Equilibrium Neglect and Political Feasibility

click here for latest version

preliminary. PLEASE DO NOT DISTRIBUTE

Bnaya Dreyfuss*

August 22, 2025

Abstract

When voters underappreciate the equilibrium effects of policies, they may rank them incorrectly, making socially optimal policies politically infeasible. I study the policy implications of *equilibrium neglect* (EN) and propose a potential remedy. I first develop a general framework of EN, and then use this framework to construct a portable and generally applicable potential remedy: using off-path contingent rebates a social planner can generate a policy that (i) implements the social optimum, (ii) does not increase the deficit—on and off the equilibrium path—and (iii) is politically feasible even in the presence of equilibrium-neglecting voters. This solution can be applied in a large class of problems. Empirically, I administer an opinion poll on congestion pricing with a large representative sample from six US metros. Respondents are overly pessimistic about traffic-reduction effects of congestion pricing, which in turn translates to strong opposition to the policy. Adding contingent compensation that pays if traffic remains high significantly increases public support, especially among potential recipients.

*Harvard University. Email: bdreyfuss@g.harvard.edu. I thank Raj Chetty, Katherine Coffman, Justin de Benedictis-Kessner, James Druckman, Benjamin Enke, Ed Glaeser, Yannai Gonczarowski, Ori Heffetz, Gabriel Kreindler, David Laibson, Shengwu Li, Matthew Rabin, Alex Rees-Jones, Joshua Schwartzstein, Jesse Shapiro, David Yang, seminar and conference participants at Harvard, BEAM 2025 and SITE 2025, for helpful feedback and suggestions. This project is supported by the Chae Family Fund, the Lab for Economic Applications and Policy (LEAP), Kenneth C. Griffin Research Fund, Institute for Quantitative Social Science (IQSS), the Center for American Political Studies (CAPS) and Time-sharing Experiments for the Social Sciences (TESS).

1 Introduction

Economists and the general public often disagree about policies. Despite years of policy advocacy, policies economists typically oppose, such as rent control and tariffs, are viewed positively in the eyes of much of the public, whereas policies economists tend to support, such as carbon taxes and congestion pricing, are viewed negatively (Caplan, 2001; Sapienza and Zingales, 2013).

There are many legitimate reasons for such disagreements: the models economists use to generate policy recommendations can be misspecified, oversimplified, or ignore aspects that people view as important. However, in many cases, there are good reasons to believe that the way people intuit about the costs and benefits of policies may lead them to the wrong conclusions. In particular, when evaluating a new policy, people often underestimate or overlook a crucial element: others' reaction to it, or its *equilibrium effects*. This tendency can distort perceptions of a policy's indirect costs or benefits, potentially leading to wrong assessment of its overall impact. I call this bias *equilibrium neglect*. Indeed, there is a large body of work, reviewed below, showing that in many contexts people fail to perform contingent reasoning (Niederle and Vespa, 2023), specifically when thinking about others (Eyster and Rabin, 2005; Dal Bó et al., 2018).

Taking this body of work as the starting point, I ask: what are the implications for policy design? In particular, can policymakers design welfare-improving policies that can win support from equilibrium-neglecting voters? I first model what it means for agents to neglect equilibrium when evaluating a new policy, and then use this model to identify a portable and generally applicable potential remedy. The remedy adds *off-path* contingent rebates to an existing policy, ensuring that if the true benefits of the policy are positive, then the agent's *perceived* benefits of the policy are also positive, even when they neglect its equilibrium effects. I then empirically test the effectiveness of this remedy in increasing public support in a representative-sample opinion poll about congestion pricing. Consistent with the theory, I find that respondents' beliefs about the effect of congestion pricing are overly pessimistic, and that pessimism predicts opposition to the policy. The theory-implied remedy significantly increases support, especially among potential recipients.

While equilibrium neglect can play a role in the political feasibility of various policies (e.g., taxation, price controls and industrial policy), I use congestion pricing for my theoretical running example and for my empirical application. Congestion pricing is particularly suitable for studying equilibrium neglect for a few reasons. First, it offers a straightforward distinction between direct effects (toll increase) and equilibrium effect (congestion reduction), both of which are unidimensional, easily observed, and precisely measured. Second, while there is consensus among economists and transportation engineers that it can increase welfare, congestion pricing is in practice notoriously

difficult to implement due to political feasibility constraints (Gu et al., 2018). Third, there is evidence (reviewed below) suggesting that EN is particularly relevant in the context of congestion pricing: support for congestion pricing after its implementation (i.e., after observing the benefits) is typically 10-20 percentage points higher compared to pre-implementation support (i.e., before observing the effects on traffic). In addition, there is evidence showing that people underappreciate the congestion-reducing effects of the policy ex ante, but strongly revise their beliefs having observed its effects.

This paper contributes to the literature on misspecified subjective models of the economy, failures in contingent reasoning, behavioral mechanism design, and the political feasibility of congestion pricing, all reviewed in Section 2.

Section 3 presents my theoretical framework. A social planner chooses a policy from a policy space. Policies shape payoffs and incentives and therefore each policy induces a different game. I model how agents perceive the costs and benefits of a policy change (relative to the status quo), i.e., the perceived benefit of moving between games within a class of games. I decompose the effect of a policy change into a *direct effect*—the effect of a policy change holding others' behavior fixed—and an *equilibrium effect*—the effect of converging to the new equilibrium, holding the policy fixed. I allow agents to underweight or cognitively discount the latter effect, which I call *equilibrium neglect* (EN). EN can be driven by explicit pessimism—believing that others may not change their behavior in response to the policy change—or by an attenuated perception of equilibrium effects, which are typically more complex, less salient and harder to predict.

In Section 3.2 I apply this framework in a simple model of congestion pricing. I show that even when congestion pricing generates a Pareto improvement, when agents are sufficiently equilibrium neglectful, it may not generate a sufficiently high level of support to make it politically feasible. This effect is driven by inframarginal drivers failing to fully realize the positive effect of congestion pricing on their commute times. I then show that adding contingent compensation completely solves this problem. In particular, a compensation that increases with the level of congestion (by, e.g., redistributing the collected revenue—which increases with the level congestion) makes equilibrium-neglectful inframarginal drivers perceive positive benefits. Since compensation increases with congestion, it effectively provides agents with insurance against bad outcomes: high congestion implies more toll-paying drivers, which increases compensation. Conversely, low compensation implies fewer toll-paying drivers and hence lower congestion. I show that a congestion-pricing policy that includes the added compensation has three desirable features: (i) it ensures positive perceived benefits for all agents and thus political feasibility, (ii) it eliminates socially suboptimal congestion, and (iii) it achieves these goals without increasing the deficit.

Theorem 1 generalizes this result, showing that the same tool can be applied in any setting with quasi-linear utilities and transfers. In particular, starting from any status

quo, and given any constrained social optimum, we can generate a policy that (i) generates the level of support that would have been achieved absent equilibrium neglect, (ii) achieves the same social outcome, and (iii) does not increase the deficit. The policy achieves these goals by introducing an off-path component to the social optimum that, while not affecting the outcome and hence the deficit, increases its perceived benefits for equilibrium-neglectful agents.

Theorem 2 shows that the contingent compensation can be constructed in a way that avoids an increase in the deficit not just *on* the equilibrium path but even *off* path. In particular, I show that even off-path, the deficit is not higher than the status quo deficit. This alleviates potential credibility concerns, as voters might treat policies involving excessively large deficits (even off path) with suspicion. It also makes the remedy effective in the presence of equilibrium-neglecting policy makers who may themselves be pessimistic about equilibrium effects, and may be hesitant to commit to large off-path deficits.

I then turn to empirically investigate equilibrium neglect and the potential remedy in the context of congestion pricing, using an opinion poll. The poll has two goals: first, documenting equilibrium neglect and mapping it to policy attitudes in a real policy context. Second, testing whether the remedy implied by the theory—contingent compensation—is effective in increasing support in practice. Ex-ante it is far from obvious that contingent compensation will have a positive effect on support: understanding off-path rebates requires agents to engage in contingent reasoning—the very skill they may lack due to EN. In addition, contingent compensation can be complex, may feel unnatural or seen as non-credible, which could negatively affect public support.

Section 4 presents the survey design. I survey a representative sample drawn NORC's AmeriSpeak probability panel. The sample consists of 2,056 adults residing in one of six major U.S. major metropolitan areas: Boston, Chicago, Dallas, Los Angeles, San Francisco, and Washington D.C. The survey has a two-stage, between-subject design. All respondents first see a standard congestion pricing policy modeled after the original $15 toll proposed by the MTA in NYC in 2024. Respondents then answer questions about their beliefs regarding congestion reduction and commute habits.

In the second stage, they are asked to state their support for one of three randomly assigned modified policies: the *EN-robust* treatment adds a "Monitoring & Assurance" component: respondents are told that traffic data will be monitored and publicly available, and that if after the implementation traffic decreases by less than 8% in a given month (compared to the baseline level of that month), workers who commute to work in the downtown area would receive a $120 monthly compensation. The *control* treatment includes only the monitoring component, without compensation. The third treatment, *toll-reduction*, lowers the toll from $15 to $9 and serves as benchmark to measure the impact of EN-robustness against a traditional policy compromise.

Section 4.3 presents the findings. First, there is strong opposition to standard congestion pricing, which in turn translates to strong opposition. About 67% of the respondents oppose the policy while only 21% support it, resulting in -46 percentage points *net support*. Respondents substantially underestimate the benefits from congestion pricing, with around 39% indicating that they expect no change in commute times, and another 20% believing that traffic conditions will *worsen*. This over-pessimism in turn translates to opposition to the policy. A one-minute increase in perceived time savings translates to about 1.4 percentage-point increase in net support. Converting beliefs into dollar units I combine perceived time savings and respondents' value of time (inferred from income) into a *perceived benefit* index for each subject. An additional dollar of perceived benefit corresponds to about 1.5 percentage points increase in net support.

