The complementarity of low taxes and pro-social guidelines when polluters have moral preferences

Caffera, Marcelo and Chávez, Carlos and Lopez, Carolina and Murphy, James J and Briozzo, Juan

Universidad de Montevideo, Universidad de Talca, University of Alaska

2025

The complementarity of low taxes and pro-social guidelines when polluters have moral preferences

Marcelo Caffera [a], Carlos Chávez [b,c], Carolina López [a], James J. Murphy [d], Juan Briozzo[a]

[a] Departamento de Economía, Universidad de Montevideo, Uruguay
[b] Escuela de Ingeniería Comercial, Facultad de Economía y Negocios, Universidad de Talca, Chile
[c] Interdisciplinary Center for Aquaculture Research, Chile
[d] Department of Economics, University of Alaska Anchorage, USA

Abstract: We present the results of a series of public-bad laboratory experiments in which we assess whether a salient message suggesting pro-social behavior with an implicit moral appeal, and a tax that is insufficient to induce the optimal level of the externality, can complement each other when implemented jointly. Our results suggest that, on average, (a) behavior is consistent with subjects having moral preferences, (b) a salient message suggesting pro-social behavior can be effective, (c) preferences are non-separable from the choice of instrument (i.e, the tax crowds-out part of the subjects´ moral preferences), and crucially, (d) the tax and the informative message do not complement each other. The tax has a greater impact on reducing the externality than the prosocial guideline, even though the tax was only half of that needed to reach the socially optimal level. Nevertheless, when implemented together, the total effect of both instruments is similar to that of the tax alone. This result is stronger for those subjects that are more "nudgeable" by the prosocial guideline. These results challenge the policy recommendation that nudges can effectively complement low taxes while awaiting the political will to raise them.

Keywords: Economic experiment, nudge, prosocial guideline, public bad, tax.

Acknowledgements: funding for this study came from Agencia Nacional de Investigación e Innovación (ANII), grants FCE_1_2019_1_156336 and POS_FCE_2020_1_1009209. Chávez acknowledges partial funding by ANID/FONDAP/1523A0007 and project Fondecyt Regular 1230266.

1 Introduction

Due in no small part to political-economy reasons, prices for pollution are frequently lower than needed to achieve the environmental goals (Cherry et a., 2014; Thaler and Sunstein, 2008). Carbon pricing is an important example of this. Implemented carbon prices are lower than the levels needed to induce an abatement of greenhouse gases (GHG) sufficient to avoid exceeding the 2ºC target of the Paris Agreement, according to recent estimates of their impact on GHG emissions (van den Bergh and Savin, 2021; Lilliestam et al., 2021). This situation has led several economists to suggest that carbon prices should be part of a broader set of instruments (Carattini et al., 2018; Blanchard, et al., 2023; Stiglitz, 2024; Sterner et al., 2024).

Several authors suggested that nudges could play a significant role in complementing prices for pollution when these are low (Thaler and Sunstein, 2008; Carlsson and Johansson-Stenman, 2019; Carlsson et al., 2021). In fact, so-called green nudges have already made their way into the environmental and energy policies (Gravert and Olsson Collentine, 2021; Carlsson et al., 2021). Nevertheless, the recommendation to complement prices for pollution with nudges still faces challenges. First, although nudge-like interventions do work on average (Mertens et al., 2022) and there is evidence that they are cost-effective (Hahn et al., 2024), their effectiveness can vary to a great degree (DellaVigna and Linos, 2022), and we know little about what type of nudges work under what conditions (Szaszi et al., 2018; 2022). Second, we know less about the complementarity of nudge-like policy interventions and economic incentives. On the one hand, comprehensive theoretical models which focus on the interactions of a nudge and an efficiency-enhancing price have yet to be developed.[1] On the other, the empirical evidence regarding the complementarity of nudges and prices on negative externalities is limited and inconclusive about the size and the direction of the synergy (Drews et al., 2020). Some of the studies show that a nudge could add to the effect of a price increase (Hernández et al., 2024; Panzone et al., 2021; Hilton et al. 2014; Mizobuchi and Takeuchi, 2013; Spraggon and Oxoby, 2010; Ambec et al., 2024). Others show that prices totally crowd out the effect of the nudge (Sudarshan, 2017; Mackay et al., 2019), or vice versa (Dolan and Metcalfe, 2015). Finally, other works find perfect complementarity (zero synergy; Fanghella, et al., 2021; Schall et al., 2016). (See section 2 for a more detailed summary of the literature). The lack of evidence on the complementarity of prices and nudges in the control of a negative externality is particularly binding in the domain of the interactions of nudges and environmental taxes. Most of this evidence corresponds to the

[1] In an early effort, Stern (1999) provides a conceptual framework, listing the factors that affect the pro-environmental behavior of consumers and its policy implications. He concludes that, due to their synergy, providing consumers with information as well as material incentives in combination may have a greater effect than the sum of their own effects. With a caveat: it is only "once incentives are large enough for consumer to take it seriously" (p. 474) that it may be more effective to invest in information than to increase the incentive.

complementarity between a tariff increase and a nudge informing households what the level of consumption of water and electricity of similar households is.

To the best of our knowledge, ours is the first paper to assess whether a nudge-like intervention could complement a low tax on an externality. Moreover, it does so in the light of a theoretical model in which economic incentives may crowd out social preferences. There is ample evidence of motivational crowding by economic incentives (Frey, 1992; Frey and Oberholzer-Gee, 1997; Frey and Stutzer, 2008; Bowles and Polania-Reyes, 2012). If economic incentives may crowd out the same social preferences that the nudge intents to trigger, moral nudges may either have no moral preferences left to trigger or recover some of the preferences crowded out by the tax. This is an empirical question. In any case, allowing for non-separability between economic incentives and social preferences is crucial to inform a regulator about the synergy between a nudge and a tax on controlling a negative externality.

To assess the degree of complementarity between a nudge and a price, we conduct a series of public bad laboratory experiments and test whether the joint effect of (a) a message informing subjects what the optimal level of a negative externality is, and (b) a tax that is insufficient to induce the optimal level of this externality, is higher or lower than the effect of the tax and the nudge standing alone. Providing information is one of the most important means of nudging people (Sunstein, 2014). Information provision can take several forms. A commonly used type of message is an injunctive one: to inform what other people are doing and communicate approval or disapproval to the person´s relative behavior. Most of the messages in the most relevant literature has this feature. Notwithstanding, the information that consumers and citizens regularly receive about the environment comes in the form of information about the state of the environment and the environmental impacts of consumption choices. In these messages, explicitly or implicitly, citizens receive tips or prescriptions on how to behave to avoid these negative impacts on the environment or on other citizens, in a significant and welfare decreasing way. An example could be the recommendations for the efficient use of heating stoves provided by local air quality offices under poor air quality conditions in urban, central-southern Chile in winter. Another example could be the "Every drop count" campaign for water conservation in India. Messages of "suggested play" have proved to be effective in the presence of heterogeneous preferences over the public good in question (Marks et al., 1999; Croson and Marks, 2001) and particularly, when it is combined with moral suasion (Dal Bó and Dal Bó, 2014). Our message is a simple, informative nudge, mimicking this situation. It does not have an explicit moral appeal. Nevertheless, it implicitly calls the subject to behave consistently with the group´s welfare. It could be considered an informative message of suggested behavior with an implicit moral appeal based on utilitarianism.

Our results show that our average subject (a) behaves consistently with having moral preferences, (b) is "nudgeable" by this type of message, (c) exhibits preferences that are non-separable from the choice of a tax as instrument (the tax crowds-out part of its moral preferences), and crucially, (d) the tax and the nudge do not complement each

other. More precisely, we cannot reject that the effect of both instruments implemented jointly is the same as that of a low tax implemented alone. Given that the effects of the two instruments when implemented separately are not zero, we say that there is a negative synergy between the two instruments. This effect is particularly observed among individuals with "stronger" baseline social preferences, who were more influenced by the nudge. This evidence is consistent with the hypothesis that the message cannot recover the moral preferences crowded out by the tax.

The paper proceeds as follows. In section 2, we present the most relevant literature. Section 3 presents the theoretical model used to conduct this study and the hypotheses that guide our research. Section 4 presents the hypotheses and section 5 the experimental design, treatments and procedures. In section 6 we present the results of the experiment. Finally, section 7 discusses the results further and in section 8 we conclude.

2 Most relevant literature

In this section, we review the most relevant literature for our work. We consider the literature that evaluates the degree of complementarity between a price and a nudge in reducing an externality generating activity when implemented simultaneously.[2] These studies differ in several features. One of these is the externality the intervention aims to control. Another feature in which these studies differ is in the type of nudge used to test the complementarity. As importantly, papers also differ in the amount and the type of economic incentive. Finally, some of the studies in this literature have an incomplete design, lacking a stand-alone price treatment (Sudarshan, 2017) or a stand-alone nudge treatment (Hilton et al., 2014; Mackay et. al., 2019; Mizobuchi and Takeuchi, 2013; Spraggon and Oxoby, 2010). Nevertheless, even if we focus only on the studies that have a complete design, no consistent evidence emerges on the complementarities of nudges and prices in the control of negative externality generating activity.

