Policy Blind Spots

Ravi Vora *

Abstract

Despite their desirable properties, carbon taxes have received low public support. Leading explanations attribute the rejection of carbon taxes to preferences and ideology. We show that voters make simple mistakes when reasoning about the consequences of such policies. We focus on leading versions of carbon taxes that are designed to draw wide public support by issuing dividend payments from collected revenues, resulting in a policy that is progressive and produces more winners than losers. Using experiments, we show that subjects have specific blind spots in reasoning about policies with taxes and dividends, even when described in simple terms and in incentivized, abstract settings that remove scope for preferences and ideologies. Based on these findings, we introduce a new policy that is isomorphic to a leading carbon tax proposal, but engineered to be robust to these cognitive frictions. In a U.S. survey conducted in Summer 2023 we find that our new policy (which we call "carbon penalty and reward") garners 30 percentage points more public support than the original proposal (known as "carbon fee and dividend"). Using independent measures, we show that this increase in support is concentrated in people who have difficulty reasoning about the policy structure.

*University of Wyoming, Department of Economics. I am grateful, first and foremost, to Ryan Oprea, Kyle Meng, and Sevgi Yuksel for their guidance with this paper, as well as to Ted Bergstrom, Chris Costello, Pedro Dal Bo, Olivier Deschenes, Erik Eyster, Daniel Martin, Kyle Meng, Antony Millner, and Klaas Van't Veld, and seminar participants at CESifo Venice Summer Institute 2023, ESA North America 2024, the Occasional Workshop 2023, the University of Wyoming, UC Santa Barbara, and the University of Melbourne, for helpful comments. Earlier versions of the paper were titled "Carbon Pricing and Bounded Reasoning."

1 Introduction

Economists have long argued that directly pricing carbon with a tax is the ideal approach to mitigate climate change. Despite being efficient and effective, carbon taxes are not common and have low support in opinion polls.[1] The leading carbon tax designs recycle the revenue collected from the tax into uniform payments. For many years, economists and policy experts expressed hope that carbon taxes, paired with payments from the revenue, would attract support from the many voters expected to financially benefit(Shultz and Halstead, 2018).[2] But public support for this Fee & Dividend policy, as it is commonly known, appears to be low in many countries, and is frequently rejected by voters most likely to benefit. The leading explanation is that ideology drives these choices (Douenne and Fabre, 2022; Mildenberger et al., 2022).

In this paper, we provide evidence of an alternative explanation. We show, by analyzing incentivized choices in laboratory experiments, that people make systematic mistakes when reasoning about the consequences of policy structures like Fee & Dividend. These experiments use abstract policy-selection tasks that lack scope for preferences and ideologies. We identify the key behavioral blind spots and use them to design a new policy that is isomorphic to Fee & Dividend but robust to key cognitive frictions identified from the lab study. We test this new policy, Penalty & Reward, in a nationally representative survey, and independently measure the voters' policy reasoning skills, with incentives, on the abstract tasks used in the lab. Despite being theoretically equivalent, Penalty & Reward is more than 30 percentage points more popular with voters than Fee & Dividend, and this gap in public support is concentrated in people who demonstrate systematic difficulties on the abstract policy-selection tasks.

[1] For appealing theoretical qualities, see Pigou (1924); Baumol and Oates (1971); Baumol (1972); Montgomery (1972); Tietenberg (1973); Nordhaus (2001); Baranzini et al. (2017); Cramton et al. (2017); Blanchard et al. (2023). For effectiveness reducing emissions at minimal cost, see Murray and Rivers (2015); Andersson (2019); Colmer et al. (2022); Metcalf (2021, 2023).

[2] The carbon fee and dividend policy has the endorsement of thousands of economists, including 28 Nobel laureates. For detail, visit econstatement.org. It is the central policy proposal for two major climate policy advocacy organizations, the bipartisan Climate Leadership Council and the Citizen's Climate Lobby. The percentage of households expected to benefit, in consumption terms, varies by study ranging from 56%(Fremstad and Paul, 2019) to 70%(Horowitz et al., 2017) in the U.S., with benefiting households belonging predominantly to lower-income and middle-income quantiles.

Cognitive costs and limitations play an increasingly important role in behavioral economics. Errors and inattention, not preferences, evidently explain a wide range of choice anomalies.[3] The role of cognitive frictions in voter behavior, even when voters seem to make imprudent choices, cannot be measured from existing data, where individual tastes and information are potentially important confounds. Studies analyzing referendums and polls on carbon pricing highlight the politicization of climate change.[4] Given the costs of assessing unfamiliar policies, it is also worth exploring the cognitive roots of any misconceptions voters might have about them.

Our experimental design isolates the core reasoning required to evaluate real-world pricing policies, but in the absence of confounds, like variations in risk, time or social preferences, or variations of information, and free of scope for ideological factors. Despite a simplified, apolitical setting, subjects frequently make reasoning mistakes, and in a systematic way that could explain policy attitudes and outcomes documented throughout the world. Distinguishing the role of cognitive frictions from preferences or beliefs is important, as it can tell us whether policies can be modified or clarified to be better received. If voters who reject Fee & Dividend despite being likely to benefit from it are revealing their true preferences – for instance, rejecting the policy on moral or political grounds – then how policy makers respond would be different than if the voters were making reasoning mistakes.

Policy choices require hypothetical thinking and sufficient attention to various costs and benefits. An individual's incentive to expend cognitive effort on policy comprehension is low, but the aggregation of these small decisions is highly consequential (Downs, 1957). This paper attempts to address this by reducing the cognitive effort required to evaluate a policy by (1) diagnosing the troublesome mechanisms of voting errors in the lab, and (2) using these errors to guide a better policy design that reduces the likelihood of evaluation errors by voters and better elicits their true

[3] On taxes, health plan selection and other consumer decisions (Abeler and Jäger, 2015; Rees-Jones and Taubinsky, 2016, 2020; Bhargava et al., 2017; Chetty et al., 2009; Lacetera et al., 2012); and on risk, time, strategic, and other decision behavior that deviates from classic choice axioms (Oprea, 2022; Enke et al., 2023a; Banovetz and Oprea, 2023; Nielsen and Rehbeck, 2022).

[4] See Mildenberger et al. (2022); Douenne and Fabre (2022); Anderson et al. (2023). The studies find strong association of policy support with political identity, and also that inaccurate initial impressions can be difficult to correct with rich supplemental information.

preferences.

Choices in the lab reveal, among many possible cognitive mistakes that might bias policy selection, one that stands out: neglect of Distributed Effects. A voter deciding whether to support a carbon tax must consider multiple effects on her welfare, and the Distributed Effect is the proposed tax's welfare impact via distribution of tax revenues. When a voter belongs to a large population, Distributed Effects are a product of aggregate behavior and independent of a voter's own consumption decisions. Neglect of Distributed Effects, or NoDE, systematically biases cost-benefit analyses of policy proposals and drives frequent rejection of privately beneficial taxes and support of privately harmful subsidies. This finding motivates a new policy design that, in a test of support using a national sample, succeeds in winning public favor.

By removing the Distributed Effects that give voters difficulty, Penalty & Reward is designed to be more transparent than Fee & Dividend. A Fee & Dividend can be understood to produce two groups: (i) net receivers, whose dividend exceeds their tax payments; and (ii) net payers, whose tax payments exceed their dividends. These two groups are separated by a threshold at the per capita level of carbon consumption, a fact that requires accurately estimating and aggregating multiple effects. Because this assessment breaks down under NoDE, we design Penalty & Reward to linearly reward below-mean consumers based on how far beneath the population average their emissions reach, and linearly penalize above-mean consumers for emissions beyond the average. Instead of everyone paying taxes and collecting dividends, Penalty & Reward conducts a single transaction. We demonstrate that when penalty and reward rates are set identical to the fee rate in Fee & Dividend, these policies are economically isometric. In a nationally representative survey, we measure attitudes toward both policies and vulnerability to cognitive biases including NoDE that could explain differences in support for the two policies.

We find dramatically higher support for Penalty & Reward than for Fee & Dividend, showing that carbon pricing can be made much more popular by reducing the scope for cognitive frictions in policy decisions. Whereas 48% of voters express support for Fee & Dividend, 83% support Penalty & Reward. Other factors, like belief in climate science, play a role in our results, but Penalty & Reward's popularity advantage over Fee & Dividend holds across the political spectrum and is strongest among participants who discounted Distributed Effects in the abstract tasks.

This study contributes to our understanding of how support forms for important economic policies. The theoretical and practical advantages of a well-designed carbon tax make the reasons for its rare adoption worth careful exploration. Studies documenting low public support of carbon taxes date back to Baron and Jurney (1993). Findings in Dresner et al. (2006), Steg et al. (2006), Kallbekken and Aasen (2010)Kallbekken and Sælen (2011),Sælen and Kallbekken (2011), Baranzini and Carattini (2017), Carattini et al. (2017) Douenne and Fabre (2022), and Dechezleprêtre et al. (2022) together suggest that worry about impact to one's household and to low earners, perception of limited effectiveness, and concerns about how revenues would be spent have been consistent factors. Green subsidies tended to be more popular in comparison.[5]

In recent years, Fee & Dividend has become the leading carbon pricing strategy because of its promising qualities. The fact that it is misunderstood is striking. After all, Fee & Dividend is designed to be among the simplest structured policies: a linear tax and a uniform lump sum offset. The perception that it is regressive (Douenne and Fabre, 2022; Mildenberger et al., 2022; Maestre-Andrés et al., 2019) contrasts with economic intuition, as well as with careful work from empirical public economists(Cronin et al., 2019; Metcalf, 2023, 2021; Fremstad and Paul, 2019; Budolfson et al., 2021; Boyce, 2018; Goulder et al., 2019). In France, Douenne and Fabre (2022) measure support for carbon Fee & Dividend at about 10%, despite the fact that 70% of households are expected to directly benefit: only 14% of households think that they will. Only 12% of expected winners who received an informational intervention positively updated their initial view of the policy. In our sample, treatments in which a Fee & Dividend proposal included supplemental sentences to assist in assessing the policy had similarly small effects, increasing the proportion supporting the policy by at most 3 percentage points.

By quantifying cognitive barriers to rational policy acceptance and designing a

[5]Interestingly, evidence in the lab and the field suggest that experience with externality pricing improves support (Janusch et al., 2021; Vora and Zappala, 2025).

policy engineered to overcome these barriers, we contribute to an emerging literature that considers political economy through a behavioral lens. Other papers leverage controlled experiments to characterize how individuals understand and select policies. Among the many important findings from this literature are that consumers prefer taxes applied to a producer than to themselves, even though the tax is passed through(Sausgruber and Tyran, 2005, 2011); policy labels matter and revenue recycling can improve policy acceptability (Kallbekken et al., 2011); subsidies are generally preferred to taxes and smaller policy magnitudes are better tolerated(Cherry et al., 2012); repetition, feedback and learning can correct aversion to corrective taxes (Cherry et al., 2014); and when choosing among games, equilibria are not especially salient (Dal Bó et al., 2018). The experimental component of our project builds on this work by attempting to identify and quantify underlying behavioral mechanisms that drive preferences for tax and subsidy policies.

Documenting strong support for Penalty & Reward suggests that people do not necessarily oppose carbon pricing, but that frames matter. A small amount of complexity can play an outsize role in the presence of cognitive and attentional constraints. Addressing these bounds to rationality in policy design can reduce gaps between people's actual priorities and the policies they support.

The paper is organized as follows. Section 2 details possible mechanisms for biased policy evaluation. Section 3 details the design and results of the Diagnostic Experiment to generate a data set of abstract problem-solving behavior that can narrow down, among possible mechanisms, a primary driver of policy selection mistakes. That primary driver, NoDE, motivates the a new policy design, Penalty & Reward, introduced in Section 4, that is isomorphic to Fee & Dividend and robust to NoDE. Section 5 details a survey experiment to elicit the popularity of Penalty & Reward and Fee & Dividend in a national sample, while gathering explanatory variables for policy support including cognitive factors. Section 6 presents the results of the survey; Section 7 elaborates on the implications of these results; and Section 8 concludes.

2 Cognitive biases in policy evaluation

In this section, we detail computational errors that can systematically bias evaluation of a policy like Fee & Dividend. We are careful to separate these from informational imperfections. A voter with perfect information about everyone's demand function can fail to use that information properly. For example, a voter may misjudge the size of what he loses in consumer surplus or what he gains from the dividend. For another example, he may fail to properly account for the private welfare impact of changes to consumption that follow a policy introduction. Ex ante, we don't know where reasoning may break down and therefore cast a broad net. Collecting data of policy selection decisions while carefully varying the scope for many potential mistakes could narrow down a list of possible errors to a mistake of highest importance in political outcomes.

To best distinguish cognitive factors from informational ones, we present a framework with perfect information: voters assess a proposal with full knowledge of the market price for carbon, their own carbon demand function, and the aggregate carbon demand function.[6] Yet assessment requires attending to the information and calculating net impact. A voter i has a demand function for carbon (i.e. fossil fuels), $q_i(p)$, and inverse demand function $p_i(q)$. Aggregate demand is $Q(p)$.[7] Let p_0 represent the price at present and p_τ represent the price after the introduction of a proposed tax τ. Voters belong to a very large population of size N. We can begin with the rational voter's estimation of the change to her own consumer surplus:

$$CS_{i,\tau} - CS_{i,0} = \int_0^{q(p_\tau)} p_i(q) - p_\tau \; dq \; - \int_0^{q(p_0)} p_i(q) - p_0 \; dq = - \int_{p_0}^{p_\tau} q_i(p) \; dp$$

Next we consider the welfare effect resulting from the introduction of a dividend,

[6] For simplicity, we abstract from demand for various fossil fuels by simply referring to a market for carbon; we assume carbon is competitively produced at price p, and thus $\tau = p_\tau - p_0$; utility is quasilinear with respect to carbon. After voting on a proposal, it is either enacted or the status quo is maintained. Then, with full knowledge of the proposal's success or failure, voters make consumption choices.

[7] Throughout the paper we assume demand functions that are continuous and nonincreasing in p.

D, paid from the tax revenue:

$$D_{i,\tau} - D_{i,0} = \frac{\tau\ Q(p_\tau)}{N}\ -\ 0$$

This results in the near-term impact of the policy, which does not include a voter's private value for the policy's reduction to climate externalities:

$$U_i^{FD} = -\int_{p_0}^{p_\tau} q_i(p)dp \quad + \quad \frac{\tau\ Q(p_\tau)}{N}$$

For now, we analyze the voter's decision problem considering only near-term consequences, which the literature points to as a key driver of negative carbon policy attitudes. For the large subset of voters unconcerned about climate change, for whatever reason, the near-term assessment represents their entire assessment problem. Later, we will discuss the inclusion of climate externalities to the assessment problem.

2.1 Assessing a Policy for the Near Term

The impact to consumer surplus of a tax (compensating variation) results in a negative **Consumer Effect**, or $-\int_{p_0}^{p_\tau} q_i(p)dp$. In contrast, the policy's impact on the distribution of tax revenue, which we will call the **Distributed Effect**, is positive, $+\frac{\tau Q(p_\tau)}{N}$. When considering a broader range of policies, Consumer and Distributed Effects can take either sign, as we will soon discuss.

A voter might estimate these two effects with cognitive biases. For example, the Consumer Effect may seem more visible, familiar, or consequential to a voter than Distributed Effects. The fact that Distributed Effects accrue to a voter passively through the aggregate behavior of other consumers may cause it to be discounted and made peripheral to a voter's focus. Behavioral economics in recent years has documented many instances of imperfect salience and difficulties with hypothetical thinking, both of which could play a role in voter assessment of the Distributed Effect. While a tax's Consumer Effect is salient and painful to all, its Distributed Effect could seem remote and difficult to quantify without experience, and a voter could misjudge the relative and absolute sizes of these two typically opposing private welfare effects.

Another key mistake that voters might make is to fail to account for how a policy introduction will change consumption. For example, a failure to reason hypothetically could cause a voter to evaluate a policy's impact as if the quantity demanded would be unaffected by the policy's introduction or cause her to account for less than the full magnitude of change.[8] We will refer to the calculation of a policy's welfare impacts, holding quantity demanded fixed, as the Direct Effect. We call the welfare adjustment that accounts for response of consumption to new prices as the Equilibrium Effect, borrowing the term from Dal Bó et al. (2018). The near-term value for Fee & Dividend:

$$U_i^{FD} = -\tau q_i(p_0) \;+\; \int_{p_0}^{p_\tau} q_i(p_0) - q_i(p)dp \quad + \;\frac{\tau\; Q(p_0)}{N} \;-\; \frac{\tau(Q(p_0) - Q(p_\tau))}{N}$$

The first term, $-\tau q_i(p_0)$, is the **Direct Effect**, the policy's impact on a voter's consumer surplus holding her consumption choices fixed. The next term, $\int_{p_0}^{p_\tau} q_i(p_0) - q_i(p)\, dp$, is the **Equilibrium Effect**, which corrects the Direct Effect to account for the voter's best consumption response. A tax raises prices, making some of the voter's carbon consumption unprofitable to continue. An Equilibrium Effect removes the losses incurred from unprofitable (and unlikely) purchases that would have occurred if the voter fixed consumption at the level prior to the policy's introduction. The Direct Effect and the Equilibrium Effect together compose the Consumer Effect. A failure of equilibrium thinking could also affect accurate estimation of the policy's Distributed Effect, which results almost entirely from others' consumption behaviors. The term $\frac{\tau\; Q(p_0)}{N}$ is the **Distributed Direct Effect**, or the dividend that would result from holding aggregate quantity demanded fixed. The term $\frac{\tau(Q(p_0)-Q(p_\tau))}{N}$ represents the **Distributed Equilibrium Effect**, the adjustment to the Distributed Effect that reflects the shift in aggregate consumption attributable to the tax.

In tax (and subsidy) policies, the elements described above are additively separa-

[8] An elastic consumer who buys 20 gallons of gas a week, when asked to consider the minimum weekly payment they would be willing to accept in exchange for a permanent $1 increase in price might reflexively (and often mistakenly) think $20.

ble. For a more detailed discussion of these four key effects, see Appendix A.1.

$$U^{\Delta\tau} = \underbrace{\textit{Direct Effect} + \textit{Equil Effect}}_{\textit{Consumer Effect}} + \underbrace{\textit{Dist Direct Effect} + \textit{Dist Equil Effect}}_{\textit{Distributed Effect}}$$

Because of the errors that we have laid out, voters could ignore or misweight the four key effects, resulting in a potentially inaccurate value, $\hat{V}$, for the policy. For a rational voter, ι and λ in Equation 1 and κ, η, ν, and θ in Equation 2 would equal 1; for a biased voter, some or all coefficients would deviate from 1.

$$\hat{U}^{\Delta\tau} = \iota \textit{ Consumer Effect} + \lambda \textit{ Distributed Effect} \tag{1}$$
$$\hat{U}^{\Delta\tau} = \kappa \textit{ Direct Eff} + \eta \textit{ Equil Eff} + \nu \textit{ Dist Direct Eff} + \theta \textit{ Dist Equil Eff} \tag{2}$$

A data set of voters' choices across a sufficiently varied set of policy scenarios would allow us to identify any such biased weighting.

There are other cognitive phenomena that may bias policy selection. A voter may find the presence of a tax or subsidy complex, wish to avoid the cognitive cost of evaluating it, and thus reject it: we call this **Policy Aversion** and its opposite, Policy Bias. A voter may be averse to deviations from the status quo, resulting in **Default Bias** (or its opposite, default aversion). Under another cognitive phenomenon, **Loss Aversion With Mental Accounts**, or LAMA, voters exaggerate the policy components that affect them negatively. This biases cost-benefit analysis by overweighting a policy's true costs. For a policy proposal with a tax (like Fee & Dividend), the Consumer Effect is negative and the Distributed Effect positive, resulting in a weight larger than 1 to the Consumer Effect, in that instance. For a subsidy proposal (or a proposal to lower a tax), the Distributed Effect is negative and would get a subjective weight larger than 1. These misweightings can lead to mistakes. We provide a deeper discussion of LAMA and its potential role in policy selection in Appendix section A.3.

The literature highlights pessimism about near-term household impact as a driver of Fee & Dividend rejection, and associates pessimism with ideology. However, if cognitive errors pertaining to Consumer Effects, Distributed Effects, Equilibrium Effects, Policy Aversion, Default Bias or LAMA play an important role, it may be possible to use that knowledge to improve the accuracy or policy perceptions.

2.2 Climate Externalities

Carbon taxes also produce private welfare impacts in the long run through the reduction in climate externalities. Unlike the calculation of immediate financial impact, it is difficult to determine whether a voter's circumstances, beliefs, or risk, time, and social preferences could rationalize her estimate of a proposal's "Climate Effect" on welfare.

We now add a term, $\phi_i(Q(p_0) - Q(p_\tau)$, to the value, where ϕ_i is a voter's private benefit from reduction to climate externalities, for each unit of carbon reduced.

$$U_i^{FD} = -\int_{p_0}^{p_\tau} q_i(p)dp \quad + \quad \frac{\tau \; Q(p_\tau)}{N} \; + \; \phi_i(Q(p_0) - Q(p_\tau))$$

Climate Effects share many attributes of Distributed Effects, and may be subject to similar discounting or neglect. To a voter, they both result (almost entirely) from the consumption choices of others and may be less salient or familiar. Regardless, for any $\tau \geq 0$ and $\phi_i \geq 0$, the term $\phi_i(Q(p_0) - Q(p_\tau) \geq 0$. Thus, when a voter rejects a financially beneficial carbon tax, the private total cost of her selection mistake is at least as large as the cost of her near-term financial mistake.

3 Diagnostic Experiment

A voter's choice can appear to conflict with the voter's best interests. Could such choices be driven by mistakes in reasoning about policy? The Diagnostic Experiment is designed to (1) study whether people make mistakes in abstract choice problems in the lab that test the reasoning required for policy selection in the real world; and (2) test which of the potential mechanisms discussed in Section 2 drive selection mistakes.

3.1 Design

We create a simple decision environment capable of generating a rich data set in which potential cognitive errors described in Section 2 can be identified and quantified.

This requires sufficient variation in economic environments and policy options. Each subject encounters 30 of these variations. For details, see Appendix B.4 and B.5.

In the decision environment, there is a fictional good, referred to as "objects." The design uses real resale values for these fictional objects to induce preferences, shown as a demand function presented to participants in the experiment as a demand function, on the left of their computer screens. Figure 1 shows an example. Participants are asked to imagine an economy in which they are the sole human buyer of the good, interacting with 1,000 robot buyers. Each robot has its own preferences, shown to subjects on the right of their screens, as in Figure 1. The 1,000 robots have identical preferences. An experimental currency unit, "francs," is used to scale the actual, dollar incentives to larger, whole numbers.[9]

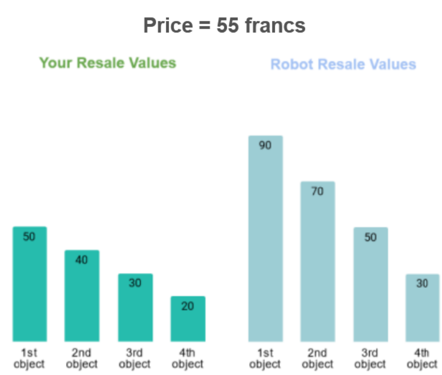


Figure 1: An example of a status-quo environment with prices and resale values.

