\theta \left(1 + W(e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1}) \right)$$

is decreasing with $a_k - b_k$ and so is the whole expression

$$\frac{c_{-1,k} - c_1}{\beta} - \theta \left(1 + W(e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1}) \right)$$

Due to the fact that $y_k^* = 0$ if $a_k - b_k$ is sufficiently large (see (40)), it follows that y_k^* is non-increasing with $a_k - b_k$, which concludes the proof. $\square$

Proof of Proposition 5

Proposition. 5. *The optimal incentive is upper-bounded, i.e.,*

$$0 \leq y_k^* < \frac{c_{-1,k} - c_1}{\beta}.$$

Proof. By Proposition 3, it suffices to show that

$$\frac{c_{-1,k} - c_1}{\beta} - \theta(1 + W(e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1})) < \frac{c_{-1,k} - c_1}{\beta} \quad (53)$$

i.e.,

$$-\theta(1 + W(e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1})) < 0 \quad (54)$$

$$\iff 1 + W(e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1}) > 0 \quad (55)$$

$$\iff W(e^{\frac{a_k - b_k}{\theta} + \frac{c_{-1,k} - c_1}{\theta\beta} - 1}) > -1 \quad (56)$$

This is true, since if the argument of the Lambert W function is positive, the Lambert W function is positive as well. $\square$

Proof of Lemma 12

Lemma.

$$GM_r^{t,InstPE} = \mathbb{E}_0[U_0(y_r^*)] - \mathbb{E}_0[U_0(y_0^*)], \quad CM_r^t = 0, \quad \forall t$$

Proof. The following equalities follow from a straightforward application of Obs. 11:

$$GM_r^{t,InstPE} = \sum_{k \in \mathcal{K}} \mathbb{P}_r^{InstPE}[Y_t = y_k^*] \mathbb{E}_r[U_0(y_k^*)] \quad (57)$$

$$- \sum_{k \in \mathcal{K}} \mathbb{P}_0^{InstPE}[Y_t = y_k^*] \mathbb{E}_r[U_0(y_k^*)] \quad (58)$$

$$= \mathbb{E}_r[U_0(y_r^*)] - \mathbb{E}_r[U_0(y_0^*)] \quad (59)$$

$$= \mathbb{E}_0[U_0(y_r^*)] - \mathbb{E}_0[U_0(y_0^*)] \quad (60)$$

and

$$CM_r^{InstPE}(y) = \mathbb{E}_0[U_0(y)] - \mathbb{E}_0[U_0(y)] = 0 \implies CM_r^t = 0. \quad \square$$

C Preference Elicitation via a Maximum-Likelihood Estimator

We now specify the deterministic profile-inference rule $\psi(\cdot)$ used in the numerical experiments. At every timeslot s , the agent chooses among all alternatives $i^s \in \mathcal{I}$ according to the multinomial-logit model in Eq. (1). For online preference elicitation, however, the regulator uses only whether the incentivized alternative 1 was selected. Define the binary random variable and its realization as

$$Z^s := \mathbf{1}\{I^s = 1\}, \quad z^s := \mathbf{1}\{i^s = 1\}. \quad (61)$$

Thus, $z^s = 1$ when the agent chooses alternative 1, whereas $z^s = 0$ aggregates all choices $i^s > 1$. This binary observation model is plausible in applications where the regulator need not collect the agent's complete choice. In particular, the data-minimization principle of the General Data Protection Regulation (GDPR) supports limiting the collected information to what is necessary for the specified purpose. Accordingly, an implementation may reveal only whether

the incentivized alternative was selected—the information needed to determine eligibility for the incentive—rather than which non-incentivized alternative was chosen. Let

$$\mathbf{h}^{t-1} = (y^1, i^1, \dots, y^{t-1}, i^{t-1})$$

be the realized history before timeslot t . Although the complete alternatives i^s appear in the history, the inference rule depends on them only through the binary observations z^s . Conditional on candidate profile k and incentive y^s , the binary observation follows

$$Z^s \mid \mathbf{H}^{s-1} = \mathbf{h}^{s-1} \sim \text{Ber}(p_k(y^s)), \quad (62)$$

where $p_k(y^s)$ is the probability of choosing alternative 1 defined in Lemma 1.

The likelihood of the binary observations induced by $\mathbf{h}^{t-1}$ under candidate profile k is therefore

$$\begin{aligned} L_k^t(\mathbf{h}^{t-1}) &:= \prod_{s=1}^{t-1} [p_k(y^s)]^{z^s} [1 - p_k(y^s)]^{1-z^s} \\ &= L_k^{t-1}(\mathbf{h}^{t-2}) [p_k(y^{t-1})]^{z^{t-1}} [1 - p_k(y^{t-1})]^{1-z^{t-1}}, \end{aligned} \quad (63)$$

with $L_k^1 = 1$. The proposed incentives do not introduce an additional factor in the likelihood because each y^s is a deterministic function of the preceding observed history.

To avoid numerical underflow, the implementation maintains the log-likelihood

$$\begin{aligned} \ln L_k^t(\mathbf{h}^{t-1}) &= \sum_{s=1}^{t-1} [z^s \ln p_k(y^s) + (1 - z^s) \ln(1 - p_k(y^s))] \\ &= \ln L_k^{t-1}(\mathbf{h}^{t-2}) + z^{t-1} \ln p_k(y^{t-1}) \\ &\quad + (1 - z^{t-1}) \ln(1 - p_k(y^{t-1})), \end{aligned} \quad (64)$$

with $\ln L_k^1 = 0$.

Let the set of maximum-likelihood profiles be

$$\mathcal{M}^t(\mathbf{h}^{t-1}) := \arg \max_{k \in \mathcal{K}} \ln L_k^t(\mathbf{h}^{t-1}). \quad (65)$$

The regulator selects, among the profiles in $\mathcal{M}^t(\mathbf{h}^{t-1})$, the profile associated with the smallest optimal incentive. If several such profiles have the same optimal incentive, the smallest profile index is selected. Consequently, the deterministic inference rule used in the numerical experiments is

$$\hat{k}^t = \psi(\mathbf{h}^{t-1}) := \min \left\{ k \in \mathcal{M}^t(\mathbf{h}^{t-1}) : y_k^* = \min_{j \in \mathcal{M}^t(\mathbf{h}^{t-1})} y_j^* \right\}. \quad (66)$$

This online Bernoulli likelihood must not be confused with the multinomial likelihood used for the offline calibration in Sect. D1. Accordingly, the Bernoulli distributions in Eq. (26) concern the binary variable Z^s , rather than the complete categorical alternative I^s .