Dolan and Metcalfe (2015) found that a social comparison decreased electricity consumption, but when coupled with a reward, the two instruments had no effect. Up in the ladder of complementarity, Schall et al. (2016) report that the joint implementation of a training course with fuel saving tips plus non-monetary rewards, had no additional effect on fuel consumption by a company's drivers, to that of the rewards alone. Hernández et al. (2024) found that a tariff increase six months after receiving a report containing

[2] Another considerable number of works compares the effect of nudges versus prices when implemented independently. These basically applies a nudge and a price to different sets of subjects, and compare the effects (Nakagawa, et al., 2022; Buckley and Lerena, 2022; Bucholz et al., 2021; Antinyan et al., 2020; My and Ouvrard, 2019; Xu et al., 2018; Ito et al., 2018; Delaney and Jacobson, 2016; Romaniuc, 2016; López et al., 2012). These studies vary in several key factors, such as the context of the test, the externality being targeted, the type of nudges, the type of prices, and the level and design of the prices being tested. Beyond these differences, general results do emerge; nudges are effective, but the effect of prices is generally higher and more persistent in time.

information and a social comparison component decreased water use by less than the sum of the effects of the two instruments applied alone. On the other hand, Panzone et al. (2021) found that the joint effect of an environmental recall and a tax was the sum of the two separate effects, suggesting perfect complementarity (zero synergy). Fanghella et al. (2021), also found perfect complementarity but in a less positive manner: they found that when implemented jointly, as well as when standing alone, a monetary reward and nudging goal setting had no effect on the consumption of electricity. Another completely different result is that of Ambec et al. (2024), who found that only the combination of traffic-light labeling and a high tax significantly reduced the carbon footprint of an average basket. A low tax combined with the labelling had no effect. Neither the traffic lights labelling alone, the high alone or the low tax alone. Finally, Maris et al. (2024) found a complementarity between an economic incentive and a message highlighting personal and environmental benefits, both of which are effective on their own on volunteering for nature restoration.

In sum, among those studies with a complete design, we have an array of heterogeneous results. Some studies showed results consistent with a negative synergy between a nudge and a price (Dolan and Metcalfe, 2015; Schall et al., 2016; Hernández et al., 2024), others showed results consisted with perfect complementarity (Panzone et al, 2021; Fanghella et al., 2021) and others showed results consistent with a positive synergy (Ambec et al. 2024; Maris et al., 2024).

3 Theoretical Framework

In this section, we present the theoretical model from which we derive the hypotheses that we test with our experiments. The setting of our model is that of a public bad: a negative externality affecting a group of producers who are at the same time the sources of the externality.[3] Producers may have moral preferences.

3.1 A moral polluter's behavior in an unregulated setting

Suppose that a source generates a quantity e of emissions of a given pollutant. Let $g(e)$ define the net economic benefits associated with generating e units of this externality and assume that $g'(e_i) > 0$ and $g''(e_i) < 0$. The aggregate level of emissions of the *n* sources is $E = \sum_{i=1}^{n} e_i$. E is a public bad. It produces a negative externality (cost) of γE $(\gamma > 0)$ to *each* of the *n* sources.

Following Levitt and List (2007) and Bowles and Polania-Reyes (2012), we assume that the utility that a source derives from emitting is additively separable in profits and a moral term $M(e_i)$ that captures her moral benefit or cost associated with the action: $U(e_i) = g(e_i) - \gamma E + M(e_i)$. According to Levitt and List (2007), $M(e_i)$ captures the

[3] The setting applies also to a group of citizens that incorrectly disposes of its waste, or another similar situation. For simplicity, we refer to producers.

desire to "do the right thing", and is affected by the size of the negative externality imposed on others, the existence of social norms or legal rules on the level of emissions, and the extent to which the action causing the externality can be scrutinized by others. Building upon Allcott and Kessler (2019), we model the moral term $M(e_i) = \mu_i[\varphi_i(m_i - e_i) + (1-\varphi_i)(s_i - e_i)] = \mu_i[\varphi_i m_i + (1-\varphi_i)s_i - e_i]$, where $0 \leq \varphi_i \leq 1$, m_i is level of emissions consistent with the producer´s personal values, and s_i is the producer´s perception of what the social norm about emissions is. The expression $\varphi m_i + (1-\varphi_i)s_i$ is a weighted average of the individual moral threshold and the perceived social norm. Finally, the parameter $\mu_i > 0$ is a moral or psychological tax for emitting. Modelled in this way, the individual moral term $M(e_i)$ captures feelings of pride/guilt arising from deviating from the "right" level of emissions for that producer.

Including this specification of the moral term, the utility function for each of the producers is:

$$U(e_i) = g(e_i) - \gamma E + \mu_i[\varphi_i m_i + (1-\varphi_i)s - e_i] \quad (1)$$

Assuming that a producer chooses e_i to maximize (1); the first order condition characterizing the choice of *e* by this moral polluter is:

$$g'(e_i) - \gamma - \mu_i = 0 \quad (2)$$

Given our assumption that $g''(e_i) < 0$, this condition is sufficient to characterize an interior optimal choice. We call this optimal choice by a moral polluter in an unregulated setting $e_i^{um}(\gamma, \mu_i)$, where the "*u*" in the superscript refers to "unregulated" and the "*m*" refers to "moral". Note than when setting $\mu_i = 0$, equation (2) characterizes the utility maximizing choice of emission by an amoral polluter, which we called $e_i^{u}(\gamma)$.

3.2 Social optimum with moral polluters

We now characterize the socially optimal distribution of emissions among a group of emitters with moral preferences. This is given by the set $(e_1, \dots, e_n)$ that solves the following social planner problem

$$\max_{\{e_1,\dots,e_n\}} \sum_{i=1}^{n} U_i = \sum_{i=1}^{n} (g_i(e_i) - \gamma E + \mu_i[\varphi_i m_i + (1-\varphi_i)s - e_i])$$

The set of first order conditions is

$$g_i'(e_i) - \gamma n - \mu_i = 0, i = 1, \dots, n \quad (3)$$

These conditions implicitly define the socially optimum individual levels of emissions with moral polluters, which we call $e_i^{mso}(\gamma, n, \mu_i)$. Assuming $g''(e_i) < 0$, two results are easy to show. First, $e_i^{mso}(\gamma, n, \mu_i)$ is lower that the socially optimum level of emissions with amoral polluters ($\mu_i = 0$), $e_i^{so}(\gamma, n)$. Second, first order conditions (2)

and (3) imply $e_i^{mso}(\gamma, n, \mu_i) < e_i^{um}(\gamma, \mu_i)$. Note that this is true even when the moral polluter fully internalizes its marginal externality $\big(\mu_i = \gamma(n-1)\big)$. The reason is that this form of morality gives rise to another social benefit or cost, additional to the public bad, as first noticed by Andreoni (1990). Alternatively put, the "impure altruistic" affects her behavior to take care of her private "warm glow" effect. This new private benefit decreases the privately chosen level of emissions with respect to the amoral producer ($e^{um} < e^{u}$), but it also decreases the socially optimum level of emissions ($e_i^{mso}(\gamma, n, \mu) < e_i^{so}(\gamma, n)$), by the same amount. In this special case when $\mu = \gamma(n-1)$, the social planner's first order condition becomes $g_i'(e_i) - \gamma - 2\gamma(n-1) = 0$, which says that the social planner, in the margin, needs to account not only for the externality $\gamma(n-1)$ itself, but also for the emitter´s moral costs of causing the externality, also $\gamma(n-1)$.

3.3 A moral polluter's response to a prosocial guideline

We now examine the response of moral polluters to a salient prosocial guideline implemented by a regulator with the objective of reducing the aggregate level of emissions from the unregulated level.

A nudge z may affect the individual´s moral utility *M* through three different mechanisms. First, they can affect the social norm s_i, such that $s_i = s_i(z)$**.** A social norm is a convention; this is what everybody expects others to believe (a normative expectation) or do (an empirical expectation) (Bicchieri and Dimant, 2022). A social norm is therefore a belief or an expectancy, and as such it can be modified by a communication of what others believe or are doing. This type of nudge has been extensively studied in the literature (see for example, the literature on the Home Energy Reports for energy conservation). Second, nudges may affect the person´s moral threshold, such that $m_i = m_i(z)$.[4] For example, a message highlighting the benefits of a healthy environment along with its current threats, may affect the person´s individual moral threshold level of emissions. A similar effect could have a prosocial guideline in the form of a message that provide citizens with tips or prescriptions on how to behave in order to avoid negatively impacting the environment, or other citizens. Notwithstanding, it is clear from equation (2) that for prosocial guidelines to affect the choice of emissions in a Levitt-List moral utility function, they must affect the moral price, μ_i. We model this effect as a shift in the moral price, $\mu_i = \mu_i^0(1 + \mathbf{1}\{z > 0\}\mu_i^z)$, where μ_i^0 is the baseline moral price of the producer and the indicator $\mathbf{1}\{z > 0\} = 1$ if $z > 0$ (a nudge is implemented). The shift parameter μ_i^z measures the effect of the nudge on the moral price. μ_i^z could be of either sign, depending on the nudge, the producer and the situation. In the case of a public bad, and a nudge that is intended to decrease the negative externality, $\mu_i^z > 0$. This could be the case of nudges of the moral suasion type, that convey information that makes the level

[4] The type of nudges that may affect only the social norm may differ from the type of nudges that may affect only the individual moral threshold. For ease, we only use a general parameter z to indicate both.

of individual wrongdoing more salient, and this increases a producer´s guilt (or shame if the action is being observed by others). It may also be the case of a prosocial guideline in the form of a salient message of suggested play with an implicit moral suasion, as the examples previously mentioned. Finally, the effect of a nudge may differ between producers. The interaction of personal traits and features of the message, such as its quality, the technology used to deliver it and its frequency, determine the "nudgeability" of the producer. Hence the subscript *i* in the shift parameter μ_i^z.