The supply of objects is inexhaustible with an exogenous base price. Any purchaser, human or robot, can buy up to four objects. Purchaser budgets are always set sufficiently high to never constrain a human or robot's capacity to buy that many objects. Subjects must complete comprehension checks to ensure they understand how to determine privately optimal purchasing behavior for purchasers in this environment.

Subjects are then informed that all purchasing decisions will be made automatically for them and for the robots to maximize each purchaser's profit from purchasing objects, taking into account price and any other relevant facts in the environment.

[9] Twenty francs exchange for $1. Resale is described as happening automatically after purchase.

Hence, no privately harmful purchasing mistake can possibly be made by any purchaser, whether human or robot.

Subjects are told that the only choice the subject (human) must make is a "Feature Choice", phrasing that is chosen to avoid evoking any public policy connotations to what is essentially a policy choice. A subject knows that she is the sole voter in this environment and that the "Feature"(Policy) will be in place and taken into account when automatic purchasing choices are made on behalf of each purchaser.

A "Feature Choice," which we refer to going forward as Policy Choice, will mirror a tax or subsidy, with a lump-sum offset that preserves the proposal's revenue neutrality.[10]

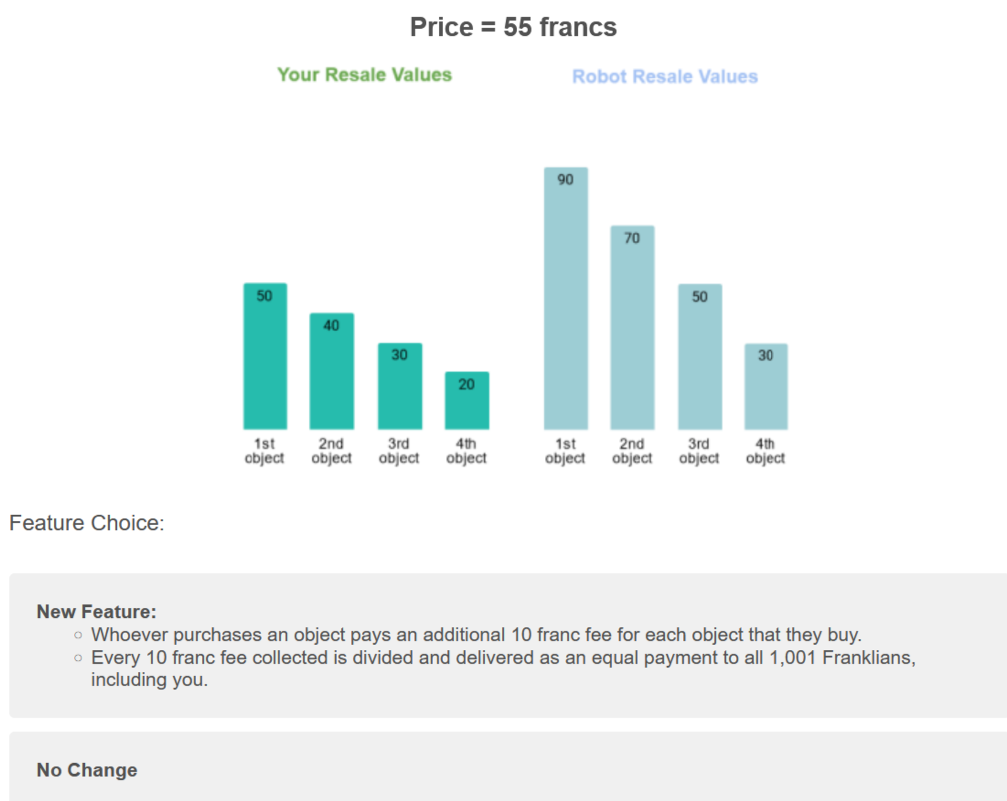


Figure 2: Consider this simple example involving a tax-and-dividend proposal. It results in no loss of consumer surplus, and a dividend gain via recycled aggregate tax revenue of 20 francs. Therefore, the optimal choice is to "Accept."

[10]Taxes are offset with positive dividends, subsidies with negative dividends. Uncertainty, temporal factors, trust and impact on other humans are absent. This simplifies policy selection into a perfect-information computation task. Verbiage of policy, politics, or pollution, too, are absent.

The Policy Choice page always comes after subjects have viewed details of the environment, and these (status-quo) details are visible while the subject makes the Policy Choice. Importantly, the Policy Choice varies by round, and does not always involve a choice of introducing a Policy. The Policy Choice can be either: (1) to accept or reject a new proposed Policy, or (2) to accept or reject a proposed reversal of an existing Policy. The value to the subject of the Policy Choice is contingent on the automatic, optimal purchases that are made subsequently on behalf of each purchaser. In essence, no purchaser, human or robot, can buy an object if that purchase is privately harmful nor can she fail to buy an object if that purchase is beneficial.

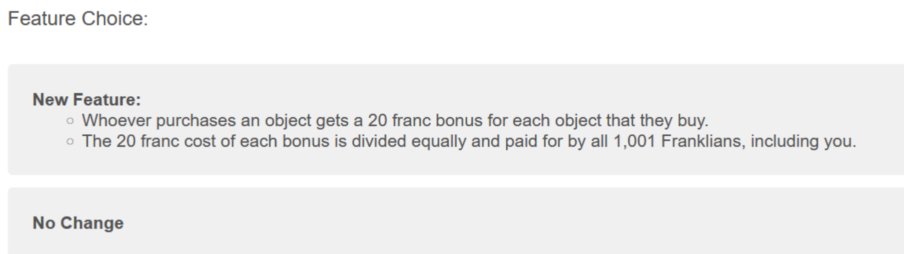
Alongside the binary Policy Choice decision, we elicit a subjective valuation for the Policy Choice decision, relative to the rejected choice, using a slider bar task. We ask subjects what it is worth to them, in francs, to get their choice, if the certain alternative is to implement the rejected choice. Subjects are told that an accurate report is in their best financial interest and that they can click a link to read details of about the incentives. The link explains the incentive compatibility and is based on the Becker-DeGroot-Marshak mechanism.[11]

The primary mistake that a subject can make is to choose a dominated option in the binary Policy Choice. A secondary mistake is to misvalue a policy. All subjects who err in the binary Policy Choice automatically err in subjective policy valuation. However, a subject who makes a wrong Policy Choice and values their choice with a high number of francs is "more wrong" than a subject who makes a wrong Policy Choice and values their choice with a smaller number of francs.

For the 30 randomly assigned rounds in which subjects participate, we vary the scope for errors pertaining to Consumer Effects, Distributed Effects, Equilibrium Effects, Distributed Equilibrium Effects, LAMA Bias, Policy Bias, and Default Bias. This is accomplished by varying parameters by round, including policy type (tax and

[11]In these details, subjects are told that their vote will be implemented with 95% probability, but that with 5% probability the rejected option will be implemented and the subject will be given a 200 franc supplement to their budget. A number between 0 and 200 will be drawn, and if that number exceeds their valuation, they keep the 200 francs and the rejected Policy Choice is implemented. If the drawn number is beneath their valuation, the value of the drawn number is deducted from their 200 francs supplement, and the subject's Policy Choice is implemented.

subsidy), policy magnitudes, status-quo conditions, prices and demand functions. Subjects are aware that one round will be randomly chosen to determine their pay. See Appendix section B.5, for full details about the experimental parameters, by round. For implementation details see Appendix section B.2.


Feature Choice:

New Feature:
- Whoever purchases an object gets a 20 franc bonus for each object that they buy.
- The 20 franc cost of each bonus is divided equally and paid for by all 1,001 Franklians, including you.

No Change

Figure 3: An example of a subsidy policy proposal.

Including subsidy proposals, an example of which is shown in Figure 3, is valuable for a number of reasons. Discounting the Distributed Effect can result in the rejection of privately optimal taxes. However, Policy Aversion and LAMA could each produce rejection of taxes as well. Recall that policy-averse voters tend to reject policies because they do not wish to assess their welfare implications whereas LAMA voters tend to reject policies because they exaggerate the costs of any policy relative to the benefits. Including subsidy proposals in many rounds allows for instances that separate the behavioral responses of subjects who discount Distributed Effects from those who are policy averse or exhibit LAMA.

3.2 Measuring behavioral biases

Using subjective valuations of policy proposals, we can estimate the weights in Equation 1, along with the other cognitive factors:

$$\begin{aligned} U_{ij} = \alpha \ + \ \beta_1 \ \textit{Consumer Effect}_j \ + \ \beta_2 \ \textit{Distributed Effect}_j \\ + \gamma_1 \ \textit{Neg. Term Magn}_j + \gamma_2 \ \textit{Policy Magn}_j + \gamma_3 \ \textit{Dev. Magn}_j + \epsilon_{ij} \end{aligned} \tag{3}$$

Similarly, we can perform a regression to assess the subjective weightings on subeffects shown in Equation 2:

$$
\begin{aligned}
U_{ij} = \alpha + \beta_1 Dir\ Eff_j + \beta_2 Equil\ Eff_j + \beta_3 Dist\ Dir\ Eff_j + \beta_4 Dist\ Equil\ Eff_j \\
+ \gamma_1\ Neg.\ Term\ Magn_j + \gamma_2\ Policy\ Magn_j + \gamma_3\ Dev.\ Magn_j + \epsilon_{ij}
\end{aligned}
\tag{4}
$$

In Equations 3 and 4, we compare the β coefficients to the rational benchmark of 1 and the γ coefficients (Negative Term Magnitude, Policy Magnitude, and Deviation Magnitude) to a rational benchmark of 0. Estimating the cognitive factors together and comparing them to their rational benchmarks will reveal which factors are more consequential. The degree of deviation from the rational benchmark corresponds proportionally to money that subjects failed to capture, in expectation.[12]

The willingness-to-pay responses allow us to measure how subjects utilize the experiment round's key parameters to form a subjective valuation for the policy proposal. Below, we assign behavioral weights to evaluation elements detailed in Section 2. Due to additive separability in the two-parameter case (Consumer Effect and Distributed Effect) and the four-parameter case (Direct Effect, Equilibrium Effect, Distributed Direct Effect, and Distributed Equilibrium Effect), expressed in Equations 1 and 2 of the previous section, the marginal impact of any of these four effects on a voter's subjective value is the behavioral weight. Therefore a linear regression of participant i's value for proposal j on the four objective effect parameters of proposal j will return an unbiased estimator of the subjective weights, given a sufficiently large set of proposals J that are linearly independent. Deviation from the rational benchmark of 1 on the estimated coefficient for the parameters represents a misweighting.[13]

Among the cognitive mechanisms that must be assessed is Loss Aversion with Mental Accounts, or LAMA, when a voter aggregates a perceived total effect of a policy move with he exaggeration to components that have negative welfare effects. It might be that the Consumer Effect of a proposal is negative, as is commonly the case with proposals to introduce a tax or remove a subsidy. It might be that the

[12]Across the rounds that subjects can encounter, some rounds lack any scope for Equilibrium Effects. In other words, the proposal does not change any consumer's best consumption response. Estimating Equation 3 solely on these rounds offers another strategy for isolating weightings on Consumer and Distributed Effects without confounding Equilibrium Effects.

[13]We cluster observations at the subject level.

Distributed Effect is negative, as is commonly the case when the proposal introduces a subsidy or removes a tax. To exaggerate costs when summing the costs and benefits of a proposal is distinct from the consistent misweighting of the Consumer Effect or Distributed Effect.[14] (Recall that NoDE is a consistent misweighting of Distributed Effects). The regression equation below can measure and control for LAMA, absorbing the systematic exaggeration of costs, by including as an explanatory parameter the magnitude of either of the two effects that represent costs in that round. This variable is called "Negative Term Magnitude," and estimating a nonzero coefficient on it represents deviation from rationality.

To capture Default Bias or Default Aversion (where subjects systematically avoid or prefer change from the status quo) the regression includes a control term that we call "Deviation Magnitude,"($\mid \tau_1 - \tau_0 \mid$) that measures in absolute terms the amount of change that is being proposed in the Policy Choice. Similarly, a term called "Policy Magnitude"($\mid \tau_1 \mid - \mid \tau_0 \mid$) is included to control for Policy Bias or Policy Aversion.[15]

Neglect of Distributed Effects can be measured nonparametrically by evaluating the binary feature choices made by subjects. Correct choices will be harder to make for subjects with NoDE when the magnitude of the Distributed Effect (DE) exceeds the magnitude of the Consumer Effect (CE). Comparing correctness rates when $\mid DE \mid > \mid CE \mid$ to when $\mid CE \mid \geq \mid DE \mid$ or the overall correctness rate can illustrate whether NoDE plays a role in policy selection.

3.3 Results and Discussion

We find that subjects frequently make mistakes in policy decisions, even in this simplified and incentivized perfect-information setting. Of binary choices, 26% were dominated. In rounds in which the magnitude of Distributed Effects exceeded the magnitude of Consumer Effects, 39.7% of binary choice responses were dominated; for rounds in which Distributed Effects were smaller than Consumer Effects, 20.7%

[14]LAMA is conditional overweighting on either the Consumer Effect and the Distributed Effect if its sign is negative.

[15]Subsidies are represented as taxes(τ) with negative values.

were dominated.[16] Thus, rounds with high scope for selection mistakes due to discounting of Distributed Effects had twice the error rate of rounds with low scope for such mistakes.

In contrast, in rounds with high scope for errors driven by discounting of Equilibrium Effects, mistake rates were only slightly higher than the average rate across all rounds. When Equilibrium Effects exceed Direct Effects, the binary choice error rate was 28%, versus 26% over all rounds. Equilibrium Effects appear to play secondary role in generating policy selection and valuation mistakes, which is confirmed in the regression results.

Table 1 distills the extent of deviation from rational benchmarks, by behavioral factor, in policy valuation responses, and is drawn from the regression results in Appendix Tables 4 and 5. It clearly demonstrates the outsize role of Distributed Effects in generating valuation errors. Each dollar of Equilibrium Effects is undervalued by 22%, on average (column 3). Loss Aversion with Mental Accounts (LAMA) mistakes generate similar financial costs as Equilibrium Effects for subjects (column 3). More than twice those losses are generated by errors undervaluing Distributed Effects, which are discounted by 59% when separate from Equilibrium Effects. An alternative way of calculating the weight on Distributed Effects, independent of Equilibrium Effects, is shown in column 1, in which the weighting on Distributed Effects is calculated in the subset of problems that lack scope for Equilibrium Effects. There, we see a similar magnitude of discounting by half (-0.500) relative to the rational benchmark.[17] Note that, in contrast, Consumer Effects and Direct Effects are valued consistently at near the rational benchmark.

On an individual subject level, the distribution of the coefficient on the Distributed

---

[16]Random guessing would produce an expected 50% correctness rate. A simple heuristic would work well for the 87% of rounds in which there are distinct high demand and low demand types for the human subject and robot counterparts. Among these questions, a heuristic in which a low-demand type rejects subsidies and accepts taxes and a high-demand type accepts subsidies and rejects taxes produces dominant binary choices 94% of the time, or 81% of the total available rounds.

[17]In column 3, we can see full neglect, on average, for Distributed Equilibrium Effects. This discount is moderately larger than the effect we might expect if the discounting of Distributed Effects and Equilibrium Effects were independent of one another ($\beta_4 < \beta_2 * \beta_3$). The welfare implications of the full neglect of this term are ambiguous in the presence of the other discounting, but would costly when voters encounter subsidy proposals. It combines two effects we are interested in independently, and therefore we focus on measures of the effects when independent of each other.

Effect reflects wide variation in Appendix Figure 8, with most subject responses reflecting a weight of between 0 and the rational benchmark of 1.

Table 1: Deviation from rational benchmarks for subjective valuations

	Deviation from rational benchmark		
	Equil Eff Absent	All Rounds	All Rounds
Consumer Effect	0.113	-0.134	
Direct Effect			0.049
Equilibrium Effect			-0.221^{*}
Distributed Effect	-0.500^{***}	-0.790^{***}	
Distributed Direct Effect			-0.590^{***}
Distributed Equil Effect			-1.011^{***}
LAMA	-0.043	-0.258^{**}	-0.225^{**}
Policy Bias	0.187	-0.200^{**}	0.042
Default Bias	-0.427	-0.227	-0.210

Notes: Columns 1 and 2 above result from estimating Equation 3 (first on the subset of rounds where Equilibrium Effects are absent, second on the full set of rounds) and are drawn from Columns 4 and 2 of Appendix Table 4. Column 3 results from Estimating Equation 4 across all rounds and is from Appendix Table 5 Column 2. Given the regression specifications, the rational benchmark for Consumer Effect and Distributed Effect and their subeffects is 1. The rational benchmark for LAMA, Policy Bias, and Default Bias is 0. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Across all specifications, the fact that Distributed Effects are weighted meaningfully lower than Consumer Effects indicates that the dominated binary choices characterized earlier in this section are to be expected. These mistakes resemble the stylized policy attitudes documented in the policy survey literature: taxes are under-favored and subsidies are over-favored (see Section 1).[18] Having removed preference confounds and controlled for alternative behavioral factors, the results suggest attentional and computational explanations have a role to play in puzzling voting behavior out in the world, most significantly via Neglect of Distributed Effects, or NoDE. For the remainder of the paper, NoDE will be the primary cognitive factor of interest.

[18]The implications of this finding likely extend beyond preference for the special case studied here, taxes and subsidies with uniformly distributed offsets as in Fee & Dividend. Neglecting Distributed Effects could be a cognitive element that underlies fiscal illusion (Puviani, 1903; Buchanan and Wagner, 2013; Dell'Anno and Mourao, 2012); consumption responses to tax changes that deviate from Ricardian equivalence (Bernheim, 1987); and preference for predistributive policies over redistributive policies (Kuziemko et al., 2023).

4 Penalty & Reward: Theoretical Properties

We design a policy, Penalty & Reward, that removes the scope for neglect of Distributed Effects, or NoDE. First, we present a behavioral model for NoDE that can explain the dominant mistake in the Diagnostic Experiment and motivates the design of Penalty & Reward. Next, we present the Penalty & Reward policy and explain why it is outcome equivalent, in theory, to Fee & Dividend. Lastly, we detail how a voter with NoDE would perceive Penalty & Reward and Fee & Dividend differently.

4.1 Neglect of Distributed Effects (NoDE)

A voter i evaluates a tax (or subsidy) proposal as follows:

$$\hat{U}_i^{\Delta\tau} \; = \; \textit{Consumer Effect} \; + \; \lambda_i \; \textit{Distributed Effect}$$

where $\lambda_i = 1$ for a rational voter, and $0 \leq \lambda_i < 1$ for a NoDE voter.

A NoDE voter would partially or completely discount the value of a dividend when evaluating a Fee & Dividend proposal.

4.2 Penalty & Reward: Definition

The Penalty & Reward policy imposes (i) a linear penalty on each unit consumed above a threshold in a period, for any consumer whose total consumption exceeds the threshold, and (ii) a linear reward on the difference between the threshold and the quantity consumed for any consumer whose total consumption is beneath the threshold. The level of the threshold can change from one time period to another.

4.3 Demand and welfare equivalence to Fee & Dividend

In this section we will show that, when the penalty and reward rates are identical to tax rates of Fee & Dividend, and when the Penalty & Reward threshold is set at the population mean level of consumption for the period, these two policies theoretically result in equivalent demand and welfare.

A consumer's utility function under a general policy that taxes carbon at rate τ and provides a lump-sum payment D is

$$U_i^{FD}(q_{carb}, q_{clean}) = u_i(q_{carb}, q_{clean}) + m_i - p_{carb}q_{carb} - p_{clean}q_{clean} \\ \underbrace{-\ \tau q_{carb}\ +\ D}_{\text{Tax \& Payment}}$$

The consumer maximizes utility with respect to two goods, q_{carb} and q_{clean}, taking endowment m_i, prices p_{carb} and p_{clean}, tax rate τ, and payment D as given. Fee & Dividend is an instance of such a policy, where D is an equal per capita share of the aggregate revenue.

Under Penalty & Reward, a consumer faces a similar maximization problem, taking m_i, p_{carb}, p_{clean}, a Penalty & Reward threshold, A, and penalty rate γ and reward rate γ as given. The consumer's utility is

$$U_i^{PR}(q_{carb}, q_{clean}) = u_i(q_{carb}, q_{clean}) + m_i - p_{carb}q_{carb} - p_{clean}q_{clean} \\ \underbrace{-\ \gamma\ (\mathbf{1}_{q_{carb}>A})(q_{carb} - A)\ +\ \gamma\ (1 - \mathbf{1}_{q_{carb}>A})(A - q_{carb})}_{\text{Penalty \& Reward}}$$

$$= U_i(q_{carb}, q_{clean}) + m_i - p_{carb}q_{carb} - p_{clean}q_{clean} \\ - \gamma q_{carb} +\ \gamma A$$

For any tax rate τ and lump-sum payment D in a Fee & Dividend regime, when the Penalty & Reward rate $\gamma = \tau$ and the Penalty & Reward threshold $A = \frac{D}{\tau}$, consumer utility functions under Penalty & Reward and Fee & Dividend are equivalent. Thus demand $(q^*_{carb}, q^*_{clean})_i = \underset{q_{carb}, q_{clean}}{argmax}\ u_i(q_{carb}, q_{clean})$ and indirect utility $V_i(m_i, p_{carb}, p_{clean})$ are equivalent under both policies.[19] [20]

[19]The equivalence of demand for each consumer implies equivalence of aggregate demand and aggregate damage from externalities. We've demonstrated this as a partial equilibrium result, taking prices as given under each policy, but it extends to general equilibrium. The policy, as framed here, would have the equivalent effect on supply as Fee & Dividend. By showing that demand is equivalent under each policy, prices, too, must be equivalent for any aggregate supply function, assuming a unique equilibrium. Also, a variation of Fee & Dividend that taxes producers instead of consumers should have the equivalent impact on equilibrium prices, demand or welfare, due to equivalence of incidence.

[20]Doing this welfare comparison using a static model requires the assumption of no lags between fees and dividends, or that such lags are accounted for by increasing dividends. If dividends can be

When $D = \tau\,\bar{q}_{carb}$, where $\bar{q}_{carb} \equiv \frac{\sum_i q_{carb,i}}{N}$, the mean level of consumption across the population, we have the Carbon Fee & Dividend policy, with its property of revenue neutrality. The threshold for the equivalent Penalty & Reward policy is $A = \bar{q}_{carb}$, which is therefore also revenue neutral. The sum of transfers for each policy can be expressed as follows:

$$\sum_i (-\ \tau q_{carb,i}\ +\ D) = \tau \sum_i (-\ q_{carb,i}\ +\ \bar{q}_{carb}) = 0$$

$$\sum_i (-\ \gamma q_{carb,i} +\ \gamma A) = \gamma \sum_i (-\ q_{carb,i} +\ \bar{q}_{carb}) = 0$$

For either of the two proposals, a policy choice might be made with uncertainty about the consumption choices of others. This uncertainty would then apply directly to the size of the dividend in Fee & Dividend. Under Penalty & Reward, this uncertainty would apply to the consumption threshold required to maintain revenue neutrality. (The threshold separates fee payers from reward receivers.) In both Fee & Dividend and Penalty & Reward, the uncertainty pertains only to the level of $\bar{q}_{carb}$, and is thus identical under both regimes. Therefore, if aggregate consumption information is imperfect, risk aversion will not result in any differences between these two policies in incentives, quantities demanded, or welfare.