D Appendix to the Case Study

D1 Dataset and Calibration Procedure

D1.1 Data The London Passenger Mode Choice (LPMC) dataset combines trip diaries from the London

Table 2: Quantities computed directly from the calibration and simulation sets. A dash indicates that the quantity is not used for that set: $\bar{d}_S$ and $\omega_{\text{bus},S}$ define the representative simulation context and are therefore computed from $\mathcal{D}_{\text{sim}}$.

Quantity	Symbol	$\mathcal{D}_{\text{cal}}$	$\mathcal{D}_{\text{sim}}$
Trips	$ S $	34,198	14,656
Travelers	$ \mathcal{N}_S $	20,577	11,709
Mean trip distance (km/trip)	$\bar{d}_S$	—	5.41
Bus-time share	$\omega_{\text{bus},S}$	—	0.65

Travel Demand Survey with the routes, travel times, public-transport fares, driving fuel costs, and congestion charges associated with walking, bike, public transport (PT), and car.⁴ The public release contains 81,086 trips collected between April 2012 and March 2015.⁵

We retain travelers aged at least 12, since they are assumed to be responsive to monetary incentives, and trips between home and work, between home and educational activities, or between home and another activity. These are the home-based work, home-based education, and home-based other purposes in the LPMC documentation⁵. Only trips made by bike, PT, or car are retained. For each trip, the following information concerning the trip or the traveler is reported: the travel time associated with each mode; out-of-pocket costs for PT and car; the number of PT interchanges; possession of a driving licence; car ownership, measured as the number of cars in the household divided by the number of people in the household; trip distance; and the bus and rail portions of PT in-vehicle time. Total PT time is the sum of access, bus, rail, and interchange time. Trips missing a quantity required by the utility specification are discarded.

Let $\mathcal{D}_{\text{cal}}$ denote the calibration set and $\mathcal{D}_{\text{sim}}$ the simulation set. Among the retained bike trips, 70% are assigned to $\mathcal{D}_{\text{cal}}$ and 30% to $\mathcal{D}_{\text{sim}}$; the same split is performed independently among PT trips and among car trips. Hence, the proportions of bike, PT, and car trips are preserved in both sets. The resulting set $\mathcal{D}_{\text{cal}}$ contains 34,198 trips made by 20,577 travelers, whereas $\mathcal{D}_{\text{sim}}$ contains 14,656 trips made by 11,709 travelers.

For a trip set $S \in \{\mathcal{D}_{\text{cal}}, \mathcal{D}_{\text{sim}}\}$, let $\mathcal{N}_S$ denote its travelers, $\bar{d}_S$ its mean trip distance, and $\omega_{\text{bus},S}$ the share of bus time in total bus-plus-rail in-vehicle time. Table 2 separates quantities computed directly from the trip records from quantities estimated by the calibration procedure.

D1.2 Profile-based choice model Let $\mathcal{K} = \{0, 1, 2\}$ be the set of possible true profiles. After calibration, the profiles are labeled so that $k = 0$ is bike-oriented, $k = 1$ is PT-

⁴Tim Hillel, Mohammed Z. E. B. Elshafie, and Ying Jin. 2018. “Recreating Passenger Mode Choice-Sets for Transport Simulation: A Case Study of London, UK.” *Proceedings of the Institution of Civil Engineers—Smart Infrastructure and Construction* 171(1):29–42. See Sec. 2.1, pp. 30–33. <https://doi.org/10.1680/jsmic.17.00018>.

⁵Tim Hillel. 2019. *London Passenger Mode Choice*. Transport and Mobility Laboratory, EPFL. See Table 1, p. 3. https://transport.epfl.ch/documents/technicalReports/CS_LPMC.pdf.

Symbol	Definition
τ	Index of a recorded trip.
$i \in \{1, 2, 3\}$	Travel alternative: $i = 1$ denotes bike, $i = 2$ public transport (PT), and $i = 3$ car.
t_τ^i	Travel time of alternative i for trip τ , in hours.
C_τ^{PT}	Out-of-pocket PT cost for trip τ .
C_τ^{car}	Out-of-pocket car cost for trip τ .
m_τ	Number of PT interchanges for trip τ .
χ_τ^{lic}	Indicator equal to 1 if the traveler making trip τ holds a driving licence, and 0 otherwise.
χ_τ^{own}	Car ownership of the traveler making trip τ , measured as the number of cars in the household divided by the number of people in the household.
ν	Common value of travel time, in pounds per hour.
ι_k	Profile- k utility penalty for each additional PT interchange.
ζ_k^{lic}	Profile- k increment in car utility associated with holding a driving licence.
ζ_k^{own}	Profile- k increment in car utility per unit of χ_τ^{own} .
ξ_k^{bike}	Intrinsic profile- k preference for bike not represented by the other trip or traveler characteristics.
ξ_k^{PT}	Intrinsic profile- k preference for PT not represented by the other trip or traveler characteristics.

Table 3: Notation used in the trip-specific systematic utilities.

oriented, and $k = 2$ is car-oriented. For every $k \in \mathcal{K}$, the systematic utility $u_k(i, y)$ introduced in the main paper is instantiated for trip τ as

$$\begin{aligned}
 u_{k,\tau}(1, y) &= \xi_k^{\text{bike}} - \nu t_\tau^{\text{bike}} + y, \\
 u_{k,\tau}(2, y) &= \xi_k^{\text{PT}} - \nu t_\tau^{\text{PT}} - C_\tau^{\text{PT}} - \iota_k m_\tau, \\
 u_{k,\tau}(3, y) &= -\nu t_\tau^{\text{car}} - C_\tau^{\text{car}} + \zeta_k^{\text{lic}} \chi_\tau^{\text{lic}} + \zeta_k^{\text{own}} \chi_\tau^{\text{own}}.
 \end{aligned} \tag{67}$$

The notation used in the trip-specific utility functions is summarized in Table 3.

For any alternative i , trip τ , profile k , and incentive y , let $\mathbb{P}_i^{u_{k,\tau}}(y)$ denote the probability in Eq. (1) evaluated with the trip-specific systematic utility function $u_{k,\tau}$ in Eq. (67). In particular, $\mathbb{P}_{i_\tau}^{u_{k,\tau}}(0)$ is the probability that profile k chooses the mode i_τ recorded for trip τ when no bike incentive is offered. The argument 0 is used during calibration because the trips in the dataset were recorded without the simulated incentive policy.

D1.3 Calibration procedure The offline calibration described in this subsection is distinct from the regulator’s online learning process in Appendix C. The offline calibration is used only to obtain behavioral parameters for the numerical experiment. Its log-likelihood ℓ_{nk}^{cal} , defined below from the LPMC trips, must therefore not be confused with the regulator’s likelihood $L_k^t(\mathbf{h}^{t-1})$ in Eqs. (63)–(64).