When nudges may alter the moral price, the social norm, and the individual moral threshold, the Levitt-List utility function of a moral producer may be written as $U_i(e_i) = g_i(e_i) - \gamma E + \mu_i^0(1 + \mu_i^z)[\varphi_i m_i(\mathrm{z}) + (1 - \varphi_i)s_i(\mathrm{z}) - e_i\,]$. In this case, the optimal choice of emissions satisfies

$$g_i'(e_i) - \gamma - \mu_i^0(1 + \mu_i^z) = 0 \tag{4}$$

It is easy to see that an effective prosocial guideline ($\mu_i^z > 0$) decreases the level of emissions of the moral "nudgeable" producer.

3.4 A moral polluter´s response to a tax on emissions

Assume the regulator sets a uniform tax *t* per unit of emissions. The utility function of the representative moral producer in this case is given by $U_i = g(e_i) - \gamma E - te_i + \mu_i^0[\varphi_i m_i + (1 - \varphi_i)s - e_i\,]$. The first-order condition that implicitly defines the level of emissions $e_i^{tm}(\gamma, \mu_i^0, t)$ that maximizes utility is

$$g_i'(e_i) - \gamma - t - \mu_i^0 = 0 \tag{5}$$

Comparing the first order condition defining the social optimum level of emissions (in (3)) with equation (5), we can conclude that the optimal tax for moral producers should be set as:

$$t^m = \gamma(n - 1) \tag{6}$$

Note that this tax is equal to the classical Pigouvian tax in the case of amoral producers. The morality of the producers does not affect the level of the optimal tax, a result obtained by Johansson (1997). This is, again, because, whether motivated by "impure" or pure altruism, morality creates a new private utility/disutility, which the producers consider when deciding how much to emit. For this reason, the marginal externality remains uninternalized.

Non-separability

The model above assumes separability (no interaction) between the emissions tax and the social preferences of those regulated by the tax. However, this is contestable. There is substantial evidence indicating that the use and the size of economic incentives

to change behaviors may crowd out people´s social preferences (see Bowles and Polania-Reyes, 2012). This crowding out of intrinsic motivation (Frey, 1992; Frey and Oberholzer-Gee, 1997) could enter our model through the moral price. The use of an economic incentive may affect the moral price of the producers through different mechanisms. When deterring negative externalities, one of such possible mechanisms is that economic incentives may trigger "moral disengagement" (Bandura, 1991). According to this interpretation, an economic incentive signals that the situation is not an ethical one but a market-like one, decreasing the moral price that a producer imposes on itself for taking a socially undesirable action. Another possibility is that a tax (as well as a regulation or appeal) may deactivate the moral norm guiding a behavior by depriving the subject of the personal satisfaction of acting according to one´s values (Schwartz, 1977). Also, an economic incentive can decrease intrinsic motivation because it can undermine people´s sense of self determination (Deci and Ryan, 2013). The issue of non-separability between the level of moral preferences and the instrument choice by the regulator is important because if an economic incentive crowds out moral motives to behave pro-socially, it may have less of an effect than expected under separability. Moreover, the incentive may be even counter-productive, reducing the targeted prosocial behavior, or increasing a negative externality (Bowles and Polania-Reyes (2012) call this "strong" crowding out).

To allow for non-separability, we follow Bowles and Polania-Reyes (2012) and postulate that $\mu_i = \mu_i^0(1 + \mathbf{1}\{t > 0\}\mu_i^t)$, where the indicator $\mathbf{1}\{t > 0\} = 1$ if $t > 0$ (a tax is implemented). The shift parameter μ_i^t measures the effect of the tax on the moral price.[5] With an emissions tax and non-separability equation (5) becomes $g_i'(e_i) - \gamma - t - \mu_i^0(1 + \mu_i^t) = 0$. Note that the Pigouvian tax in (6) is still optimal. Nevertheless, if the tax crowds out social preferences $\mu_i^t < 0$, we have that $\mu_i^0(1 + \mu_i^t) \leq \mu_i^0$, and as commented above, the level of emissions with which producers respond to the tax is higher than under separability ($e_i^{tm}(\gamma, \mu_i^0, t)$).

3.5 A tax and a salient prosocial guideline

We now consider the possibility that a regulatory agency uses both a tax and a nudge (in the form of a prosocial guideline) jointly. Following the previous discussion, in such a case, the individual utility function is given by:

$$U_i(e_i) = g(e_i) - \gamma E - te_i + \mu_i^0(1 + \mu_i^z + \mu_i^t)[\varphi_i m_i(z) + (1 - \varphi_i)s_i(\mathrm{z}) - e_i\,]$$

An individual's choice of emission in this case, $e^{tz}((\gamma, \mu_i^z, \mu_i^t, t)$ satisfies

[5] We do not distinguish between the categorical effect of the tax (due to the presence of the tax, whatever the value) and the marginal effect (due to the level of the tax), as Bowles and Polania-Reyes did. The reason is that in our experiments we treat subjects with only one level for the tax, and therefore, we are not able to disentangle the categorical effect from the marginal effect. Our μ_i^t captures both effects.

$$g'(e_i) - \gamma - t - \mu_i^0(1 + \mu_i^z + \mu_i^t) = 0 \quad (7)$$

Given the lack of theory and the lack of conclusive empirical evidence, our model is silent with respect to the possible complementarity or substitutability of these two instruments. Therefore, it is also silent with respect to whether the resulting level of emissions is higher or lower than that with which the sources respond to a prosocial guideline and tax alone.

4 Hypotheses

Based on the theoretical framework presented in the preceding section, we now present the hypotheses that we test with our laboratory experiment (For the corresponding statistical hypotheses and proof, see Appendix A).

Hypothesis 1 (Morality): *The average producer behaves as if it has "moral preferences".*

Hypothesis 2 (Nudgeability): *A prosocial guideline reduces the average level of emissions with respect to the baseline level.*

Hypothesis 3 (Non-separability): *The average producer behave as if it has non-separable preferences.*

Hypothesis 4 (Complementarity): *the joint implementation of a tax and a prosocial guideline reduces the average level of emissions with respect to both the level under the tax and the level under the guideline implemented in isolation.*

Although we formally test the four hypotheses, the first three serve as building blocks to validate our model, which provides a theoretical framework to interpret Hypothesis 4, our hypothesis of interest.

5 Experimental Design and Procedures

In this section, we present the experimental design, treatments, and the procedures we used to implement our experiments.

5.1 The experiment

We framed the experiment as a neutral production decision of an unspecified good q, the production of which generates economic benefits for its producer. Every subject had a production capacity of up to 10 units (whole numbers). The schedule of marginal benefits is presented in Table 1 and is the same for every producer throughout the experiments. Each individual decides how many units of the unspecified good to produce. Starting from a baseline situation without regulation, we study the effectiveness of three policy interventions: a uniform unit tax on production, a salient message informing what

the aggregate profit-maximizing level of production is (a nudge in the form of a prosocial guideline), and the combination of the tax and the guideline.

Table 1. Marginal benefits per unit of production (Ur $)

Unit of production	Marginal benefits
1	$30
2	$22
3	$18
4	$14
5	$11
6	$9
7	$7
8	$6
9	$5
10	$4

Apart from generating economic benefits for its producer, each unit of production generates a public bad, affecting the producer of the unit and the rest of the producers in its group. Each group consist of five producers. We model the public bad as a linear function of the aggregate level of production, γQ, where $\gamma > 0$ is a constant parameter capturing the marginal (and average) value of the damage that each unit of production generates to each of the 5 members of the group, and $Q = \sum_{i=1}^{5} q_i$ is the group level of production (q_i is the production level of individual *i*). Consequently, the level of economic benefits that a producer *i* obtains from producing q_i units of this good, are given by the profits obtained when producing q_i units of the good (according to Table 1), minus the value of the public bad γQ. In our experiments, we set $\gamma = \$\ 2$ (two Uruguayan pesos).

5.2 Treatments and theoretical benchmarks

We implemented the following treatments:

Baseline: In this treatment, subjects decide freely and uncoordinatedly the number of units that each one wants to produce in each round. With the chosen parameterization, producing 10 units is a dominant strategy for those interested in maximizing profits (amoral subjects). Nevertheless, given the public bad, if all end up producing 10 units, the individual profit is $ 26 (Uruguayan pesos), while if the 5 subjects in the group produce 5 units each, every subject earns $ 45, the maximum possible. Therefore, while 10 units is the Nash equilibrium individual level of production, 5 units is the aggregate-profit-maximizing level.

Low tax: The second treatment is a uniform tax per unit of production. We set the level of the tax to $ 5 per unit. Looking at Table 1 and recalling that the producer has an additional $ 2 self-imposed cost due to the public bad, it is easy to see that an amoral

profit-maximizer producer faced with this tax would choose to produce 6 or 7 units. This level of production is higher than the level of production that maximizes the aggregate profits of the group (5 units per individual). The reason is that the $ 5 tax is lower than the $ 8 tax that is needed for producers to fully internalize the externality. Our choice of a lower tax of $ 5 is consistent with our motivation to study the complementarity of nudges and taxes on negative externalities, when taxes are low due to political-economy reasons.