4.4 Behavioral Distinctions in How Policies Are Perceived

A voter must evaluate how a proposed policy will change his or her welfare, to determine whether and how intensely to support the proposal. Proposals can have Consumer Effects, which are more direct and transparent to voters, and Distributed Effects, which are less direct and opaque to NoDE voters. Below we revisit the framework from Section 2 for policy valuation.

As shown earlier in this section, the unbiased value of both the Penalty & Reward and Fee & Dividend policies are equivalent. However, Penalty & Reward consists of

paid arbitrarily frequently, for instance, we have no need to consider how a lag between first-period tax payments and first-period dividend could result in a welfare difference than if tax payments and dividends occurred simultaneously. In the survey, we address this by offering an initial prepaid expected dividend, which should remove any potential advantage of Penalty & Reward.

only transparent effects, lacking a mechanism that produces any Distributed Effect. Therefore, a NoDE voter evaluating a Penalty/Reward policy would come to the same value as an equivalent non-NoDE counterpart. We denote the subjective value of a NoDE voter i with $\tilde{V}_i$.

$$U_i^{FD} = U_i^{PR} = \tilde{U}_i^{PR} = -\left(\int_{p_0}^{p_\tau} q_i(p)dp\right) \ + \ \frac{\tau \ Q(p_\tau)}{N} \ + \ \phi_i(Q(p_0) - Q(p_\tau))$$

The first term represents the change in consumer surplus, with $q_i(p)$ as the individual demand function for carbon. The middle term is the dividend, with $Q(p)$ representing the aggregate demand function for carbon. The final term is the private benefit from the change in the externality, with ϕ_i representing an individual's private damage per unit damage. Prices p_0 and p_τ correspond to effective prices of carbon without taxes and with tax rate τ, respectively.

In comparison to the transparency of Penalty & Reward, Fee & Divdend consisting of a direct, familiar Consumer Effect and a relatively opaque Distributed Effect. As a result, aggregating these two effects when evaluating Fee & Dividend, a NoDE voter discounts the Distributed Effect.

$$\tilde{U}_i^{FD} = -\left(\int_{p_0}^{p_\tau} q_i(p)dp\right) \ + \ \lambda_i\frac{\tau \ Q(p_\tau)}{N} \ + \ \phi_i(Q(p_0) - Q(p_\tau))$$

While the difference in the value of these two policies is 0, a NoDE voter, with $0 \leq \lambda_i < 1$, assuming that both the tax rate and resulting aggregate consumption quantity is greater than zero.

$$U_i^{PR} - U_i^{FD} = 0$$

$$\tilde{U}_i^{PR} - \tilde{U}_i^{FD} = \tau(1 - \lambda_i)(\bar{Q}(p_\tau)) > 0$$

An alternative way to describe this is that under both policies there is an actual and perceived consumption threshold above which you lose from the policy, and below which you gain from the policy. In reality, the threshold is the same under both policies. This threshold is transparent under Penalty & Reward. But under Fee & Dividend, it is opaque; for subjects with NoDE, the subjective threshold separating fee payers from reward receivers is lower, suggesting that more consumers will lose from the policy than is actually the case.

There are two alternative or additional cognitive mechanisms, False Correlation and Mental Budgeting that could result in greater support for Penalty & Reward than for Fee & Dividend that we test in the survey and are able to control for in the evaluation of our primary NoDE hypothesis. We provide background on these mistakes and detail the theoretical implications in Appendix A.4.

5 Survey-Experiment Design

The survey-experiment has two main purposes: (1) to assess whether people indeed more strongly support Penalty & Reward than Fee & Dividend; and (2) to understand whether the cognitive mechanism of NoDE can indeed help to explain any difference of popularity between the two policies.

Accomplishing this requires survey participants both to assess policies and to participate in incentivized tasks akin to those shown in the Diagnostic Experiment. We designed the survey to also include questions that test for alternative or additional mechanisms and to measure other voter characteristics that might drive policy attitudes.

We will detail the three pieces of the survey: **Policy Elicitation**, in which participants rate the two carbon policies; **Abstract Policy Task**, which tests for NoDE; and a **Questions** section that includes markers for two additional cognitive biases and other relevant factors. Afterward, we detail supplemental treatments, along with delivery and sample characteristics.

5.1 Policy Elicitation

Half of subjects see the policy section first. The order in which the two policies are displayed is randomized by individual, and that order remains constant throughout the survey, to avoid any confusion.

Please read the following policies completely. You will answer several questions about both of them.

> **Carbon fees and rebate payments:** The government would tax greenhouse gas emissions by $100 per ton. This is roughly equivalent to a $1 tax per gallon of gasoline. The government gathers the money collected from the new emissions tax and uses all of the money to send every living resident an equal rebate payment. Before the first week of the tax, one estimated rebate payment is paid to all people in advance. Then, at the end of each week, the government SSsends a rebate payment that evenly divides the week's tax collection.

> **Carbon penalties and rewards:** The government would tax greenhouse gas emissions by $100 per ton only on a person's emissions that exceed the average individual level for the week. If your emissions are above average, then the higher you emit, the more tax you pay. Also, the government rewards $100 per ton for how much below average a person emits. If your emissions are below average, the less you emit, the bigger your reward. Taxes and rewards are roughly equivalent to $1 per gallon of gasoline and paid weekly.

Participants then provide four ratings for each policy: **Overall Support**; **Emissions**, or how successful a policy will be in curbing climate emissions; **Your Finances**, or how the policy will affect their household financially; and **Low Income**, or how the policy will affect low income individuals. (For full text of each rating prompt, see Appendix Section C.1.1.) The order of these questions is randomized (as shown in Appendix Figure 10). For each rating question, subjects respond using a slider bar with an integer from -100 to +100, and the text of the policy is always visible. (For an image, see Appendix Figure 11.)

5.2 Abstract Policy Task

Using decision environments from the Diagnostic Experiment, we measure NoDE for each survey participant. Half of subjects encounter this section first. In each of 15 rounds, participants are offered a choice: to adopt a proposal or to maintain the status quo. A proposal can take one of two forms: (1) accept or reject a new proposed policy, or (2) accept or reject a proposed reversal of an existing policy.[21]

Every subject encounters the same 15 rounds, but in an order randomized by subject. The 15 rounds vary in policy type (tax or subsidy), status quo (policy or no policy), and optimal choice (accept or reject). Most importantly, the policies vary in the magnitude of the change in Consumer Effect and the Distributed Effect. This is accomplished by varying parameters across the 15 rounds, such as prices, policy proposals, demand functions assigned to the participant, and demand functions assigned to her robot counterparts.[22]

To minimize the survey length and difficulty, we do not elicit willingness to pay as we did in the lab experiment. To measure neglect of Distributed Effects, we first calculate the coefficients on the Consumer Effect and the Distributed Effect when they are used jointly as regressors, with controls, on each participant's 15 binary outcome choices.[23] Then for each participant a calculate the difference between these coefficients.[24]

$$\mathbf{1}_{ij}^{Accept} = \ \beta_{0,i} + \beta_{1,i}\ Consumer\ Effect_j \ + \ \beta_{2,i}\ Distributed\ Effect_j + X_j'\gamma_i + \epsilon_{ij}$$

$$NoDE\ Score_i = (\beta_{1,i} - \beta_{2,i}) \ * \ \kappa\ (a\ scaling\ constant)$$

That is, an individual's NoDE Score is the difference between the weights they

[21] Participants earn a \$ 1 bonus if they make the better of the policy choices in a randomly drawn round. Consistent with the Diagnostic Experiment, the Abstract Policy Task never uses the word "policy," and always uses the word "feature."

[22] For a detailed look at all 15 of the rounds, please see Appendix C.3.

[23] Controls include terms for LAMA, Policy Aversion and Default Bias, shown in Section 3.2

[24] We multiply this difference in coefficients for each subject by a fixed scaling constant $\kappa = 40$ to allow our regression coefficients to not be large.

place on the Consumer Effect and the Distributed Effect, as estimated from a linear probability model of their binary choices.

5.3 Questions section

Finally, subjects complete a questionnaire. All subjects encounter the questionnaire after the Policy Elicitation and Abstract Policy Task sections. The questionnaire includes state of residence, a demographic trait that Prolific does not provide.

There are six questions to assess factors that might predict support for redistributive climate change policies broadly: political orientation score, belief in climate science, faith in how government spends money, impact of a carbon tax on own job income, belief in appropriateness of redistribution from high earners to low earners, and personal emissions relative to the population average. See Appendix section C.1.2 for more details.

Included are two questions to test alternative or additional mechanisms that would produce differences in support between Fee & Dividend and Penalty & Reward: Mental Budgeting and False Correlation mistakes. For theoretical background, see Appendix Section A.4. With the Fee & Dividend policy displayed on the page, the Mental Budgeting question asks: "*Do you agree with the following statement? Because carbon taxes in the "Carbon fees and rebate payments" policy make fossil fuels more expensive than before, the payments that people get back from the policy will be spent on their fossil fuel budget.*," with Agree as 1 and Disgaree as 0.

Similarly, the False Correlation question asks: "*True or False? Under the policy, I should not change my behavior because the more fossil fuel tax I pay, the more I'll get back in payments.*," with True as 1 and False as 0.

5.4 Supplemental treatments

The elements discussed thus far comprise the main study. But we test variations in the Fee & Dividend policy text that add information meant to improve receptivity and have been proposed or tested in the literature.[25] The goal is to compare support

[25]Informational interventions like these have had modest or disappointing results in the past (Mildenberger et al., 2022; Douenne and Fabre, 2022).

for Penalty & Reward to support of these "assisted" versions of Fee & Dividend. [26]

The first treatment, which we refer to as COMPUTE, tries to clarify the likely financial impact of Fee & Dividend by appending this statement to the end of the policy description: *Individuals who emit below the average level will receive a higher rebate payment than what they paid in emission taxes. Individuals who emit above the average level will spend more in taxes than they will receive in the rebate payment.*

The second treatment, which we refer to as EXPERT, tries to clarify the impact of Fee & Dividend by instead adding this sentence to the standard Fee & Dividend description: *Experts say that most below-average emitters will benefit by gaining money from this arrangement and that most above-average emitters will lose money.*

5.5 Robustness to single/sequential policy presentation

Simultaneously providing two policies simultaneously offers participants an opportunity to convey relative preference or indifference between Fee & Dividend and Penalty & Reward. Policy proposals can also be presented alone, without alternative options, and for that reason we conduct a robustness study in which support is measured one policy at a time. Participants convey their support for a randomly assigned policy from the pair, either Fee & Dividend or Penalty & Reward. (While reading and answering rating questions about it, participants are unaware of an alternative policy they will be asked to rate subsequently.) A between-subjects comparison of ratings for first-shown policies is equivalent to a study in which each subject was shown only one policy.

5.6 Delivery and sample characteristics

The surveys are delivered online to 678 United States participants, 18 or older, via Prolific. Participants know nothing about the survey's topic prior to beginning it. Because political representativeness is potentially important, subjects are enrolled in a manner that delivers 40% Republicans, 40% Democrats, and 20% independents

[26] Each of the two treatments enrolled 50 participants, were designed after the collection of the baseline study results, and were not pre-registered.

for the entire study and its subsets, including each treatment and robustness sample. The final sample is representative of national age, gender, race and geographic characteristics.[27]

As measured by their Abstract Policy Task performance, participants are much likelier to show NoDE than not, and there was substantial heterogeneity. The extent of bias varies widely by individual as shown in the distribution of NoDE Scores in Appendix Figure 12. Notably, there is no meaningful correlation between NoDE Score and political orientation.[28] Overall, participants answered 56% of the binary choice questions correctly, with a 46% correctness rate when the Distributed Effect's magnitude exceeded the Consumer Effect's magnitude and a 68% correctness rate when the Consumer Effect's magnitude exceeded the Distributed Effect's magnitude. The structure of these outcomes resembles results of the Diagnostic Experiment. Approximately 13 percent of participants answered the False Correlation question incorrectly, and 60 percent answered the mental budgeting question incorrectly.

6 Survey-Experiment Results

In the Baseline study, participants overwhelmingly favored Penalty & Reward, with 71 percent of participants rating it higher than Fee & Dividend and 17 percent rating it lower. Users were asked to rate their opposition or support for a policy on a -100 to +100 scale. Their numerical response is referred to throughout as "support intensity."

[27]Prolific provides age, sex, race and party affiliation, though some participants decline to provide each detail. We supplement these details with our state-of-residence question. As shown in Appendix Figure 12, there is broad distribution of ages, with mean 42.7. Survey distribution occurred at times of day that enabled daytime participation in both Eastern, Central and Pacific time zones. There was broad geographic participation, with the 43 most populous states represented. 47.0% of subjects identify as male, and of the subjects who provided race, 63.1% identified as White, compared to 59.3% at the population level. In addition 13.8% described themselves as students, and 11.1% as unemployed. Of the 347 enrolled in the main study and two supplemental treatments the base pay rate was $4. Upon entering the study did they learned of the possibility of earning a $1 bonus through the Abstract Policy Task. Participants took on average 22 minutes and earned $4.54, on average. For the shorter robustness study, 331 subjects spent about 5 minutes and earned $1.

[28]The Pearson correlation coefficient (-0.07) between NoDE Score and Left-Right Score is not statistically meaningful (p-value of 0.167) and indicates slightly lower NoDE bias on the right. This suggests that even if voters on the right are averse to real taxes, they are very slightly likelier to support an abstract-setting tax when it is optimal than those on the left.

Figures 4 and 5 offer a deeper look at support for the policies across participants and across subgroups. Figure 4 shows that individuals across the board tend to favor Penalty & Reward over Fee & Dividend, with an average Penalty & Reward Premium – or the difference in support intensity for the two isomorphic policies – of 49 points.[29] The effect is greater among individuals showing more pronounced NoDE in the Abstract Policy Task. The Penalty & Reward Premium is 33 points larger for individuals with a NoDE score above the median (0.17) than for individuals with scores below it.

A average high-NoDE individual's rating for Fee & Dividend is -13.4, or 19.3 points below the rating given by the average low-NoDE individual. Meanwhile, the average support for Fee & Dividend among those on the left is 28.2 points higher than average support on the right. For Penalty & Reward the left-right gap is 24 points. The 19.3-point gap in support for Fee & Dividend between high-NoDE and low-NoDE individuals is more than two-thirds the size of the gap in support for Fee & Dividend between right- and left-leaning individuals. This is just one piece of evidence that cognitive factors may deserve as much attention as political identity in driving support for these policies. On average, participants on the political right support Penalty & Reward to a greater extent than participants on the left support Fee & Dividend, by a margin of 16.5 points. The difference is highly significant (Wilcoxon rank sum p-value: 0.0044).

For Figure 5, we transform baseline support intensities into a binary variable of support or non-support.[30] Of the subgroups presented, Fee & Dividend enjoys majority support only among the left-leaning and low-NoDE groups, while Penalty & Reward is supported by a solid majority of every subgroup. The patterns in the proportion measure are reflective of the mean intensity scores in Figure 4.[31]

[29]We conduct a robustness exercise on Policy Elictation, in which we present 331 participants with policies sequentially. We compared the ratings of the policy shown first, the results of which are shown in Table 12. Broadly speaking, the gap in support for the two policies is robust but moderated in size. We elaborate further in the Appendix with Table 12.

[30]We define support as providing a rating of 5 or more on the -100 to +100 scale.

[31]To formally test hypotheses, we conduct Wilcoxon rank sum tests shown in Table 11. The first column tests our pre-registered hypothesis that Penalty & Reward would be supported higher than Fee & Dividend, and the results are strongly significant. The second column is one of several tests for our second pre-registered hypothesis, that the Penalty & Reward Premium (the difference in

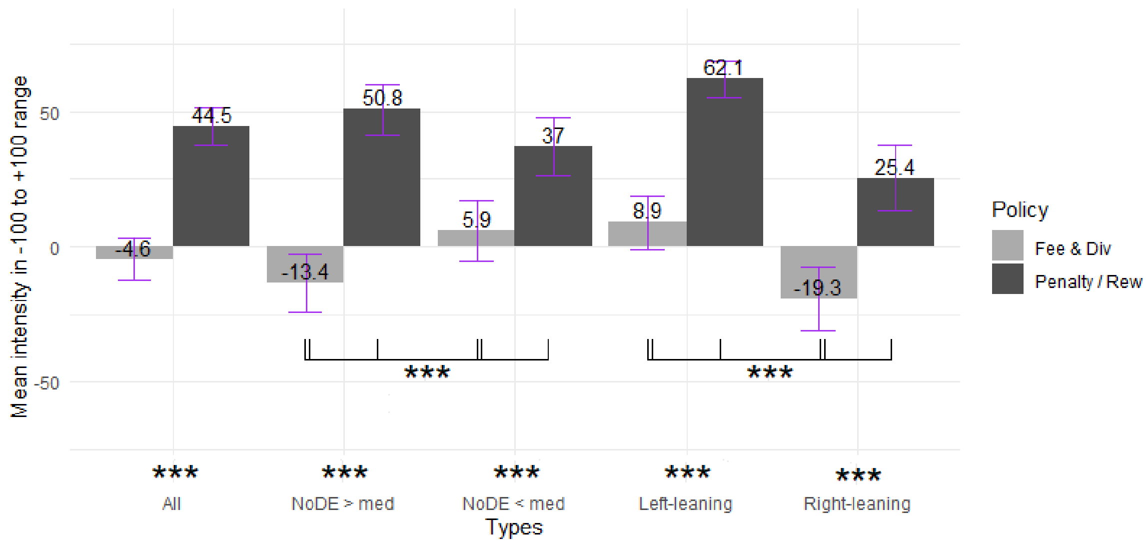


Figure 4: Mean support for policies by group in the baseline study, shown with 95 percent confidence intervals.

To test the effect of the cognitive factors on the difference of support between Penalty & Reward and Fee & Dividend, we employ the following regression:

$$\begin{aligned} Penalty\ Reward\ Premium_i = \beta_0 \ + \ &\beta_1 NoDE\ Score_i \\ &+ \beta_2 \mathbf{1}(Correl\ Error)_i + \beta_3 \mathbf{1}(Mental\ Budget)_i + \beta_4 \mathbf{1}(Exp\ First)_i \\ &+ \beta_5 \mathbf{1}(Compute\ Prompt)_i + \beta_6 \mathbf{1}(Expert\ Prompt)_i + X_i'\gamma + \epsilon_i \end{aligned} \tag{5}$$

The results shown in Table 2 suggest that cognitive factors play an important role in the Penalty & Reward Premium. Neglect of Distributed Effects (NoDE), as well as the Mental Budgeting mistakes play a consistent predictive role in the hypothesized direction. (Recall that a subject making a mental budgeting mistake believes that dividends will be allocated to carbon-intensive consumption.) A full point of additional bias, as measured in the NoDE, score would be expected to increase the Penalty & Reward Premium by an estimated 30 to 50 points, depending on the

support compared to Fee & Dividend) would be higher for above-median NoDE-score types than for for below-median types. The outcome of this test, too, is strongly significant. As mentioned earlier, we also find that the higher distribution of support for Penalty & Reward among right-leaning participants than the support of Fee & Dividend among left-leaning participants is statistically significant.

Table 2: Effect of Behavioral Factors on Differences in Policy Support

	Dependent variable: Penalty & Reward Premium		
	Support for P&R minus support for F&D		
	(1)	(2)	(3)
Constant	35.767***	4.109	−12.877
	(8.245)	(21.035)	(35.353)
NoDE Score	35.310**	29.689**	48.185***
	(14.664)	(15.028)	(17.279)
False Correlation error	14.812	16.090	19.612
	(12.267)	(11.896)	(14.010)
Mental budgeting error	16.179**	15.609**	17.880**
	(7.462)	(7.487)	(8.798)
Survey order: experiment first	−13.988*	−14.349*	−18.597*
	(7.966)	(8.092)	(9.740)
"Compute" prompt	−13.092	−16.164*	−2.544
	(9.823)	(9.793)	(11.489)
"Expert" prompt	−4.498	−0.869	1.624
	(10.248)	(11.038)	(13.486)
Controls:			
Support factors		X	X
Demographic/employment			X
Observations	347	347	253
Mean outcome	45.176	45.176	48.091
R^2	0.067	0.097	0.154
Adjusted R^2	0.050	0.062	0.089

Notes: The table above results from estimating Equation 5, which predicts Penalty & Reward premium, i.e. each subject's support for Penalty & Reward minus their support for Fee & Dividend. If Fee & Dividend is favored, the premium is negative. Column 1 estimates over the full sample with the key cognitive explanatory factors. "Compute" prompt and "Expert" prompt represent the information treatments meant to enhance Fee & Dividend that a subset of participants received. Column 2 includes a set of controls derived from answers to survey questions that would predict support for policies in general (Relative emissions, Belief in climate change, Left-right score, Belief job benefits from carbon tax, Belief job hurt from carbon tax, Trust government use of revenue, and Support for redistribution). Column 3 includes a set of demographic controls (Age, Male, Black, Student, and Unemployed) on the large subset of participants for whom all demographic details were available.Extended table: Table 8. *p<0.1; **p<0.05; ***p<0.01.

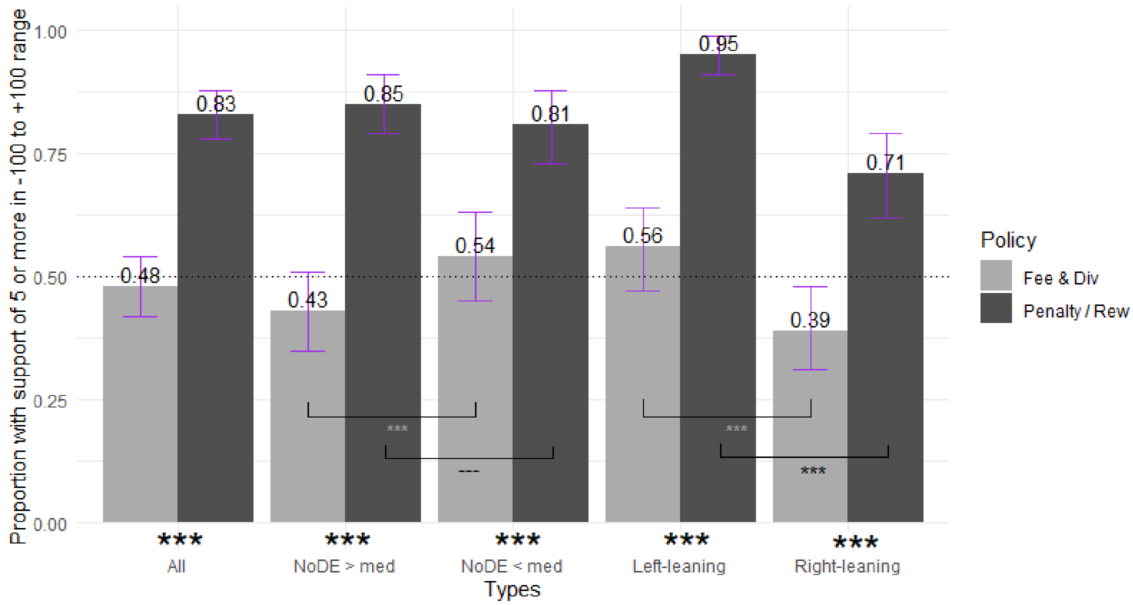


Figure 5: Proportion in support of policies by group in the baseline study, shown with 95 percent confidence intervals. Note that "bias" refers to NoDE score.

specification.[32] An increase in the NoDE score by one standard deviation (0.3 points) is estimated to increase the Penalty & Reward Premium by 0.2 standard deviations (14.4 points).