Let $\mathcal{N}_{\text{cal}}$ be the travelers represented in $\mathcal{D}_{\text{cal}}$, let $\mathcal{T}_n^{\text{cal}}$ be the set of calibration trips made by traveler $n \in \mathcal{N}_{\text{cal}}$, and

let i_τ be the mode chosen on trip τ . For parameter vector Θ , the offline calibration log-likelihood of traveler n under candidate true profile k is

$$\ell_{nk}^{\text{cal}}(\Theta) = \sum_{\tau \in \mathcal{T}_n^{\text{cal}}} \log \mathbb{P}_{i_\tau}^{u_{k,\tau}}(0). \quad (68)$$

The trip and traveler quantities in Eq. (67) are fixed inputs. The parameters to be calibrated are

$$\Theta = \left(\nu, \theta, \{ \xi_k^{\text{bike}}, \xi_k^{\text{PT}}, \iota_k, \zeta_k^{\text{lic}}, \zeta_k^{\text{own}} \}_{k \in \mathcal{K}} \right), \quad (69)$$

where θ is the Gumbel scale in Eq. (1).

The initial profile-membership probabilities are based on each traveler's mode shares. Let μ_{nk} be the fraction of trips in $\mathcal{T}_n^{\text{cal}}$ made using the mode associated with profile k : bike for $k = 0$, PT for $k = 1$, and car for $k = 2$. Before enforcing equal total membership across profiles, the initial profile-membership probabilities are

$$\tilde{q}_{nk}^{(0)} = \frac{\exp(\eta_{\text{init}} \mu_{nk})}{\sum_{j \in \mathcal{K}} \exp(\eta_{\text{init}} \mu_{nj})}, \quad \eta_{\text{init}} = 2. \quad (70)$$

The mode-share scaling parameter η_{init} Controls how strongly observed mode shares influence the initial profile probabilities. A larger value gives more probability to profiles associated with frequently used modes; 0 would give equal probability to every profile.. We heuristically set $\eta_{\text{init}} = 2$.

At calibration step $d \geq 1$, the parameters $\Theta^{(d)}$ determine the pre-adjustment profile-membership probabilities

$$\tilde{q}_{nk}^{(d)} = \frac{\exp(\ell_{nk}^{\text{cal}}(\Theta^{(d)}))}{\sum_{j \in \mathcal{K}} \exp(\ell_{nj}^{\text{cal}}(\Theta^{(d)}))}. \quad (71)$$

The superscript (0) in Eq. (70) denotes initialization from empirical mode shares, whereas the superscript (d) in Eq. (71) denotes calibration step d . The tilde identifies probabilities before imposing the profile-size condition in Eq. (73).

For every $d \in \{0, \dots, N_{\text{step}}\}$, adjustment parameters $\lambda_k^{(d)} \in \mathbb{R}$ are selected so that

$$q_{nk}^{(d)} = \frac{\tilde{q}_{nk}^{(d)} \exp(\lambda_k^{(d)})}{\sum_{j \in \mathcal{K}} \tilde{q}_{nj}^{(d)} \exp(\lambda_j^{(d)})}, \quad (72)$$

$$\sum_{n \in \mathcal{N}_{\text{cal}}} q_{nk}^{(d)} = \frac{|\mathcal{N}_{\text{cal}}|}{|\mathcal{K}|}, \quad k \in \mathcal{K}. \quad (73)$$

The quantity $q_{nk}^{(d)}$ is the membership probability assigned to traveler n and profile k at calibration step d . Therefore, $\sum_{n \in \mathcal{N}_{\text{cal}}} q_{nk}^{(d)}$ is the fraction of agents associated to profile k within the calibration set, and Eq. (73) sets this quantity to one third of $|\mathcal{N}_{\text{cal}}|$ for every $k \in \mathcal{K}$. We do so to prevent calibration to favor one profile over the others.⁶ Observe that the adjustment parameters $\lambda_k^{(d)}$ shift the log-membership scores in Eq. (72). Adding the same constant to every $\lambda_k^{(d)}$

⁶We also run calibration imposing that the fraction of trips (and not of agents) be 1/3 per each profile. Final result did not change significantly and are thus omitted.

leaves the probabilities unchanged; hence, we set $\lambda_0^{(d)} = 0$ and solve the two equations in Eq. (73) corresponding to $k = 1$ and $k = 2$ for $\lambda_1^{(d)}$ and $\lambda_2^{(d)}$.

For fixed profile-membership probabilities $q_{nk}^{(d-1)}$, we find the parameters Θ that minimize

$$\mathcal{J}(\Theta; q^{(d-1)}) = \mathcal{J}_{\text{NLL}}(\Theta; q^{(d-1)}) + \rho \mathcal{R}(\Theta), \quad (74)$$

$$\mathcal{J}_{\text{NLL}}(\Theta; q^{(d-1)}) = -\frac{1}{|\mathcal{D}_{\text{cal}}|} \sum_{n \in \mathcal{N}_{\text{cal}}} \sum_{\tau \in \mathcal{T}_n^{\text{cal}}} \sum_{k \in \mathcal{K}} q_{nk}^{(d-1)} \log \mathbb{P}_{i_\tau}^{u_{k,\tau}}(0). \quad (75)$$

Here, $|\mathcal{D}_{\text{cal}}|$ is the number of calibration trips and $\mathcal{J}_{\text{NLL}}$ in Eq. (75) is the average negative log-likelihood of the modes chosen in each recorded trip, weighted by the profile-membership probabilities $q_{nk}^{(d-1)}$. The second term in Eq. (74) is the regularization term, with

$$\begin{aligned} \mathcal{R}(\Theta) = \sum_{k \in \mathcal{K}} \big[& (\xi_k^{\text{bike}} - \bar{\xi}^{\text{bike}})^2 + (\xi_k^{\text{PT}} - \bar{\xi}^{\text{PT}})^2 \\ & + (\iota_k - \bar{\iota})^2 + (\zeta_k^{\text{lic}} - \bar{\zeta}^{\text{lic}})^2 \\ & + (\zeta_k^{\text{own}} - \bar{\zeta}^{\text{own}})^2 \big], \end{aligned} \quad (76)$$

where each barred quantity is the mean of the corresponding coefficient over $k \in \mathcal{K}$. Because both the traveler memberships q_{nk} and the profile-specific coefficients in Eq. (69) are estimated, markedly different coefficient vectors can produce similar values of the weighted likelihood in Eq. (75). The regularization term in Eq. (76) favors the less extreme of such parameterizations by penalizing cross-profile dispersion. The coefficient $\rho = 0.01$ gives this term a small weight relative to the average negative log-likelihood. The common value of time ν and common Gumbel scale θ are calibrated but not regularized because each has a single value shared by all profiles; neither contributes to cross-profile dispersion.