Prosocial guideline: Our nudge consists of a salient message informing the subjects that the level of individual production that maximizes the group's profits is 5 units. The message that appeared on the decision screen was the following:

> *"The individual production level that maximizes the group's profits is 5 units. To choose an individual production level higher than 5 means that the aggregated profits of the group would be lower than when choosing a production level of 5."*

Note that the message suggests sources to emit 5 units ($e_i^{so}(\gamma, n)$). One could argue that in the light of the model above, a social planner should suggest $e_i^{mso}(\gamma, n, \mu) < e_i^{so}(\gamma, n)$. Therefore, the message implicitly assumes that the social planner assumes that sources are amoral profit maximizers. Nevertheless, note further that it is not necessary to assume that the social planner is as sophisticated as to consider moral prices in its problem to test our hypotheses. In addition, it would have complicated the experiments and the tests unnecessarily.[6]

Instructions informed subjects that "(t)he income per unit produced is the same for all participants". Therefore, our message cannot provide information to rational, attentive, and capable subjects. Moreover, although this guideline provides information to producers that are inattentive or cognitively limited, it may not affect their behavior either, if this information only helps these subjects to form their belief of the social norm (s_i) or to update their private moral threshold (m_i). In order to alter behavior, the guideline in the message must affect the moral price. Nevertheless, shame cannot be a mechanism at play because individual decisions were not revealed to the other players in the group, or the experimenter. On the other hand, feelings such as guilt may be a mechanism at play. Of course, we are unable to provide a theoretical benchmark of the level of individual production with which the subjects will respond when facing such a nudge, since we do not observe the individuals' moral price.

[6] The sophisticated regulator could have estimated the average baseline moral price μ_i^0 in a similar fashion that we do below but *before* implementing the message and altering the guideline consistently. This, nevertheless, would have possibly translated in the implementation of messages with different suggested levels of production for different groups, which would have decreased the power of our tests.

Low tax + pro-social guideline: Lastly, we include a treatment in which we implement the low tax and the pro-social guideline jointly. In this case, the message reads

"The individual production level that maximizes the group's profits plus the tax revenues is 5 units. To choose an individual production level higher than 5 means that the aggregated profits of the group would be lower than when choosing a production level of 5."

The message adds a reference to the tax revenues. This modification is necessary, given that a tax decreases private production profits. A full rebate of the tax revenues among the five subjects, according to some rules, would have made the modification of the message unnecessary. Nevertheless, actual rebates are not full rebates. At least, revenues have to finance the implementation of the public bad control program, including tax collection and administration cost. In addition, revenues could finance environmental education campaigns, restoration of habitats, defense measures (adaptation in the case of climate change) and/or technology adoption. Implementing any other rebate different than a full one in our experiments would have needed a similar message or a message with a reference to profits after taxes.

5.3 Procedures

We conducted computer-based lab experiments at the Experimental Economics Laboratory of the University of Montevideo (UM). We recruited the participants via email invitations sent to university students in Montevideo. Invitations to programmed sessions to those registered for the experiments were administered through ORSEE.

The day of the session, the experimenter and an assistant received the subjects showing up. Participants were randomly assigned to groups of five. A maximum of six groups of five participated in a given session. Each session began with the instructions of the game. (A full transcript of the whole set of instructions are included in Appendix B). Instructions were played from a previously recorded audio, accompanied by a Power Point presentation highlighting the main points and illustrating the tasks and screens (see Appendix C). After playing the instructions, the experimenter answered the remaining questions. After these questions, the subjects had two practice rounds.

Each session started with the baseline treatment, followed by a second treatment consisting in a policy intervention to control the externality. The intervention was one of the three treatments discussed above: the "low" tax, the prosocial guideline, or both. The total number of rounds per session was 10, equally divided between the two treatments (baseline + intervention).

After finishing the 10 rounds of the experiment, subjects answered a questionnaire. Questions sought information about the participants' socio-demographic characteristics, about their pro-environmental attitudes, their religious beliefs, political orientation, beliefs about other people's attitudes and beliefs, and motives behind their

decisions in the experiment. The complete version of the questionnaire is in Appendix D, at the end of the Power Point presentation.

An important feature in our experiments was that the payment procedure preserved the confidentiality of the participants' decisions. When the activity finished, the experimenter left the room, and the assistant, who stayed outside the room during the entire session, entered the room, extracted the information on the participants´ earnings from the server, but not the information about how they played, and paid each participant in private. The instructions explained this payment procedure and underscored to participants that with this procedure, nobody (neither the experimenter, nor the assistant) could know what decisions they make in the experiment. In addition to the earnings from the exercises, participants were paid $ 150 for showing up on time for the experiment.

5.4 Participant´s characteristics

In total, we conducted 18 experimental sessions, recruiting 200 subjects, in 40 groups of 5 participants each. (See Table 2).

Table 2. Subjects per treatment

Treatment	Groups	Subjects
Baseline	40	200
Low Tax	14	70
Prosocial guideline	13	65
Low Tax + Prosocial guideline	13	65

Most of the recruited subjects were between 18 and 21 years old. Forty-six percent (46%) were females. Almost all the subjects were undergraduate students from the University of the Republic (74%), or from the University of Montevideo (21%). Fifty-seven percent (57%) of the subjects majored in economics. Seventy-seven percent (77%) were pursuing a degree within the STEM/ECON fields. Subjects reported household income levels that are well distributed across the different income ranges, with nearly equal numbers in each range. Appendix E presents more detailed descriptive statistics on the characteristics of the participants, based on their responses to the questionnaire.

6 Results

In this section, we present the results of our work. First, we present the descriptive statistics of the outcome variable in each treatment. We then present the results of the parametric and non-parametric tests of our hypotheses. Finally, we present a regression analysis, as an additional test of our hypotheses.

6.1 Descriptive statistics

Table 3 presents the descriptive statistics of our outcome variable, the individual production level per round, by treatment. Table 3 also presents the corresponding theoretical benchmarks (expected values) of this variable, depending on the assumed morality of the producer.

Table 3. Descriptive statistics and theoretical benchmarks for the per- round individual level of production (*q*), by treatment

	Baseline	Prosocial Guideline	Low Tax	Low Tax + Prosocial guideline
Statistics				
Mean	7.45	7.13	6.31	6.43
Median	8.00	7.00	6.00	6.00
Std. Deviation	2.32	2.19	2.11	2.04
Observations	1,000	325	350	325
Theoretical benchmarks				
Amoral producer	10	10	6 - 7	6 - 7
Moral producer				
with separability	<10	< 10	<6	<6
without separability	<10	<10	? (*)	? (*)
Group profit maximizer	5	5	5	5

(*): the subject with non-separable preferences may respond to a tax with a level of emissions that could be lower or higher than 6-7, depending on the degree in which the tax affects social preferences (μ^t). In the case of the Low tax + Prosocial guideline treatment, the response depends also on the synergy between the two instruments.

Pending the results of the statistical tests that we present below, one can informally note first that subjects behave consistently with having moral preferences. Second, that the prosocial guideline does not look very effective, on average. Nevertheless, it does if we observe the median level of production. Third, the low tax seems to be effective and somewhat more than the prosocial guideline. Finally, both instruments, when implemented jointly, do not seem to be more effective than tax alone. In the following subsections we subject these observations to formal tests.

6.2 Parametric and non-parametric tests

6.2.1 Hypothesis 1 (Morality)

To evaluate our first hypothesis, concerning the morality of subjects, we test whether the average level of production in the baseline treatment is equal to the amoral

individual profit maximizing level, against the alternative hypothesis that is lower. According to their marginal benefits of production presented in Table 1, an amoral, profit – maximizing producer should produce 10 units of the good. Formally stated then, our null hypothesis is $\boldsymbol{H_0}$ (No moral preferences): $\bar{q}_{baseline} = 10$ and our alternative hypothesis is $\boldsymbol{H_1}$(Moral preferences): $\bar{q}_{baseline} < 10$, where $\bar{q}$ is the corresponding average level of production. As seen in Table 3, we have: $\bar{q}_{baseline} = 7.45$. The result of a t-test is presented in the first line of Table 4. According to this test, we can reject the null in favor of the alternative (t-statistic: -34.69; p-value: 0.0000). Therefore, the behavior of the subjects is consistent with them having moral preferences in the form of equation (1) ($\mu_i^0 > 0$).

6.2.2 Hypothesis 2 (Nudgeability)

To evaluate our second hypothesis, we test whether, consistently with moral preferences being affected by prosocial guidelines ($\mu_i^z > 0$), a salient message that informs subjects what the social optimum level of emissions is, reduces the subjects' levels of emissions with respect to the baseline level. To test Hypothesis 2, we performed three different types of tests: two non-parametric tests (the Wilcoxon rank-sum test (also known as the Mann-Whitney U test) and the median test) and a t-test. The results of these tests are presented in lines 2 to 4 of Table 4.

In the t-test (line 2), we test whether the average level of production in the prosocial guideline treatment ($\bar{q}_{guideline} = 7.13$) is the same as or higher than the average level of emissions in the baseline treatment ($\bar{q}_{baseline} = 7.45$), against the alternative hypothesis that it is lower. According to the t-test, we should reject H_0 in favor of H_1 (t-statistic: 2.12; p-value: 0.0171). In other words, the prosocial guideline is effective in reducing the average level of production. The observed difference in the levels of average production between the two treatments (0.32 units, a 4.3% decrease), has a standard deviation of 0.15 (95% confidence interval: [0.023, 0.598]).

The results of the non-parametric tests are consistent with the result of the t-test. According to the Wilcoxon rank-sum test (line 3, Table 4), we should reject the null that the two samples of production levels come from the same distribution (p-value= 0.0055). Likewise, the value of the Pearson chi-squared statistics of the median test (9.69) indicates that we should reject the hypothesis that the median production level in the baseline treatment is equal to the median production level in the prosocial guideline treatment (p-value = 0.002).