The study included information treatments intended to help participants reason about the private welfare impact of a policy. The regression in Table 2 suggests that the information treatments (which added words to the policy text) do not appear to play a significant role in reducing (or increasing) the Penalty & Reward Premium. Interestingly, doing the Abstract Policy Task prior to expressing one's views on the two real-world policies reduces the Penalty & Reward Premium, although with weak statistical significance across specifications. However, the magnitude of the effect is comparable to the better of the two information treatments. Experience with thinking about the policy selection abstractly was helpful in reducing an irrational gap in support between the policies. This suggests that the formation of intuition from experience can improve performance to a greater extent than instruction via information treatments, in line with recent findings (Esponda et al., 2020). It is important to note that participants in this study did not receive any feedback about

[32]The range for NoDE scores observed in the sample is 1.5 points.

their performance on the Abstract Policy Task until the entire survey was complete.

Additional support of hypothesized mechanisms are shown in Table 3, which creates Factor Premiums, or differences in the ratings of policies across the three factors discussed earlier in Section 5.1 (Emissions, Your Finances, and Low Income). NoDE is associated with a higher factor premium in the relative rating for Penalty & Reward in the two factors we would expect to be most affected, (1) the effect on your own finances and (2) the effect on low income individuals. Based on the theoretical analysis in Section 4.4, these two channels concerning who wins and loses financially appear to be clarified by Penalty & Reward's design. Most survey participants believe themselves below-average emitters of carbon, shown in Figure 12, and therefore we would expect them to respond positively to Penalty & Reward concerning the financial impact to their own household.[33]

Analyzing the effect of controls on the outcome variable, which is visible in the extended version of Table 2, Appendix Table 8, we see no strong effects of control variables on Penalty & Reward Premium, apart from belief in climate science, which is strongly and positively predictive of the premium. This may result from the possibility that the expected predictive effect of belief in climate science for Fee & Dividend support is masked by a distaste for the policy, indicated in Appendix Table 9. Because participants believing in climate science overwhelmingly expressed support for Penalty/Reward, climate science beliefs are predictive of the gap in support between these two policies.

In Appendix Table 9, NoDE is strongly and negatively associated with support for Fee & Dividend policy, as hypothesized, with little significant effect on support for Penalty & Reward. In line with theoretical predictions, one's own emissions relative to the average are strongly and negatively associated with support for either policy, as is the belief that one's job will be negatively affected by a carbon tax.

Political orientation is not meaningfully predictive of support for either policy in the presence of the other regressors. Leaning right has a slight positive association

[33] In analyzing the predictiveness of rating factors on overall policy support, we see in Appendix Table 10 that, in both policies, the effectiveness in reducing emissions receives the highest weighting. Impact on one's own finances follows next, with impact on low-income persons weighed least of the three.

Table 3: Baseline Study: Factor Premium for Penalty & Reward

	Dependent variable: *Factor premium of P&R over F&D*		
	Emissions	Your Finances	Low Income
	(1)	(2)	(3)
Constant	28.713	1.516	-48.417^{*}
	(25.267)	(20.742)	(27.440)
NoDE Score	10.222	29.958^{**}	32.225^{**}
	(14.008)	(14.470)	(14.341)
False Correlation error	9.146	16.616	33.415^{**}
	(12.069)	(11.896)	(12.904)
Mental budgeting error	17.085^{**}	3.786	0.767
	(8.514)	(8.757)	(10.296)
Survey order: experiment first	−2.848	1.602	−2.642
	(8.871)	(8.357)	(9.718)
Controls:			
Support factors	X	X	X
Demographic/employment	X	X	X
Observations	248	248	248
R^2	0.049	0.043	0.080
Adjusted R^2	0.008	0.002	0.041

Notes: The table above estimates the effect of cognitive bias on the Factor Premium for Penalty Reward across three factors. The first of the factors, "Emissions" stems from a rating given to each of the two policies based on their effectiveness at reducing emissions. The Emissions factor premium is Emissions rating given to Penalty & Reward minus the Emissions rating given to Fee & Dividend. The second factor, "Your Finances," stems from ratings to each policy based on its impact to your personal household. The third factor, "Low Income" stems from each policy's impact on low earners. See Appendix Section C.1.1 for more detail on the factor questions. The controls include the Support factors (Relative emissions, Belief in climate change, Left-right score, Belief job hurt from carbon tax, Trust government use of revenue, and Support for redistribution) and Demographic/employment factors (Age, Male, Black, Student, and Unemployed) $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

with support for either policy in Appendix Table 9, when controlling for belief in climate science, cognitive predictors, demographics and responses to the other questions included in the survey. The belief in climate science variable appears to be capturing the portion of variation in policy support that is commonly attributed to political orientation. When belief in climate science is removed as a variable from the regression, political orientation is predictive of policy support. Belief in climate science has become highly politicized and is strongly correlated with left-right score in the sample.[34] Still, the results imply that a conservative who believes in climate science is no less likely to support carbon pricing policies than would a liberal who believes in climate science, when controlling for cognitive and other factors.

It appears that accounting for cognitive bias can help reduce the impact of politicization of carbon pricing. In Figure 6, we can see that among participants on the right, for whom support for Fee & Dividend is weak, support is weakest among high-NoDE subset. This same group saw a far sharper increase in support for Penalty & Reward than their low-NoDE counterparts. We see a very similar effect among left-leaning participants.

Considering the proportions favoring the policy using the binary support measure, the political gap narrows dramatically, with all political affiliations supporting Penalty & Reward, at the mean. Only the left-leaning subsample supported Fee & Dividend. But remarkably, there is a sharp increase in the intensity of support among left-leaning individuals for Penalty & Reward. While the literature notes that support for carbon pricing typically splits along party lines, the results here reveal very tepid support for carbon taxes on the left. Lowering the cognitive costs, for both political affiliations, appears to be a more important effect than the moderate gap between the two political groups, which largely remains in the support of the new policy as well.

Lastly, Figure 7 shows that Fee & Dividend has a higher density of support ratings near zero, indicating uncertainty or indifference, than does Penalty & Reward. A policy with support that is less clear-cut due to its complexity may be more vulnerable

[34]The Pearson correlation coefficient is -0.65 and is extremely statistically significant, with (-0.58,-0.71) as the 95% confidence interval.

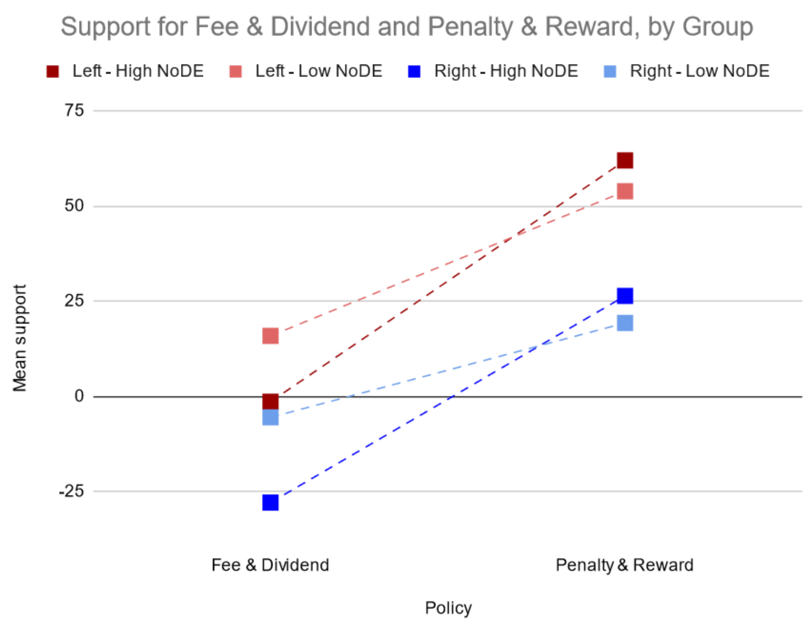


Figure 6: Comparison of mean policy support by left-leaning voters and right-leaning voters, each separated into high- and low-NoDE subsets.

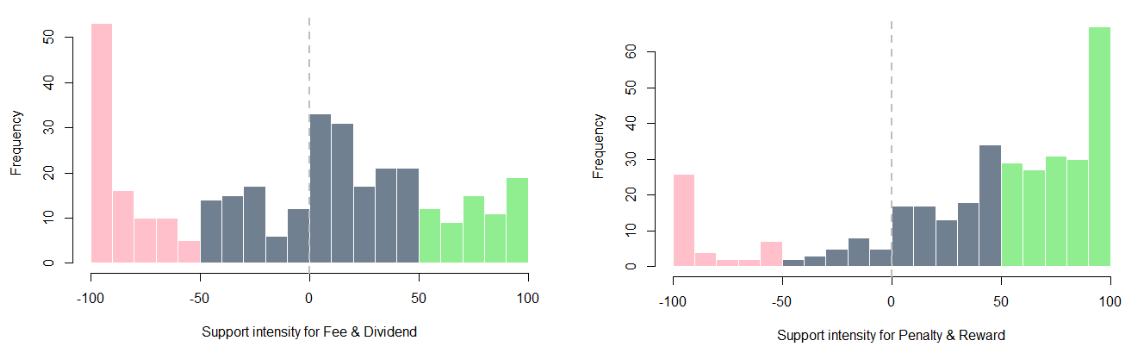


Figure 7: Distribution of Overall Rating for Fee & Dividend and Penalty & Reward.

to misleading characterizations by vested interests.

7 Discussion

This study examined the role of cognitive frictions in carbon pricing support from multiple angles: a diagnostic test of such frictions, applied to lab subjects engaging with an abstract policy; a test of a new policy designed to sidestep a key bias; and a test of the association of problem-solving ability with policy attitudes. The combination conveys a clear message: Neglect of Distributed Effects is a cognitive/attentional phenomenon and it affects how people choose real policies. For voters with constraints on attention or computation, Fee & Dividend's opacity lowers support for it. This magnitude of this phenomenon may be larger in the public discourse, where a noisy information environment, influence from vested interests, and the lack of extrinsic incentives to evaluate costs and benefits can result in outcomes like the 10% support in France for a carbon tax and rebate, shown in Douenne and Fabre (2022). Accurate initial impressions and intuitions may be crucial, given that informational interventions are not necessarily well-absorbed. This study shows that policy design informed by behavioral theory and evidence can produce better public intuitions about efficacy. When we lessen the burden of choosing well, public choices appear to respond.

Much of the literature about the public acceptability of carbon pricing concludes that political orientation drives opinion. This is not exactly what we observe in the analysis of the survey data. Many on the left reject Fee & Dividend, perhaps because they perceive it to be regressive. And even though, for both policies, support is stronger on the left than on the right, the explanatory power of political orientation disappears when controlling for cognitive factors, belief in climate change, and other covariates. For studies that assess the role of partisanship in support for environmental policy, it is worth considering presenting policies (or creating policies), so that all of the important information can be easily and quickly integrated. Our results suggest that the ability to form a clear intuition about how a beneficial policy works gives rise to strongly favorable opinions, which could be more resistant to misleading messaging.

Separating the explanatory power of political identity from cognitive factors is important. A recent paper, Kotchen (2022), advocates labeling corrective tax proposals with a ratio of the value of the externality reduction to the revenue collected by the policy. This is in part to address weak support on the right, who may not view government use of tax revenues as desirable. The paper suggests considering corrective taxes only with large ratios (greater than 1). A consequence of proposing only taxes with ratios greater than one would be to tax the use of coal, but not of other fossil fuels. But the ratio lacks a clear, objective welfare interpretation. And if revenues are returned as dividends, then it is even more difficult to rationalize poor public support in the Kotchen (2022) framework.

We show that voter discounting of the Distributed Effects of corrective tax proposals make it more likely that policies with smaller revenue collection would be better supported, but also that this is rooted in cognitive factors and not solely by preferences or beliefs about the purpose, prudence or efficiency of government spending. Using Abstract Policy Tasks, we can rule out the role played by concerns about prudence, efficiency, or trustworthiness of institutions. If real-world judgments about policy effects result from cognitive constraints, we would not necessarily want to enshrine mistaken notions of policy effects into the kinds of policies we promote, prioritize, or enact. It is possible that we can do better than that with thoughtful policy design.

By avoiding taxing and rebating everyone, Penalty & Reward may alleviate political penalties applied to Pigouvian taxes for the scale of their resulting revenue (Kotchen, 2022). As Fee & Dividend and similarly structured policies are presently framed, a major portion of the tax money collected makes an unnecessary round trip to government and back, acquiring a subjective discount along the way, based on our findings. By removing the round trip, Penalty & Reward slashes the scale of collections and mechanically increases the ratio of damage reduction to revenue that Kotchen (2022) describes, potentially boosting political support.

Going forward, analysis of real-world climate policy support could be extended to non-carbon-pricing approaches, like subsidies or command-and-control policies. The Diagnostic Experiment results suggest a bias favoring subsidies among participants, and it could be useful to measure if NoDE is predictive of support for environmental subsidies over taxes. Subsidies, despite being less efficient, appear to be widely popular in the world.[35] Our results and behavioral model suggest that opacity and complexity have the ability to affect support for certain policies. It would be interesting if the Distributed Effect-related opacity about how subsidies are paid for were important to its popularity and could somehow be reversed. Understanding what features of policies can be adjusted to make their true effects more intuitive to the public is an important, underexplored area of research.

Among the arguments emphasized by many advocates for Fee & Dividend policy is that that its progressive redistribution of money would fulfill certain social justice goals, independent of its positive effects of mitigating climate change.[36] A majority of households, mainly with lower incomes, are expected to be able to increase consumption due to Fee & Dividend's immediate financial benefits.[37] Yet, Fee & Dividend proposals offer its beneficiaries a salient but small cost, and an opaque but large benefit, and we have shown how this fosters bad policy selection and can aggregate into harms for most and benefits for a few. If policy designs can be made easier to understand, the public may consider the use of a simplified designs like Penalty & Reward more ethical.[38] The cost of understanding a difficult policy may cause an expected beneficiary to misperceive its welfare impacts, to avoid engaging with it and take cues from their party, or to abstain from voting because of confusion.[39] As a

[35]Clean subsidies, when implemented with broader consumption taxes, can be shown equivalent to pollution taxes (Fullerton and Mohr, 2003; Fullerton and Wolverton, 2005). But such two-part policies may not prove popular, nor is it realistic to assume that a government can have perfect information of what behaviors and potential innovations to actively subsidize.

[36]This is highlighted in academic papers (Klenert et al., 2018; Budolfson et al., 2021); public statements from economists and public policy experts Akerlof et al. (2019); Center (2008); and statements of advocacy groups, such as the Climate Leadership Council and the Citizen's Climate Lobby.

[37]The top 10 percent of earners in the U.S. contribute 40 percent of the total emissions, and the average person among households in the top 0.1% of emitters in the U.S. contributed 57 times the emissions of the average person in the lowest decile (Starr et al., 2023a,b).

[38]Ambuehl et al. (2022) explores how in employment and consumer choice settings incentives disproportionately attract people whose learning costs are higher. For example, a physically risky job with a high wage has a transparent incentive with an opaque cost. Voters constrained in attention or computation optimally use less information before making a decision, but are more likely to be disappointed in the end than more discerning, cognitively advantaged types. Most people find this dynamic unethical (Ambuehl and Ockenfels, 2017).

[39]Enke et al. (2023b) demonstrates that a lack of confidence resulting from cognitive difficulty can

result, popular measures of support may not reflect the true preferences of the public. The voter's welfare and aggregate welfare can be hampered if policy makers ignore the cognitive costs of policy evaluation.

Finally, the primary purpose of the Penalty & Reward policy frame in this study is as a tool to better understand voter preferences. But in Appendix section D, we discuss strategies for literal and approximate implementations of Penalty & Reward to mitigate negative externalities.

8 Conclusion

The fact that voters' rationality is bounded requires us to move away from the assumption that voters always express their true preferences. The incentives to thinking carefully about policy tend to be quite low for many people, and so it is important to understand the cognitively demanding elements of evaluation when designing policies. Environmental policies can be especially complex. Corrective taxes involve many moving parts, including changes to consumer surplus, usage of revenue, and reductions in externalities. This study identifies and measures mechanisms that suggests voters do not perfectly attend to each of these parts. Yet this does not necessarily mean societies must abandon useful tools like carbon taxes. The failure of citizens to adopt carbon prices has shifted the tone of environmental economics in recent years, with many taking as given that relatively efficient policies have little hope of manifesting, due to public disapproval. Attention has shifted to "second-best" approaches.[40] Much attention has been given to polarization of beliefs or attitudes, for which little can be done by intervention.

The findings of this study suggest that some of this pessimism may be misplaced. Appealing to voters' financial self-interest is indeed a way to potentially bridge a

often cause individuals to select out of a participatory democratic process. An equitable climate policy that benefits all but a few can fail to realize if its beneficiaries happen to be more constrained, less confident, and thus more inclined to select out of participation.

[40]Many papers assessing "second-best" approaches include, as motivation, the observation that carbon taxes have failed to gain political momentum; economists have been scolded by academics of other disciplines for their attachment to carbon pricing (Kopp, 2022); and economists themselves have blogged or tweeted about moving on from their ideal tool of carbon taxes (Ho, 2021).

divide between worldviews; however, to be successful, it must take into account the high sensitivity of the public to complexity.

References

Abeler, J. and Jäger, S. (2015). Complex tax incentives. *American Economic Journal: Economic Policy*, 7(3):1–28.

Akerlof, G., Aumann, R., Deaton, A., Diamond, P., Engle, R., Fama, E., Hansen, L., Hart, O., Holmström, B., Kahneman, D., et al. (2019). Economists' statement on carbon dividends. *Wall Street Journal*, 16.

Ambuehl, S. and Ockenfels, A. (2017). The ethics of incentivizing the uninformed: A vignette study. *American Economic Review*, 107(5):91–95.

Ambuehl, S., Ockenfels, A., and Stewart, C. (2022). Who opts in? composition effects and disappointment from participation payments. *Review of Economics and Statistics*, pages 1–45.

Anderson, S., Marinescu, I., and Shor, B. (2023). Can pigou at the polls stop us melting the poles? *Journal of the Association of Environmental and Resource Economists*, 10(4):903–945.

Andersson, J. J. (2019). Carbon taxes and co2 emissions: Sweden as a case study. *American Economic Journal: Economic Policy*, 11(4):1–30.

Banovetz, J. and Oprea, R. (2023). Complexity and procedural choice. *American Economic Journal: Microeconomics*, 15(2):384–413.

Baranzini, A. and Carattini, S. (2017). Effectiveness, earmarking and labeling: testing the acceptability of carbon taxes with survey data. *Environmental Economics and Policy Studies*, 19:197–227.

Baranzini, A., Van den Bergh, J. C., Carattini, S., Howarth, R. B., Padilla, E., and Roca, J. (2017). Carbon pricing in climate policy: seven reasons, complementary instruments, and political economy considerations. *Wiley Interdisciplinary Reviews: Climate Change*, 8(4):e462.

Baron, J. and Jurney, J. (1993). Norms against voting for coerced reform. *Journal of Personality and Social Psychology*, 64(3):347.

Baumol, W. J. (1972). On taxation and the control of externalities. *The American Economic Review*, 62(3):307–322.

Baumol, W. J. and Oates, W. E. (1971). The use of standards and prices for protection of the environment. In *The Economics of Environment: Papers from Four Nations*, pages 53–65. Springer.

Bernheim, B. D. (1987). Ricardian equivalence: An evaluation of theory and evidence. *NBER macroeconomics annual*, 2:263–304.

Bhargava, S., Loewenstein, G., and Sydnor, J. (2017). Choose to lose: Health plan choices from a menu with dominated option. *The Quarterly Journal of Economics*, 132(3):1319–1372.

Blanchard, O., Gollier, C., and Tirole, J. (2023). The portfolio of economic policies needed to fight climate change. *Annual Review of Economics*, 15(1):689–722.

Boyce, J. K. (2018). Carbon pricing: effectiveness and equity. *Ecological Economics*, 150:52–61.

Buchanan, J. M. and Wagner, R. E. (2013). *Fiscal responsibility in constitutional democracy*, volume 1. Springer Science & Business Media.

Budolfson, M., Dennig, F., Errickson, F., Feindt, S., Ferranna, M., Fleurbaey, M., Klenert, D., Kornek, U., Kuruc, K., Méjean, A., et al. (2021). Protecting the poor with a carbon tax and equal per capita dividend. *Nature Climate Change*, 11(12):1025–1026.

Carattini, S., Baranzini, A., Thalmann, P., Varone, F., and Vöhringer, F. (2017). Green taxes in a post-paris world: are millions of nays inevitable? *Environmental and Resource Economics*, 68:97–128.

Center, C. T. (2008). Pricing carbon efficiently and equitably. *Carbon Tax Center, May*, 23.

Cherry, T. L., Kallbekken, S., and Kroll, S. (2012). The acceptability of efficiency-enhancing environmental taxes, subsidies and regulation: An experimental investigation. *Environmental Science & Policy*, 16:90–96.

Cherry, T. L., Kallbekken, S., and Kroll, S. (2014). The impact of trial runs on the acceptability of environmental taxes: Experimental evidence. *Resource and Energy Economics*, 38:84–95.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American economic review*, 99(4):1145–1177.

Colmer, J., Martin, R., Muûls, M., and Wagner, U. J. (2022). Does pricing carbon mitigate climate change? firm-level evidence from the european union emissions trading scheme.

Cramton, P., MacKay, D. J., Ockenfels, A., and Stoft, S. (2017). *Global carbon pricing: the path to climate cooperation*. The MIT Press.

Cronin, J. A., Fullerton, D., and Sexton, S. (2019). Vertical and horizontal redistributions from a carbon tax and rebate. *Journal of the Association of Environmental and Resource Economists*, 6(S1):S169–S208.

Dal Bó, E., Dal Bó, P., and Eyster, E. (2018). The demand for bad policy when voters underappreciate equilibrium effects. *The Review of Economic Studies*, 85(2):964–998.

Dechezleprêtre, A., Fabre, A., Kruse, T., Planterose, B., Chico, A. S., and Stantcheva, S. (2022). Fighting climate change: International attitudes toward climate policies. Technical report, National Bureau of Economic Research.

Dell'Anno, R. and Mourao, P. (2012). Fiscal illusion around the world: an analysis using the structural equation approach. *Public finance review*, 40(2):270–299.

Douenne, T. and Fabre, A. (2022). Yellow vests, pessimistic beliefs, and carbon tax aversion. *American Economic Journal: Economic Policy*, 14(1):81–110.

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of political economy*, 65(2):135–150.

Dresner, S., Dunne, L., Clinch, P., and Beuermann, C. (2006). Social and political responses to ecological tax reform in europe: an introduction to the special issue. *Energy policy*, 34(8):895–904.

Enke, B., Graeber, T., and Oprea, R. (2023a). Complexity and time. Technical report, National Bureau of Economic Research.

Enke, B., Graeber, T., and Oprea, R. (2023b). Confidence, self-selection, and bias in the aggregate. *American Economic Review*, 113(7):1933–1966.

Esponda, I., Vespa, E., and Yuksel, S. (2020). Mental models and learning: The case of base-rate neglect. Technical report, Tech. rep.

Farhi, E. and Gabaix, X. (2019). Online appendix for "optimal taxation with behavioral agents".

Fremstad, A. and Paul, M. (2019). The impact of a carbon tax on inequality. *Ecological Economics*, 163:88–97.