Looking at the 2nd-stage support across treatments, I find a significant effect of contingent compensation on support, concentrated among potential recipients. Among respondents who work in the downtown area ($N = 184$), i.e., those who would receive the compensation if congestion remains high, net support in EN-robust increases from baseline by 48 percentage points more than in control, and 42 percentage points more than in toll reduction (p-value of the diff-in-diff estimates = 0.0001–0.006). Looking at the whole sample ($N = 2,056$), I find a more modest, yet highly significant treatment effect of EN-robust (diff-in-diff estimate = 11.3–17.3, p-value = 0.003–0.000). This suggests that as predicted by the theory, contingent compensation is effective in increasing public support by alleviating the concerns of those who are negatively affected, making it a potentially effective tool of increasing policy support, especially compared to conventional policy adjustments such as toll reduction, which, unlike contingent compensation, is both less effective and generates lower revenue.

I conclude in Section 5.

2 Related Literature

It is widely acknowledged that people's models of the economy are often misspecified. Most relevant to our paper, Dal Bó et al. (2018) experimentally shows that people underappreciate equilibrium effects, which causally makes them support the suboptimal Prisoner's Dilemma over Harmony Game. Other work modeling and estimating people's potentially misspecified subjective models of the economy includes Andre et al. (2022); Stantcheva (2022); Chinoy et al. (2023); and Binetti et al. (2024). This paper contributes to this literature by constructively using a particular formal model of misperception to generate and empirically test a potential policy remedy.

More generally, our paper is related to the large body of work on failures in contingent reasoning, reviewed in detail in Niederle and Vespa (2023). In particular, failing to understand how actions and incentives are related has been shown to be particularly im-

portant, both theoretically (Eyster and Rabin, 2005; Cohen and Li, 2023) and empirically (e.g., Esponda and Vespa, 2014).

Similar to how these documented empirical regularities have been modeled and then applied to characterize different mechanisms by their strategic simplicity (e.g., Li, 2017; Bó and Hakimov, 2023; and Pycia and Troyan, 2023), we use a general framework to derive a generally applicable remedy for equilibrium neglect and test it empirically. While these papers are focused on people's behavior *within* a mechanism, we are interested in people's preferences *over* different mechanisms or policies.

Last, by using congestion pricing as our main application, we also contribute to the large literature on public acceptability of congestion pricing, reviewed in Gu et al. (2018). This literature consistently finds support for congestion pricing to be higher after its implementation, and that people are overly pessimistic about the congestion-reducing effects of such policies, strongly revising their beliefs post-implementation (Schuitema et al., 2010). Both findings are consistent with equilibrium neglect.

In terms of remedies, while the perceived effectiveness of the policy in reducing congestion has been shown to strongly predict support, informational treatments are found ineffective (Baranzini et al., 2021). In contrast, real changes to the policy have been shown to generate support. In particular, a reduction in the size of the toll (Baranzini et al., 2021), as well as earmarking revenues for specific purposes (Jaensirisak et al., 2005), are both generally associated with higher levels of support. Our proposed remedy addresses equilibrium neglect not by attempting to change people's beliefs, but by introducing a real change to the policy. However, rather than decreasing the toll, which would affect its effectiveness, our theory-driven remedy involves a particular earmarking by transparently re-investing the revenue in order to increase the perceived benefits for equilibrium-neglecting voters.

3 Theory

3.1 Framework

In this section, I present a framework capturing equilibrium neglect.

3.1.1 Preliminaries

There is a set of N of agents $n \in N$. Each agent $n \in N$ chooses an action a_n from an action space A_n. I denote the action profile by $a = \{a_n\}_{n \in N}$, and the set of possible action profiles by $A \equiv \prod_{n \in N} A_n$. For each player $n \in N$ there is a *policy function* $T_n \in \mathbb{T}_n$ chosen by the social planner. The policy function that maps action profiles into a consequence for the player: $T_n : A \to C_n$. We may think of T_n as capturing taxes, subsidies, public

goods and infrastructure, or any regulation a government might impose. A profile of policy functions, which I simply call "policy," is denoted by $T = \{T_n\}_{n \in N}$. The set of all possible policies is denoted by $\mathbb{T}$. Agents have preferences over action profiles and their own consequences (but not directly over policies), and so player n's utility is given by $u_n : A \times C_n \to \mathbb{R}$.

A set of players N, action profiles A, utility functions $u = \{u_n\}_{n \in N}$ combined with a set of possible policies $\mathbb{T}$ jointly define a class of games governed by T. The framework is aimed at capturing the perceived effect of a policy change, i.e., moving across games within the class. It will be convenient to denote the agent's utility function in the game induced by the policy T by $v_n(\cdot \mid T) \equiv u_n(\cdot, T_n(\cdot))$.

A policy T is *anonymous* if $T_n(a) = T_{n'}(a)$ for all $n, n' \in N$ and for all $a \in A$. In what follows I largely focus on anonymous policy. Abusing notation, when there is no ambiguity anonymous policies will appear with no index and $T(a)$ will denote the a single policy. A policy function is *unilateral* if it depends only the agent's own action, i.e., $T_n(a_n, a_{-n}) = T_n(a_n)$. In contrast, a policy function is *multilateral* if T_n also depends on a_{-n}. The distinction between unilateral and multilateral policies will become important in constructing the remedy later in the section.

For simplicity, I restrict attention to settings with a unique pure Nash Equilibrium for any $T \in \mathbb{T}$ (as is the case in all congestion games and, more generally, all potential games). The Nash equilibrium under policy T is denoted by $a^*(T)$. The best response for agent n to others' action profile a_{-n} and a policy function T_n is denoted by $a_n^*(a_{-n}, T_n)$.

3.1.2 Perceived Benefits

To formalize equilibrium neglect and how it shapes political feasibility, we need to formalize how agents perceive the costs and benefits of a policy change. Throughout, let T^0 denote the status-quo policy and T some newly proposed policy. For any policy pair (T^0, T), define the *direct effect* for agent n

$$D_n(T, T^0) \;\equiv\; v_n\left((\hat{a}_n,\, a_{-n}^*(T^0)) \mid T\right) \;-\; v_n\left(a^*(T^0) \mid T^0\right),$$

and the *equilibrium effect*

$$E_n(T, T^0) \;\equiv\; v_n\left(a^*(T) \mid T\right) \;-\; v_n\left((\hat{a}_n,\, a_{-n}^*(T^0)) \mid T\right),$$

where $\hat{a}_n \;=\; a_n^*\left(a_{-n}^*(T^0), T_n\right)$ is n's best response to T holding others behavior at the status-quo equilibrium.

The *total gain* from moving the policy from T^0 to T can therefore be written as

$$v_n\left(a^*(T) \mid T\right) - v_n\left(a^*(T^0) \mid T^0\right) \;=\; D_n(T, T^0) \;+\; E_n(T, T^0).$$

This expression decomposes the effect of a policy change into two parts: the *direct effect* of a policy change, holding the behavior of others fixed at the status quo, and the *equilibrium effect*, departing from the status-quo equilibrium to a new equilibrium—i.e., the effect of *others'* behavior change on the agent's utility—holding the policy fixed.

We use this decomposition to define the *perceived* benefit from a policy change:

Definition 1 (Perceived benefits)**.** *Let $\eta \in [0,1]$. The η-perceived benefit of of switching from T^0 to T for agent n is*

$$\delta_n(T, T^0; \eta) = D_n(T, T^0) + \eta\, E_n(T, T^0).$$

The perceived benefit allows the equilibrium effect to be cognitively discounted by $\eta \leq 1$. This discount factor has a few potential interpretations. First, while the direct effect is typically straightforward to calculate, predicting the equilibrium effect is much more difficult. For example, it typically requires the agent to hold an explicit belief about the distribution of preferences in the population. It is therefore discounted by η.[1]

A slightly different interpretation of η is probabilistic: we can think of it as representing the probability that the agent assigns to the economy converging to the new equilibrium as opposed to the event where the policy changes and yet others' behavior remains fixed.

In the most extreme case, $\eta = 0$ the voter only perceives (or believes in) the direct effect, and completely neglects (or misbelieves in-) changes in utility coming from others' behavioral reactions. Conversely, the $\eta = 1$ case embeds standard models in which voters perfectly perceive equilibrium effects.[2]

3.1.3 Political Feasibility

To the extent that $\eta < 1$ causes voters to rank policies incorrectly, EN can have direct policy implications by making socially desirable policies politically infeasible.

The following definition formalizes political feasibility as a function of perceived benefits for voters:

Definition 2. *Let T^0 be a status-quo policy, let $\{\eta_n\}_{n\in N}$, and let $\beta \in [0,1]$. A policy T is politically feasible under the status quo T^0 with respect to the threshold β given η if the share of players for whom $\delta_n(T', T; \eta_n) \geq 0$ is at least β.*

[1] Notice that the agent's own optimization is included in the direct effect. While one can also model own-optimization neglect on top of equilibrium neglect, for local changes this effect is negligible (by the envelope theorem).

[2] Notice that this framework can also capture *over*-weighting of equilibrium effects when $\eta > 1$. While this form of bias can certainly be important in policy discourse (and is perhaps common even, or maybe especially, among economists), I leave it for future work.

Two particularly interesting cases for feasibility are $\beta = 1$, i.e., the case where only Pareto improvements are politically feasible (e.g., because a vocal minority has enough power to maintain the status quo), and $\beta = \frac{1}{2}$, i.e., the median voter case.

3.1.4 Quasilinear Utility, Taxation and Deficit

While myframework can in principle deal with a wide range of policy tools such as infrastructure investment, zoning laws, and bans,[3] in what follows I focus on monetary consequences and quasilinear utility, i.e., I assume $\mathbb{T} = \{T : A \times N \to \mathbb{R}\}$ and $v_n(a \mid T) = w_n(a) + T_n(a)$.

This allows us to use the size of the deficit as a natural way to constrain and rank policies within $\mathbb{T}$. The *deficit* of policy T is given by $D_T(a^*(T)) = \sum_n T_n(a^*(T))$ or the equivalent integral in a continuum setting. Notice that $D_T(a^*(T^*)$ is the *on-path* deficit. Later in this section I discuss *off-path* deficits. Last, we say that T is (on-path) budget-balanced if $D_T(a^*(T)) \leq 0$.

3.2 Application: Congestion pricing

3.2.1 Setup

In this section, I outline mymain application of EN: congestion pricing. There is a unit continuum of agents and a single road. All agents decide between driving ($a_n = 1$) and riding the train ($a_n = 0$).