Table 4: Results of parametric and non-parametric tests of the hypotheses

Hypothesis		Test	H_0	H_1	Statistic	p-value
1	Morality	t-test	$\bar{q}_{baseline} = 10$	$\bar{q}_{baseline} < 10$	-34.69	0.0000
2	Nudgeability	t-test	$\bar{q}_{guideline} \geq \bar{q}_{baseline}$	$\bar{q}_{guideline} < \bar{q}_{baseline}$	2.12	0.0171
		Rank-sum	$\overline{q}_{baseline} - \overline{q}_{guideline} = 0$	$\overline{q}_{baseline} - \overline{q}_{guideline} \neq 0$	2.78	0.0055
		Median	$q_{baseline}^{median} = q_{guideline}^{median}$	$q_{baseline}^{median} \neq q_{guideline}^{median}$	9.69	0.002
3	Tax	t-test	$\overline{q}_{tax} \geq \overline{q}_{baseline}$	$\overline{q}_{tax} < \overline{q}_{baseline}$	8.05	0.0000
		Rank-sum	$\overline{q}_{baseline} - \overline{q}_{tax} = 0$	$\overline{q}_{baseline} - \overline{q}_{tax} \neq 0$	8.29	0.000
		Median	$q_{baseline}^{median} = q_{tax}^{median}$	$q_{baseline}^{median} \neq q_{tax}^{median}$	66.61	0.000
	Separability	t-test	$\overline{q}_{tax} \leq q_{tax}^{separability} = 5$	$\overline{q}_{tax} > q_{tax}^{separability} = 5$	11.55	0.0000
4	Complementarity	t-test	$\overline{q}_{tax+guideline} \geq \overline{q}_{tax}$	$\overline{q}_{tax+guideline} < \overline{q}_{tax}$	0.76	0.7765
		Rank-sum	$\overline{q}_{tax+guideline} - \overline{q}_{tax} = 0$	$\overline{q}_{tax+guideline} - \overline{q}_{tax} \neq 0$	-0.070	0.9445
		median	$q_{tax+guideline}^{median} = q_{tax}^{median}$	$q_{tax+guideline}^{median} \neq q_{tax}^{median}$	1.9507	0.163

In sum, we conclude that a prosocial guideline informing subjects what the group´s profit maximizer level of production is, reduces the average level of production with respect to the baseline level. This is consistent with at least some of the subjects exhibiting moral preferences and these being affected by the guideline ($\mu_i^z > 0$). This result is consistent with Antinyan et al. (2020), who find that, similarly to us, informing participants about the joint welfare-maximizing consumption bundle changes their behavior by increasing the moral price and psychological cost of generating the externality.

6.2.3 Hypothesis 3 (Separability)

To test our third hypothesis, concerning the separability between the use of economic incentives and moral preferences, we first test whether the low tax is effective in reducing the level of production, with respect to the baseline. That is, we test whether the average level of production in the tax treatment ($\bar{q}_{tax} = 6.31$) is the same as or higher than that of the baseline treatment ($\bar{q}_{baseline} = 7.45$), against the alternative hypothesis that is lower. According to the t-test presented in line five of Table 4, we should reject H_0 in favor of H_1 (t-statistic: 8.05; p-value: 0.0000). In other words, the tax effectively reduces the average level of production of subjects. The observed difference (1.14 units, a 15.2% decrease) has a standard deviation of 0.14 units (95% confidence interval: [0.86, 1.414]). This result is confirmed by the non-parametric tests. According to the rank-sum test (line 6 of Table 4), we must reject the null hypothesis that the sample mean of the levels of production in the baseline treatment come from the same distribution as those coming from the tax treatment (p-value: 0.000). Consistently, according to the median test, we must reject also the null hypothesis that the median level of production in the baseline treatment (8 units) is equal to the median level in the tax treatment (6 units) (p-value: 0.000).

Having tested that the tax is effective, we now test whether the decrease in production levels from 7.45 units to 6.31 units caused by the $ 5 tax is consistent with the separability assumption. More formally, we test $H_0: \overline{q}_{tax} \leq q_{tax}^{separability}$ against the alternative that is higher, with $q_{tax}^{separability}$ being the solution to equation 5 ($g_i'(e_i) - \gamma - t - \mu_i^0 = 0$). To solve for $q_{tax}^{separability}$ in (5), we need to estimate, and later to substitute for μ_i^0, $\gamma = 2$, $t = 5$. To estimate μ_i^0, we use Equation (2) ($g'(q_i) - \gamma - \mu_i^0 = 0$) to solve for the value of μ_i^0 that is consistent with the average level of production in the baseline treatment. To do it, we fit a continuous function to the discrete values of the sources' marginal benefits of production (in Table 1). The fitted function is $g´(q) = 0.3295q^2 - 6.3159q + 34.65$. Substituting q for $\bar{q}_{baseline} = 7.45$ and $\gamma = 2$, we obtain $\bar{\mu}^0 \cong 3.9$. In other words, knowing that it produces a public bad, each additional unit of production costs an average of $ 3.9 in the form of a moral price to its producer.

Now we can use $\bar{\mu}^0 = 3.9$, $\gamma = 2$, $t = 5$, and the fitted $g'(e)$ above in equation (5) to obtain $q_{tax}^{separability} = 5$.[7] We saw that the sources, on average, respond to a \$ 5 tax with a level of production of 6.31 units. This is 26% higher than that predicted level for a producer with separable preferences. According to the result of the t-statistic for this test (11.55; in line 8 in Table 4), we should reject the null hypothesis in favor of the alternative that $\overline{q}_{tax} = 6.31 > q_{tax}^{separability} = 5$ (p-value: 0.0000). This result is consistent with the subjects having non-separability between their moral price and the use of a tax to control the externalities ($\mu_i^t < 0$).

6.2.4 Hypothesis 4 (Complementarity)

With the former three tests we have tested a model upon which to stand to test the degree of complementarity (or substitutability) between a low tax and a prosocial guideline as policy instruments to reduce negative externalities. Before proceeding, a clarification. We say the tax and the nudge are complements when the effect of both instruments jointly implemented is higher than the higher of the two effects when both instruments are implemented in isolation. Consistently, we say they are not complements when the effect is lower.[8] Note that complementarity includes negative synergy; the joint effect could be lower than the sum of the two separate effects.

We test the null hypothesis of no complementarity between the two instruments, against the alternative of some degree of complementarity. More formally, we test $\boldsymbol{H_0}$: $\bar{q}_{tax+guideline} \geq \min(\bar{q}_{tax}, \bar{q}_{guideline})$, against $\boldsymbol{H_1}$: $\bar{q}_{tax+guideline} < \min(\bar{q}_{tax}, \bar{q}_{guideline})$. According to our results, we have $\min(\bar{q}_{tax}, \bar{q}_{guideline}) = \bar{q}_{tax}$.

As it can be seen in the last three lines of Table 4, the results of the t-test and the non-parametric tests indicate that we cannot reject that the sample mean level of production of the "tax + prosocial guideline" treatment (6.43) is higher than or equal to that of the tax alone ($\bar{q}_{tax} = 6.31$ units). Therefore, our experiments show results consistent with the hypothesis that there is no complementarity between the two instruments. And we cannot discard a negative synergy between the two instruments.

This result invalidates the model with a moral price $\mu_i^0(1 + \mu_i^z + \mu_i^t)$ in favor of a model with a moral price of the form $\mu_i^0(1 + \mu_i^z + \mu_i^t + \mu_i^{tz})$, with the latter term ($\mu_i^{tz}$) capturing the negative synergy between the two instruments. Moreover, our result suggests that $\mu_i^{tz} \sim -\mu_i^z$; the negative synergy between the two instruments completely crowds out the effect of the prosocial guideline.

---

[7] Note that this value is equal to individual level of production that maximizes the group´s profit. In other words, subjects with moral preferences that are separable from the choice of instruments by the social planner need not be taxed optimally to produce the level of an externality that maximizes the group´s profits.

[8] What Drews et al. (2020) calls "strong negative synergy".

6.3 Regressions

To complement the results of the parametric and nonparametric tests previously presented, we carried out two random-effects linear panel regression estimations in which our outcome variable is the level of production of subject *i* in round *t* (q_{it}). In the first specification, the covariates are the set of treatment indicator variables. The second specification is $q_{i,t} = \beta_0 + \beta_1 Prosocial\ guideline_{i,t} + \beta_2 Tax_{i,t} + \beta_3 (Prosocial\ guideline + Tax)_{i,t} + \boldsymbol{X}_i + \varepsilon_{i,t}$, where $\boldsymbol{X}_i$ is a vector of socio-economic characteristics. These include the university the subject attends (public/private), the subject the student is majoring in (STEM/ECON vs others) and the declared household income. These variables were the least balanced among those unbalanced characteristics reported in the questionnaire. Including other unbalanced candidates, such as political ideology and environmental attitudes and behaviors, do not change the results.

The results in Table 5 show that the three interventions are effective in decreasing the average individual production level, compared to the Baseline. They also confirm the results obtained by our parametric and non-parametric tests, regarding our hypotheses 1 and 2. If we look at Specification 1, we can see that the estimate of the constant is identical to the $\bar{q}_{baseline}$ (7.45) presented in Table 3. Moreover, $\bar{q}_{baseline} = 7.45 < 10$ (p-value = 0.0000). This result is consistent with subjects having moral social preferences of the Levitt and List form, $(\mu^0 > 0)$, our Hypothesis 1. If we turn to Hypothesis 2 (Nudgeability), the $\hat{\beta}_{guideline} = -0.47$ is statistically different from zero, which indicates that $\bar{q}_{guideline} < \bar{q}_{baseline}$. This is consistent with subjects being nudged by the prosocial guideline through a shift in their baseline moral price ($\mu^z > 0$).