Fullerton, D. and Mohr, R. D. (2003). Suggested subsidies are sub-optimal unless combined with an output tax. *Contributions in Economic Analysis & Policy*, 2(1):1–20.

Fullerton, D. and Wolverton, A. (2005). The two-part instrument in a second-best world. *Journal of Public Economics*, 89(9-10):1961–1975.

Goulder, L. H., Hafstead, M. A., Kim, G., and Long, X. (2019). Impacts of a carbon tax across us household income groups: What are the equity-efficiency trade-offs? *Journal of Public Economics*, 175:44–64.

Hastings, J. and Shapiro, J. M. (2012). Mental accounting and consumer choice: Evidence from commodity price shocks. Technical report, National Bureau of Economic Research.

Ho, B. (2021). Time to give up on a carbon tax? https://news.climate.columbia.edu/2021/04/26/retire-carbon-tax/.

Horowitz, J. K., Cronin, J.-A., Hawkins, H., Konda, L., and Yuskavage, A. (2017). *Methodology for analyzing a carbon tax.*

Janusch, N., Kroll, S., Goemans, C., Cherry, T. L., and Kallbekken, S. (2021). Learning to accept welfare-enhancing policies: an experimental investigation of congestion pricing. *Experimental Economics*, 24:59–86.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, 5(1):193–206.

Kallbekken, S. and Aasen, M. (2010). The demand for earmarking: Results from a focus group study. *Ecological economics*, 69(11):2183–2190.

Kallbekken, S., Kroll, S., and Cherry, T. L. (2011). Do you not like pigou, or do you not understand him? tax aversion and revenue recycling in the lab. *Journal of Environmental Economics and Management*, 62(1):53–64.

Kallbekken, S. and Sælen, H. (2011). Public acceptance for environmental taxes: Self-interest, environmental and distributional concerns. *Energy Policy*, 39(5):2966–2973.

Klenert, D., Mattauch, L., Combet, E., Edenhofer, O., Hepburn, C., Rafaty, R., and Stern, N. (2018). Making carbon pricing work for citizens. *Nature Climate Change*, 8(8):669–677.

Kopp, B. (2022). Tweet from aug. 7, 2022: Environmental economists should probably take a moment of humble reflection... https://www.twitter.com/.

Kotchen, M. (2022). Taxing externalities: Revenue vs. welfare gains with an application to us carbon taxes. Technical report, National Bureau of Economic Research.

Kuziemko, I., Marx, N. L., and Naidu, S. (2023). “compensate the losers?” economic policy and partisan realignment in the us. Technical report, National Bureau of Economic Research.

Lacetera, N., Pope, D. G., and Sydnor, J. R. (2012). Heuristic thinking and limited attention in the car market. *American Economic Review*, 102(5):2206–2236.

Lohmann, P. M., Gsottbauer, E., Doherty, A., and Kontoleon, A. (2022). Do carbon footprint labels promote climatarian diets? evidence from a large-scale field experiment. *Journal of Environmental Economics and Management*, 114:102693.

Maestre-Andrés, S., Drews, S., and van den Bergh, J. (2019). Perceived fairness and public acceptability of carbon pricing: a review of the literature. *Climate policy*, 19(9):1186–1204.

Metcalf, G. E. (2021). Carbon taxes in theory and practice. *Annual Review of Resource Economics*, 13:245–265.

Metcalf, G. E. (2023). Five myths about carbon pricing. Technical report, National Bureau of Economic Research.

Mildenberger, M., Lachapelle, E., Harrison, K., and Stadelmann-Steffen, I. (2022). Limited impacts of carbon tax rebate programmes on public support for carbon pricing. *Nature climate change*, 12(2):141–147.

Montgomery, W. D. (1972). Markets in licenses and efficient pollution control programs. *Journal of economic theory*, 5(3):395–418.

Muller, L., Lacroix, A., and Ruffieux, B. (2019). Environmental labelling and consumption changes: A food choice experiment. *Environmental and resource economics*, 73:871–897.

Murray, B. and Rivers, N. (2015). British columbia's revenue-neutral carbon tax: A review of the latest "grand experiment" in environmental policy. *Energy Policy*, 86:674–683.

Nielsen, K. and Rehbeck, J. (2022). When choices are mistakes. *American Economic Review*, 112(7):2237–2268.

Nordhaus, W. D. (2001). Global warming economics.

Oprea, R. (2022). Simplicity equivalents. *Working Paper.*

Pigou, A. C. (1924). *The Economics of Welfare.* Macmillan.

Puviani, A. (1903). *Teoria Dell'illusione Finanziaria.* Milan: Remo Sandon.

Rees-Jones, A. and Taubinsky, D. (2016). *Heuristic perceptions of the income tax: Evidence and implications for debiasing.* National Bureau of Economic Research.

Rees-Jones, A. and Taubinsky, D. (2020). Measuring "schmeduling". *The Review of Economic Studies*, 87(5):2399–2438.

Rivers, N. and Schaufele, B. (2015). Salience of carbon taxes in the gasoline market. *Journal of Environmental Economics and management*, 74:23–36.

Rivers, N. and Shaffer, B. (2022). How do households spend carbon tax rebates? *Available at SSRN 4054023.*

Sælen, H. and Kallbekken, S. (2011). A choice experiment on fuel taxation and earmarking in norway. *Ecological Economics*, 70(11):2181–2190.

Sausgruber, R. and Tyran, J.-R. (2005). Testing the mill hypothesis of fiscal illusion. *Public choice*, 122(1):39–68.

Sausgruber, R. and Tyran, J.-R. (2011). Are we taxing ourselves?: How deliberation and experience shape voting on taxes. *Journal of Public Economics*, 95(1-2):164–176.

Shultz, G. P. and Halstead, T. (2018). The dividend advantage. *Climate Leadership Council.*

Starr, J., Nicolson, C., Ash, M., Markowitz, E. M., and Moran, D. (2023a). Assessing us consumers' carbon footprints reveals outsized impact of the top 1%. *Ecological Economics*, 205:107698.

Starr, J., Nicolson, C., Ash, M., Markowitz, E. M., and Moran, D. (2023b). Income-based us household carbon footprints (1990–2019) offer new insights on emissions inequality and climate finance. *PLOS Climate*, 2(8):e0000190.

Steg, L., Dreijerink, L., and Abrahamse, W. (2006). Why are energy policies acceptable and effective? *Environment and behavior*, 38(1):92–111.

Thaler, R. H. (1990). Anomalies: Saving, fungibility, and mental accounts. *Journal of economic perspectives*, 4(1):193–205.

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral decision making*, 12(3):183–206.

Tietenberg, T. H. (1973). Controlling pollution by price and standard systems: a general equilibrium analysis. *The Swedish Journal of Economics*, pages 193–203.

Vora, R. and Zappala, G. (2025). Endogenous green preferences. CESifo Working Paper Series 11857, CESifo.

Yamazaki, A. (2017). Jobs and climate policy: Evidence from british columbia's revenue-neutral carbon tax. *Journal of Environmental Economics and Management*, 83:197–216.

A Appendix 1: Theory Supplement

A.1 General Framework

Choosing a pricing policy as a voter-consumer requires thinking of two stages recursively:

1. Voter expresses policy preference.

2. Consumers make consumption decisions given a policy.

All policy proposals and status quo policy conditions correspond to a value, τ, which when zero implies no policy, when positive implies a tax, and when negative implies a subsidy. Voters value the policy shift $\Delta\tau$, when forming a view of a proposal and the strength of that view. For a rational voter:

$$U_i^{\Delta\tau} = U_i(\tau_1) - U_i(\tau_0)$$

When $U_i(\Delta\tau \mid \tau_0) = U_i(\tau_1) - U_i(\tau_0) > 0$, the voter supports the proposal, when $U_i(\Delta\tau \mid \tau_0) < 0$, we support the status quo, and otherwise we are indifferent. In this paper we consider $U_i(\Delta\tau \mid \tau_0)$ our primary object of interest (i.e. how do subjects sign it and perceive its magnitude relative to the rational full attention benchmark?)[41]

$$U_i(\Delta\tau \mid \tau_0) = \pi_i(c_i(\tau_1)) - \pi_i(c_i(\tau_0)) + \sum_{j\in J} (\pi_i(c_j(\tau_1)) - \pi_i(c_j(\tau_0)))$$

Here, $\pi()$ is an argument representing net gain or loss to voter-consumer i from actions from consumer i and any consumer j from the set of J consumers. Because actions are in dominant strategies and a function of revealed policy outcomes, there

[41] You might ask, 'Why have the object of interest be the value of a shift $U_i(\Delta\tau \mid \tau_0)$, and not simply one's comprehensive welfare under the new policy; $U_i(\tau_1)$?' For two important reasons: (1) In a political context, people are asked about preferences about a policy change, but not to explicitly quantify their welfare at present or under a proposal. So, $U_i(\tau_0)$ and $U_i(\tau_1)$ may be rather strange objects. Furthermore, (2) both in the field and in the lab, asking people to estimate their before and after welfare is leading them to a kind of calculation that they may not typically make when considering a proposal. In general, we don't want to assume $U_i^{\Delta\tau} = U_i(\tau_1) - U_i(\tau_0)$.

are no interaction terms between the private net gain of subsequent consumption choice of any two agents. Hence we can represent the gain or loss from one's own consumption choices and the consumption choices of others as additively separable.

We define the portion of your change in welfare that stems from your own consumption choices c_i, in essence $\pi_i(c_i(\tau_1)) \; - \; \pi_i(c_i(\tau_0))$ as a Consumer Effect or *CE*. The portion of your change in welfare that stems from the consumption of others c_j, i.e. $\sum_{j \in J} (\pi_i(c_j(\tau_1)) \; - \; \pi_i(c_j(\tau_0)))$ is Distributed Effect, or *DE* (or additionally an Externality Effect, *EE*, depending on the circumstance. If the consumption of others has no bearing on your welfare, $U_i(\Delta\tau \mid \tau_0) = CE_i((\tau_1, \tau_0))$.

Now generalizing to a behavioral voter:

$$\hat{U}_i^{\Delta\tau} = \iota \; \overbrace{(\pi_i(c_i(\tau_1)) \; - \; \pi_i(c_i(\tau_0)))}^{\textit{Consumer Effect } (CE)} + \lambda \overbrace{\sum_{j \in J} \; (\pi_i(c_j(\tau_1)) \; - \; \pi_i(c_j(\tau_0)))}^{\textit{Distributed Effect } (DE)}$$

Above, ι and λ are weights that might bias a voter on a proposed policy. (Each weight would equal 1 for an unbiased individual.) The CE-DE structure comes from the two natural components of the problem, but the weights could also be interpreted 'as if.'

To explore the implications of neglect to the Distributed Effects of subsidy and tax policies that we've discussed, we define the policy variable τ as a number that would be added the market price p of a purchasable object. Therefore when $\tau > 0$, it represents a tax, and when $\tau < 0$ it represents a subsidy.

Subsequent to policy outcome, one's own actions are purchase choices. For any given private value for goods consumed, the net benefit from one's optimal purchases (i.e. personal consumer surplus) can only be the same or lower when policy moves in the direction of taxation or desubsidization. The effective price under the new policy $\tilde{p}_1 = \tau_1 + p$ is greater than than $\tilde{p}_0 = \tau_0 + p$. Therefore, when $\tau_1 - \tau_0 > 0$, $CE \leq 0$. Analagously, under direction of subsidization $\tau_1 - \tau_0 < 0$, $CE \geq 0$.

For the analysis of Distributed Effects, let us define x as the cumulative demand function for all other consumers.[42] $\gamma(\tau_1\ x(\tau_1))$ represents one's share of the cumulative revenue or debt generated by the new policy, as $\gamma(\tau_0\ x(\tau_0))$ represents the same for the status quo. Then $DE = \gamma\left[(\tau_1\ x(\tau_1)) - (\tau_0\ x(\tau_0))\right]$.[43]

Let us consider both the proposed introduction of a new policy and the proposed removal of an existing policy. First, if introducing a tax, $\tau_0 = 0$ and $\tau_1 > 0$. Demand is nonnegative $x() \geq 0$, and therefore $\tau_1 x(\tau_1) \geq 0$. Because $\tau_0 x(\tau_0) = 0$, we know $\tau_1 x(\tau_1) \geq \tau_0 x(\tau_0)$ and that therefore $DE \geq 0$. If introducing a subsidy, $\tau_0 = 0$ and $\tau_1 < 0$. Therefore, $\tau_1 x(\tau_1) \leq 0$. Because $\tau_0 x(\tau_0) = 0$, $\tau_1 x(\tau_1) \leq \tau_0 x(\tau_0)$ and $DE \leq 0$. If removing an existing tax, $\tau_0 > 0$ and $\tau_1 = 0$. Therefore, $\tau_1 x(\tau_1) \leq \tau_0 x(\tau_0)$ and $DE \leq 0$. If removing an existing subsidy, $\tau_0 < 0$ and $\tau_1 = 0$. Therefore, $\tau_1 x(\tau_1) \geq \tau_0 x(\tau_0)$ and $DE \geq 0$.

Returning to our complete value function, let's consider what happens when introducing a new tax or removing an existing subsidy. Based on the analysis above, in both of these cases $CE \leq 0$ and $DE \geq 0$. If $\iota > \lambda$ for a biased individual,

$$\hat{U}_i(\Delta\tau \mid \tau_0) = \iota CE(\tau_1, \tau_0) + \lambda DE(\tau_1, \tau_0) \leq U_i(\Delta\tau \mid \tau_0) = CE(\tau_1, \tau_0) + DE(\tau_1, \tau_0)$$

Furthermore, as long as either $CE \neq 0$ or $DE \neq 0$, then:

$$\hat{U}_i(\Delta\tau \mid \tau_0) = \iota CE(\tau_1, \tau_0) + \lambda DE(\tau_1, \tau_0) < U_i(\tau_1, \tau_0) = CE(\tau_1, \tau_0) + DE(\tau_1, \tau_0)$$

Analogously, when removing an existing tax, or introducing a new subsidy, $CE \geq 0$ and $DE \geq 0$. So when $\iota > \lambda$:

$$\hat{U}_i(\Delta\tau \mid \tau_0) = \iota CE(\tau_1, \tau_0) + \lambda DE(\tau_1, \tau_0) \geq U_i(\Delta\tau \mid \tau_0) = CE(\tau_1, \tau_0) + DE(\tau_1, \tau_0)$$

[42] To formally define the cumulative demand function x() for a set J of rational consumers, given policy τ, price p, and a set of object values L_j for a consumer j
$x(\tau) = \sum_{j \in J} \left(\sum_{l \in L_j} I_{(l > p + \tau\omega)} \right)$ where $\omega = 1 - \frac{1}{N}$. N represents the population size and as it increases, ω approaches 1.

[43] $\gamma > 0$. For example, with equal dividend, $\gamma = \frac{1}{N}$ where $N = \mid J \mid + 1$

When either $CE \neq 0$ or $DE \neq 0$, then:

$$\hat{U}_i(\Delta\tau \mid \tau_0) = \iota CE(\tau_1, \tau_0) + \lambda DE(\tau_1, \tau_0) > U_i(\Delta\tau \mid \tau_0) = CE(\tau_1, \tau_0) + DE(\tau_1, \tau_0)$$

A.2 Equilibrium Effects

In Dal Bó et al. (2018), which explores Equilibrium Effects and shows evidence that individuals do not fully incorporate how a change in policy should affect subsequent behavior, it may be useful to decompose CE and DE even further along a direct v. equilibrium (static v. dynamic) dimension. CE can be decomposed into a Direct Effect and an Equilibrium Effect, , and DE can be decomposed into Distributed Direct Effect and a Distributed Equilibrium Effect.

$$U_i(\Delta\tau \mid \tau_0) = \textit{Direct Eff} + \textit{Equilibrium Eff} + \textit{Distrib Direct Eff} + \textit{Distrib Equil Eff}$$

$$\textit{Direct} \quad \textit{Effect} = \pi_i(c_i(\tau_0), \tau_1) - \pi_i(c_i(\tau_0), \tau_0)$$

The Direct Effect, or the static part of the Consumer Effect, represents the net gain or loss from a proposed policy attributable to your own consumption choices, assuming that you would not make any change to your consumption choice in response to the enactment of the proposed policy.

$$\textit{Equilibrium} \quad \textit{effect} = \pi_i(c_i(\tau_1), \tau_1) - \pi_i(c_i(\tau_0), \tau_1)$$

The Equilibrium Effect, or the dynamic part of the Consumer Effect, represents the net gain or loss from a proposed policy attributable only to the change in your own consumption choices in response to the enactment of the proposed policy.

$$\textit{Distributed Direct Effect} = \sum_{j \in J} \left(\pi_i(c_j(\tau_0), \tau_1) - \pi_i(c_j(\tau_0), \tau_0)\right)$$

The Distributed Direct Effect, or static part of the Distributed Effect, represents the net gain or loss to you from a proposed policy that is attributable to consumption choices of others, assuming that others would not make any change to consumption choices in response to the enactment of the proposed policy.

$$\textit{Distributed Equilibrium Effect} = \sum_{j \in J} \left(\pi_i(c_j(\tau_1), \tau_1) - \pi_i(c_j(\tau_0), \tau_1)\right)$$

The Distributed Equilibrium Effect represents the net gain or loss from a proposed policy attributable only to the change in others' consumption choices in response to the enactment of the proposed policy.

$$CE = \textit{Direct Effect} + \textit{Equil Effect} = \pi_i(c_i(\tau_1), \tau_1) - \pi_i(c_i(\tau_0), \tau_0)$$

$$DE = \textit{Dist Direct Effect} \ + \ \textit{Dist Equil Effect} = \sum_{j \in J} \left(\pi_i(c_j(\tau_1), \tau_1) - \pi_i(c_j(\tau_0), \tau_0)\right)$$

Now let's consider some potential bias terms on computing the shift in welfare resulting from a shift in policy:

$$\hat{U}_i^{\Delta\tau} = \kappa \textit{Direct} + \eta \textit{Equilibrium} + \nu \textit{Distrib Direct} + \theta \textit{Distrib Equil}$$

A.3 Loss Aversion With Mental Accounts

If subjects are irrationally tax averse in real-world and abstract settings, there is an alternative explanation to NoDE. This explanation, Loss Aversion With Mental Accounts, or LAMA, cannot be rationalized by behavioral risk preferences, as there is no uncertainty, only computational constraints.[44] If deciding on whether to support the introduction of a tax, a behavioral voter may value the policy as follows:

$$\hat{U}^{\Delta\tau} = \begin{cases} CE(\tau_1,\tau_0) + DE(\tau_1,\tau_0), & \text{if } CE(\tau_1,\tau_0) \geq 0 \text{ and } DE(\tau_1,\tau_0) \geq 0 \\ \alpha CE(\tau_1,\tau_0) + DE(\tau_1,\tau_0), & \text{if } CE(\tau_1,\tau_0) < 0 \text{ and } DE(\tau_1,\tau_0) \geq 0 \\ CE(\tau_1,\tau_0) + \alpha DE(\tau_1,\tau_0), & \text{if } CE(\tau_1,\tau_0) \geq 0 \text{ and } DE(\tau_1,\tau_0) < 0 \\ \alpha CE(\tau_1,\tau_0) + \alpha DE(\tau_1,\tau_0), & \text{if } CE(\tau_1,\tau_0) < 0 \text{ and } DE(\tau_1,\tau_0) < 0 \end{cases}$$

Here, a voter is not biased in evaluation of DE, but in either of CE or DE, con-

[44] Oprea (2022) shows subjects make similar aggregate errors in deterministic and lottery settings.

ditional on taking a negative value. This impedes an accurate calculation of the net effect of the two policies, as evident in past work (Kahneman et al., 1991; Thaler, 1999). Because for tax introductions, $CE \leq 0$ and $DE \geq 0$, the effect of LAMA can result in suboptimal rejection of taxes relative to a no-policy reference point. By forcing subjects to evaluate subsidy proposals ($\tau < 0$) in certain rounds, where $CE > 0$ and $DE < 0$, we are able to evaluate NoDE, while controlling for LAMA.

A.4 Penalty & Reward Premium: Other mechanisms

There are two alternative or additional behavioral mechanisms that could result in greater support for Penalty & Reward than for Fee & Dividend that we test in the survey and can ultimately control for in the evaluation of the primary NoDE hypothesis. **False correlation** is a misperception that the dividend under Fee & Dividend policy is strongly correlated with one's own total carbon taxes paid. A heuristic known as **mental budgeting**, when applied to evaluation of Fee & Dividend, may cause individuals to choose policies as if dividends will be allocated by individuals to the taxed consumption category.

Both of these cognitive mechanisms could have effects on policy perception through the "effectiveness in reducing emissions" channel. First, a misperception of correlation between fees and dividends that is rectified by Penalty & Reward. Then, a novel application of a well-established behavioral heuristic, mental budgeting, the scope of which is also reduced in Penalty & Reward.

A.4.1 False correlation

A potential misperception concerning Fee & Dividend is that the dividend is strongly correlated with total carbon taxes that an individual pays. There are at least three ways this mistake could arise:

1. A perception of homogeneity of consumer behavior, i.e. decisions are made as if all consumers act as one or mirror each others' behavior.

2. A misunderstanding of the policy that an individual dividend is a repayment of exactly the individual taxes paid.

3. An exaggerated sense of the influence of one's own tax payments on the size of one's personal dividend.

Consider the following formula for a subjective dividend:

$$Subjective\ Dividend = \alpha\ \tau_{carb}\ q_i^{carb} + (1-\alpha)\ \tau_{carb}\ \bar{q}_{carb}$$

When $\alpha = 0$ an individual is unmistaken and when $\alpha = 1$, they are fully mistaken. The effect of a positive α is to effectively shrink the size of τ. For a fully mistaken individual thinking they will receive the entirety of their paid taxes back in the form of a dividend, it's as if the tax does not exist, and the effective price is the original market price.

In reality, with enough people in the population, there is nearly no correlation ($\frac{1}{N}$) between an individual's consumption choices and their personal dividend. But it may be difficult for people to recognize this. For example, suppose the ex ante perception of a policy proposal corresponds to $\alpha = \frac{1}{2}$. The perceived effectiveness of reducing consumption of the bad would be equivalent to if a tax of $\frac{\tau}{2}$ were implemented, instead of τ. This would make offering a dividend seem counterproductive to the goal, like tying one hand behind one's back.[45]

Penalty & Reward has no dividend, and therefore there is no potential to form a false correlation. By directly taxing or rewarding, the step of computing and comparing the incentive effects of two pieces, fees and dividends, is absent.

A.4.2 Mental Budgeting

Descriptions of Fee & Dividend sometimes provoke the following skeptical response: *If you give the carbon tax revenue back to people, they will just spend it on fossil fuels. How will that accomplish anything?*(Rivers and Shaffer, 2022). Some of these people would prefer to see the money spent on green projects or other forms of government spending.

[45] This is the kind of mistake that could be corrected through experience. However, an unfortunate feature of policy selection is an intrinsic difficulty to learn through experience. Aversion to a policy due to a misperception is not easily corrected when learning is endogenous.

The intuition that dividends will be allocated to the problematic consumption category appears related to concept with a long history in behavioral economics, mental budgeting. Consumers often use fixed mental budgets for categories of consumption.[46] Meant to help manage the complexity of allocating income to many spending categories, this disciplining technique can be counterproductive and in violation of the classic consumer optimization problem.