The calibration has $N_{\text{step}} = 4$ steps. At calibration step d , we first initialize Θ at $\Theta^{(d-1)}$; for $d = 1$, the initial values $\Theta^{(0)}$ in Table 5 are used. We then find $\Theta^{(d)}$ that minimizes Eq. (74), holding $q_{nk}^{(d-1)}$ fixed and enforcing the bounds in Table 5. To do so, we use the analytic gradient of loss (74) with respect to parameters Θ and the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm for bound-constrained optimization (L-BFGS-B).⁷ The resulting $\Theta^{(d)}$ is substituted in Eqs. (68) and (71); the adjustment in Eqs. (72)–(73) then gives $q_{nk}^{(d)}$.

Profile 0 is required to satisfy

$$a_0 - b_0 \geq \pounds 0.15. \quad (77)$$

This condition represents a profile for which the zero-incentive bike utility is slightly larger than the inclusive utility of the non-bike alternatives in the representative travel context. The value $\pounds 0.15$ is a small separation margin rather than

⁷Richard H. Byrd, Peihuang Lu, Jorge Nocedal, and Ciyu Zhu. 1995. “A Limited Memory Algorithm for Bound Constrained Optimization.” *SIAM Journal on Scientific Computing* 16(5):1190–1208. See Sec. 2, pp. 1193–1198. <https://doi.org/10.1137/0916069>.

Table 4: Settings of the offline calibration procedure.

Quantity	Symbol	Value
Number of profiles	$ \mathcal{K} $	3
Mode-share scaling in Eq. (70)	η_{init}	2
Calibration steps	N_{step}	4
Maximum L-BFGS-B iterations per step	N_{opt}	80
Function tolerance	ε_{fun}	10^{-10}
Gradient tolerance	$\varepsilon_{\text{grad}}$	10^{-5}
Maximum line-search steps	N_{line}	30
Regularization coefficient	ρ	0.01
Profile-0 margin in Eq. (77)	$a_0 - b_0$	at least £0.15

 Table 5: Initialization, admissible ranges, and calibrated values. Profile-indexed values are ordered as $k = 0, 1, 2$. The calibrated profile-specific coefficients enter the simulations through the constructed utilities a_k and b_k reported in Table 6.

Parameter	$\Theta^{(0)}$	Range	Calibrated
ξ_k^{bike} : intrinsic bike preference (£)	(5, 1, -1)	[-15, 15]	through a_k
ξ_k^{PT} : intrinsic PT preference (£)	(0, 2, -1)	[-15, 15]	through b_k
ν : value of time (£/h)	10	[3, 30]	10.48
ι_k : PT-interchange penalty (£/interchange)	(1, 1, 1)	[0, 8]	through b_k
ζ_k^{lic} : driving-licence utility increment (£)	(2, 2, 4)	[0, 10]	through b_k
ζ_k^{own} : car-ownership utility increment (£/unit)	(2, 2, 4)	[0, 10]	through b_k
θ : Gumbel scale (£)	3.50	[1, 12]	3.68
a_k : constructed bike utility (£)	-	-	(6.12, -6.78, -6.59)
b_k : constructed non-bike utility (£)	-	-	(5.76, 4.94, 5.35)
$a_k - b_k$: constructed utility difference (£)	-	Eq. (77)	(0.36, -11.73, -11.93)

an empirical estimate. At the calibrated scale $\hat{\theta} = £3.68$, the boundary in Eq. (77) implies, by Eq. (2),

$$p_0(0) \geq \sigma\left(\frac{0.15}{3.68}\right) = 51.02\%. \quad (78)$$

Thus, the condition requires only a 1.02-percentage-point preference above indifference between bike and the aggregate non-bike option. It identifies a strictly bike-oriented profile without imposing a large behavioral separation. The calibrated value is $\hat{a}_0 - \hat{b}_0 = £0.36$, which is above the boundary in Eq. (77). The quantities a_0 and b_0 are defined as in Eq. (2), using the profile-weighted trip attributes at the current calibration step. After estimating $\Theta^{(d)}$ and again after obtaining $q_{nk}^{(d)}$, the bike constant is updated according to

$$\xi_0^{\text{bike}} \leftarrow \xi_0^{\text{bike}} + \max\{0, £0.15 - (a_0 - b_0)\}, \quad (79)$$

which restores Eq. (77) whenever that condition is violated.

The mixture negative log-likelihood used to evaluate de-

scriptive fit is

$$\mathcal{J}_{\text{mix}}(\Theta) = -\frac{1}{|\mathcal{D}_{\text{cal}}|} \sum_{n \in \mathcal{N}_{\text{cal}}} \log \left[\sum_{k \in \mathcal{K}} \frac{1}{|\mathcal{K}|} \exp(\ell_{nk}^{\text{cal}}(\Theta)) \right]. \quad (80)$$

A nat is the unit of log loss obtained when the natural logarithm is used. The fitted value of $\mathcal{J}_{\text{mix}}$ in Eq. (80) is 0.67 nat/trip. Uniform choice gives 1.10 nat/trip, and assigning each mode its aggregate frequency in $\mathcal{D}_{\text{cal}}$ gives 0.83 nat/trip. For a model M with average negative log-likelihood $\mathcal{J}_M$, define the geometric mean probability assigned to the modes chosen in the recorded trips as

$$\bar{p}_{\text{geom}}^M = \exp(-\mathcal{J}_M). \quad (81)$$

Using the unrounded losses gives

$$(\bar{p}_{\text{geom}}^{\text{fit}}, \bar{p}_{\text{geom}}^{\text{shares}}, \bar{p}_{\text{geom}}^{\text{uniform}}) = (51.12\%, 43.74\%, 33.33\%). \quad (82)$$

The fitted profile model therefore assigns 7.38 percentage points more geometric-mean probability to the realized choices than the aggregate-share baseline and 17.79 percentage points more than uniform choice. Hence, the profile-specific utilities capture systematic variation in mode choice beyond the aggregate frequencies of the three modes.