Riding the train always takes l minutes; in contrast, driving time depends on the share of agents who choose to drive. I assume that the cost of congestion is given by a simple step function: the road has some capacity $\alpha < 1$, and driving takes $l-(b-c\cdot\mathbf{1}(\gamma(a) > \alpha))$ where b is the time saving from uncongested driving, $\gamma(a)$ is the share of road users and c is the cost of congestion. Slightly abusing notation, I denote $c(a) = c \cdot \mathbf{1}(\gamma(a) > \alpha)$.

A *standard* congestion pricing policy is action-based and anonymous, and can be written as $T_n(a) = k - p \cdot a_n$, i.e., an unconditional lump-sum transfer and a price for using the road. I denote a standard congestion pricing policy by $(p, k) \in \mathbb{R}^2$.

Agents have quasi linear utility from time and money:

$$v_n(a \mid T) = I_n - \theta_n(l - a_n(b - c(a))) + T_n(a),$$

where I_n is income and θ_n captures the value of time for agent n. For simplicity, I assume $\theta_n \sim U(0, 1)$ and index agents by their type.

[3] While the action space for each agent does not depend on the policy, myframework can capture a ban on some action a_n as choosing for all agents a policy function $T_n((a_n, a_{-n})) = c_n$ for all a_{-n} such that $u(\cdot, c_n)$ is arbitrarily low.

To simplify the presentation but WLOG I assume $I_n = 0$ for all n and $l = 0$. To avoid equilibrium-multiplicity issues, I assume $c < b$.[4] It is therefore straightforward to see that absent of congestion pricing, i.e., if $(p, k) = (0, 0) \equiv T^0$, all agents drive (i.e., $\gamma(a^*(T^0) = 1)$ and the road is congested. We will treat this policy as the status quo.

3.2.2 Political Feasibility of Congestion Pricing

Congestion is socially suboptimal if total welfare is lower under the status quo, unpriced equilibrium than in the case where the road is used exactly at capacity and only the top α types drive, i.e., if $\int_0^1 \tilde{\theta}(b-c)d\tilde{\theta} \leq \int_{1-\alpha}^1 \tilde{\theta}bd\tilde{\theta}$. A sufficient condition for this to hold is $\frac{b-c}{b} \leq \alpha$. The sufficient condition holds if the road capacity is high enough (so that the social cost of eliminating congestion is not too high) or if the cost of congestion is high enough (so that eliminating congestion is sufficiently socially beneficial).

I first establish the baseline result: in this simplified setting, congestion pricing combined with lump-sum transfers maintaining budget balancedness can generate a Pareto improvement:

Proposition 1. *There exists a congestion pricing policy $T^* = (p^*, k^*)$ that satisfies the following:*

1. *T^* eliminates congestion*
2. *T^* is budget-balanced*
3. *T^* Pareto dominates T^0*

The fact that T^* Pareto dominates T^0 implies $\delta_\theta(T^*, T^0; 1) \geq 0$ for all θ. In other words, absent of equilibrium neglect, T^* is always politically feasible under T^0:

Corollary 1. *Let $\eta = 1$. Then T^* is politically feasible for any $\beta \in [0, 1]$.*

What happens when $\eta < 1$? Since much of the gains (and all of the surplus) from congestion pricing come from reduction in equilibrium congestion, lowering the value of η can narrow the set of feasibility thresholds such that the policy is politically feasible:

Proposition 2. *T^* is politically feasible under T^0 iff $\beta \leq \overline{\beta}(\eta)$, where*

1. *$\overline{\beta}(\eta)$ is weakly increasing in η*
2. *$\overline{\beta}(1) = 1$, $\overline{\beta}(0) = \min\{\frac{\alpha(1-\alpha)b}{b-c}, 1\}$*

Notice that this result implies that in some cases even Pareto improvements are politically infeasible. For example, if $\frac{b-c}{b} > \frac{1}{2}$ we have $\overline{\beta}(0) < \frac{1}{2}$—a Pareto improvement that cannot achieve a simple majority.

[4] If $c > b$, the equilibrium determines only the share of drivers riding the train but not their type.

3.3 EN-robust Optimal Congestion Pricing

The fundamental challenge for achieving political feasibility is to make agents who neglect equilibrium effects support a policy whose benefits come from equilibrium effects. Indeed, in this application, equilibrium neglect can make a Pareto improvement infeasible. Moreover, since k^* is the highest possible transfer that keeps the budget balanced, there is no other standard pricing policy that achieves the no-congestion optimum and is politically feasible without running a deficit.

A natural candidate for a remedy is the use of subsidies. For example the social planner can subsidize the use of public transportation instead of charging a price for driving. However, subsidies are costly, and depending on the social cost of congestion and the degree of equilibrium neglect, the necessary expansion in the program's budget may induce its own set of feasibility issues.

Our objective is therefore to find a policy satisfying the following desiderata: (i) eliminates congestion, (ii) keeps the budget balanced and (iii) is politically feasible in the presence of equilibrium-neglecting agents. More generally, given a (possibly constrained) social optimum T^* that induces an equilibrium a^*, we are looking for a policy $\tilde{T}$ that satisfies: (i) *outcome equivalence*, i.e., $a^*(\tilde{T}) = a^*$, (ii) *deficit equivalence*, i.e., $D_{\tilde{T}}(a^*) = D_{T^*}(a^*)$, and (iii) *EN-robustness*, i.e., if T^* is feasible with a threshold β absent of equilibrium neglect, $\tilde{T}$ is feasible with the same threshold β under any degree of equilibrium neglect. When there is no ambiguity, we will often refer to EN-robust policies satisfying outcome and budget equivalence simply as "EN-robust policies," while referring to the baseline standard policy as "EN-sensitive."

As the analysis above shows, in the context of congestion pricing, a standard—i.e., anonymous and action-based—congestion pricing policy satisfying these conditions may not exist. What about other types of policies?

Relaxing anonymity can typically help increase political feasibility by allowing transfers targeted at agents whose direct losses from the policy change are the highest. However, the practical use of such policies may be limited, since they require the planner to condition the policies on types, which are typically assumed to be private information.

Instead, the remedy relies on the use of profile-based policies. Intuitively, EN constrains feasibility through agents' mistaken beliefs on others' actions. Profile-based policies allow us to condition payoffs on *others'* actions, and therefore can guarantee the agents benefits even under their mistaken perceptions about others'. Unlike relaxing anonymity, profile-based policies do not require any additional information.

The social planner can propose a profile-based anonymous policy conditioning price on congestion: $T_n(a) = k^* - a_n \cdot p^* \cdot \mathbf{1}(\gamma(a) \leq \alpha)$. Such policy may be intuitively appealing as it effectively ensures toll-paying agents a refund in the event that congestion persists—precisely the event that equilibrium-neglecting voters perceive as likely. Notice that this

policy also satisfies deficit equivalence, since without congestion, the deficit is still zero. However, crucially, this policy does *not* satisfy outcome equivalence. In particular, it is easy to show that if congestion is an equilibrium under T^0, it is also an equilibrium under T. In other words, relative to T^*, T satisfies deficit equivalence and EN-robustness, but not outcome equivalence.

Offering a compensation for congestion for agents who drive does not satisfy outcome equivalence because it changes driving incentives. Therefore, in order to achieve political feasibility while maintaining outcome and deficit equivalence, we need to compensate for congestion while keeping driving incentives unchanged. This implies conditioning the *transfer*, rather than the price, on congestion, thus providing compensation for equilibrium-neglecting voters while keeping driving incentives constant:

Proposition 3. *Let $T^* = (p^*, k^*)$ be the standard optimal congestion pricing policy. Then the policy $\tilde{T}_n(a) = p^* \cdot \gamma(a) - a_n \cdot p^*$ is outcome and deficit equivalent to T^*, and is politically feasible under T^0 for any $\beta \in [0, 1]$ and any $\eta > 0$.*

This proposition shows that the social planner can guarantee the trifecta of outcome equivalence, deficit equivalence (relative to T^*), and EN robustness using a non-standard albeit quite natural and intuitive policy tool: *explicit* redistribution that conditions the payout on the share of agents driving. Redistribution achieves political feasibility by effectively providing insurance for voters: when congestion is high, so is compensation, guaranteeing benefits to drivers and non-drivers in all contingencies.

Notice that by construction, this policy keeps the budget balanced both on- and off path: since the size of the compensation always equals the size of the revenue, the government is never giving away more than it collects, independent of the action profile. This makes explicit redistribution not only natural, but also credible, since it does not increase the deficit relative to the status quo, in any contingency.

3.4 EN Robustness: a General Remedy

The previous section outlines how EN can make a socially desirable policy politically infeasible even when it generates benefits for all voters. It then shows how the social planner can generate an alternative policy that achieves political feasibility even under EN, without chaining the outcome or deficit. Of course, the EN feasibility constraint extends beyond the simple example we analyzed. Does the remedy I proposed generalize as well?

The following theorem shows that the same logic generalizes to *any* situation captured by myframework:

Theorem 1. *Let T^0 be a status quo policy, let $\beta \in [0, 1]$, and let T^* be an anonymous action-based policy. If T^* is politically feasible under T^0 with respect to the threshold β*

assuming $\eta_n = 1$ for all n, then there exists an anonymous (profile-based) policy $\tilde{T}$ that is outcome and deficit equivalent to T^*, and is politically feasible with respect to β for all $\eta \in [0,1]^{|N|}$.

The theorem shows that, starting from any status quo, for any action-based policy that is politically feasible assuming $\eta = 1$, there exists an EN-robust policy that achieves the same social outcome and deficit. In other words, given any constrained optimum that achieves standard political feasibility, we can generate at no additional cost an equivalent policy that completely eliminates potential EN constraints that may have prevented its implementation.

The proof of the theorem is constructive and therefore provides a recipe for constructing an EN-robust implementation. In essence, it takes the original policy and adds an *off-path* rebate component: a contingent compensation paid only when others' behavior remains unchanged. Since the additional component only pays off the equilibrium path and since the policies are outcome equivalent, the deficit remains unchanged.

3.5 EN Robustness and Off-path Deficit

While relying on off-path rebates never increases the deficit *on path*, it may increase the deficit off path, which has potential downsides. For example, voters might view a policy promising excessively large compensation as less credible and may cause them to suspect whether the social planner will follow through. Moreover, even when the social planner can credibly commit to such policies, they themselves might suffer from equilibrium neglect, or may be skeptic of the equilibrium predictions. Therefore, the off-path rebates would be constructed in a way that prevents running large deficits, even in off-path contingencies.