A version of mental budgeting that we explore here involves voters or policy evaluators imagining that consumers would deposit the dividend in the mental budget allocated to the taxed category, for example, gasoline. Perhaps the dividend, as a compensatory measure, is perceived by some as buffering consumers from making necessary changes to their consumption patterns. They think consumption choices are made strictly in terms of category budgets, without regard to prices. Then, they perceive that consumers treat their regular dividend as "gas money."

This notion is well represented by Terrence Corcoran, a Canadian financial newspaper columnist, who argued that the rebate "negates the incentive" to cut emissions fostered by the tax. Corcoran's argument is cited in Rivers and Shaffer (2022), which investigates empirically how carbon dividends in British Columbia have been spent. What they discover is in sharp contrast to Corcoran's view: relatively little of dividends are spent on fossil fuels. Related work details how the Fee & Dividend tax policy in British Columbia was more effective than assumed, given its modest size, and that consumers responded to taxes far more than would be expected from evidence of responses to market price changes (Rivers and Schaufele, 2015; Murray and Rivers, 2015), findings that concur with Andersson (2019) , which evaluates a standard carbon tax in Sweden.[47]

Because Fee & Dividend has had few adopters, there is a lack of empirical research on this matter beyond British Columbia. Arguably the notion that dividends would be spent disproportionately on the taxed category is unlikely, and even less likely to

[46]Mental budgeting is part of a broader literature called mental accounting. For a deeper look: Thaler (1990, 1999); Hastings and Shapiro (2012); Farhi and Gabaix (2019).

[47]Murray and Rivers (2015) and Yamazaki (2017) also provide evidence of a negligible or slightly positive effect on the aggregate economy and employment. While a majority of the public opposed the policy ahead of time, after three years a majority of the public supported the policy.

persist over time.

To illustrate how a mental budgets notion would impact perceived effectiveness, we formally model it below.

In the absence of any policy:

$$\dot{q}_i^{carb} = q_i^{carb}(p_{carb},\ p_{clean},\ w_i)$$

$$\dot{q}_i^{clean} = q_i^{clean}(p_{carb},\ p_{clean},\ w_i)$$

In optimal response to a carbon Fee & Dividend policy :

$$\hat{q}_i^{carb} = q_i^{carb}(p_{carb} + \tau_{carb},\ p_{clean},\ w_i + \tau_{carb} * \bar{q}_{carb})$$

$$\hat{q}_i^{clean} = q_i^{clean}(p_{carb} + \tau_{carb},\ p_{clean},\ w_i + \tau_{carb} * \bar{q}_{carb})$$

Perceived response to a proposed carbon Fee & Dividend policy:

$$\tilde{q}_i^{carb} = q_i^{carb}(p_{carb} + \tau_{carb},\ p_{clean},\ w_i) + \frac{\tau_{carb} * \bar{q}_{carb}}{p_{carb} + \tau_{carb}}$$

$$\tilde{q}_i^{clean} = q_i^{clean}(p_{carb} + \tau_{carb},\ p_{clean},\ w_i)$$

Making the assumptions that $\frac{dx}{dw} > 0$ and $\frac{dx}{dp} < 0$, and that the optimal bundle under Fee & Dividend is an interior solution, we can conclude the following:

$$\hat{q}_i^{carb} < \tilde{q}_i^{carb}$$

$$\hat{q}_i^{clean} > \tilde{q}_i^{clean} > \dot{q}_i^{clean}$$

Aggregating the optimal responses we would find that:

$$\hat{Q}_i^{carb} < \dot{Q}_i^{carb}$$

$$\hat{Q}_i^{carb} < \tilde{Q}_i^{carb}$$

$$\hat{Q}_i^{clean} > \tilde{Q}_i^{clean} > \dot{Q}_i^{clean}$$

To illustrate further, let's consider a Cobb-Douglas case of imperfect substitutes:

$$U(q_{carb}, q_{clean}) = q_{clean}^{\alpha} q_{carb}^{1-\alpha} \qquad q_{clean}p_{clean} + q_{carb}p_{carb} = w$$

Since $q^*_{clean} = \alpha \frac{w}{p_{clean}}$ and $q^*_{carb} = (1-\alpha)\frac{w}{p_{carb}}$, the Cobb Douglas case is a special one. Here, the use of mental budgets, in this case αw and $(1-\alpha)w$, turn out to cause no deviation from optimality for any uncompensated changes in price, e.g., a tax. Units purchased of the taxed category will drop, but the budget in the numerator remains.

$$q^{*tax}_{clean} = \frac{\alpha w}{p_{clean}}$$

$$q^{*tax}_{carb} = \frac{(1-\alpha)w}{p_{carb} + \tau_{carb}}$$

But the dividend, which we label D, can cause a mistaken view for those thinking with mental budgets. The optimal bundle now becomes $q^{*fd}_{clean} = \frac{\alpha(w+D)}{p_{clean}}$ and $q^{*fd}_{carb} = \frac{(1-\alpha)(w+D)}{p_{carb}+\tau_{carb}}$.

How much carbon emissions would this reduce relative to no policy? (Here we use the superscript "sq" to denote the status quo in which there was no policy.)

$$\sum_i q^{*sq}_{i,carb} = \frac{(1-\alpha)\sum_i w}{p_{carb}}$$

$$\sum_i q^{*fd}_{i,carb} = \frac{(1-\alpha)(\sum_i w_i + \sum_i D)}{p_{carb} + \tau_{carb}} = \frac{(1-\alpha)(\sum_i w_i + \tau_{carb}\sum_i q^{*fd}_{i,carb})}{p_{carb} + \tau_{carb}}$$

$$\sum_i q^{*fd}_{i,carb} = \frac{(1-\alpha)\sum_i w}{p_{carb} + \tau_{carb} + \tau_{carb}(\alpha - 1)} = \frac{(1-\alpha)\sum_i w}{p_{carb} + \alpha\tau_{carb}} < \frac{(1-\alpha)\sum_i w}{p_{carb}} = \sum_i q^{*sq}_{i,carb}$$

But a mental-budgeting-afflicted consumer's bundle is:

$$\tilde{q}_{clean} = \frac{\alpha w}{p_{clean}}$$

$$\tilde{q}_{carb} = \frac{(1-\alpha)w + D}{p_{carb} + \tau_{carb}}$$

$$\tilde{q}_{clean} - q^*_{clean} = -\alpha D$$

$$\tilde{q}_{carb} - q^*_{carb} = \alpha D$$

Aggregated:

$$\sum_i \tilde{q}^{*fd}_{i,carb} = \frac{(1-\alpha)\sum_i w_i + \tau_{carb}\sum_i \tilde{q}^{*fd}_{i,carb}}{p_{carb} + \tau_{carb}}$$

$$\sum_i \tilde{q}^{*fd}_{i,carb} = \frac{(1-\alpha)\sum_i w_i}{p_{carb}} = \sum_i q^{*sq}_{i,carb}$$

The mental-budget affected voter overestimates a consumer's carbon emissions in Fee & Dividend by the portion of the dividend that would optimally be spent on clean consumption. And so we see under the Cobb-Douglas conditions, the mental budget notion turns Fee & Dividend from an emission reducer to a policy with precisely no effect on emissions.

Penalty & Reward addresses this by not having a dividend. Consider a relatively high carbon consumer. They are taxed above the population mean threshold and given no compensatory funds to allocate or misallocate. Consider a relatively low consumer who receives a payment. This payment is not compensation for a tax, because for the low consumer, no tax exists. Therefore to allocate the reward payment to additional fossil fuel consumption does not seem logical nor an application of a known behavioral phenomenon. In fact, consumers may be averse to spending rewards on fossil fuels at all since such a choice will reduce the size of the next reward.

B Appendix 2: Diagnostic Experiment

B.1 Instruction summary

- Think of each round in isolation: a randomly drawn one determines your pay.
- In any round: you coexist with 1000 robots, and all 1,001 of you are each

provided a menu of 4 objects and a 420 franc budget, always sufficient to buy either 1, 2, 3 or all 4 objects on the menu.[48]

- The only value to you of buying an object is its resale value. On each object you buy, your profit (or loss) is the resale value minus the price you paid. These rules also apply to the robots, whose resale values may differ from yours.
- Your choice is whether to accept or reject a proposal to introduce or remove a feature within the round that might help or hurt you. (See Figure 2 for an example.) Your decision about the feature is implemented, and we then automatically select the number of objects to buy that maximize your profit, given the feature decision you made and do the same for the robots, as well. You are paid your final total of francs, each worth \$0.05, for a random round.
- Next, we want you to assess the value of your choice, in francs, relative to the choice you rejected. Within the round, how many francs are you willing to exchange for your choice? Your incentives are to be completely honest and accurate here.
- More detail: 1 in 20 subjects will be randomly selected and receive 200 extra francs. If you are one of these subjects you must decide how many of these 200 francs you are willing to pay to receive your choice. A number will be drawn between 1 and 200. If the number of francs you were willing to offer is above the randomly drawn number, you will get your choice and we will deduct the value of the randomly drawn number, in francs from your final total. If the randomly drawn number between 1 and 200 is above the number of francs you are willing to offer, then you keep the 200 francs but your payout from purchasing objects will be using the Feature Choice that you rejected, not the one that you selected.

B.2 Implementation details

Subjects spent approximately 60 minutes in UC Santa Barbara's experimental economics laboratory on an incentivized experiment delivered via Qualtrics and earned

[48]Francs are the experiment's currency units to facilitate arithmetic. 20 francs = \$1.

on average \$21.86. There was no time limit given, and some subjects spent as much as 90 minutes. For the diagnostic study detailed here, 56 subjects participated.

B.3 Results Supplement

Table 4: Regressions from Diagnostic Experiment with Consumer and Distributed Effects

	Dependent variable:			
	Valuation of feature proposal			
	(1)	(2)	(3)	(4)
Constant	−3.416	5.062	−2.564	5.501
	(3.033)	(5.949)	(3.465)	(6.844)
Consumer Effect	0.978	0.866	1.155	1.113
	(0.117)	(0.119)	(0.168)	(0.202)
Distributed Effect	0.314***	0.210***	0.533***	0.500***
	(0.062)	(0.073)	(0.113)	(0.158)
Negative term magnitude (LAMA)		−0.258**		−0.043
		(0.103)		(0.173)
Policy magnitude (Policy Bias/Aversion)		0.200*		0.187
		(0.119)		(0.141)
Deviation magnitude (Default Bias/Aversion)		−0.227		−0.427
		(0.262)		(0.402)
Analysis on Subset With No Equilibrium Effect			X	X
Observations	1,680	1,680	632	632
Clusters	56	56	56	56

Notes: Result from estimating Equation 3 across all rounds in columns 1-2 and in the subset of rounds where Equilibrium Effects are absent in columns 3-4. Given the regression specifications, the rational benchmark for Consumer Effect, and Distributed Effect are 1. Rational benchmark for Negative term magnitude, Policy magnitude, and Deviation magnitude are 0. Negative coefficient for Negative term magnitude implies LAMA. *p<0.1; **p<0.05; ***p<0.01.

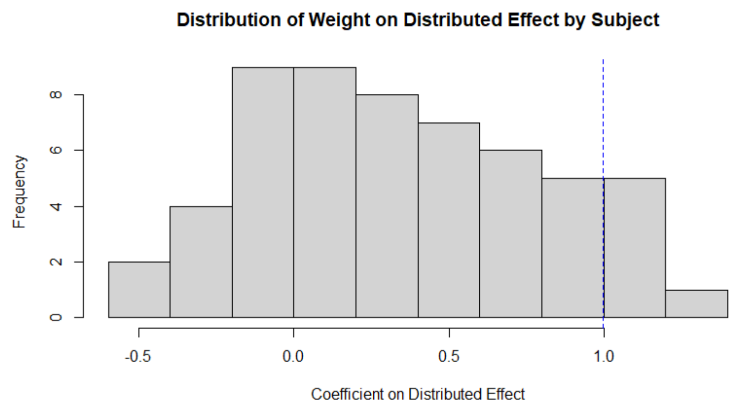


Figure 8: The distribution by individual lab subject of coefficient on Distributed Effect from regression in Equation 3.

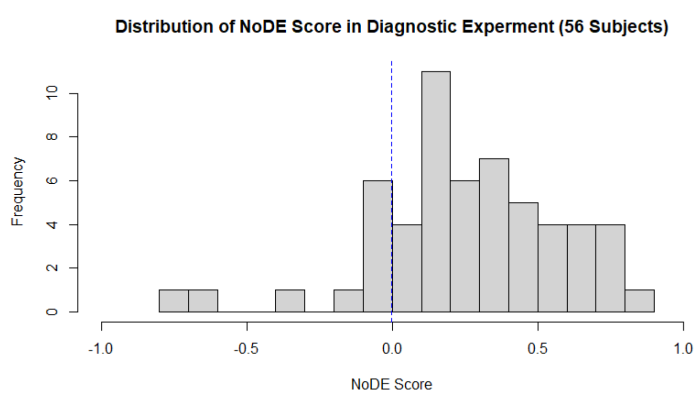


Figure 9: Distribution of subject NoDE Score applying methodology in Section 5.2 to Diagnostic Experiment data. To compare with survey participants' NoDE Score distribution, see Figure 12, lower right.

Table 5: Regressions from Diagnostic Experiment With Subeffects

	Dependent variable:	
	Valuation of feature proposal	
	(1)	(2)
Constant	−4.133	3.943
	(3.299)	(5.720)
Direct Effect	1.142	1.049
	(0.151)	(0.151)
Equilibrium Effect	0.809*	0.779*
	(0.126)	(0.151)
Distributed Direct Effect	0.523***	0.410***
	(0.103)	(0.112)
Distributed Equilibrium Effect	0.093***	−0.011***
	(0.065)	(0.069)
Negative term magnitude (LAMA)		−0.225**
		(0.101)
Policy magnitude (Policy Bias/Aversion)		0.042
		(0.105)
Deviation magnitude (Default Bias/Aversion)		−0.210
		(0.283)
Observations	1,680	1,680
Clusters	56	56

Notes: Result from estimating Equation 4 across all rounds. Given the regression specifications, the rational benchmark for Direct Effect, Equilibrium Effect, Distributed Direct Effect, and Distributed Equilibrium Effect are 1. Rational benchmark for Negative term magnitude, Policy magnitude, and Deviation magnitude are 0. Negative coefficient for Negative term magnitude implies LAMA. *p<0.1; **p<0.05; ***p<0.01.

B.4 Instructions and Comprehension Check

In this experiment, you will experience several **rounds** of choice tasks**.**

In **each round** you must imagine yourself as a resident of a country, in which the residents consist only of **you** and **1000 robots**. You are not playing with the other humans in the room today. Each of you is playing in your own decision environment that consist of just you and your own set of 1,000 robots.

In **each round**, the residents (all 1,001 of you) will each be provided a 420 Franc budget and menus of **4 objects**. (For reference, $1 = 20 Francs).

After you are done, one round is **randomly drawn**, and the choices you had made there will be used to determine your payment. This means you will **not** be able to save money from your budget from one round to the next.

Here is a first look at the process you will encounter in each round:

PART 1: SELECTING THE FEATURES OF THE ROUND **THIS WILL BE DONE FOR EACH ROUND**

FEATURE CHOICE: **Choice A** — **CHANGE** **(Examples might include: Add a New Feature or Remove an Old Feature)** **OR** **Choice B** — **NO CHANGE** **(This keeps conditions of The Starting Environment)** **STEP 1** **Selecting Your Chosen Option**	**Next** **Quantify how much MORE valuable to you is** **Your Chosen Option** **than** **The Other Option** **STEP 2** **Comparing Choices' Values**

If you pick "Choice A" it is Your Chosen Option; "Choice B" is The Other Option.

If you pick "Choice B" it is Your Chosen Option; "Choice A" is The Other Option.

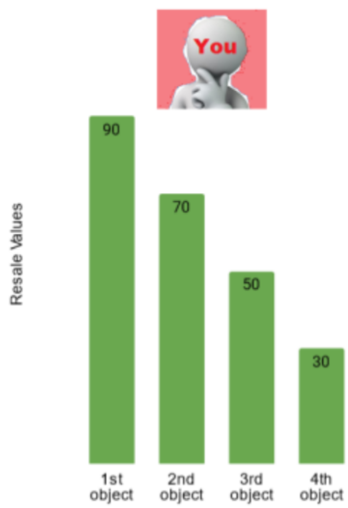
Any participant can buy 0-4 objects in their menu **in every round**, but they can also choose to buy **none**. Your budget of 420 Francs is always enough or more than enough to buy all the objects on your menu. But you may not want to buy them all.

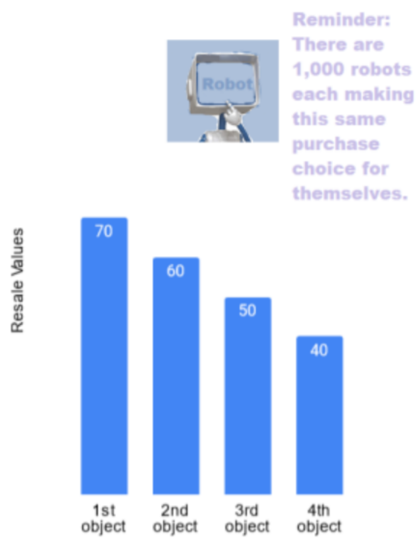
In the round that is selected for payment, whatever you don't spend on objects is money that you keep. But becuse only one round matters, money is not moved between rounds.

There are **unlimited** objects. Even if every robot buys all 4 objects on its menu, it has **no** effect on your ability to buy any or all of the objects on your menu.

Each object carries a **resale value**: this is the only worth of the object to you.
For each object that you buy, you will experience either a **profit or loss** of the **resale value** of that object **minus** the **price** you paid. Reselling is done for you automatically at the end.

EXAMPLE:





Above is only an example of resale values. They will vary in all of the different rounds you will experience.

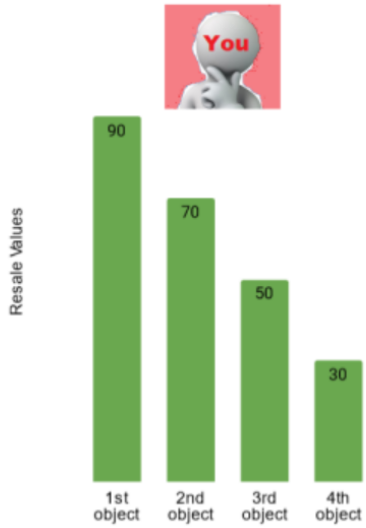
The same rules that apply to you also apply to each of the 1000 robots. But the robots could each have **different resale values** for their 1st, 2nd, 3rd and 4th object **than you**.

The things that are in **common** between you and the robots is the **price and special rules that we call features** (more about features later).

When it is time to make purchase choices, all objects, whether on your menu or on the robots menu, will carry the same price.

The **robots**, like you, would make purchase choices to **maximize their profit**. They are not concerned about helping or hurting you, just as you are not concerned about helping or hurting a robot, which is nothing more than some lines of software programming.

So what will a purchase decision look like? Let's continue with the **example** above:



Price = 55 francs

The price shown is for **each** object that you buy. Let's try it out.

Purchase Choice:

Buy 1st Object

Buy 1st and 2nd Object

Buy 1st, 2nd and 3rd Object

Buy 1st, 2nd, 3rd and 4th Object

Buy no object

Comprehension check: How much did you earn from the previous decision?

- 20 francs
- 50 francs
- 160 francs
- 195 francs
- 220 francs
- 240 francs

→

As you can see, once you buy the 1st object, which has the highest resale value, you can't buy it again. The next object to consider is 2nd object.

All robots will have the same resale values as each other, but their resale values will often be different than yours.

Step 2:
In every round, you will be asked about whether you want to **Add or Remove an added feature** to your universe.

Choose what is best for you. In the one round that we randomly select to determine your payment, there is a 95 percent chance we will automatically use **your choice** to determine whether have an added feature or don't have it. The robots don't get a vote. So your choice of feature is highly likely to determine the environment in which the purchase decision (that matters for your pay) will ultimately be made.

Let's revisit our **example**, now with a **choice** to make about a new **feature**. From now on, resale values will be visible by clicking on labeled **buttons**.

It is your choice whether to use these buttons. It is not required.

Feature Choice:

New Feature:
- Whoever purchases an object gets a 20 franc **bonus** for each object that they buy.
- The **cost** of each 20 franc bonus is divided equally and paid by the 1,001 residents of the country.

No Change

Price = 55 francs

Your Resale Values | Robot Resale Values

Assume this round will determine your payment.

Next, suppose we give you an extra 200 francs ($10).
But, instead of giving you the choice you just selected, the choice that you just **rejected** becomes the environment in which you and robots must make purchasing decisions.

Of the 200 francs you were just given, what **portion** are you **willing to trade** to get back to **Your Choice**. You keep the remaining francs.

In other words, what is it worth to you to get your choice vs. the other option.

(Honesty is the best for you here. If you'd like us to prove why, you can click the button below)

More Info

"More info" button opens following text explanation of willingness to pay mechanism:

This method ensures that the optimal answer is your honest opinion of how much more, in francs, you like one choice than the other.

Put a number above your true personal value?
You might overspend for the choice you do like.

Put a number below your true personal value?
You're likelier to get the choice you don't like.

How it works:

When you select between, for example, "Feature X" or "No Change":

There is a 95 out of 100 chance that you'll gete your choice

There is a 5 out of 100 chance that you'll get the other option that you rejected, plus 200 francs.

After you decide your value for your original choice compared to the alternative by using the slider bar, we save the number and call it Y. Next, I randomly draw a number between 1 and 200 and call it X.

If Y > X , then we give you your original choice and deduct only X, which is less than your value Y, so it's a good deal for you still.

If X > Y, then we deduct nothing, and you remain in the environment of the rejected choice, and you keep the 200 francs.

This may seem complicated, but you can ignore the details and trust that it's designed to incentivize you to put your actual true maximum willingness to pay for your choice over the alternative, instead of underreporting or overreporting that value.

Very Important: Please Keep in mind when you make feature choices, that your **final purchase choices** for the round and the robots' final purchase for the round, would be made **after the feature settings are revealed**. In other words, you would know exactly what the feature and prices are.

Finally, to make things easier, after the features are set (based very likely on your selections), we make the optimal purchase decision for you, given all the best and current information (features, prices, etc.). In other words, given the feature that you select, we will make sure your purchase choice is the best one possible.

→

Here are some comprehension questions:

In this experiment, you maximize your pay by:

Making feature choices to beating the robots.

Making feature choices that are best for your finances.

Making feature choices and purchase choices that are good for you and good for robots.

I use the slider bar to ...

Express my subjective value for my choice in the Purchase Choice segment.

Express my subjective value for my choice in the Feature Choice segment.

Express my subjective value for how much **more** **I value** my choice **than I value** the rejected choice in the Feature Choice segment.

Suppose a specific round is selected for payment. If I didn't spend all of my budget on objects ...

That money I didn't spend disappears and nobody gets it.

That leftover money gets carried over to the next round

I get to keep the money I didn't spend and add it to profit I might have made from any objects that I bought in that round. For the one round randomly selected for my pay, this total amount will determine my $ earnings.

When considering a feature choice, what should I do?

Consider potential benefits only

Consider potential costs only

Consider both potential costs and benefits by carefully reading every line in the description.