D1.4 Calibrated parameter values, profile construction and context parameters Let $\mathcal{N}_{\text{sim}}$ be the travelers represented in $\mathcal{D}_{\text{sim}}$, and let $\mathcal{T}_n^{\text{sim}}$ be the trips made by traveler $n \in \mathcal{N}_{\text{sim}}$ in that set. The calibrated parameters $\hat{\Theta}$ give

$$\ell_{nk}^{\text{sim}}(\hat{\Theta}) = \sum_{\tau \in \mathcal{T}_n^{\text{sim}}} \log \mathbb{P}_{i_\tau}^{u_k, \tau}(0). \quad (83)$$

The simulation-set membership probabilities q_{nk}^{sim} are obtained from Eq. (71) by replacing $\ell_{nk}^{\text{cal}}(\Theta^{(d)})$ with $\ell_{nk}^{\text{sim}}(\hat{\Theta})$. The profile-size adjustment in Eqs. (72)–(73) is not applied to $\mathcal{D}_{\text{sim}}$. The share of profile k in the simulation set $\mathcal{D}_{\text{sim}}$ is

$$\pi_k = \frac{1}{|\mathcal{N}_{\text{sim}}|} \sum_{n \in \mathcal{N}_{\text{sim}}} q_{nk}^{\text{sim}}. \quad (84)$$

For each trip $\tau \in \mathcal{D}_{\text{sim}}$ and alternative i , define

$$\mathbf{x}_{\tau i} = (\mathbb{I}_{\{i=1\}}, \mathbb{I}_{\{i=2\}}, t_\tau^i, C_\tau^i, m_\tau \mathbb{I}_{\{i=2\}}, \chi_\tau^{\text{lic}} \mathbb{I}_{\{i=3\}}, \chi_\tau^{\text{own}} \mathbb{I}_{\{i=3\}})^\top, \quad (85)$$

where $C_\tau^1 = 0$, $C_\tau^2 = C_\tau^{\text{PT}}$, and $C_\tau^3 = C_\tau^{\text{car}}$. Thus, $\mathbf{x}_{\tau i}$ contains the alternative indicators, alternative-specific travel time, out-of-pocket cost, number of PT interchanges, driving-licence indicator, and household car ownership entering Eq. (67). Let $n(\tau)$ denote the traveler who made trip τ . For every profile-alternative pair (k, i) , the representative attribute vector is

$$\mathbf{x}_{i,k} = \frac{\sum_{\tau \in \mathcal{D}_{\text{sim}}} q_{n(\tau)k}^{\text{sim}} \mathbf{x}_{\tau i}}{\sum_{\tau \in \mathcal{D}_{\text{sim}}} q_{n(\tau)k}^{\text{sim}}}. \quad (86)$$

The weight $q_{n(\tau)k}^{\text{sim}}$ gives greater influence to trips made by travelers whose profile-membership probability $q_{n(\tau)k}^{\text{sim}}$ is

Table 6: Calibrated parameter values and profile-level context quantities.

	Common quantity		Value	Unit
	Value of time $\hat{\nu}$		10.48	£/h
	Gumbel scale $\hat{\theta}$		3.68	£

k	π_k (%)	$\mathbb{P}_1^{u_k}(0)$ (%)	$\mathbb{P}_2^{u_k}(0)$ (%)	$\mathbb{P}_3^{u_k}(0)$ (%)	a_k (£)	b_k (£)
0	21.26	52.44	19.12	28.44	6.12	5.76
1	39.36	3.96	52.08	43.96	-6.78	4.94
2	39.38	3.75	43.10	53.15	-6.59	5.35

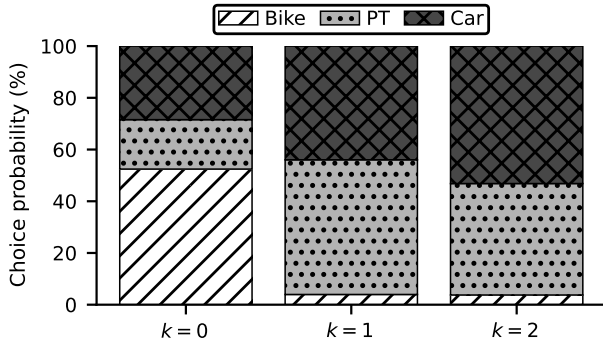


Figure 3: Zero-incentive mode probabilities $\mathbb{P}_i^{u_k}(0)$ for the three calibrated profiles $k \in \mathcal{K}$. Alternatives $i = 1, 2, 3$ denote bike, PT, and car, respectively.

higher. This ensures that the representative attribute vector $\mathbf{x}_{i,k}$ is more strongly affected by profile memberships inferred with higher confidence. Substituting $\mathbf{x}_{i,k}$ and $\hat{\Theta}$ into Eq. (67) determines the profile-level systematic utilities $u_k(i, 0)$. Profiles are relabeled so that $a_0 - b_0 \geq a_1 - b_1 \geq a_2 - b_2$, consistently with the ordering in the main paper.

Table 6 reports the common calibrated parameters $\hat{\nu}$ and $\hat{\theta}$, the profile shares π_k , and the profile-level quantities obtained from Eqs. (84)–(86). The zero-incentive mode probabilities $\mathbb{P}_i^{u_k}(0)$, $i \in \{1, 2, 3\}$, are evaluated using Eq. (1) with the utilities $u_k(i, 0)$.

For clarity, before rounding, the representative profile-level quantities used throughout the case study are

$$\{a_k\}_{k=0}^2 = (6.1178, -6.7834, -6.5886),$$

$$\{b_k\}_{k=0}^2 = (5.7590, 4.9421, 5.3451),$$

and therefore

$$\{a_k - b_k\}_{k=0}^2 = (0.3588, -11.7255, -11.9337).$$

D1.5 Externalities and calibration of the policy weight

β Table 7 reports the externally specified quantities. Greenhouse-gas (GHG) factors are expressed as carbon-dioxide equivalent (CO_2e), because the official conversion factors aggregate the climate effects of the covered green-

house gases into an equivalent mass of CO_2 .⁸ Bike is assigned zero direct operational emissions. The monetary carbon value $V_{\text{CO}_2\text{e}}$ follows UK government appraisal guidance.⁹ The road externalities C_{cong} and C_{other} are taken from the Institute for Fiscal Studies assessment of motoring external costs.¹⁰ The cycling payment rate R_{Cic} is taken from a field experiment.¹¹ The exchange rate $X_{\text{NOK/GBP}}$ is the 2021 annual average reported by DNB Markets.¹²

Using the direct simulation-set quantities $\bar{d}_{\mathcal{D}_{\text{sim}}}$ and $\omega_{\text{bus}, \mathcal{D}_{\text{sim}}}$ in Table 2, the operational emissions of a representative PT trip and a representative car trip are

$$E_{\text{PT}}^{\text{trip}} = \bar{d}_{\mathcal{D}_{\text{sim}}} [\omega_{\text{bus}, \mathcal{D}_{\text{sim}}} E_{\text{bus}} + (1 - \omega_{\text{bus}, \mathcal{D}_{\text{sim}}}) E_{\text{rail}}], \quad (87)$$

$$E_{\text{car}}^{\text{trip}} = \bar{d}_{\mathcal{D}_{\text{sim}}} E_{\text{car}}. \quad (88)$$

Thus, emissions from car trips $E_{\text{car}}^{\text{trip}}$ are obtained by multiplying the official average-car vehicle-kilometre factor E_{car} by the average kilometres traveled $\bar{d}_{\mathcal{D}_{\text{sim}}}$ computed from the simulation set $\mathcal{D}_{\text{sim}}$. PT emissions $E_{\text{PT}}^{\text{trip}}$ use the same distance and a bus/rail factor weighted by the bus-time share $\omega_{\text{bus}, \mathcal{D}_{\text{sim}}}$.