The econometric analysis does not allow us to test for the non-separability hypothesis directly. Nevertheless, since we observe in Table 5 that the estimate of the constant is identical to the $\bar{q}_{baseline}$ (7.45) presented in Table 3, we could use our estimate $\bar{\mu}_i^0 = 3.9$ and obtain, as before, $q_{tax}^{separability} = 5$. At the same time, according to our regression results, $\bar{q}_{tax} = \widehat{Constant} + \hat{\beta}_{tax} = 7.45 - 1.02 = 6.43$. Since we reject that this value is equal to 5 in favor of the alternative that is higher (p value = 0.0000), the econometric analysis confirms the evidence favoring the non-separability hypothesis that we obtained with the tests.

Table 5. Linear Random Effect regression results

Dependent variable: level of production	Specification 1	Specification 2
Prosocial guideline	-0.47***	-0.50***
	(0.13)	(0.13)
Tax	-1.02***	-1.00***
	(0.13)	(0.13)
Prosocial guideline and Tax	-0.98***	-0.97***
	(0.13)	(0.13)
Constant	7.45***	7.92***
	(0.11)	(0.33)
Controls	No	Yes
Chi-squared	126.51	150.16
N	2000	2000
Hypothesis Tests		
Prosocial guideline vs Tax	0.56***	0.50***
	(.177)	(.178)
Tax vs Prosocial guideline and Tax	-0.04	-0.02
	(.177)	(.178)
Prosocial guideline vs Prosocial guideline and Tax	0.52***	0.47***
	(.181)	(.18)

Notes: This table presents estimates from random-effect GLS regressions, estimated on a panel data structure grouped in two levels: a concatenation of session and subject as the group identifier, and round as the time identifier. In each column, the dependent variable is the production level chosen by subject i of group g, in round t. There is one independent variable for each treatment, recalling that a subject only participated in one of them. Each treatment is represented by a dummy variable, which is 1 if subject *i* has faced that treatment, and 0 otherwise. The constant on each estimation reflects the average production level in baseline. Column 1 reflects the regression results without controlling for participants characteristics and Column 2 the results but controlling for several covariates where we identified imbalances between treatments, which are income, field of studies and University. The last 3 rows exhibit the t-tests performed for the difference between treatment effects. Significance levels are indicated as * 0.10 ** 0.05 *** 0.01

The estimations we report in Table 5 also confirm the results obtained for our main hypothesis, Hypothesis 4 (complementarity). We reject the hypothesis that $\hat{\beta}_{Tax} = \hat{\beta}_{Guideline}$ in favor of the alternative that $|\hat{\beta}_{Tax}| > |\hat{\beta}_{Guideline}|$ (p-value = 0.002). We also reject the hypothesis that $|\hat{\beta}_{Tax+Guideline}| = |\hat{\beta}_{Guideline}|$ against the alternative that $|\hat{\beta}_{Tax+Guideline}| > |\hat{\beta}_{Guideline}|$ (p-value is 0.004). Nevertheless, we cannot reject the hypothesis that $\hat{\beta}_{Tax+Guideline} = \hat{\beta}_{Tax}$, (the associated p-value is 0.84). We, therefore, have that $|\hat{\beta}_{Guideline}| < |\hat{\beta}_{Tax}| \approx |\hat{\beta}_{Tax+Guideline}|$. This is the same result obtained with the tests reported in Table 4. Namely, the impact of jointly implementing the two instruments is not different to that of implementing the tax alone. Tested in this manner, our results are still consistent with the hypothesis that the tax and the prosocial guideline are not complements.

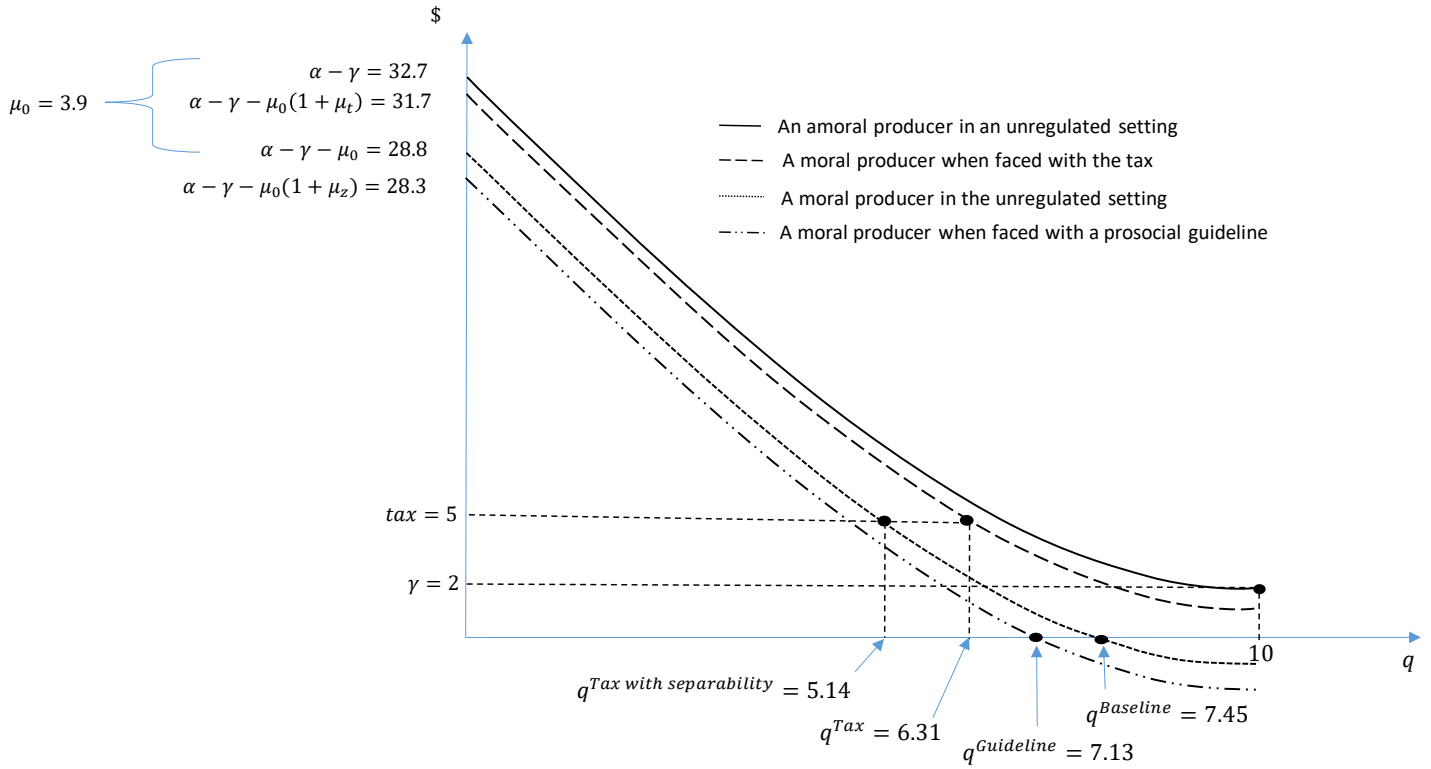
Specification 2 shows that the absolute and relative effects of the different treatments are robust to adding covariates based on unbalanced, self-reported characteristics of the subjects in our sample. Moreover, the coefficients associated with these different characteristics we control for are not statistically significant at the 5% significance level (not shown, see Appendix F for more detail). Thus, we can conclude that despite having an unbalanced sample, this does not affect the conclusions drawn from our previously presented results.

To sum up, our results indicate, first, that the average and median subjects in our sample behave consistently with (a) exhibiting baseline social preferences in the form of a moral price (b) being "nudgeable" by a salient message in the form of a prosocial guideline, and (c) having social preferences that are not separable from the instrument use (tax or nudge). Finally, with polluters exhibiting such moral preferences, we find no evidence of complementarity between these two instruments. Implementing the prosocial guideline and the tax jointly have no additional effect to that of the tax when implemented alone or it may even have a lower effect, consistent with negative synergy between the two instruments.

Figure 1 illustrates our results graphically. The upper solid line illustrates the marginal profits for an amoral producer in an unregulated setting. This is the fitted marginal profits line $g_i'(q_i)$ net of the self-imposed cost via de the public bad (γ). The dotted line illustrates the marginal benefits of a moral producer in the unregulated setting. It reflects a downward vertical shift from the profit curve of an amoral producer equal to $\bar{\mu}^0 = 3.9$, the estimated moral price. This line also shows that such a moral producer chooses to produce $q^{baseline} = 7.45$ instead of the profit maximizer level of 10 units. The dashed-dotted line illustrates the marginal benefits of this moral producer when faced with a prosocial guideline. This line reflects an additional downward vertical shift of $-\mu_i^0 \mu_i^z$ with respect to the marginal benefits loci of a moral producer in the unregulated setting, where the estimated $\bar{\mu}^z = 0.12$, implying a ≈3% increase in the average moral price. The additional downward shift reflects the effect of the prosocial guideline on the baseline moral price of the producer and explains the chosen level of production as a response to the prosocial guideline ($q^{guideline} = 7.13$; a 4.3% reduction with respect to the unregulated level.