B.5 Experimental Round Catalog

The following tables detail experimental parameters for each of the potential rounds that subjects could encounter. All subjects encountered the first eight rounds, which are the simplest in structure and referred to as benchmark rounds, though they did not necessarily encounter these first. Of the remaining 29 rounds, 22 were selected at random for each subject. The order of the total 30 rounds was random.

Table 6: 37 Problems: Problem parameters

numb	dh1	dh2	dh3	dh4	dr1	dr2	dr3	dr4	p	sP	nP
1	90	60	30	0	30	20	10	0	55	0	-20
2	90	60	30	0	50	40	30	20	95	0	-40
3	50	40	30	20	110	80	50	20	55	0	20
4	40	30	20	10	60	50	40	30	65	0	-20
5	90	60	30	0	60	50	40	30	65	0	20
6	60	50	40	30	40	30	20	10	65	0	-20
7	50	40	30	20	90	70	50	30	55	0	10
8	40	30	20	10	50	40	30	20	65	0	-20
9	90	60	30	0	30	20	10	0	55	-20	0
10	50	40	30	20	110	80	50	20	55	20	0
11	40	30	20	10	70	60	50	40	25	50	40
12	40	30	20	10	60	50	40	30	25	40	20
13	90	60	30	0	60	50	40	30	65	20	0
14	50	40	30	20	30	20	10	0	35	20	0
15	50	40	30	20	90	70	50	30	55	10	0
16	90	60	30	0	60	40	20	0	45	0	10
17	90	60	30	0	120	90	60	30	65	0	20
18	90	60	30	0	120	110	100	90	85	0	-20
19	90	80	70	60	60	50	40	30	75	0	20
20	90	80	70	60	60	50	40	30	85	0	-20
21	90	80	70	60	50	40	30	20	85	0	-30
22	90	60	30	0	40	30	20	10	55	0	-20
23	50	40	30	20	90	60	30	0	55	0	-20
24	50	40	30	20	50	40	30	20	55	0	-20
25	70	50	30	10	70	60	50	40	75	0	-30
26	70	60	50	40	70	50	30	10	75	0	-20
27	40	30	20	10	70	60	50	40	45	0	20
28	40	30	20	10	60	50	40	30	45	0	20
29	30	20	10	0	60	50	40	30	55	0	-20
30	90	60	30	0	60	40	20	0	45	10	0
31	90	60	30	0	120	90	60	30	65	20	0
32	90	60	30	0	120	110	100	90	85	-20	0
33	50	40	30	20	40	30	20	10	25	30	10
34	50	40	30	20	70	60	50	40	75	-20	-10
35	70	50	30	10	40	30	20	10	25	20	10
36	20	20	20	20	30	30	30	30	55	-30	-20
37	20	20	20	20	30	30	30	30	45	-20	-10

Note: "dh1" represents the induced demand(d) of the human(s) for the first(1) object. Similarly "dr2" represents the induced demand(d) for any robot(r) for its second(2) object. "p" represents the price. "sP" represents the policy setting in the status quo, where a negative number represents a subsidy and a positive number a tax. "nP" represents the proposed policy setting.

Table 7: 37 Problems: Problem outputs

numb	sqW	CE	EE	DDE	DEE	net	final	DM	PM
1	40	40	0	0	0	40	80	20	20
2	0	0	40	0	0	40	40	40	40
3	0	0	0	40	0	40	40	20	20
4	0	0	0	0	-40	-40	-40	20	20
5	25	-20	0	0	0	-20	5	20	20
6	0	0	20	0	0	20	20	20	20
7	0	0	0	20	0	20	20	10	10
8	0	0	0	0	-20	-20	-20	20	20
9	80	-40	0	0	0	-40	40	20	-20
10	40	0	0	-40	0	-40	0	20	-20
11	0	0	0	0	40	40	40	10	-10
12	0	0	0	0	40	40	40	20	-20
13	5	20	0	0	0	20	25	20	-20
14	0	0	20	0	0	20	20	20	-20
15	20	0	0	-20	0	-20	0	10	-10
16	60	-20	0	10	0	-10	50	10	10
17	25	-20	0	40	0	20	45	20	20
18	5	20	0	-80	0	-60	-55	20	20
19	20	-40	20	0	0	-20	0	20	20
20	5	20	20	0	0	40	45	20	20
21	5	30	45	0	0	75	80	30	30
22	40	40	0	0	-20	20	60	20	20
23	0	0	20	-40	0	-20	-20	20	20
24	0	0	20	0	-40	-20	-20	20	20
25	0	0	30	0	-90	-60	-60	30	30
26	0	0	20	0	-20	0	0	20	20
27	0	0	0	60	-40	20	20	20	20
28	0	0	0	40	-40	0	0	20	20
29	0	0	0	-20	-40	-60	-60	20	20
30	50	20	0	-10	0	10	60	10	-10
31	45	20	0	-40	0	-20	25	20	-20
32	-55	-20	0	80	0	60	5	20	-20
33	0	0	20	0	10	30	30	20	-20
34	-40	0	0	20	10	30	-10	10	-10
35	30	20	0	0	10	30	60	10	-10
36	-120	0	0	40	80	120	0	10	-10
37	-80	0	0	40	40	80	0	10	-10

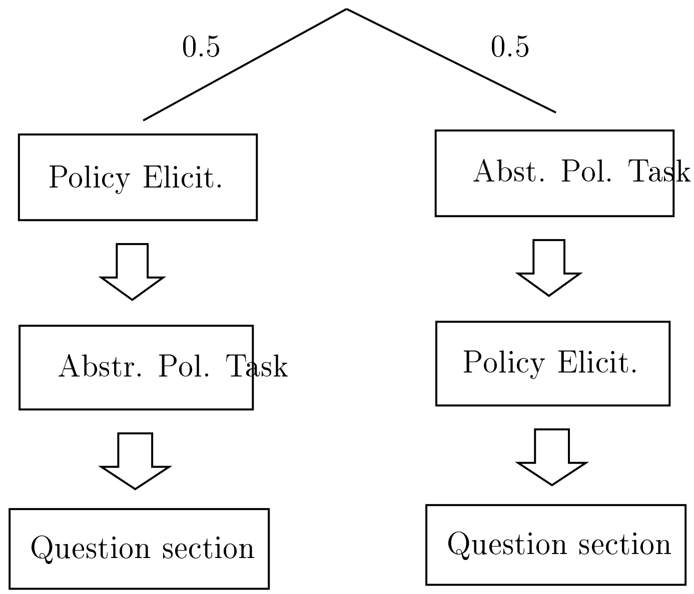
Note: The remaining variables are not shown to subjects and are calculated outputs from the inputs of the rounds in the previous table. They include "sqW" for status quo welfare, 'CE' for Consumer Effect, 'EE' for equilibrium Effect, 'DDE' for Distributed Direct Effect, and 'DEE' for Distributed Equilibrium Effect, "net" the value of the proposal, "final" the final welfare given acceptance of the proposal, and "DM" and "PM" representing the deviation magnitude term and the policy magnitude term.

C Appendix 3: Survey Experiment

C.1 Implementation details

C.1.1 Policy section

The survey consists of three main sections: Policy Elicitation, Policy Lab Task, and Questions section. One of two possible orders is randomly assigned, as shown below.



Every subject encounters all four of the rating questions for each of the two policies. The four prompts, whose sequences are represented in Figure 10, read:

- Overall support: "For each policy, provide your overall degree of support or opposition."
- Impact on emissions rating: "Please rate each policy, solely on its effectiveness for reducing emissions."
- Financial impact on you rating: "Please rate each policy, solely on its financial impact to your household."
- Financial impact on low inc. rating: "Please rate each policy solely on its financial impact on low income individuals."

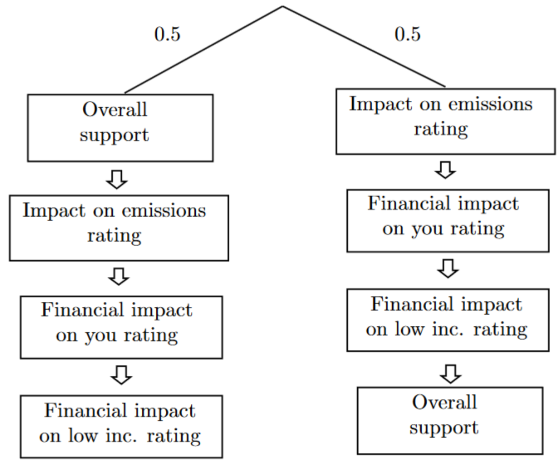


Figure 10: Policy question order: Participants are then randomly assigned to one of two question orders.

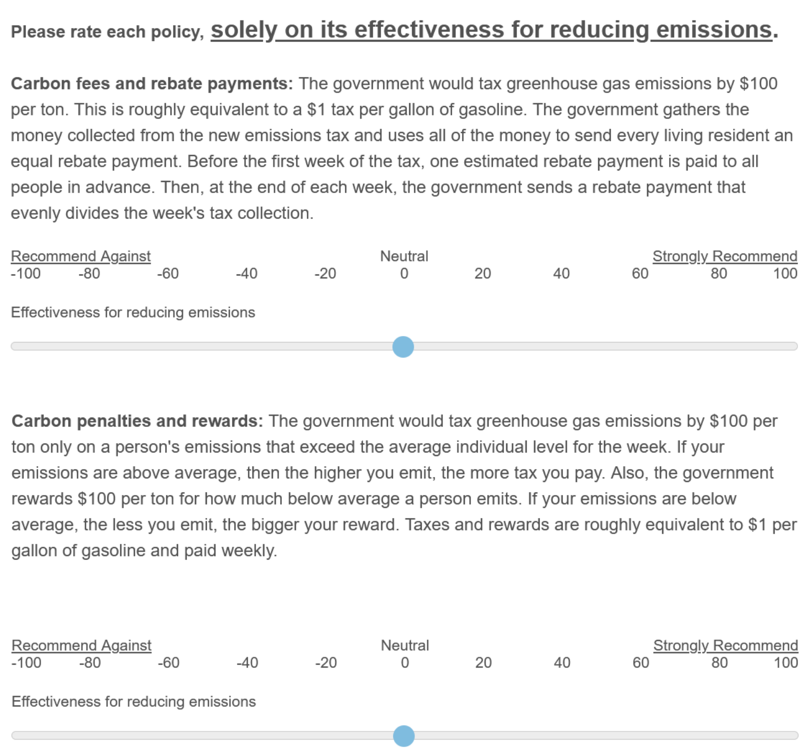


Figure 11: The layout for the rating question pertaining to effectiveness of the policy for reducing emissions. This is only one example of the factor questions that complement the overall support rating. All four questions share this layout.

C.1.2 Question Section

What follows is a list of questions that were consistently included in the exit questionnaire sections of each survey distribution. (Age, sex, party affiliation, employment and student status were provided directly through Prolific's screening questionnaire for the participants who agreed to provide and release those details.)

State of residence was asked to monitor the geographic diversity and representativeness of the sample:

In what state do you currently reside? The drop-down menu included 50 states, the District of Columbia and Puerto Rico.

The next panel of six questions measure predicted factors for support for a carbon pricing policy like Fee & Dividend or Penalty & Reward. For the six questions, we had no ex ante reason to think that any of these factors would be predictive of a difference in support between the two policies. In theory, political background, climate beliefs, personal emissions, and the other factors should affect support for the two policies identically.

Here is a 9-point scale on which the political views that people might hold are arranged from very liberal (left) to very conservative (right). Where would you place yourself on this scale? In the numerical score, higher values correspond to further right politically. This response allows a finer and more current measure of partisanship than the party affiliation variable provided through Prolific.

How confident are you in the following statement? Climate change is real and caused by human activity. A slider bar is provided ranging from 0 (Not confident at all) to 100 (Entirely confident).

How much faith and trust do you place in how the government spends money? Responses range on a scale of 1 (Not well at all) to 5 (Extremely well).

What do you guess the effect of a carbon tax on your income (from your job or other sources) would be? Choices are "Positive," "Negative" or "No effect," and the order in which these options are shown is randomized. In the analysis we transform this question into two dummy variables for "Positive" and "Negative."

It is appropriate/inappropriate for the government to sometimes redistribute some resources, including money, from the wealthy, high-earning to others, such as the poor or middle class. A slider bar ranging from -100 (Totally inappropriate) to +100 (Totally appropriate) measures the response.

Consider the average level of carbon emissions in the country. How much higher or lower do you think your emissions total is, relative to the average level? By what percent? We combine the response to the first question of "Higher" or "Lower" with the open-ended percent response to form a numerical composite variable, where responses can range from -1 (implying 100 percent less than the average level, or zero emissions) to +10 (Implying 1,000 percent, or ten times the average level of emissions).

Finally, two questions each measure additional behavioral factors that could theoretically result in difference of support between the two policies.

Do you agree with the following statement? Because carbon taxes in the "Carbon fees and rebate payments" policy make fossil fuels more expensive than before, the payments that people get back from the policy will be spent on their fossil fuel budget. (Participants are shown the text of the "Carbon fees and rebate payments" policy in case the need a reminder.) Choices are "Yes," "No," and "I don't know," which we transform into a dummy variable measuring 1 if the response was either "Yes" or "I don't know." This question is a measure of the application of mental budgeting heuristic detailed in the previous section.

Subjects are shown the text of "Carbon fees and rebate payments," and asked: **True or False? Under the policy, I should not change my behavior because the more fossil fuel tax I pay, the more I'll get back in payments.** If subjects choose "True," a dummy variable takes the value of 1. This question measures whether subjects misunderstand that the dividend in a Fee & Dividend policy is essentially uncorrelated with the taxes paid.

C.2 Results Supplement

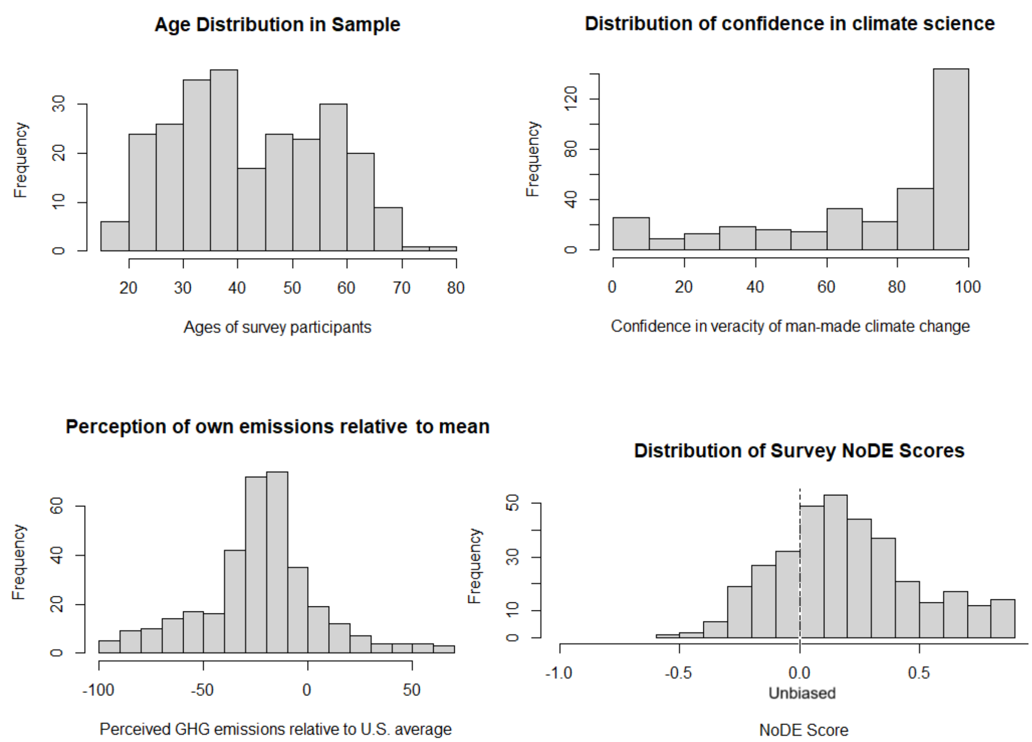


Figure 12: Some descriptive properties that characterize the sample. More participants believe themselves to be below-average emitters than above-average emitters. By far, more subjects demonstrate NoDE, or low weighting to Distributed Effects, in the incentivized, Abstract Policy Task than high weighting to Distributed Effects.

Table 8: Effect of Beh. Factors on Diff. in Support, Extended to Show Controls

	Dep. Var.: Penalty & Reward Premium		
	(1)	(2)	(3)
Constant	35.767***	4.109	−12.877
	(8.245)	(21.035)	(35.353)
NoDE Score	35.310**	29.689**	48.185***
	(14.664)	(15.028)	(17.279)
Correlation error	14.812	16.090	19.612
	(12.267)	(11.896)	(14.010)
Mental budgeting error	16.179**	15.609**	17.880**
	(7.462)	(7.487)	(8.798)
Survey order: experiment first	−13.988*	−14.349*	−18.597*
	(7.966)	(8.092)	(9.740)
"Compute" prompt	−13.092	−16.164*	−2.544
	(9.823)	(9.793)	(11.489)
"Expert" prompt	−4.498	−0.869	1.624
	(10.248)	(11.038)	(13.486)
Relative emissions		0.058	0.094
		(0.134)	(0.148)
Belief in climate change science		0.535***	0.557**
		(0.185)	(0.223)
Left right score		1.241	3.269
		(1.874)	(2.472)
Believe job benefits from carbon tax		−11.910	−6.947
		(10.259)	(11.245)
Believe job hurt from carbon tax		3.153	−3.898
		(9.784)	(11.156)
Trust government use of revenue		−4.206	−2.183
		(4.960)	(5.976)
Support for redistribution		−0.044	−0.027
		(0.096)	(0.117)
Black			−11.620
			(13.795)
Student			16.843
			(11.383)
Unemployed			−13.027
			(12.736)
Male			3.307
			(9.581)
Age			−0.008
			(0.385)
Observations	347	347	253

Notes: The table above results from estimating Equation 5 on the survey participant responses. The dependent variable is the Penalty & Reward premium, i.e. the difference in each subject's support for Penalty & Reward versus Fee & Dividend. *p<0.1; **p<0.05; ***p<0.01.

Table 9: Baseline Study: Policy Support/Opposition

	Dependent variable: Policy Rating (-100 to +100)			
	Fee & Dividend		Penalty & Reward	
	(1)	(2)	(3)	(4)
Constant	−23.447	−31.599	−23.902	−34.249
	(20.523)	(30.245)	(18.101)	(25.061)
NoDE Score	-19.792^{*}	-28.567^{**}	0.790	8.071
	(11.679)	(13.946)	(11.467)	(11.846)
Correlation error	-24.112^{**}	−18.302	1.069	3.891
	(10.294)	(13.307)	(8.689)	(10.250)
Mental budgeting error	−9.151	−6.292	2.057	6.481
	(6.887)	(8.746)	(6.275)	(6.848)
Relative emissions	-0.329^{***}	-0.346^{**}	-0.317^{***}	-0.371^{***}
	(0.121)	(0.139)	(0.113)	(0.120)
Belief in climate change science	0.241	0.260	0.813^{***}	0.817^{***}
	(0.176)	(0.201)	(0.164)	(0.171)
Left right score	−1.608	−1.596	1.198	2.454
	(1.786)	(2.650)	(1.295)	(1.587)
Believe job benefits from carbon tax	10.166	8.970	−2.843	0.342
	(9.971)	(11.709)	(7.399)	(8.359)
Believe job hurt from carbon tax	-19.210^{**}	−15.477	-17.656^{**}	-18.300^{**}
	(8.563)	(9.496)	(7.009)	(7.502)
Trust government use of revenue	7.947	7.930	3.380	8.368^{*}
	(4.911)	(6.170)	(3.822)	(4.375)
Support for redistribution	0.087	0.082	0.064	0.084
	(0.084)	(0.105)	(0.068)	(0.075)
Survey order: experiment first	7.629	16.452^{*}	−9.783	−5.293
	(7.180)	(8.793)	(6.368)	(7.120)
Black		25.485^{*}		10.187
		(13.232)		(7.535)
Student		−6.352		1.209
		(10.805)		(7.606)
Unemployed		22.917^{*}		1.708
		(12.209)		(7.291)
Male		−9.638		−7.363
		(9.766)		(7.493)
Age		0.071		−0.199
		(0.304)		(0.307)
Observations	248	184	248	184
R^2	0.257	0.282	0.341	0.416
Adjusted R^2	0.223	0.213	0.310	0.360

Notes: The table above results from estimating a set of explanatory variables on elicitation of support for each policy, Fee & Dividend in columns 1 and 2, and Penalty & Reward in columns 3 and 4. This estimation is on the baseline participants not receiving a supplemental information treatment. *p<0.1; **p<0.05; ***p<0.01.

Table 10: Predictiveness of Impact factors on overall policy support rating

	Dependent variable: Overall Support (-100 to +100)	
	Fee & Dividend	Penalty & Reward
	(1)	(2)
Constant	-5.274^{**}	3.470
	(2.078)	(3.179)
Emissions impact rating (-100 to +100)	0.398^{***}	0.430^{***}
	(0.058)	(0.075)
Your Finances impact rating (-100 to +100)	0.290^{***}	0.341^{***}
	(0.072)	(0.087)
Low Income impact rating	0.246^{***}	0.202^{***}
	(0.061)	(0.068)
Observations	347	347
R^2	0.608	0.597
Adjusted R^2	0.604	0.594

Notes: The Emissions impact rating was solicited by asking for both policies,"Please rate each policy, solely on its effectiveness for reducing emissions." Financial impact rating asked: "Please rate each policy, solely on its financial impact to your household." And Low Income impact rating: "Please rate each policy solely on its financial impact on low income individuals." *p<0.1; **p<0.05; ***p<0.01.

Table 11: Wilcoxon Rank Sum Test With Continuity Correction (p-value)

Treatment	P&R = F&D	Diff(HNoDE)=Diff(LNoDE)	Diff(Left)=Diff(Right)	P&R(Right)=F&D(Left)
BASELINE	$< 2e-16^{***}$	0.0002^{***}	0.0367^{**}	0.0044^{***}
T:COMPUTE	0.0154^{**}	0.0338^{**}	0.3429	0.151
T:EXPERT	0.0006^{***}	0.1897	0.1992	0.5762

The p-values, $< 2e-16^{***}$ and 0.0002^{***} correspond to Hypotheses 1 (Penalty & Reward is better supported than Fee & Dividend) and Hypothesis 2 (Penalty Reward Premium is higher among high NoDE than among low NoDE subjects), respectively.

Table 12: Robustness to Single Policy Presentation (Between Subjects)

	Dependent variable:
	Support for Policy
Constant (Fee & Dividend)	-18.814^{***}
	(6.580)
Fee & Dividend With "Compute" Treatment	10.814
	(13.719)
Fee & Dividend With "Expert" Treatment	13.510
	(10.962)
Penalty & Reward	37.826^{***}
	(8.159)
Observations	331

Notes: In Table 12, above, subjects either encountered a basic Fee & Dividend policy (represented as the constant term in the regression), the "Compute" assisted Fee & Dividend policy (represented as the first treatment dummy in the regression), the "Expert" assisted Fee & Dividend policy (the second treatment dummy), or Penalty & Reward (the third treatment dummy). Higher support for Penalty & Reward is clearly robust to sequential policy presentation, but the magnitude of the gap in support intensity is a more modest 38 points than the 49 point gap observed in the simultaneous-presentation sample discussed in Section 6. While not statistically significant, the effect magnitude of information treatments is modestly larger in the robustness sample. *p<0.1; **p<0.05; ***p<0.01

C.3 Full Survey

Now, you will be asked to consider some policies.