We define an *off path* deficit as $D_T(a)$ where $a \neq a^*(T)$ (notice that there are many such deficits). I will now show that we can generate a policy that is deficit equivalent to either the new policy or the status quo, but doing so requires knowledge of the status quo action.

Let T^0 be some status-quo policy and let $a^0 = a^*(T^0)$. We say that a policy T is an anonymous past-action based (PA based) if T_n depends on a^0_n. As it turns out, knowing the agent's status quo action is a way to guarantee a form of off-path deficit equivalence:

Theorem 2. *Let T^0 be a status quo policy and let $a^0 = a^*(T^0)$, let $\beta \in [0,1]$, and let T^* be an anonymous action-based policy. If T^* is politically feasible under T^0 with respect to the threshold β assuming $\eta_n = 1$ for all n, then there exists an anonymous PA (profile-based) policy $\tilde{T}$ that is outcome and deficit equivalent to T^* and is politically feasible with respect to β for any $\eta \in [0,1]$. Moreover, $\tilde{T}$ satisfies $D_{\tilde{T}}(a) = D_{T^*}(a)$ for all $a \in A \setminus a^0$, and $D_{\tilde{T}}(a^0) = D_{T^0}(a^0)$.*

In other words, allowing the policy to depend on agents' status-quo actions, we can restrict the off-path deficit in the contingency where players do not change their behavior– the only contingency where $\tilde{T}$ and T^* diverge—to be the same as the status quo deficit. In other words, the EN-robust policy achieves the desired social outcome while ensuring that even off path, the deficit does not increase relative to the status quo.

3.5.1 Discussion

Every EN-sensitive socially optimal action-based policy has an outcome and deficit equivalent EN-robust counterpart. As shown in the proof of Theorems 1 and 2, to construct such policies, the social planner adds a sufficiently large contingent compensation that pays when other players' behavior is the same as the status quo. In order to make the policy deficit-equivalent to the status quo policy under the status-quo profile, this compensation needs to depend on past actions. In particular, it is given by

$$T_n^0(a^0) - T^*(a_n^0),$$

i.e., the difference between what the agent pays under the status quo, and what the agent pays under the new policy if their behavior remains the same. Notice that this sum only depends on the agents' actions, not their type.[5] Intuitively, this difference serves as an upper bound on the losses from the direct effect, but aggregating it over all agents results in the same deficit as in the status quo.

Off-path payments can operate through different psychological channels. First, they might simply change perceived utility: if equilibrium neglect is driven by an explicit belief that with some probability the status quo does not change, then off-path payments ensure that the agent utility is higher even in this contingency. Second, they can operate through salience: when agents mistakenly overweight direct effects (relative to equilibrium effects), off-path payments ensure that this effect is sufficiently positive. Last, while I do not explicitly model this channel, off path payments can operate through persuasion: on top of increasing utility from the direct effect, promising an off-path payment can credibly signal to potentially skeptic agents that the social planner is confident in their equilibrium prediction. Therefore, I conjecture that adding the contingent component to the policy can increase support by directly affecting η, making agents less pessimistic about potential equilibrium effects.

[5] In fact, the social planner does not need to know the player's status-quo action, since the payment only depends on what the agent paid in the status quo, and what they would pay if they did not change their behavior, which in some cases (such as in mycongestion-pricing application) may be a strictly weaker requirement.

4 Survey

To test the theory, I use on a policy opinion poll, focusing on congestion pricing. As discussed in the introduction, congestion pricing is particularly suitable for studying equilibrium neglect and testing EN-robustness, for a few reasons. First, congestion pricing has a natural and clean distinction between the direct effect (toll increase) and equilibrium effect (congestion reduction). Second, convergence to the new equilibrium is typically fast, significantly reducing timing-related complications.[6] Third, the main outcome, traffic volume/travel time, has a single dimension, further reducing potential complications in the definition of the action space A. Last, outcomes can be easily and transparently measured, making contingent payments simpler and easier to credibly commit to.

The survey has two main purposes, testing two main components of the theory. The first is descriptive, testing for the presence of equilibrium neglect (in the form of congestion-reduction pessimism) in an ecologically important policy setting and showing how it shapes policy stances in a representative sample. Second, testing the remedy suggested by the theory, and showing whether contingent compensation indeed increases support.

4.1 Design

The survey has two-stage, between-subject design, where subjects first indicate support for a standard congestion pricing policy, answer beliefs and habits questions, and then in the second stage indicate support for a randomly assigned modified congestion pricing policy. These two stages correspond to the two goals stated above: directly documenting equilibrium neglect, and then testing EN robustness.

4.1.1 First stage: baseline support, habits, beliefs

The experimental materials are available in the Online Appendix. At the beginning of the survey, all respondents are presented with a standard congestion pricing policy in the major downtown of their metro area. According the policy, modeled after the originally proposed NYC Central Business District Tolling Program, vehicles entering the downtown area during peak hours would be charged a $15 daily toll, and the revenue would be used for transportation improvements.

Respondents are presented with the arguments made by supporters (citing research from other cities suggesting it can cut traffic by about 15–20% and reduce travel times by 10–15 minutes) and opponents (saying that the policy is an unfair burden that will not meaningfully reduce congestion). Adding supporters argument has two main reasons. First, it is meant to eliminate trivial information or attention frictions which can

[6] As a potential game, it also has a unique pure-strategy Nash equilibrium.

presumably be corrected by simple informational treatments (rather than the involved policy change I am proposing). Second, it provides respondents with a natural traffic-reduction benchmark which is exploited in the EN-robust condition in the second stage. Opponents' arguments are presented to avoid one-sided presentation, which can bias the results.[7] Respondents then indicate their level of support for the policy on a five-point scale: Strongly/Somewhat Support, Neither Support nor Oppose, or Somewhat/Strongly Oppose.

Respondents then answer questions in two order-randomized blocks. Block 1 contains two questions about commute habits: (i) the location of their workplace, and (ii) how frequently they drive into the downtown area. The first question is meant to identify downtown workers—the recipient population of the contingent compensation in the EN-robust condition. The second question is used in constructing the perceived benefit index described below.

Block 2 elicits beliefs about the effectiveness of congestion pricing and their level of certainty in their prediction. The question in this block is randomized: half of the respondents are asked to predict how congestion pricing "would affect driving time from [their] home to downtown [city]." Respondents select from nine options organized in 5-minute increments, ranging from 'Driving time would decrease by more than 15 minutes' to 'Driving time would not change' to 'Driving time would decrease by more than 15 minutes.'

The other half of the sample is asked to predict the effect congestion pricing had in NYC. This question asks respondents to estimate post-implementation travel times through the Holland Tunnel during rush hour. Respondents are told the pre-toll average commute was 23 minutes (compared to 8 minutes with no traffic), and select from seven 5-minute increments ranging from '10 minutes or less' to '36 minutes or more.' All respondents then report their level of certainty (0% to 100%) in their prediction (Enke and Graeber, 2023).

4.1.2 Second-stage elicitation

Respondents then see a modified congestion pricing policy and are asked to state their support on the same scale. The randomly assigned policy is identical to the policy presented at baseline, with one of the following modifications, where the modification is highlighted.

EN-robust Condition (40% of the sample) The policy adds a "Monitoring & Assurance" clause: "Traffic levels would be monitored and published monthly. If monthly

[7] A balanced presentation in policy questions is standard practice (American Association for Public Opinion Research, 2023), meant to prevent framing/experimenter demand effects which have been shown to persist under one-sided presentation of arguments (Schuman and Presser, 1996).

traffic decreases by less than 8% (about half the typical 15–20% reduction) all workers who commute to work in the downtown area would receive $120 tax credit for that month, regardless of how they commute."

Control Condition (40% of the sample) : The policy only includes the "monitoring" part without the contingent compensation ("Traffic levels would be monitored and published monthly").

Toll Reduction Condition (20%): The policy has a reduced toll from $15 to $9.

Discussion Proposition 2 from the theory section identifies status-quo commuters, those who currently drive into the tolled area, as the recipients of the contingent compensation in budget-balanced EN-robust policies. Since conditioning on past behavior may be impractical and may seem unnatural, the contingent compensation in the EN-robust condition targets downtown workers, an approximation meant to identify those who would be most directly affected by the policy. While not all downtown workers drive, and not all frequent drivers work downtown, this population segment has many regular commuters into the area and represents a natural, clearly defined, administratively feasible group for implementing the compensation.[8]

The control condition absorbs potential framing and repeated elicitation effects as well as the pure informational effect from monitoring. Therefore, comparing the support in both conditions (controlling for baseline support) isolates the effect of the compensation itself. In that sense, this comparison is a lower bound on the overall effect of introducing compensation into an existing congestion pricing policy in practice, which will likely include these other forces.

Last, the toll-reduction condition is meant to serve as a benchmark. In the face of strong opposition, a natural reaction by a politician/policymaker is to lower the toll to appease the opposition. This condition is therefore included to test off-path compensation against a standard adjustment margin. Notice that unlike off-path compensation, starting from any constrained optimal toll, this solution leads to worse equilibrium outcomes *and* lower revenue and/or higher deficit.

Parameter Selection The key parameters are based on the NYC tolling program:

Peak-Hour Tolls The $15 baseline toll reflects the original proposal for New York City's congestion pricing program. Toll reduction to $9 was selected based on NYC's revised proposal announced in 2024, which was adjusted in response to opposition. These

[8]The mean monthly trips for respondents who work downtown ($N = 184$) is 8.825 (median = 6) compared to the rest of the sample ($N = 1863$) whose mean monthly trips is 2.055 (median = 0.5), p-value < 0.000.

provide two focal benchmarks that respondents can evaluate and allow us to compare our contingent compensation approach against an actual recently implemented policy compromise.

Projected Traffic Reduction The expected traffic reduction with a $15 toll is derived from the Metropolitan Transportation Authority's environmental assessment of the congestion pricing program. The 8% threshold for compensation represents approximately half the projected reduction.

Monthly Compensation The $120 monthly compensation amount was calculated to ensure budget balance, even if congestion remains high. The calculation takes into account average daily revenue from toll; administrative costs and exemptions (estimated at 50% of gross revenue), and the ratio between total commuters and the number of workers who commute to work in the downtown area. Further details are in the Online Appendix.