The per-trip externalities c_1 , c_{PT} , and c_{car} are

$$c_1 = 0,$$

$$c_{\text{PT}} = \frac{E_{\text{PT}}^{\text{trip}} V_{\text{CO}_2\text{e}}}{10^6}, \quad (89)$$

$$c_{\text{car}} = \frac{E_{\text{car}}^{\text{trip}} V_{\text{CO}_2\text{e}}}{10^6} + \bar{d}_{\mathcal{D}_{\text{sim}}} (C_{\text{cong}} + C_{\text{other}}). \quad (90)$$

Here, $V_{\text{CO}_2\text{e}}$ is the monetary value assigned to one tonne of CO_2e , and 10^6 converts grams to tonnes. The car externality c_{car} adds the carbon cost, the marginal congestion cost C_{cong} of an additional vehicle-kilometre on principal London roads, and the remaining road costs C_{other} . For profile

⁸UK Department for Energy Security and Net Zero. 2026. *Greenhouse Gas Reporting: Conversion Factors 2026*. See methodology Secs. 1 and 5 and the flat-file worksheet “Factors by Category.” <https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2026>.

⁹UK Department for Energy Security and Net Zero and UK Department for Business, Energy and Industrial Strategy. 2021. *Valuation of Greenhouse Gas Emissions: For Policy Appraisal and Evaluation*. See Annex 1. <https://www.gov.uk/government/publications/valuing-greenhouse-gas-emissions-in-policy-appraisal/valuation-of-greenhouse-gas-emissions-for-policy-appraisal-and-evaluation>.

¹⁰Stuart Adam and Rebekah Stroud. 2019. “A Road Map for Motoring Taxation.” In *The IFS Green Budget: October 2019*. Institute for Fiscal Studies, London. See Sec. 9.2 and Tables 9.1–9.2. <https://ifs.org.uk/books/road-map-motoring-taxation>.

¹¹Alice Ciccone, Aslak Fyhri, and Hanne B. Sundfør. 2021. “Using Behavioral Insights to Incentivize Cycling: Results from a Field Experiment.” *Journal of Economic Behavior & Organization* 188:1035–1058. See Sec. 3.1, pp. 1038–1039. <https://doi.org/10.1016/j.jebo.2021.06.011>.

¹²DNB Markets. 2021. *Monthly and Yearly Averages for Major Currencies 2021*. Annual-average GBP/NOK mid-rate; the table states that the reported values average daily mid-rates. <https://www.dnb.no/en/markets/fx-and-ir/fx-rates/monthly-yearly-averages/average-rates-2021?la=EN>.

Table 7: Externally specified inputs. The reference column identifies the precise source location.

Input	Symbol	Value	Reference
London local-bus emissions	E_{bus}	63.60 gCO ₂ e/passenger-km	Category “Local bus, London” and methodology in [Footnote 8 - Sec. 5, pp. 74–76.]
National-rail emissions	E_{rail}	30.92 gCO ₂ e/passenger-km	Category “National rail” and methodology in [Footnote 8 - Sec. 5, p. 78.]
Average-car emissions	E_{car}	165.91 gCO ₂ e/vehicle-km	Category “Average car” and methodology in [Footnote 8 - Sec. 5, pp. 46–69].
Central carbon value	$V_{\text{CO}_2\text{e}}$	£264/tCO ₂ e	Central 2026 value in [Footnote 9 - Annex 1]
Congestion externality	C_{cong}	£0.81/vehicle-km	Marginal congestion cost of one additional vehicle-kilometre on principal roads in London, estimated in [Footnote 10 Sec. 9.2, Table 9.2].
Other car externalities	C_{other}	£0.03/vehicle-km	Sum of accident, local-air-pollution, noise, and infrastructure costs, estimated in [Footnote 10 Sec. 9.2, Table 9.2].
Reference biking incentive	R_{Cic}	2.00 NOK/cycled-km	Pay-per-kilometre treatment, tested in a real world experiment - see [Footnote 11 Sec. 3.1, pp. 1038–1039]. We use this value only to set the value of β , as discussed in Eqs. (92)–(95). We do not use it directly as the incentives proposed, which we instead calculate via Prop. 3.
NOK/GBP exchange rate	$X_{\text{NOK/GBP}}$	11.83 NOK/£	2021 annual-average GBP rate [Footnote 12].

k , the expected externality conditional on not choosing bike, $c_{-1,k}$, is

$$c_{-1,k} = \mathbb{P}_k(2 \mid i \neq 1)_{\text{CPT}} + \mathbb{P}_k(3 \mid i \neq 1)_{\text{Ccar}}. \quad (91)$$

The generalized social-cost objective $\tilde{f}^v(y)$ in Eq. (5) weights expected incentive expenditure by β . To set the policy weight β , we anchor the optimal incentive of the car-oriented profile $k = 2$ to an incentive implemented in a cycling field experiment. Ciccone, Fyhri, and Sundf r paid participants $R_{\text{Cic}} = 2.00$ Norwegian kroner (NOK) for each kilometre cycled in their pay-per-kilometre treatment¹¹. Applying this rate to the mean trip distance $\bar{d}_{\mathcal{D}_{\text{sim}}} = 5.41$ km/trip gives

$$y_{\text{Cic}}^{\text{NOK}} = R_{\text{Cic}} \bar{d}_{\mathcal{D}_{\text{sim}}} = 10.81 \text{ NOK/bike trip}. \quad (92)$$

The annual-average exchange rate $X_{\text{NOK/GBP}} = 11.83$ NOK/£ converts this amount to

$$y_{\text{Cic}} = \frac{y_{\text{Cic}}^{\text{NOK}}}{X_{\text{NOK/GBP}}} = \text{£}0.91 \text{ per bike trip}. \quad (93)$$