Please read the following policies completely. You will answer several questions about both of them.

Carbon fees and rebate payments: The government would tax greenhouse gas emissions by $100 per ton. This is roughly equivalent to a $1 tax per gallon of gasoline. The government gathers the money collected from the new emissions tax and uses all of the money to send every living resident an equal rebate payment. Before the first week of the tax, one estimated rebate payment is paid to all people in advance. Then, at the end of each week, the government sends a rebate payment that evenly divides the week's tax collection.

Carbon penalties and rewards: The government would tax greenhouse gas emissions by $100 per ton only on a person's emissions that exceed the average individual level for the week. If your emissions are above average, then the higher you emit, the more tax you pay. Also, the government rewards $100 per ton for how much below average a person emits. If your emissions are below average, the less you emit, the bigger your reward. Taxes and rewards are roughly equivalent to $1 per gallon of gasoline and paid weekly.

For each policy, provide your overall degree of support or opposition.

Carbon fees and rebate payments: The government would tax greenhouse gas emissions by $100 per ton. This is roughly equivalent to a $1 tax per gallon of gasoline. The government gathers the money collected from the new emissions tax and uses all of the money to send every living resident an equal rebate payment. Before the first week of the tax, one estimated rebate payment is paid to all people in advance. Then, at the end of each week, the government sends a rebate payment that evenly divides the week's tax collection.

Recommend Against | Neutral | Strongly Recommend

-100 -80 -60 -40 -20 0 20 40 60 80 100

Overall rating

Carbon penalties and rewards: The government would tax greenhouse gas emissions by $100 per ton only on a person's emissions that exceed the average individual level for the week. If your emissions are above average, then the higher you emit, the more tax you pay. Also, the government rewards $100 per ton for how much below average a person emits. If your emissions are below average, the less you emit, the bigger your reward. Taxes and rewards are roughly equivalent to $1 per gallon of gasoline and paid weekly.

Recommend Against | Neutral | Strongly Recommend

-100 -80 -60 -40 -20 0 20 40 60 80 100

Overall rating

Please rate each policy, <u>solely on its effectiveness for reducing emissions</u>.

Carbon fees and rebate payments: The government would tax greenhouse gas emissions by $100 per ton. This is roughly equivalent to a $1 tax per gallon of gasoline. The government gathers the money collected from the new emissions tax and uses all of the money to send every living resident an equal rebate payment. Before the first week of the tax, one estimated rebate payment is paid to all people in advance. Then, at the end of each week, the government sends a rebate payment that evenly divides the week's tax collection.

Recommend Against | Neutral | Strongly Recommend
-100 -80 -60 -40 -20 0 20 40 60 80 100

Effectiveness for reducing emissions

Carbon penalties and rewards: The government would tax greenhouse gas emissions by $100 per ton only on a person's emissions that exceed the average individual level for the week. If your emissions are above average, then the higher you emit, the more tax you pay. Also, the government rewards $100 per ton for how much below average a person emits. If your emissions are below average, the less you emit, the bigger your reward. Taxes and rewards are roughly equivalent to $1 per gallon of gasoline and paid weekly.

Recommend Against | Neutral | Strongly Recommend
-100 -80 -60 -40 -20 0 20 40 60 80 100

Effectiveness for reducing emissions

Please rate each policy, <u>solely on its financial impact to your household.</u>

Carbon fees and rebate payments: The government would tax greenhouse gas emissions by $100 per ton. This is roughly equivalent to a $1 tax per gallon of gasoline. The government gathers the money collected from the new emissions tax and uses all of the money to send every living resident an equal rebate payment. Before the first week of the tax, one estimated rebate payment is paid to all people in advance. Then, at the end of each week, the government sends a rebate payment that evenly divides the week's tax collection.

Recommend Against | Neutral | Strongly Recommend
-100 -80 -60 -40 -20 0 20 40 60 80 100

Financial impact to your household

Carbon penalties and rewards: The government would tax greenhouse gas emissions by $100 per ton only on a person's emissions that exceed the average individual level for the week. If your emissions are above average, then the higher you emit, the more tax you pay. Also, the government rewards $100 per ton for how much below average a person emits. If your emissions are below average, the less you emit, the bigger your reward. Taxes and rewards are roughly equivalent to $1 per gallon of gasoline and paid weekly.

Recommend Against | Neutral | Strongly Recommend
-100 -80 -60 -40 -20 0 20 40 60 80 100

Financial impact to your household

Please rate each policy, <u>solely on its financial impact to low-income individuals.</u>

Carbon fees and rebate payments: The government would tax greenhouse gas emissions by $100 per ton. This is roughly equivalent to a $1 tax per gallon of gasoline. The government gathers the money collected from the new emissions tax and uses all of the money to send every living resident an equal rebate payment. Before the first week of the tax, one estimated rebate payment is paid to all people in advance. Then, at the end of each week, the government sends a rebate payment that evenly divides the week's tax collection.

Recommend Against | Neutral | Strongly Recommend
-100 -80 -60 -40 -20 0 20 40 60 80 100

Financial impact to low-income people

Carbon penalties and rewards: The government would tax greenhouse gas emissions by $100 per ton only on a person's emissions that exceed the average individual level for the week. If your emissions are above average, then the higher you emit, the more tax you pay. Also, the government rewards $100 per ton for how much below average a person emits. If your emissions are below average, the less you emit, the bigger your reward. Taxes and rewards are roughly equivalent to $1 per gallon of gasoline and paid weekly.

Recommend Against | Neutral | Strongly Recommend

-100 -80 -60 -40 -20 0 20 40 60 80 100

Financial impact to low-income people

Now comes a section in which you can earn some extra real money by choosing well.

You will experience 15 independent **rounds.**
(Nothing from one round can affect anything in another round.)

Only one round, which is randomly selected, will be used to determine if you will earn a bonus payment.

In **each round**, you and 1,000 others are called Franklians.
Money is measured in units called francs.

The 1,000 other Franklians are **robots**.

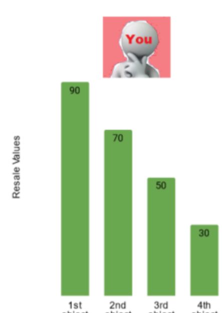
All 1,001 Franklians will each be provided
a 500 franc budget and the option to buy some objects.

Any Franklian (you and the 1,000 robots)
can buy 0, 1, 2, 3, or 4 objects **in any round.**

Everyone's budget is always enough
to buy up to the limit of 4 objects,
and the supply of objects never runs out.

Why might you buy an object?
Each object carries a **resale value**,
which is the only worth of the object to you.

AN EXAMPLE OF YOUR RESALE VALUES:



Price = 55 francs

These resale values can vary between rounds.

Your **profit (or loss)** from **buying** any object
is the **resale value** of that object
minus the **price** you paid.

Reselling is done for you automatically at the end of the round.
The money from reselling any objects that you may have bought
plus unspent money from your budget will go into your **Final Money**.

Each object has the same price. In the example, it's 55 francs per object.

Let's try a purchase decision, using the resale values shown above.

Purchase Choice:

- Buy 1st Object
- Buy 1st and 2nd Object
- Buy 1st, 2nd and 3rd Object
- Buy 1st, 2nd, 3rd and 4th Object
- Buy no object

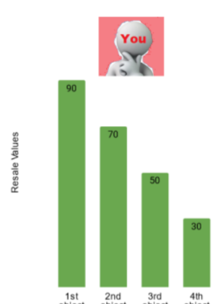
Comprehension check: How much did you earn from the previous decision?

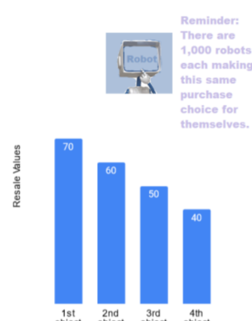
- 20 francs
- 50 francs
- 160 francs
- 195 francs
- 220 francs
- 240 francs

Great! So now you know what buying
(or not buying) objects is like.

The same rules that apply to you
also apply to each of the 1000 robots,
who can also buy objects and affect
their own **Final Money**.

But the robots could (and probably do) have **different resale values**
for their own 1st, 2nd, 3rd and 4th objects than you do.
As a result they may often make different purchase choices than you.





Each robot, like you, would make purchase choices
to maximize their own **Final Money**.
They are not interested in helping or hurting you.

The things that is in common between you
and all of the the robots is the **price per object**
and special rules called **Features.**

A Feature changes the cost of buying an object,
and can affect your **Final Money** in many ways.
You should read all lines of a Feature carefully.

In every round, you will have to make a **Feature Choice**:
Whether you want to **Add** or **Remove** a **Feature**.

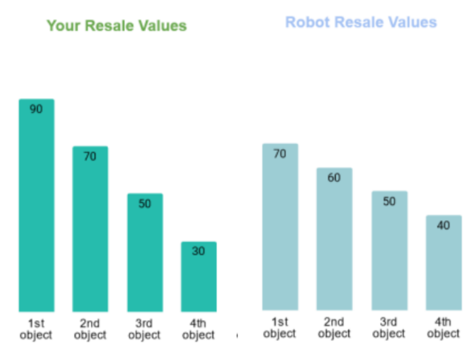
Your choice applies just for that round.

The Feature Choice is the decision on which you will be judged.
Once you decide about the Feature, your choice will be the law for that round.

Then, we will automatically make the best purchase choice
of 0,1,2,3, or 4 objects for you and all of the robots.
(We will only purchase an object, for you or for a robot,
if the resale value of that object is a bigger number
than the total cost of buying an object.)

Let's revisit our example:

Price = 55 francs



Feature Choice:

New Feature:
- Whoever purchases an object gets a 20 franc bonus for each object that they buy.
- The 20 franc cost of each bonus is divided equally and paid for by all 1,001 Franklians, including you.

No Change

Note that the Feature had two lines,
both of which would ultimately affect your **Final Money**.

If you are randomly selected to be potentially paid a bonus,
we will select **one round** at random.
Your pay is completely decided by your **Feature Choice** in the selected round.

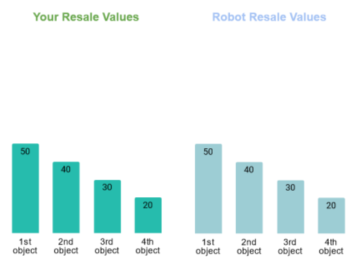
If you make a Feature Choice that results
in a larger amount of **Final Money** than the other choice,
you will earn a **$1 bonus**.

Note: in any round, the subject first sees the initial scenario alone, then must click "Next" to move to a slide where the initial scenario is accompanied by the Feature Choice being offered in the round. To avoid unnecessary duplication, the images below do not include the initial slide, only the Feature Choice slide, which includes the content of that round's initial slide (resale values, price, and any other initial features.)

Now the real series of rounds begins.

Here is a new scenario.

Price = 55 francs



Feature Choice:

New Feature:
- Whoever purchases an object gets a 20 franc bonus for each object that they buy.
- The 20 franc cost of each bonus is divided equally and paid for by all 1,001 Franklians, including you.

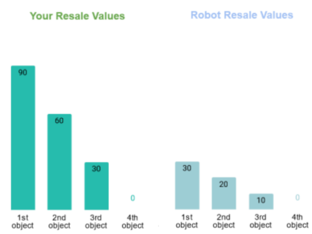
No Change

Here is a new scenario.

Price = 55 francs

Existing Feature:

- Whoever purchases an object gets a 20 franc bonus for each object that they buy.
- The 20 franc cost of each bonus is divided equally and paid for by all 1,001 Franklians, including you.



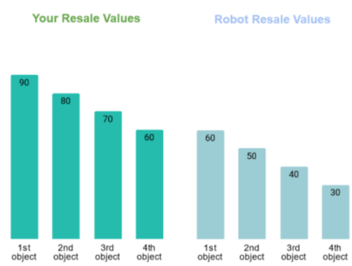
Feature Choice:

Remove 'Existing Feature'

No Change

Here is a new scenario.

Price = 75 francs



Feature Choice:

New Feature:
- Whoever purchases an object pays an additional 20 franc fee for each object that they buy.
- Every 20 franc fee collected is divided and delivered as an equal payment to all 1,001 Franklians, including you.

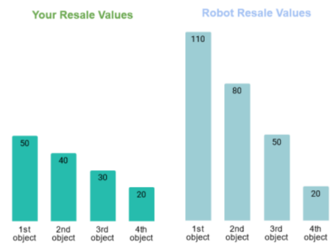
No Change

Here is a new scenario.

Price = 55 francs

Existing Feature:

- Whoever purchases an object pays an additional 20 franc fee for each object that they buy.
- Every 20 franc fee collected is divided and delivered as an equal payment to all 1,001 Franklians, including you.



Feature Choice:

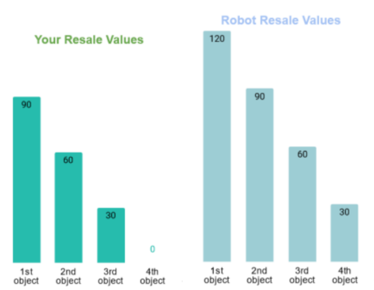
- Remove 'Existing Feature'
- No Change

Here is a new scenario.

Price = 65 francs

Existing Feature:

- Whoever purchases an object pays an additional 20 franc fee for each object that they buy.
- Every 20 franc fee collected is divided and delivered as an equal payment to all 1,001 Franklians, including you.

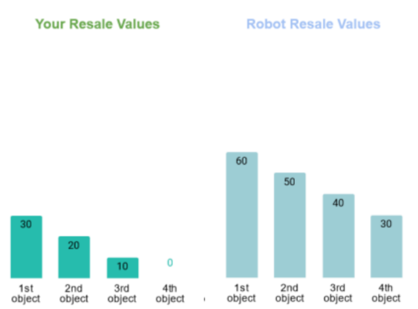


Feature Choice:

- Remove 'Existing Feature'
- No Change

Here is a new scenario.

Price = 55 francs



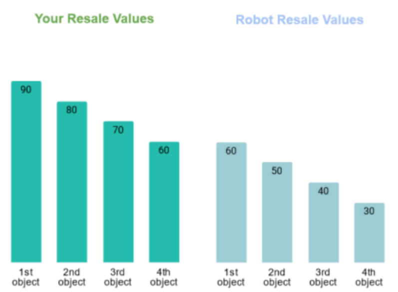
Feature Choice:

New Feature:
- Whoever purchases an object gets a 20 franc bonus for each object that they buy.
- The 20 franc cost of each bonus is divided equally and paid for by all 1,001 Franklians, including you.

No Change

Here is a new scenario.

Price = 85 francs



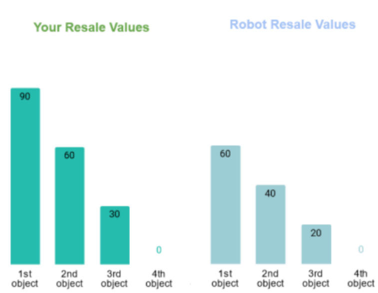
Feature Choice:

New Feature:
- Whoever purchases an object gets a 20 franc bonus for each object that they buy.
- The 20 franc cost of each bonus is divided equally and paid for by all 1,001 Franklians, including you.

No Change

Here is a new scenario.

Price = 45 francs



Feature Choice:

New Feature:
- Whoever purchases an object pays an additional 10 franc fee for each object that they buy.
- Every 10 franc fee collected is divided and delivered as an equal payment to all 1,001 Franklians, including you.

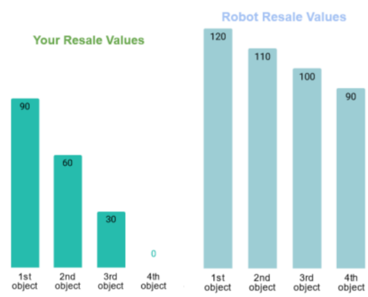
No Change

Here is a new scenario.

Price = 85 francs

Existing Feature:

- Whoever purchases an object gets a 20 franc bonus for each object that they buy.
- The 20 franc cost of each bonus is divided equally and paid for by all 1,001 Franklians, including you.



Feature Choice:

Remove 'Existing Feature'

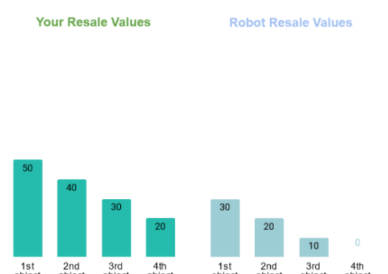
No Change

Here is a new scenario.

Price = 35 francs

Existing Feature:

- Whoever purchases an object pays an additional 20 franc fee for each object that they buy.
- Every 20 franc fee collected is divided and delivered as an equal payment to all 1,001 Franklians, including you.



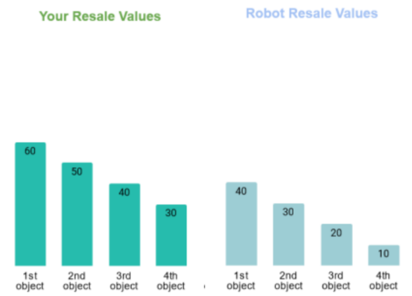
Feature Choice:

Remove 'Existing Feature'

No Change

Here is a new scenario.

Price = 65 francs



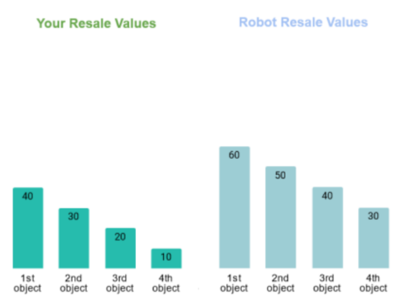
Feature Choice:

New Feature:
- Whoever purchases an object gets a 20 franc bonus for each object that they buy.
- The 20 franc cost of each bonus is divided equally and paid for by all 1,001 Franklians, including you.

No Change

Here is a new scenario.

Price = 65 francs



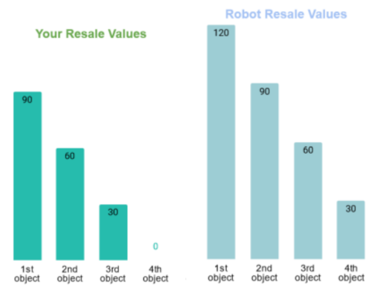
Feature Choice:

New Feature:
- Whoever purchases an object gets a 20 franc bonus for each object that they buy.
- The 20 franc cost of each bonus is divided equally and paid for by all 1,001 Franklians, including you.

No Change

Here is a new scenario.

Price = 65 francs



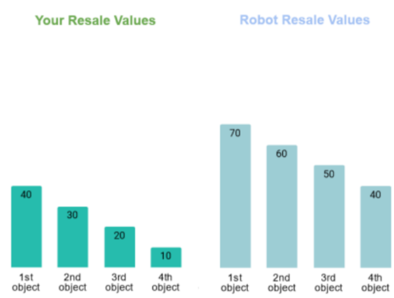
Feature Choice:

New Feature:
- Whoever purchases an object pays an additional 20 franc fee for each object that they buy.
- Every 20 franc fee collected is divided and delivered as an equal payment to all 1,001 Franklians, including you.

No Change

Here is a new scenario.

Price = 45 francs



Feature Choice:

New Feature:
- Whoever purchases an object pays an additional 20 franc fee for each object that they buy.
- Every 20 franc fee collected is divided and delivered as an equal payment to all 1,001 Franklians, including you.

No Change

In which state do you currently reside?

Here is a 9-point scale on which the political views that people might hold are arranged from very liberal (left) to very conservative (right). Where would you place yourself on this scale?

Very liberal								Very conservative
1	2	3	4	5	6	7	8	9

Political Ideology

How confident are you in the following statement? Climate change is real and caused by human activity.

0 10 20 30 40 50 60 70 80 90 100

Your confidence level

How much faith and trust do you place in how the government spends money?

- Not well at all
- Slightly well
- Moderately well
- Very well
- Extremely well

Consider the average level of carbon emissions in the country. How much higher or lower do you think **your** emissions total is, relative to the average level?

- Higher
- Lower

By what percent?

Do you agree with the following statement?
Because carbon taxes in the "Carbon fees and rebate payments" policy make fossil fuels more expensive than before, the payments that people get back from the policy will be spent on their fossil fuel budget.

- Agree
- Disagree
- I don't know

True or False?
Under the policy, I should not change my behavior because the more fossil fuel tax I pay, the more I'll get back in payments.

- True
- False

What do you guess the effect of a carbon tax on your income (from your job or other sources) would be:

- Positive
- Negative
- No major effect

It is **appropriate/inappropriate** for the government to sometimes redistribute some resources, including money, from the wealthy, high-earning to others, such as the poor or middle class.

Totally inappropriate Totally appropriate
-100 -80 -60 -40 -20 0 20 40 60 80 100

D Appendix 4: Implementing Penalty & Reward

When a single corporation or administrator oversees the usage of a the good or service, as is the case for local toll roads, electricity, and water providers, administering congestion and conservation pricing using Penalty & Reward is straightforward. With records of individual usage, no additional data gathering is required. Penalty & Reward can offer a popular, transparent way to accomplish the goal of reducing usage at peak times by helping people see that (1) the goal of the policy is not to "price-gouge" or obtain revenue, and (2) they may stand to gain financially from the policy.

With carbon, the consumer does not usually transact with a sole producer, making observability more difficult. For this reason, Fee & Dividend, with taxes levied at the production of emissions, is considered ideal for implementation.[49] Still, a version of Penalty & Reward can be implemented using carbon credits.[50] Every consumer is issued a quantity of carbon credits determined by the population average consumption level in the nearest prior period. These credits can be thought of as coupons that exempt purchases from the carbon tax at the time of purchase or redeemable at any time for equivalent cash value.[51] This implementation of Penalty & Reward is not game-able. In Penalty & Reward, the credits are based on the population mean consumption in a prior period, a figure no sole consumer can meaningfully influence.

As an example, consider the taxation of transportation emissions through gasoline taxes. Consumers could be issued a card that they use to exempt themselves, up to a point, from taxes on gasoline purchases. Using the card is unnecessary and merely an ex ante acceptability tool, as credits can also be immediately redeemed as cash for their tax savings value. Credits, a form of cash, can be consolidated between consumers. These credits are, in effect, the dividend payment of the prior period.

The approach could be further enhanced by informing consumers of the carbon

[49] There may be disadvantages to this approach, however, given that tax salience to consumers can deliver outsize reductions in externalities relative to commensurate effects of passed-through prices alone. Evidence can be found in Andersson (2019) and Murray and Rivers (2015).

[50] To distinguish from cap-and-trade, a credit's price is fixed, not quantity.

[51] In the language of Fee & Dividend, these coupons are essentially dividends - they are redeemable for the equivalent amount of money. But perceptually they can also be more readily perceived as buffering consumers from taxes, up to a threshold.

emissions embedded in broader categories of purchases, as consumers appear to be unaware and respond to this information when provided (Muller et al., 2019; Lohmann et al., 2022). If all purchases accounted for embedded emissions, the tax and environmental costs would be more salient and it would be possible to use carbon credits in a wider range of consumer settings.

Finally, Penalty & Reward can be used as a more effective way of communicating a Fee & Dividend policy. Often policies, including tax policies, are communicated in a way that does not convey literally the somewhat more opaque and complex implementation mechanism (Abeler and Jäger, 2015). Communicating Penalty & Reward, while implementing Fee & Dividend, may not be considered misleading, given the fact that they are isometric and a short experience window could clarify this fact for individuals who doubt it.