4.1.3 2nd-stage belief elicitation

To test whether the treatment has an effect on beliefs, after stating support for the modified policy, respondents answer a second question about their beliefs in congestion reduction. Those who answered a question about NYC in the first stage now answer about their own own commute, and vice versa. Therefore, each respondent answers both questions (own commute and NYC traffic reduction), in randomized order.

4.2 Predictions

The primary outcome of interest is *net support*—the percentage supporting minus the percentage opposing the policy, representing political feasibility. The model makes three main predictions. First, the model predicts a non-trivial amount of equilibrium neglect, manifested in overly pessimistic beliefs about the effect of congestion pricing on commute time. Second, the model predicts that those pessimistic beliefs in turn predict opposition to the policy. Last, the model predict that contingent compensation increases net support. Notice that absent of altruistic motives, contingent compensation should only affect the support of potential recipients. Therefore, we should expect this sub-population to be more reactive to treatment compared to the general population. As discussed before, the comparison to the toll reduction condition serves as a benchmark, measuring the effectiveness of our contingent compensation approach relative to a more standard policy tool. Last, while the theory does not predict that the treatment will change beliefs, this is a plausible psychological mechanism: contingent compensation may signal confidence in

the policy, and therefore may make people more optimistic about its congestion-reducing effect. My design allows for testing of belief shifts as a mechanism for increase in support.

4.3 Results

4.4 Survey Administration and Sample

The poll was fielded by NORC's AmeriSpeak® Panel in April 2025.[9] The sample consists of a total of 2,056 adults residing in one the following Metropolitan Statistical Areas: Boston–Cambridge–Newton, MA-NH ($N = 161$); Chicago–Naperville–Elgin, IL-IN ($N = 590$); Los Angeles-Long Beach-Anaheim, CA ($N = 613$); Washington-Arlington-Alexandria, DC-VA-MD-WV ($N = 209$); Dallas-Fort Worth-Arlington, TX ($N = 211$); and San Francisco-Oakland-Berkeley, CA ($N = 272$).[10]

4.4.1 First Stage Support and Beliefs

In this section I first outline baseline support, quantify the degree of pessimism, and then show that these beliefs strongly predict oppositive.

Starting with net support, I find large opposition to the standard congestion pricing policy. As shown in Figure 1, 67% of respondents oppose congestion pricing, while only 21% support it; the remaining 13% chose "neither support nor oppose." This translates to a net support of -46 percentage points.

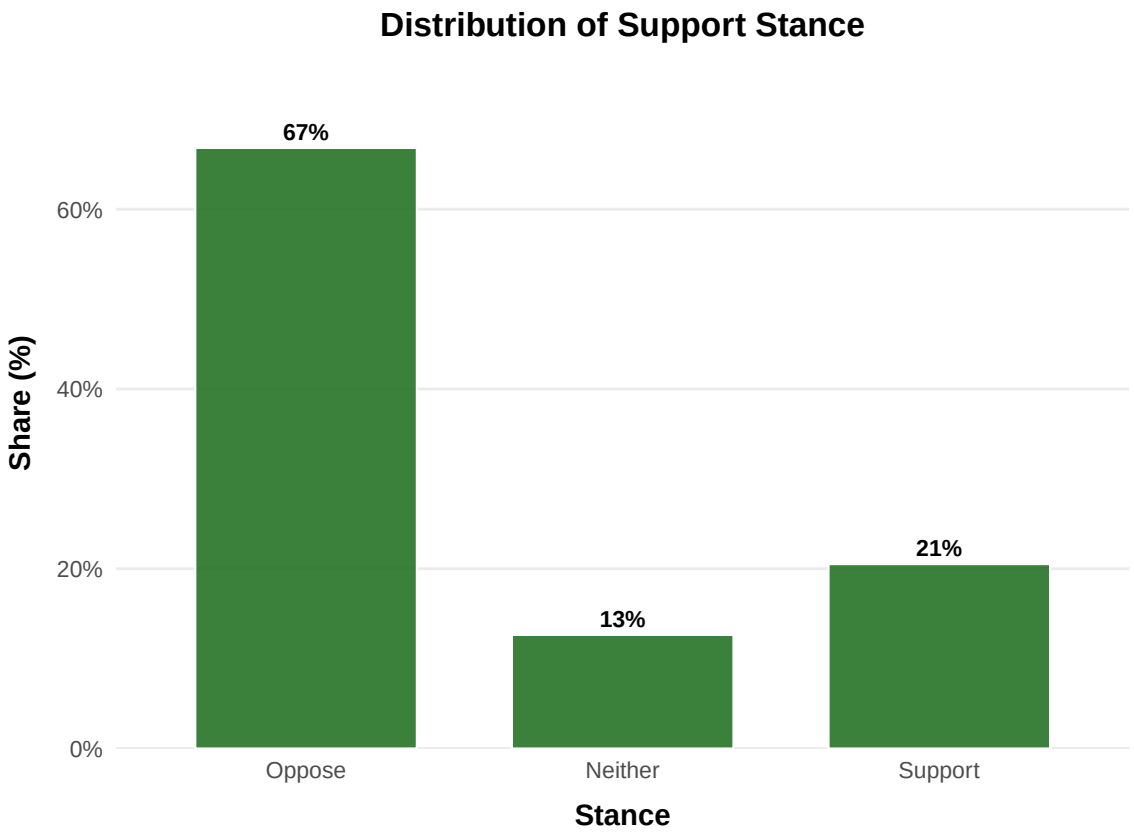
Next, to quantify the degree of over-pessimism in the data, I compare first-stage belief elicitation an objective measure of time savings. Figure 2 panel (a) plots the distribution of responses to the question about NYC. It shows that while the post-congestion pricing travel time through Holland tunnel is around 11–12 minutes, 65% of respondents estimate it to be 16-20 minutes or more. Notably, almost 33% of respondents estimate it at 21–25 minutes or more, indicating little to no improvement, or even a worsening in traffic conditions compared to the 23 minutes baseline.

Beliefs about traffic reduction in NYC have a clear interpretation, and can be easily compared to a verifiable ground truth. However, they may be only weakly correlated with support for congestion pricing if respondents believe that traffic reduction where they live would be different from NYC. In fact, information about "other cities" is already included

[9] AmeriSpeak® is a probability-based, nationally representative household panel operated by NORC at the University of Chicago. Households are randomly selected using area probability and address-based sampling from the NORC National Sample Frame, covering approximately 97% of the U.S. household population, excluding only P.O. Box addresses, some unlisted USPS addresses, and certain new constructions.

[10] These six MSAs were selected based on NORC's feasibility constraints, and in order to represent diverse geographic regions across the United States with the high congestion. According to the 2024 INRIX Global Traffic Scorecard, Chicago, Los Angeles, and Boston rank 2nd, 3rd, and 4th respectively for congestion in the United States, with Washington DC (9th), San Francisco (15th), and Dallas (20th) following in the rankings (INRIX and Pishue, 2024).

Figure 1: Distribution of Support Stance



Note: Distribution of support for $15 congestion toll in the sample.

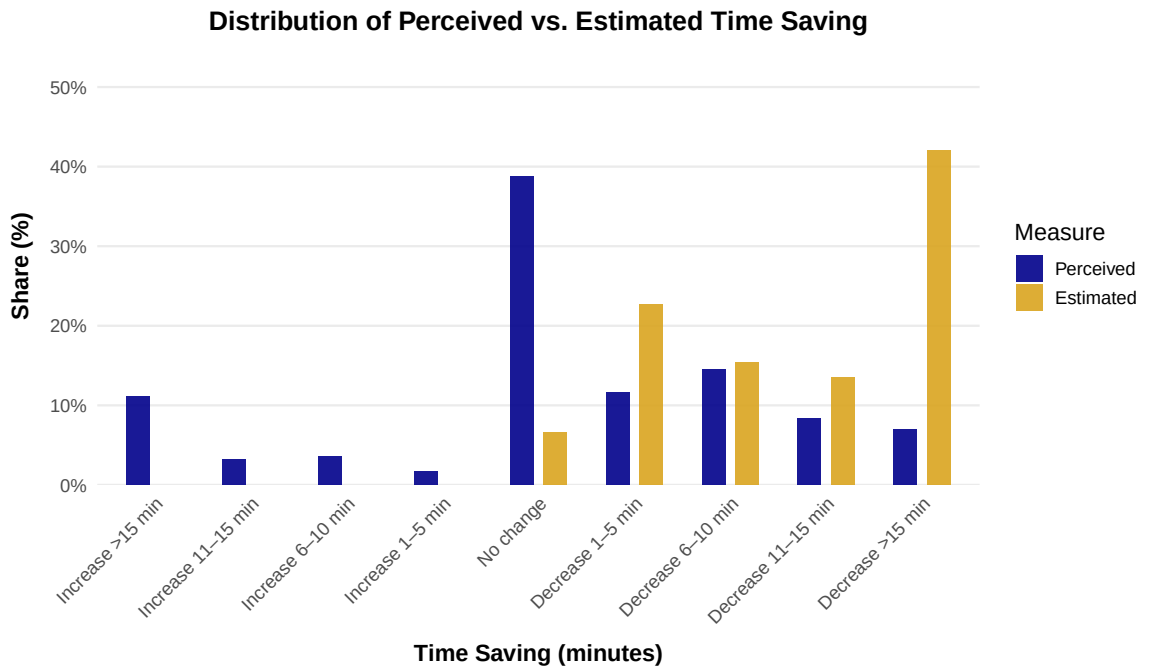
in the baseline elicitation as a part of supporters' arguments ("Supporters cite research from other cities suggesting it can cut traffic by about 15–20% and reduce travel times by 10–15 minutes.") It is therefore a strong test of equilibrium neglect, but at the same time, a potentially suboptimal measure of attitudes-relevant beliefs.