Let $p_2(y) := \mathbb{P}_1^{u_2}(y)$ and $\Delta c_2 := c_{-1,2} - c_1$. For profile $k = 2$, differentiating the objective $\tilde{f}^{u_2}(y)$ in Eq. (5), and using $p'_2(y) = p_2(y)[1 - p_2(y)]/\theta$ from Eq. (2), gives

$$\begin{aligned} 0 &= p'_2(y)(c_1 - c_{-1,2} + \beta y) + \beta p_2(y) \\ &= \frac{p_2(y)}{\theta} \{ -[1 - p_2(y)]\Delta c_2 + \beta[\theta + (1 - p_2(y))y] \}. \end{aligned} \quad (94)$$

We impose $y_2^* = y_{\text{Cic}}$ in Eq. (94); solving for β yields

$$\beta = \frac{[1 - p_2(y_{\text{Cic}})](c_{-1,2} - c_1)}{\theta + [1 - p_2(y_{\text{Cic}})]y_{\text{Cic}}}. \quad (95)$$

The values issued from the calibration are the common scale $\hat{\theta} = \text{£}3.68$ and the profile-2 parameters entering $u_2(i, y)$ in Eq. (67). The payment y_{Cic} is instead constructed in

Eqs. (92)–(93). Substituting $\hat{\Theta}$, $\mathbf{x}_{i,2}$, and y_{Cic} into Eq. (67), and then evaluating Eq. (1), gives $p_2(y_{\text{Cic}}) = 4.76\%$. Thus, $p_2(y_{\text{Cic}})$ is a post-calibration model evaluation, not a calibrated parameter. The context construction in Eqs. (89)–(91) gives $c_{-1,2} = \text{£}2.67$ and $c_1 = 0$. Equation (95) therefore gives $\beta = 0.56$.

D1.6 Optimal incentives and induced mode choices

After calibration, the profile-specific behavioral parameters $\{\zeta_k^{\text{bike}}, \zeta_k^{\text{PT}}, \iota_k, \zeta_k^{\text{lic}}, \zeta_k^{\text{own}}\}_{k \in \mathcal{K}}$, the common parameters $\hat{v}$ and $\hat{\theta}$, and the representative attributes $\mathbf{x}_{i,k}$ determine the systematic utilities $u_k(i, 0)$ via Eq. (67). The externalities c_1 and $c_{-1,k}$ and the policy weight β then determine the optimal incentive y_k^* through Prop. 3. For each profile $k \in \mathcal{K}$ and mode $i \in \{1, 2, 3\}$, define

$$P_{ik}^0 := \mathbb{P}_i^{u_k}(0), \quad P_{ik}^* := \mathbb{P}_i^{u_k}(y_k^*). \quad (96)$$

The probabilities P_{ik}^* represent the choices generated by the calibrated profile k if the regulator offers y_k^* . Table 8 reports P_{ik}^0 , P_{ik}^* , and the resulting change in bike probability $\Delta P_{1k} := P_{1k}^* - P_{1k}^0$.

D1.7 Monte Carlo simulation The experiment consists of $N_{\text{MC}} = 100$ independent Monte Carlo replications, each with horizon $T = 18$ timeslots. We select $T = 18$ to examine the initial learning phase, in which the regulator has fewer than twenty choices from which to infer the adopted profile.

For replication $m \in \{1, \dots, N_{\text{MC}}\}$ and timeslot $t \in \{1, \dots, T\}$, a common random number $U_{m,t} \sim \text{Unif}(0, 1)$ is generated. The honest and manipulative mode choices are obtained by applying the inverse cumulative distribution of their respective choice probabilities $\mathbb{P}_i^{u_k}(y^t)$ in Eq. (1) to the same $U_{m,t}$. Pairing the same $U_{m,t}$ removes variation caused solely by using different random numbers when the two trajectories are compared.

Under InstPE, declaring profile r immediately determines the offered incentive y_r^* , while choices continue to follow the

Table 8: Profile-specific optimal incentives and induced mode choices. For $i \in \{1, 2, 3\}$, denoting bike, PT, and car, $P_{ik}^0 = \mathbb{P}_i^{u_k}(0)$ and $P_{ik}^* = \mathbb{P}_i^{u_k}(y_k^*)$. The change $\Delta P_{1k} = P_{1k}^* - P_{1k}^0$ is reported in percentage points.

k	π_k (%)	P_{1k}^0 (%)	P_{2k}^0 (%)	P_{3k}^0 (%)	$a_k - b_k$ (£)	$c_{-1,k}$ (£/trip)	y_k^* (£/bike trip)	P_{1k}^* (%)	ΔP_{1k} (pp)
0	21.26	52.44	19.12	28.44	0.36	2.88	0	52.44	0
1	39.36	3.96	52.08	43.96	-11.73	2.22	0.15	4.11	0.15
2	39.38	3.75	43.10	53.15	-11.93	2.67	0.91	4.76	1.01

Table 9: Settings of the Monte Carlo experiment.

Quantity	Symbol	Value
Simulation horizon	T	18 timeslots
Monte Carlo replications	N_{MC}	100
Confidence level	p_{CI}	0.95
Common random numbers	$U_{m,t}$	paired trajectories

true profile-0 distribution. Under RealPE, an agent adopting profile r chooses among bike, public transport, and car according to the multinomial-logit probabilities of profile r . Recall that, for online profile inference, the regulator uses only the binary observation

$$z^s = \mathbf{1}\{i^s = 1\},$$

so that a bike choice is recorded as $z^s = 1$, while both public transport and car choices are recorded as $z^s = 0$. Before timeslot t , the regulator evaluates the Bernoulli log-likelihoods $\ln L_k^t(\mathbf{h}^{t-1})$ defined in Eq. (64) and selects the inferred profile $\hat{k}^t$ according to Eq. (66). If several profiles attain the same maximum log-likelihood, the profile with the smallest optimal incentive y_k^* is selected; if a tie remains, the smallest profile index is selected.

Let $\Delta U_{r,m}^{s,PE}$ denote the realized manipulation advantage in timeslot s of replication m under $PE \in \{\text{InstPE}, \text{RealPE}\}$. The estimated average manipulation advantage through timeslot t is

$$\widehat{\Delta U}_{r,t}^{PE} = \frac{1}{N_{MC}} \sum_{m=1}^{N_{MC}} \frac{1}{t} \sum_{s=1}^t \Delta U_{r,m}^{s,PE}. \quad (97)$$

The 95% confidence interval for $\widehat{\Delta U}_{r,t}^{PE}$ is computed from the N_{MC} replication-specific time averages using a Student- t critical value. Figure 4 shows $\widehat{\Delta U}_{r,t}^{PE}$ over the complete horizon.