Finally, the long-dashed line in Figure 1 illustrates the marginal benefits of the moral producer when faced with the tax. The curve lies above that depicting the marginal benefits of the producer in the unregulated setting, illustrating that social preferences are not separable from the choice of a tax as an internality controlling mechanism. More precisely, the line is a shift upward equal $\mu_i^0 \mu_i^t$ ($\bar{\mu}^t = -0.77$) with respect to the marginal benefits of the moral producer in the unregulated setting. Consequently, the moral producer with separable preferences responds to the tax with $q^{tax} = 6.31$, while it would have responded with $q^{Tax\ with\ separability} = 5.14$ in the case of having separable preferences. According to the results of our experiments, the marginal benefits of the moral producer when faced with the tax and the guideline at the same time

Figure 1: Marginal Benefits and production choices of producers with different social preferences



Note: The figure summarizes our experimental findings by illustrating the marginal benefit functions under different social preference structures. In the horizontal axis we measure the level of production and in the vertical axis the marginal benefits in Uruguayan pesos ($), according to our design and findings. The solid line labeled "*An amoral producer in an unregulated* setting" illustrates the fitted marginal profit function net of the is the value of the self-imposed marginal environmental cost ($\gamma = \$2$). The dotted line, labeled "A moral producer in an unregulated setting" is a downward of the former. It reflects a downward vertical shift from the profit curve of an amoral producer equal to $\bar{\mu}^0 = 3.9$, the estimated moral price, and the consistent choice of production $q^{Baseline} = 7.45$. The dashed-dotted line labelled "*A moral producer when faced with a pro-social guideline"* illustrates the marginal benefits of this moral producer when faced with a prosocial guideline. It reflects an additional downward vertical shift of $-\mu_i^0\mu_i^z$, where $\bar{\mu}^z = 0.12$ and the chosen level of production $q^{Guideline} = 7.13$. Finally, the long-dashed line labelled "*A moral producer when faced with a tax"* illustrates the marginal benefits of a moral producer with non-separable preferences under a $5 tax, with $\bar{\mu}^t = -0.77$ yielding a production level of $q^{Tax} = 6.31$. A subject with separable preferences would respond to the same tax with $e^{Tax\ with\ separability} = 5.14$.

7 Discussion

In what follows, we perform an explanatory analysis that could help shed light towards the task of unraveling the mechanisms behind the effects of taxes and pro-social guidelines on preferences. To do it, we use the subjects´ responses to the questionnaire to classify individuals into different groups, based on some of their self-reported characteristics, and we explore how these characteristics are associated with the observed effects.

First, we divide participants into two categories. The first one, STEM/ECON, includes students majoring in Science, Technology, Engineering, Mathematics, Economics (including related fields such as accountancy, business administration, finance), and Data Science. The second group comprises students majoring in Communications, Architecture, Psychology or Social Work. The rationale behind this classification is that there is considerable evidence that students majoring in Economics and Engineering, either as a result of indoctrination or self-selection, are more inclined to maximize private benefits, in line with the "homo-economicus" model (see Frey and Meier, 2003; Rubinstein, 2006; Frank et al, 1996). In our framework, this would be translated into STEM/ECON students might have different moral prices and be less responsive to the guideline than those in other fields. To examine this, we run separate random effects regressions for both groups. Table G.1 in Appendix G presents the results: Column 1 reports estimates for the STEM/ECON group (n = 156), Column 2 for the non-STEM/ECON group (n = 56), and Column 3 for the pooled sample with indicator variables. The dependent variable in all regressions is subject i in round t (q_{it})

The first observation is that Non-STEM/ECON majors produced, on average, 0.62 fewer units than STEM/ECON majors in the baseline (Column 3), suggesting stronger baseline moral preferences (higher moral price μ^0) among the former. In addition, those majoring in Non-STEM/ECON area are more "nudgeable": they reduce output by 1 unit on average when nudged, compared to just 0.3 units among STEM/ECON students (Column 2). In contrast, both groups respond similarly to the stand-alone tax, each reducing output by 1 unit. Very interestingly their reactions differ under joint implementation of the tax and nudge. Among Non-STEM/ECON students, the combined effect is a 0.45-unit reduction (about half the effect of either instrument alone), statistically significant only at the 10% level. The negative synergy between both instruments when implement jointly is higher for Non-STEM/ECON than for the overall sample (see Table 5). For STEM/ECON students, on the other hand, the difference between the effect of the tax alone and that of the combined instruments is 0.18 and is not statistically significant. Moreover, we cannot reject that the combined effect equals the sum of the individual effects, 1.22=1.04+0.29 (p-value=0.651). This result is novel: the negative synergy of a tax and a nudge could be stronger for those subjects that are more nudgeable. For these subjects, combining a tax and a nudge could result in a very negative policy option. On the other hand, for those subjects that have lower social preferences and are less nudgeable, combining a tax and a prosocial guideline could make a

difference. Although the effect of the prosocial guideline is relatively smaller for more profit-oriented subjects, the effect of adding the guideline to a tax for such subjects is 0.18 units, on average; an increase of 17% from the effect of the tax alone.

Assuming that the classification of students in STEM/ECON and the rest reflects a broader distinction between more and less profit-oriented individuals, we can conclude that the effect of jointly implementing a price and a nudge on an externality will depend on the distribution of this trait in the population. While such a policy may harness the power of both instruments in a population of profit-oriented individuals, this complementarity can turn into a negative synergy if the share of less profit-oriented subjects is sufficiently large. In intermediate cases, such as in our sample, the result lies between both extremes as shown in Table 5: for the average subject, the tax entirely crowds out the effect of the prosocial guideline.

A similar conclusion arises when classifying participants by their self-reported pro-sociality. Using responses to six Likert-scale questions designed to assess prosocial behavior (Appendix D, Questions 18–23), we assigned scores from 0 ("never") to 4 ("always") to each answer. Summing the scores across questions and dividing by the maximum total (24) yields a *pro-sociality index* ranging from 0 to 1, where higher values indicate stronger prosocial preferences. Based on this index, subjects were classified into "Less Prosocial" (bottom 33%), "Moderately Prosocial" (middle 33%), and "Very Prosocial" (top 33%). Out of 200 subjects, 35 fell into the "Less Prosocial" group, 140 into the "Moderately Prosocial", and 25 into the "Very Prosocial". We estimate separate random effects regressions for each group; Table G.2 in Appendix G presents the results. As in Table G.1, the dependent variable is subject i's production in round t (q_{it}), with Columns 1–3 reporting group-specific estimates and Column 4 the pooled model including group dummies. Regarding the complementarity between both instruments, results again suggest heterogeneity across groups. All groups react similarly to the tax, decreasing production by about 1 unit (interaction terms in Column 4 are not significant). In contrast, the pro-social guideline is far more effective among the "Very Prosocial": it reduces output by 1.77 units (Column 3), and its differential effect relative to the "Less Prosocial" group is statistically significant (1.22 units; p-value= 0.024). Finally, for the "Very Prosocial" group, the joint implementation of the tax and the nudge reduces production by 1.27 units—approximately 48% of the sum of the individual effects. This difference is statistically significant (p-value = 0.047, CI: [−2.7412, −0.0181]), indicating clear negative synergy between the instruments. In contrast, for the "Less Prosocial" group, the combined effect is a 1.06-unit reduction—roughly 66% of the sum of the separate effects—but the difference is not statistically significant (p-value = 0.285, CI: [−1.5427, 0.4531]). Similarly, for the "Moderately Prosocial" group, which comprises the majority of subjects, the joint effect is a 0.93-unit decrease, about 69% of the summed individual effects (1.34 = 0.29 + 1.05). The 0.41-unit gap is also not statistically significant (p-value = 0.122, CI: [−0.9216, 0.1091]). Overall, only the "Very Prosocial" group exhibits statistically significant negative synergy.

In sum, two results emerge from this exercise. One is that we cannot reject the hypothesis that the message affects subjects with higher levels of pro social preferences more, while the tax affects subjects with different social preferences evenly. Second, when combined, the price and the guideline jointly implemented make little difference with respect to the tax alone for those with lower social preferences, while for those with "stronger" social preferences, they show a negative synergy.

8 Conclusions

In this work, we assess the degree of complementarity between (a) a salient prosocial guideline informing subjects what the aggregate-profit-maximizing level of a negative externality is, and (b) a tax that is insufficient to induce this level, in a set of experiments that seek to mimic a local public bad situation in which socially disconnected individuals contribute to a pollution problem.

Using both parametric and non-parametric tests, along with econometric analysis, we find evidence that supports the following hypotheses. First, subjects exhibit baseline moral preferences. Second, subjects react to a salient message in the form of a prosocial guideline with an implicit moral appeal by decreasing their baseline level of production in a manner consistent with the hypothesis that such a message increase their moral price. Moreover, the prosocial guideline is more effective in subjects with higher social preferences. Third, subjects exhibit social preferences that are non-separable from the type of instrument chosen by the regulator, and a relatively low tax partially crowds out baseline moral preferences. This evidence is consistent with that found in previous works.

Finally, we find that there is no complementarity between a low tax and a prosocial guideline. Moreover, they may exhibit a negative synergy when jointly implemented. For the average subject, we cannot reject that the effect of both instruments jointly implemented is equal or lower than the effect of the tax when implemented alone. The negative synergy is stronger for those types of subjects with higher social preferences. These results challenge the policy recommendation that nudges can effectively complement low taxes while we wait for the political will to increase taxes to develop.

References

Allcott, H. and Kessler, J. B. (2019). "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons." *American Economic Journal: Applied Economics*, 11 (1): 236-76.

Ambec, S., Andersson, H., Cezera, S., Kanay, A., Ouvrard, B., Panzone, L. and Simon, S. (2024). "Taxing and nudging to reduce carbon emissions: Results from an online shopping experiment." *Working Paper*

Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, *100*(401), 464–477.

Antinyan, A., Horváth, G., & Jia, M. (2020). Curbing the consumption of positional goods: Behavioral interventions versus taxation. *Journal of Economic Behavior & Organization*, 179, 1-21.

Bandura, A. (1991). Social cognitive theory of moral thought and action. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development* (Vol. 1, pp. 45-103). Hillsdale, NJ: Erlbaum.

Bicchieri, C., & Dimant, E. (2022). Nudging with care: The risks and benefits of social information. *Public choice*, 191(3-4), 443-464.