The other half of the sample is asked to predict the change in their *own* commute times if congestion pricing is implemented in their metro area. This measure may be more hypothetical and less concrete, but is more likely to reflect the relevant belief for the attitude toward the policy. I compare these beliefs to an estimated ground truth: for each respondent, I construct an estimated time-saving measure using the following formula:

$$\text{Estimated rush-hour time saving} = \alpha(\text{current rush-hour time} - \text{current off-peak time})$$

Travel times are obtained by feeding respondents' zip codes into Google Maps Distance Matrix API and calculating the commute times to a focal point downtown. As I am still in the process of obtaining NYC traffic data, I temporarily set $\alpha = \frac{2}{3}$ based on preliminary

Figure 2: Perceived vs. Estimated Time Saving



Notes: Perceived time savings from congestion pricing (blue bars) vs. estimated time savings (gold bars). Estimated time saving is calculated as $\alpha \cdot$ (current rush-hour time − current off-peak time), where $\alpha = 0.6$ and travel times are obtained from zip codes using Google Maps API. Perceived time saving is elicited by asking respondents how congestion pricing would affect their driving time.

analysis; this number will be updated once I get access to more comprehensive traffic data. I refer to this measure as the *calculated* time savings, in contrast to the participant's *perceived* time saving, which is based on their response to the question.

The distribution of calculated time savings (Figure 2 panel (b); yellow bars) shows that most respondents are predicted to experience significant time reductions, with the largest share (42%) estimated to save more than 15 minutes per trip. Only about 7% of respondents would experience a change in travel time that is less than one minute. However, perceived time savings (Figure 2; blue bars) are again consistent with strong equilibrium neglect among a solid majority of respondents: About 39% believe congestion pricing would result in no change at all to travel times, and another 20% believe travel times would *increase*, whereas less than 30% believe that travel time would decrease by 6 minutes or more.

Notably, 39% of respondents select the option "Driving time would not change," which is unlikely to be affected by the choice of the scale (15-minute increase to 15-minute decrease) or misunderstanding (e.g., mistaking "increase" for "decrease"). These beliefs are also unlikely to be driven by confusion: limiting the sample to respondents with below-median cognitive uncertainty (i.e., those reporting certainty in their prediction

$> 50\%$; $N = 460$), beliefs are *more* pessimistic, with 51% of respondents predicting no change, and another 20% predicting an increase in travel time.

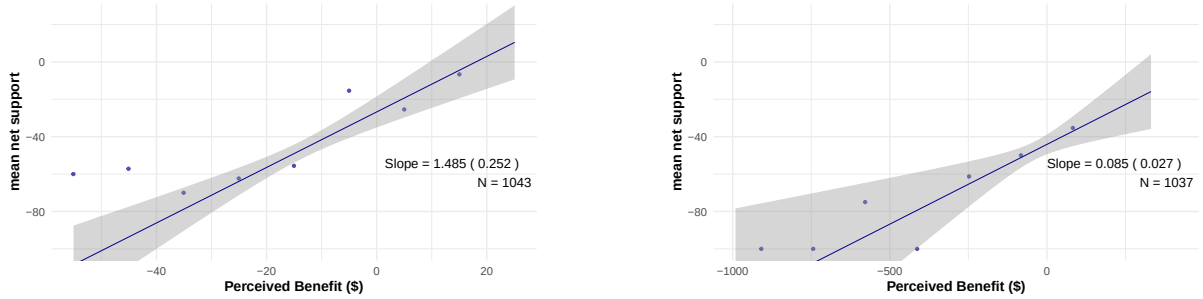
To quantify how beliefs translate into preferences and then to policy attitudes, I construct a per-trip perceived benefit index for each respondent:

$$\tilde{\delta}_{trip} = \theta_n \times \text{perceived time saving} - \$15,$$

where θ_n represents the value of time (VOT), set at 50% of hourly household income following US DOT guidelines Belenky (2011), and perceived time saving is the belief about respondents' own commute. This index maps to $\delta_n(T, T^0; \eta)$ from the theory section with the subtle difference that rather than plugging in the true effect and discounting by η, I directly plug in a potentially attenuated effect. To take into account trip frequency, I construct $\tilde{\delta}_{month}$ which multiplies per-trip benefit by the number of monthly driving trips reported by the respondent.

Figure 3 plots net support against these perceived benefit indices. It shows a positive relationship between net support and all of them. For example, panel (a) shows that each additional perceived per-trip dollar would result in an increase of almost 1.5 net support percentage points (SE = 0.252). Panel (b) does the same for monthly benefit, showing that, similarly, a \$30 increase in perceived monthly benefit would result in a 2.55 net support percentage point increase. Figure 4 plots net support on beliefs directly, providing a direct comparison between calculated vs. perceived time savings in predicting support. It shows that while elicited beliefs strongly predict net support, and each perceived one-minute saving increases net support by 1.38 (SE = 0.241), calculated time savings are essentially uncorrelated with support (coefficient -0.129, SE = 0.101). Therefore, respondents who live in congestion-prone areas and would typically benefit from the time saving associated with congestion pricing are not more likely to support congestion pricing.

Figure 3: Perceived Benefit and Net Support

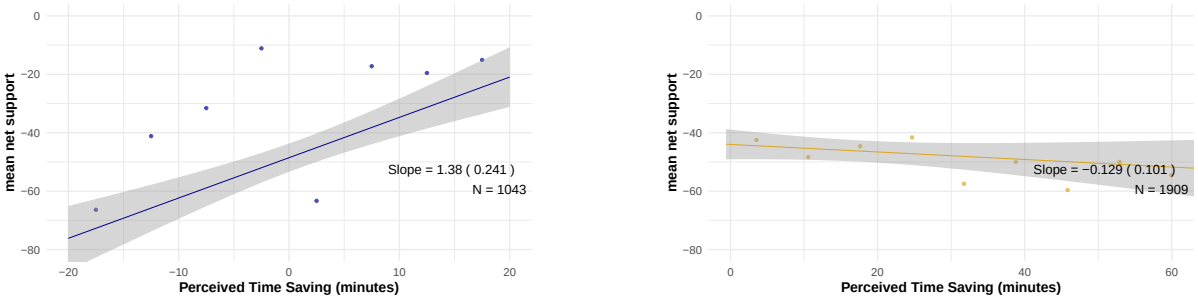


(a) Per trip (all)

(b) Monthly (all)

Note: Binned scatterplots of perceived benefits from congestion pricing on policy net support. Perceived benefit is calculated as $-\$15 + \theta_n \times$ perceived time saving, where θ_n represents the value of time (50% of hourly household income).

Figure 4: Time Saving and Net Support



(a) Perceived (b) Calculated

Note: Binned scatterplots of perceived and calculated time savings from congestion pricing on policy net support. Perceived time saving is elicited in the survey. Calculated time savings is calculated as $\frac{2}{3}$ of the difference between current rush-hour travel time and off-peak travel time from the respondent's home to downtown.

To summarize, these findings suggest that (1) there is substantial equilibrium neglect in my data, manifested in overly pessimistic beliefs about the effect of the policy on travel time, and (2) these over-pessimistic beliefs translate to opposition to the policy.

4.4.2 Second Stage: The Effect of Contingent Compensation

Having shown that pessimism is prevalent and predicts opposition to congestion pricing, I now turn to the remedy: are policies that include contingent compensation indeed robust to equilibrium neglect?

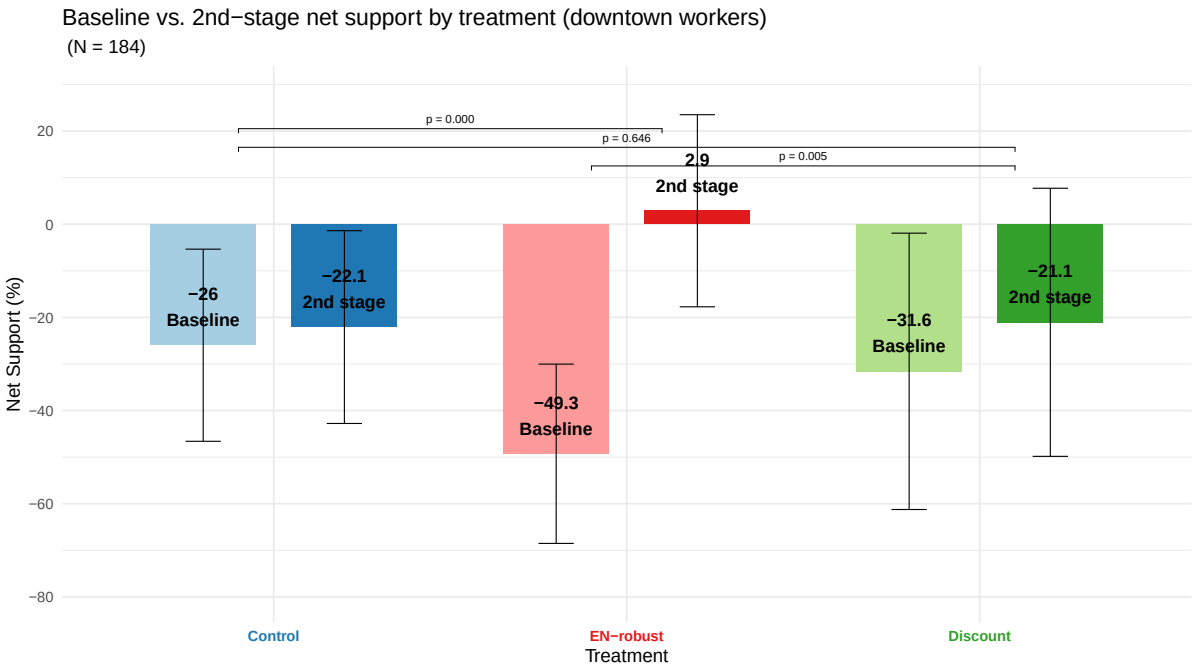
I first look at the (off-path) treated group: downtown workers. As the potential recipients of the compensation, I expect this group to be most affected by the EN-robustness. Other respondents should only respond to the treatment to the extent that they are altruistic or have other fairness concerns.

Panel (a) in Figure 5 compares the treatments among downtown workers. It shows that the EN-robust treatment leads to a 52.2 percentage point increase in net support (from −49.3% to 2.9%), effectively neutralizing the strong initial opposition. In contrast, the control condition shows a small increase (+3.9 percentage points), while the toll reduction condition shows a slightly larger increase (+10.5 percentage points). The difference in increase of the EN-robust condition is highly statistically significant compared to both the control condition ($p = 0.000$) and the toll reduction condition ($p = 0.006$).