The time-averaged gain from manipulation and cost of manipulation are obtained by replacing $\Delta U_{r,m}^{s,PE}$ in Eq. (97) with $GM_{r,m}^{s,PE}$ and $CM_{r,m}^{s,PE}$, respectively. Their difference equals the manipulation advantage by Eq. (12). Figure 2a and Table 10 report the resulting decomposition at $T = 18$.

For each preference-elicitation model PE and simulated profile r , define the analytic long-run average manipulation

Table 10: Manipulation results at $T = 18$. The estimate $\widehat{\Delta U}_{r,T}^{PE}$ and its 95% interval are in pounds per timeslot; cumulative quantities are in pounds.

PE	r	$\widehat{\Delta U}_{r,T}^{PE}$	95% interval
InstPE	1	0.08	[0.08, 0.09]
InstPE	2	0.55	[0.53, 0.57]
RealPE	1	-0.16	[-0.18, -0.15]
RealPE	2	-0.16	[-0.17, -0.15]

PE	r	$\sum_{s=1}^T GM_r^s$	$\sum_{s=1}^T CM_r^s$	$\sum_{s=1}^T \Delta U_r^s$
InstPE	1	1.47	0	1.47
InstPE	2	10.04	0	10.04
RealPE	1	0.36	3.34	-2.99
RealPE	2	0.49	3.36	-2.87

advantage and its finite-horizon discrepancy as

$$\overline{\Delta U}_{r,\infty}^{PE} := \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t \Delta U_r^{s,PE}, \quad (98)$$

$$\Gamma_{r,T}^{PE} := \left| \widehat{\Delta U}_{r,T}^{PE} - \overline{\Delta U}_{r,\infty}^{PE} \right|. \quad (99)$$

For InstPE with $r = 1, 2$, the values of $\overline{\Delta U}_{r,\infty}^{\text{InstPE}}$ are 0.08 and 0.56 £/timeslot; for RealPE with $r = 1, 2$, the values of $\overline{\Delta U}_{r,\infty}^{\text{RealPE}}$ are -0.17 and -0.13 £/timeslot. The corresponding discrepancies $\Gamma_{r,T}^{PE}$ at $T = 18$ are 0.00, 0.00, 0.00, and 0.03 £/timeslot. Thus, the finite-horizon estimates have the same sign as all four analytic long-run averages; the largest remaining finite-horizon discrepancy is the RealPE case with $r = 2$.

D2 Social-Impact Relevance

The proposed framework is intended as a pre-deployment stress test for personalized incentive policies. Before implementing such a policy, a regulator can calibrate behavioral profiles from observed choices, compute the incentives associated with these profiles, and evaluate whether the additional incentive that an agent may obtain through manipulation compensates for the resulting Cost of Manipulation. The theoretical bounds provide an analytical robustness certificate, while the simulation procedure evaluates the same mechanism under calibrated behavioral heterogeneity and stochastic choices.

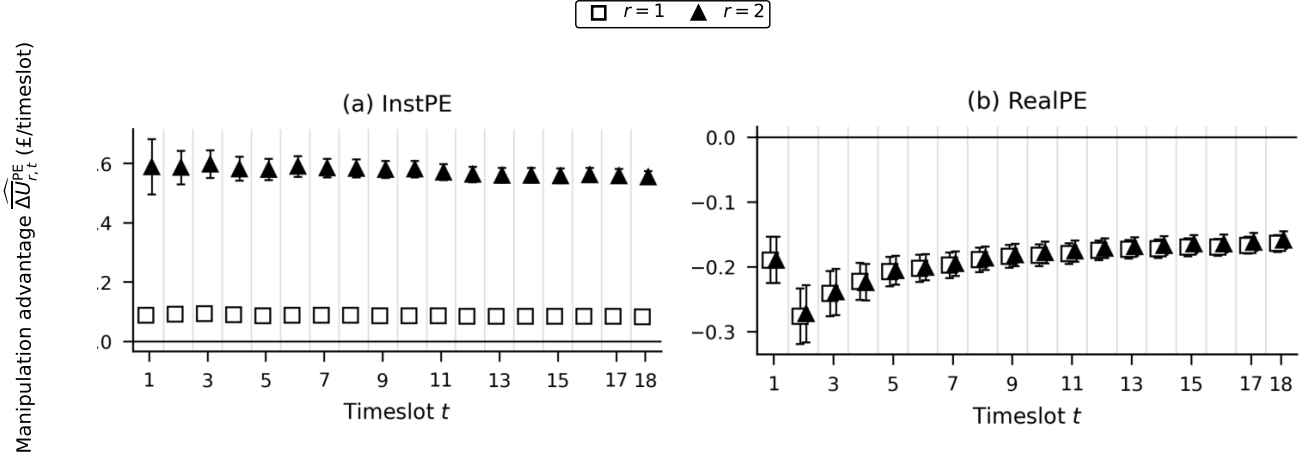


Figure 4: Estimated time-averaged manipulation advantage $\widehat{\Delta U}_{r,t}^{\text{PE}}$ from Eq. (97). Vertical intervals are 95% Student- t confidence intervals across the Monte Carlo replications.

The London case study provides empirical grounding for this evaluation. The choice model and behavioral profiles are calibrated from observed travel-mode choices, after which the interaction between preference elicitation, personalized incentives, and strategic behavior is simulated under controlled and reproducible conditions. This design isolates the mechanism studied in the paper: whether repeatedly imitating another behavioral profile remains advantageous once the regulator learns from the agent’s choices. The results show that the conclusion obtained under instantaneous preference elicitation can be reversed when this learning process is explicitly represented. In the calibrated instance, realistic preference elicitation makes the Cost of Manipulation exceed the Gain from Manipulation while the incentive policy continues to increase the share of sustainable choices.

This use of calibrated simulation is appropriate for policy screening: it allows candidate policies to be rejected or revised before exposing individuals or public budgets to an insufficiently robust intervention. It also makes the assumptions governing the policy explicit and auditable. In particular, the regulator applies a deterministic and pre-announced mapping from inferred profiles to incentives, and the preference-elicitation procedure can operate on the binary information required to determine whether the incentivized alternative was selected, rather than requiring the complete observation of every non-incentivized choice.

The analysis therefore supports a concrete deployment workflow: calibrate the behavioral model for the population of interest, evaluate manipulation incentives under realistic preference elicitation, and deploy the policy only when the resulting manipulation advantage remains non-positive over the relevant interaction horizon. When this condition is not satisfied, the same framework identifies the profiles and time horizons for which the policy should be redesigned. The contribution is thus not tied to a single transport intervention: it provides a reproducible procedure for assessing the strategic robustness of personalized sustainability policies before implementation.