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

Bowles, S., and S. Polania-Reyes (2012). "Economic incentives and social preferences: substitutes or complements?" *Journal of Economic Literature*, 50(2), 368-425.

Bucholz, M., & Musshoff, O. (2021). Tax or green nudge? An experimental analysis of pesticide policies in Germany. *European Review of Agricultural Economics*, 48(4), 940-982.

Buckley, P., & Llerena, D. (2022). Nudges and peak pricing: A common pool resource energy conservation experiment. *Journal of Behavioral and Experimental Economics*, 101, 101928.

Carattini, S., Carvalho, M., & Fankhauser, S. (2018). Overcoming public resistance to carbon taxes. *Wiley Interdisciplinary Reviews: Climate Change*, *9*(5), e531.

Carlsson, F. and O. Johansson-Stenman (2019). "Optimal Prosocial Nudging." *Working Paper in Economics No 757, Department of Economics, University of Gothenburg.*

Carlsson, F., Gravert, C., Johansson-Stenman, O., & Kurz, V. (2021). The use of green nudges as an environmental policy instrument. *Review of Environmental Economics and Policy*, *15*(2), 216-237

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

Croson, R, and Marks, M. (2001). The Effect of Recommended Contributions in the Voluntary Provision of Public Goods. *Economic Inquiry, 39, issue 2, p. 238-49*

Dal Bó, E. and P. Dal Bó (2014). ""Do the right thing:" The effects of moral suasion on cooperation." *Journal of Public Economics* 117: 28-38.

Deci, E. L., & Ryan, R. M. (2013). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.

Delaney, J., & Jacobson, S. (2016). Payments or persuasion: common pool resource management with price and non-price measures. *Environmental and Resource Economics*, *65*, 747-772.

DellaVigna, S., & Linos, E. (2022). RCTs to scale: Comprehensive evidence from two nudge units. *Econometrica*, 90(1), 81-116.

Dolan, P., & Metcalfe, R. (2015). Neighbors, knowledge, and nuggets: two natural field experiments on the role of incentives on energy conservation. *Becker Friedman Institute for Research in Economics Working Paper*, (2589269).

Drews S., Exadaktylos F., van den Bergh J C. (2020). Assessing synergy of incentives and nudges in the energy policy mix. *Energy Policy, Volume 144, 111605, ISSN 0301-4215.*

Fanghella, V., Ploner, M., & Tavoni, M. (2021). Energy saving in a simulated environment: An online experiment of the interplay between nudges and financial incentives. *Journal of Behavioral and Experimental Economics*, 93, 101709.

Frank, R. H., Gilovich, T., & Regan, D. T. (1993). Does studying economics inhibit cooperation? *Journal of Economic Perspectives*, *7*(2), 159-171.

Frey, B.S. (1992). Tertium Datur: Pricing, Regulating and Intrinsic Motivation. *Kyklos*, 45(2), 161-184.

Frey, B. S., & Meier, S. (2003). Are political economists selfish and indoctrinated? Evidence from a natural experiment. *Economic Inquiry*, *41*(3), 448-462.

Frey, B. S., & Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *The American economic review*, *87*(4), 746-755.

Frey, B. S., and A. Stutzer (2008). "Environmental Morale and Motivation." In *The Cambridge Handbook of Psychology and Economic Behaviour*, edited by Alan Lewis, 406-28. Cambridge and New York: Cambridge University Press.

Gravert, C., & Olsson Collentine, L. (2021). When nudges aren't enough: Norms, incentives and habit formation in public transport usage. *Journal of Economic Behavior & Organization*, 190, 1–14. https://doi.org/10.1016/j.jebo.2021.07.012

Hahn, R. W. and Hendren, N. and Metcalfe, R. and Sprung-Keyser, B., 2024. "A Welfare Analysis of Policies Impacting Climate Change". NBER Working Paper No. w32728, Available at SSRN: https://ssrn.com/abstract=4908545

Hernández, F., Jaime, M., and F. Vásquez (2024). Nudges versus prices: Lessons and challenges from a water-savings program. *Energy Economics* 134: 107546.

Hilton, D., L. Charalambides, C. Demarque, L. Waroquier and C. Raux (2014), "A Tax Can Nudge: The Impact of an Environmentally Motivated Bonus/Malus Fiscal System on Transport Preferences," *Journal of Economic Psychology*, 42, 17-27.

Ito, Koichiro, Takanori Ida, and Makoto Tanaka. 2018. "Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand." *American Economic Journal: Economic Policy*, 10 (1): 240-6

Johansson, O. (1997). Optimal Pigovian taxes under altruism. *Land Economics*, 297-308.

Levitt, S. and J. List (2007). "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?" *Journal of Economic Perspectives* 21(2): 153-174.

Lilliestam, J., Patt, A., & Bersalli, G. (2021). The effect of carbon pricing on technological change for full energy decarbonization: A review of empirical ex-post evidence. *Wiley Interdisciplinary Reviews: Climate Change*, *12*(1), e681.

López, M. C., J. Murphy, J. Spraggon and J. Stranlund (2012). "Comparing the Effectiveness of Regulation and Pro-Social Emotions to Enhance Cooperation: Experimental Evidence from Fishing Communities in Colombia." *Economic Inquiry* 50(1): 131-142.

Mackay, M., Yamazaki, S., Jennings, S., Sibly, H., van Putten, I. E., & Emery, T. J. (2019). The influence of nudges on compliance behaviour in recreational fisheries: a laboratory experiment. *ICES Journal of Marine Science*, *77*(6), 2319-2332.

Maris, R., Dorner, Z., & Carlsson, F. (2024). Nudges and Monetary Incentives: A Green Partnership? *Working Papers in Economics 842, University of Gothenburg, Department of Economics.*

Marks, M.B., Schansberg, D.E. and Croson, R.T.A. (1999) 'Using Suggested Contributions in Fundraising for Public Good', *Nonprofit Management & Leadership, 9(4), p. 369. doi:10.1002/nml.9403.*

Mertens, S., Herberz, M., Hahnel, U. J., & Brosch, T. (2022). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proceedings of the National Academy of Sciences*, 119(1), e2107346118.

Mizobuchi, K., & Takeuchi, K. (2013). The influences of financial and non-financial factors on energy-saving behaviour: A field experiment in Japan. *Energy Policy*, 63, 775-787.

My B. and B. Ouvrard (2019). "Nudge and Tax in an Environmental Public Goods Experiment: Does Environmental Sensitivity Matter?" *Resource and Energy Economics*, 55, 24-48.

Nakagawa, M., Lefebvre, M., & Stenger, A. (2022). Long-lasting effects of incentives and social preference: A public goods experiment. *Plos one*, *17*(8), e0273014.

Panzone, L., A. Ulph, D. Zizzo, D. Hilton and A. Clear (2021). "The impact of environmental recall and carbon taxation on the carbon footprint of supermarket shopping." *Journal of Environmental Economics and Management* 102137.

Romaniuc, R. (2016). What makes law to change behavior? An experimental study. *Review of Law & Economics*, 12(2), 447-475.

Rubinstein, A. (2006). A Sceptic's Comment on the Study of Economics. *The Economic Journal*, *116*(510), C1-C9.

Schall, D. L., Wolf, M., & Mohnen, A. (2016). Do effects of theoretical training and rewards for energy-efficient behavior persist over time and interact? A natural field experiment on eco-driving in a company fleet. *Energy Policy*, *97*, 291-300.

Schwartz, S. H. (1977). Normative influences on altruism. In *Advances in experimental social psychology* (Vol. 10, pp. 221-279). Academic Press.

Spraggon J, J. Oxoby R (2010). Ambient-Based Policy Instruments: The Role of Recommendations and Presentation. *Agricultural and Resource Economics Review*. 2010;39(2):262-274.

Stern, P.C. (1999). Information, Incentives, and Proenvironmental Consumer Behavior. *Journal of Consumer Policy* 22, 461–478.

Sterner, T., & Coria, J. (2013). *Policy instruments for environmental and natural resource management*. Routledge.

Sterner, T., Ewald, J., & Sterner, E. (2024). Economists and the climate. *Journal of Behavioral and Experimental Economics*, *109*, 102158.

Stiglitz, J., Barrett, S., & Kaufman, N. (2024). How Economics Can Tackle the 'Wicked Problem'of Climate Change.

Sudarshan A. (2017). Nudges in the marketplace: The response of household electricity consumption to information and monetary incentives. *Journal of Economic Behavior & Organization, Volume 134, Pages 320-335, ISSN 0167-2681.*

Sunstein, Cass and Thaler, Richard. (2008). *NUDGE: Improving Decisions About Health, Wealth, and Happiness.* Penguin Books.

Sunstein, Cass. (2014). Nudging: a very short guide. *The Handbook of Privacy Studies*, 173-180.

Szaszi, B., Palinkas, A., Palfi, B., Szollosi, A., & Aczel, B. (2018). A systematic scoping review of the choice architecture movement: Toward understanding when and why nudges work. *Journal of Behavioral Decision Making*, 31(3), 355-366.

Szaszi, B., Higney, A., Charlton, A., Gelman, A., Ziano, I., Aczel, B., ... & Tipton, E. (2022). No reason to expect large and consistent effects of nudge interventions. *Proceedings of the National Academy of Sciences*, 119(31), e2200732119.

van den Bergh, J., and Savin, I. (2021). Impact of carbon pricing on low-carbon innovation and deep decarbonisation: controversies and path forward. *Environmental and Resource Economics*, *80*(4), 705-715.

Xu, L., Ling, M., & Wu, Y. (2018). Economic incentive and social influence to overcome household waste separation dilemma: A field intervention study. Waste management, 77, 522-531.