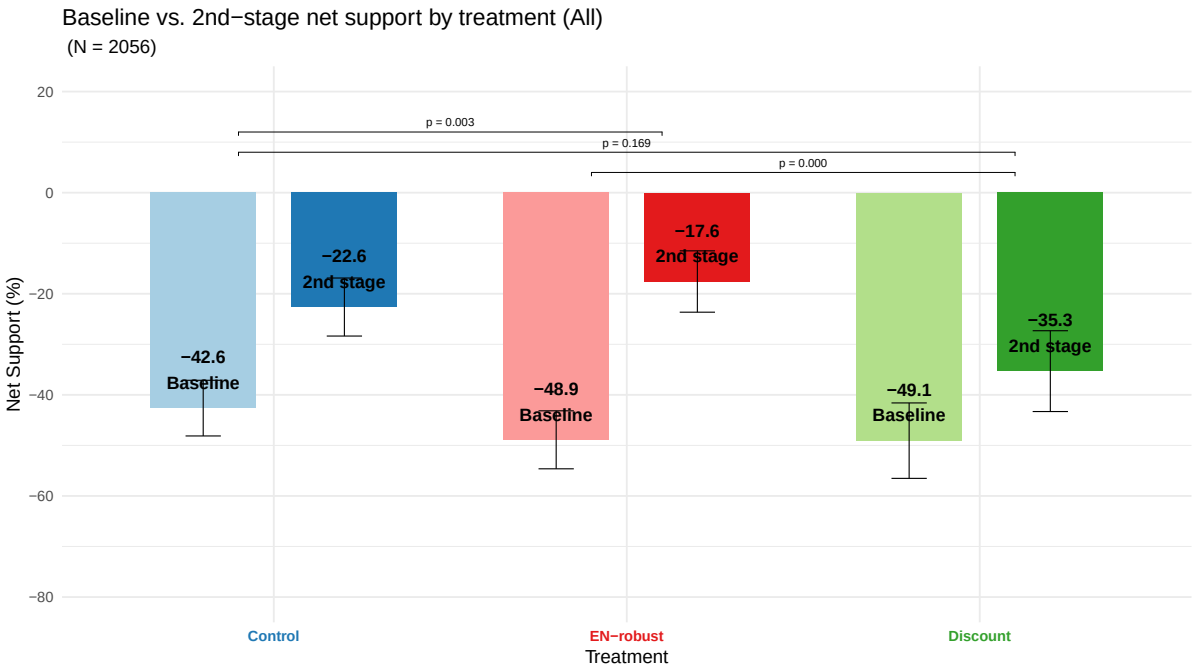
These results suggest that contingent compensation specifically targets the concerns of those who would potentially lose from the policy and creates a meaningful increase in their support. The small improvement in the toll reduction condition suggests that simply lowering the toll amount may not be very effective, at least among downtown workers.

Turning to the effect on the entire sample (Figure 5, panel (b)), while support increases

Figure 5: Effect of Treatment on Support for Congestion Pricing



(a) Downtown Workers ("Treatment on Treated")



(b) All Respondents (Average Treatment Effect)

Note: Treatment effect on net support increase. Net support is defined as the percentage supporting minus the percentage opposing. Panel (a) shows results for downtown workers only ($N = 184$). Panel (b) shows results for the entire sample ($N = 2056$). For each treatment, the light-colored bar represents baseline net support (first-stage elicitation), and the dark-colored bar represents net support after treatment (second-stage elicitation). The p-values above the bars are from tests on the difference in movement across treatments.

in all treatment conditions from baseline to second-stage elicitation, the EN-robust treatment shows the largest gain, 11.3 percentage points higher than the gain in the control condition ($p = 0.003$), and 17.5 percentage points higher than the toll reduction condition ($p = 0.000$). The improvement in the control condition likely reflects the effect of monitoring alone or some framing effects. The toll reduction condition shows the weakest improvement, suggesting that the effect of reducing the toll is similar in magnitude to simpler changes in transparency.

Contingent compensation increases support mainly among potential recipients. The fact that the increase is concentrated among potential recipients suggests that, consistent with the theory, the increase in support is driven by relieving a real concern about the outcome, rather than mere framing effects. It also suggests that fairness concerns and altruism play a minor role in this sample, since compensating downtown workers has a smaller effect on the general population.[11] Taken together, these findings support the third prediction of the model: contingent compensation significantly increases support for congestion pricing, especially among potential recipients. Furthermore, a more standard policy compromise in the form of a lower the toll achieves lower increase in support while sacrificing effectiveness and revenue.

5 Conclusion

TBD...

A Proofs

A.1 Proposition 1

Proof. Let $(p^*, k^*) = (b(1-\alpha), \alpha \cdot b(1-\alpha))$.

To see part (1), notice that without congestion, the share of agents who drive given a price p is given by $1 - \frac{p}{b}$. Therefore, plugging in p^* indeed implies no congestion. In other words, under p^* the equilibrium is such that only drivers with $\theta \in [1-\alpha, 1]$ drive.

Moving on to part (2), notice that since $\gamma(a^*((p^*, k^*))) = \alpha$, the revenue is given by $\alpha p^* = \alpha b(1-\alpha) = k^*$ which implies $D_{T^*}(a^*(T^*)) = 0$.

To show that T^* Pareto dominates T^0, we focus on the agent indifferent between

[11] Excluding downtown workers from the analysis (leaving only those who would *not* benefit from the compensation) shows a smaller and statistically weaker difference between the EN-robust and control conditions (7.7 percentage points difference, $p = 0.059$).

driving and riding the train: $\tilde{\theta} \equiv 1 - \alpha$. We have

$$\begin{aligned} v_{\tilde{\theta}}(a^*(T^*) \mid T^*) &= \underbrace{b(1-\alpha)\alpha}_{\text{transfer}} + \underbrace{(1-\alpha)b}_{\text{time}} - \underbrace{b(1-\alpha)}_{\text{price}} \\ &= b(1-\alpha)\alpha \\ &\geq v_{\tilde{\theta}}(a^*(T^0) \mid T^0) = (1-\alpha)(b-c) \iff \alpha \geq \frac{b-c}{b}, \end{aligned}$$

which we assumed holds as a sufficient condition for congestion being suboptimal. For all higher types, $\theta \in [1-\alpha, 1]$, we have

$$\frac{\partial v_{\tilde{\theta}}(a^*(T^*) \mid T^*)}{\partial \theta} = b > \frac{\partial v_{\tilde{\theta}}(a^*(T^*) \mid T^*)}{\partial \theta} = b - c,$$

and therefore the inequality holds for them as well.

For all lower types, $\theta \in [0, 1-\alpha]$, we have $\frac{\partial v_{\tilde{\theta}}(a^*(T^*)|T^*)}{\partial \theta} = 0 < \frac{\partial v_{\theta}(a^*(T^0)|T^0)}{\partial \theta} = b - c$, and therefore they all strictly gain from moving from T^0 to T^*.

□

A.2 Proposition 2

Proof. First part: For a given policy T, $v_\theta(a \mid T)$ is always weakly decreasing in the share of road users (i.e., if $\gamma(a') > \gamma(a)$ then $v_\theta(a \mid T) \geq v_\theta(a' \mid T)$). Since $\gamma(a^*(T^*)) = \alpha < \gamma(a^*(T^0)) = 1$, we have $v_\theta(a^*(T^*) \mid T^*) - v_\theta(a^*(T^0) \mid T^*) \geq 0$, which implies $\frac{\partial \delta_\theta(T^*, T^0; \eta)}{\partial \eta} \geq 0$. Therefore, the share of agents for which $\delta_\theta(T^*, T^0; \eta) \geq 0$ is increasing in η.

Moving on to the second part, when $\eta = 1$, Proposition 1 applies. When $\eta = 0$, we have two types: perceived non-drivers, and perceived drivers. These are the types that are (not) driving when the price is p^* and *the road is congested*. Solving for the perceived marginal (the type indifferent between driving and riding the train, when the price is p^* and the road is congested) yields $\hat{\theta} = \frac{b(1-\alpha)}{b-c}$.

For $\theta \in [0, \hat{\theta}]$ we have $\delta_\theta(T^*, T^0; 0) = \underbrace{b(1-\alpha)\alpha}_{k^*} - \underbrace{\theta(b-c)}_{\text{time loss}}$. Solving $\delta_\theta(T^*, T^0; 0) = 0$ for θ yields $\theta^* = \frac{b(1-\alpha)\alpha}{b-c}$, which implies that for $\theta \in [0, \min\{\theta^*, 1\}]$ we have $\delta_\theta(T^*, T^0; 0) \geq 0$, and therefore for $\eta = 0$, the feasibility threshold is given by $\min\{\frac{b(1-\alpha)\alpha}{b-c}, 1\}$, as needed.

Notice that for perceived drivers, the perceived benefit is given by $\underbrace{b(1-\alpha)\alpha}_{k^*} - \underbrace{b(1-\alpha)}_{p^*}$ which is always negative. □

A.3 Proposition 3

Proof. Notice that since $\tilde{T}$ and T^* only differ in the component that is independent of a_n, by quasi linearity we can write $v_\theta(a \mid \tilde{T})) = v_\theta(a \mid T^*) + h(a_{-n})$. Therefore if a_n^* is

the unique maximizer of $v_\theta(a^*(T^*) \mid T^*))$, it is also the unique maximizer of $v_\theta(a \mid \tilde{T}))$, showing outcome equivalence of $\tilde{T}$ and T^*. Since the transfer is equal to the sum collected, it is budget-balanced (and in particular, deficit equivalent) by construction. To show political feasibility, notice that

$$\delta_\theta(\tilde{T}, T^0; 0) = p^* \cdot \underbrace{\gamma(a^*(T^0))}_{=1} - \min\{\theta(b-c), p^*\} \geq 0$$

As shown in the proof of Proposition 2, we have $\frac{\partial \delta_\theta(T^*, T^0; \eta)}{\partial \eta} \geq 0$, which ensures $\delta_\theta(\tilde{T}, T^0; \eta) \geq 0$ for all $\eta > 0$. □

A.4 Theorems 1 and 2

Proof. Let $v_n(a \mid T) = w_n(a) + T_n(a)$. Denote $a^*(T^*) = a^*(T^*)$ and $a(T^0) = a^0$. Since T^* is politically feasible under $\eta = 1$, we have $w_n(a^*) + \hat{T}(a_n^*) - w_n(a^0) - T^0(a^0) \geq 0$ for at least β share of the population.

Let $\overline{a}_n = \arg\max_{a_n} w(a_n, a_{-n}^0) + T^*(a_n)$. We have

$$\delta_n(T^*, T^0; \eta) = w_n(\overline{a}_n, a_{-n}^0) + T^*(\overline{a}_n) - w_n(a^0) - T^0(a^0) + \eta[w_n(a^*) + \hat{T}(a_n^*) - w_n(\overline{a}_n, a_{-n}^0) - \hat{T}(\overline{a}_n)]$$

Let $\tilde{T}(a) = T^*(a_n) + \tilde{f}(a_{-n})$ where $\tilde{f}(a_{-n}) = k_n \cdot \mathbf{1}\{a_{-n} = a_{-n}^0\}$. We characterize k below in the case of Theorem 1 and 2. For now we emphasize that k_n does not depend on agent n's actions. Notice that by quasi-linearity, the best response is the same under T^* and $\tilde{T}$.

Take some agent for which $\delta_n(T^*, T^0; 1) \geq 0$. Notice that for any $k \geq 0$, we have $\delta_n(\tilde{T}, T^0; \eta) \geq \delta_n(T^*, T^0; \eta)$. There are two possible cases: the equilibrium effect can be either positive or negative. If the equilibrium effect is (weakly) negative, then $\delta_n(T^*, T^0; \eta)$ is (weakly) decreasing in η. Therefore we have

$$\delta_n(\tilde{T}, T^0; \eta) \geq \delta_n(T^*, T^0; \eta) \geq \delta_n(\hat{T}, T^0; 1) \geq 0,$$

and if T^* is politically feasible under T^0, so is $\tilde{T}$.

If the equilibrium effect is positive, then $\delta_n(\hat{T}, T^0; \eta)$ is increasing in η, and so it suffices to look at the case where $\eta = 0$. We have

$$\delta_n(\tilde{T}, T^0; 0) = w_n(\overline{a}_n, a_{-n}^0) + \hat{T}(\overline{a}_n) + \tilde{f}(a_{-n}^0) - w_n(a^0) - T^0(a^0)$$

Therefore, for any

$$k \geq w_n(a^0) + T^0(a^0) - w_n(\overline{a}_n, a_{-n}^0) - \hat{T}(\overline{a}_n), \tag{1}$$

the perceived benefit is positive, i.e., k_n must be equal to or larger than the direct effect. We can therefore choose for all n

$$k_n = k = \max_m w_m(a^0) + T^0(a^0) - w_m(\overline{a}_m, a^0_{-m}) - \hat{T}(\overline{a}_m)$$

which only requires knowledge of the support.

To prove theorem 2, notice that since $\overline{a}_n$ is a best response, we must have

$$w_n(\overline{a}_n, a^0_{-n}) + T^*(\overline{a}_n) \geq w_n(a^0) + T^*(a^0_n).$$

Let $\tilde{T}(a^0, a) = T^*(a_n) + k_n(a^0_n) \cdot \mathbf{1}\{a_{-n} = a^0_{-n}\}$ where $k_n(a^0_n) = T^0_n(a^0) - T^*(a^0_n)$. Notice that whether k_n is paid or not only depends on other players' current actions, and its size only depends on past actions, which ensures outcome equivalence.

To show political feasibility, we use the same argument as before and notice that

$$T^0_n(a^0) - T^*(a^0_n) = w_n(a^0) + T^0(a^0) - w_n(a^0) - T^*(a^0_n) \geq w_n(a^0) + T^0(a^0) - w_n(\overline{a}_n, a^0_{-n}) - T^*(\overline{a}_n)$$

Last, to characterize deficit, for any $a \in A \setminus a^0$ we have $D_{\tilde{T}}(a) = D_{T^*}(a)$ by construction. For a^0 we have

$$D_{\tilde{T}}(a^0) = \sum_{n \in N} T^*(a^0_n) + k_n(a^0) = \sum_{n \in N} T^0_n(a^0) = D_{T^0}(a^0)$$

as needed. □

References

American Association for Public Opinion Research. Question Wording. Technical report, American Association for Public Opinion Research, 2023. URL https://aapor.org/wp-content/uploads/2023/01/Question-Wording.pdf.

Peter Andre, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart. Subjective Models of the Macroeconomy: Evidence From Experts and Representative Samples. *The Review of Economic Studies*, 89(6):2958–2991, November 2022. ISSN 0034-6527. doi: 10.1093/restud/rdac008. URL https://doi.org/10.1093/restud/rdac008.

Andrea Baranzini, Stefano Carattini, and Linda Tesauro. Designing Effective and Acceptable Road Pricing Schemes: Evidence from the Geneva Congestion Charge. *Environmental and Resource Economics*, 79(3):417–482, July 2021. ISSN 1573-1502. doi: 10.1007/s10640-021-00564-y. URL https://doi.org/10.1007/s10640-021-00564-y.

Peter Belenky. Revised departmental guidance on valuation of travel time in economic analysis. Technical report, US Department of Transportation, Washington, DC, 2011.

Alberto Binetti, Francesco Nuzzi, and Stefanie Stantcheva. People's Understanding of Inflation, May 2024. URL https://www-nber-org.ezp-prod1.hul.harvard.edu/papers/w32497.

Inácio Bó and Rustamdjan Hakimov. Pick-an-Object Mechanisms. *Management Science*, September 2023. ISSN 0025-1909. doi: 10.1287/mnsc.2023.4908. URL https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2023.4908. Publisher: INFORMS.

Bryan Caplan. What Makes People Think Like Economists? Evidence on Economic Cognition from the "Survey of Americans and Economists on the Economy". *The Journal of Law and Economics*, 44(2):395–426, October 2001. ISSN 0022-2186. doi: 10.1086/322812. URL https://www.journals.uchicago.edu/doi/10.1086/322812. Publisher: The University of Chicago Press.

Sahil Chinoy, Nathan Nunn, Sandra Sequeira, and Stefanie Stantcheva. Zero-Sum Thinking and the Roots of U.S. Political Divides, September 2023. URL https://www.nber.org/papers/w31688.

Shani Cohen and Shengwu Li. Sequential Cursed Equilibrium. Paper, arXiv.org, August 2023. URL https://econpapers.repec.org/paper/arxpapers/2212.06025.htm.

Ernesto Dal Bó, Pedro Dal Bó, and Erik Eyster. The Demand for Bad Policy when Voters Underappreciate Equilibrium Effects. *The Review of Economic Studies*, 85(2):964–998, April 2018. ISSN 0034-6527. doi: 10.1093/restud/rdx031. URL https://doi.org/10.1093/restud/rdx031.

Benjamin Enke and Thomas Graeber. Cognitive Uncertainty*. *The Quarterly Journal of Economics*, 138(4):2021–2067, November 2023. ISSN 0033-5533. doi: 10.1093/qje/qjad025. URL https://doi.org/10.1093/qje/qjad025.

Ignacio Esponda and Emanuel Vespa. Hypothetical Thinking and Information Extraction in the Laboratory. *American Economic Journal: Microeconomics*, 6(4):180–202, November 2014. ISSN 1945-7669. doi: 10.1257/mic.6.4.180. URL https://www.aeaweb.org/articles?id=10.1257/mic.6.4.180.

Erik Eyster and Matthew Rabin. Cursed Equilibrium. *Econometrica*, 73(5):1623–1672, 2005. ISSN 1468-0262. doi: 10.1111/j.1468-0262.2005.00631.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2005.00631.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0262.2005.00631.x.

Ziyuan Gu, Zhiyuan Liu, Qixiu Cheng, and Meead Saberi. Congestion pricing practices and public acceptance: A review of evidence. *Case Studies on Transport Policy*, 6 (1):94–101, March 2018. ISSN 2213-624X. doi: 10.1016/j.cstp.2018.01.004. URL https://www.sciencedirect.com/science/article/pii/S2213624X16300608.

INRIX and Bob Pishue. INRIX 2024 Global Traffic Scorecard, 2024. URL https://inrix.com/scorecard/.

S. Jaensirisak, M. Wardman, and A. D. May. Explaining Variations in Public Acceptability of Road Pricing Schemes. *Journal of Transport Economics and Policy (JTEP)*, 39(2):127–154, May 2005.

Shengwu Li. Obviously Strategy-Proof Mechanisms. *American Economic Review*, 107 (11):3257–87, November 2017. doi: 10.1257/aer.20160425. URL http://www.aeaweb.org/articles?id=10.1257/aer.20160425.

Muriel Niederle and Emanuel Vespa. Cognitive Limitations: Failures of Contingent Thinking. *Annual Review of Economics*, 15(Volume 15, 2023):307–328, September 2023. ISSN 1941-1383, 1941-1391. doi: 10.1146/annurev-economics-091622-124733. URL https://www.annualreviews.org/content/journals/10.1146/annurev-economics-091622-124733. Publisher: Annual Reviews.

Marek Pycia and Peter Troyan. A Theory of Simplicity in Games and Mechanism Design. *Econometrica*, 91(4):1495–1526, 2023. ISSN 1468-0262. doi: 10.3982/ECTA16310. URL https://onlinelibrary-wiley-com.ezp-prod1.hul.harvard.edu/doi/abs/10.3982/ECTA16310. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA16310.

Paola Sapienza and Luigi Zingales. Economic Experts versus Average Americans. *American Economic Review*, 103(3):636–642, May 2013. ISSN 0002-8282. doi: 10.1257/aer.103.3.636. URL https://www.aeaweb.org/articles?id=10.1257/aer.103.3.636.

Geertje Schuitema, Linda Steg, and Sonja Forward. Explaining differences in acceptability before and acceptance after the implementation of a congestion charge in Stockholm. *Transportation Research Part A: Policy and Practice*, 44(2):99–109, February 2010. ISSN 0965-8564. doi: 10.1016/j.tra.2009.11.005. URL https://www.sciencedirect.com/science/article/pii/S0965856409001256.

Howard Schuman and Stanley Presser. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. SAGE, March 1996. ISBN 978-0-7619-0359-8. Google-Books-ID: Je640UKqNaYC.

Stefanie Stantcheva. Understanding of Trade, May 2022. URL https://www.nber.org/papers/w30040.