The effect of low Prices *plus* informative Nudges on externalities[1]

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

[a] Departamento de Economía, Universidad de Montevideo, Uruguay
[b] 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

[1] We gratefully acknowledge financial support provided by the Agencia Nacional de Investigación e Innovación (ANII) - Fondo Clemente Estable - Uruguay, under Project FCE_1_2019_1_156336. Chávez also gratefully acknowledges partial financial support from ANID-Chile, under project Fondecyt Regular No. 1230266 and by INCAR through ANID/FONDAP/1523A00074. We thank Juan Briozzo for his superb work as research assistant. Also, Alvaro Pérez and Eduardo Cancela, who provided help in the programming and assistance.

The effect of low Prices *plus* informative Nudges on externalities

Abstract: We present the results of a series of laboratory experiments designed to study the effect of different policies to control negative externalities. The experiments consisted of a regulatory environment in which participants faced a uniform tax that is lower than optimal, a nudge, or a combination of a tax and a nudge. We consider a public bad with impacts among either the experiments participants or third parties who were not participants in the experiments. We test whether a descriptive message whose objective is to nudge behaviors towards the internalization of a public bad, implemented together with a tax that is insufficient to induce the optimal level of the externality, adds to the effect of the tax in reducing the negative externality. Our results suggest that the tax is more effective in reducing the level of production than the message. Also, adding a nudge (message) to the tax does not have an additional effect (in both, public bad inside and outside the lab scenarios). Finally, the combined effect of tax and message could be larger under the public bad outside the lab than in the public bad inside the lab scenario.

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

JEL Codes: C91, L51, Q58.

1 Introduction

In their popular book "Nudge", Thaler and Sunstein (2008) state that: "*... the most important step in dealing with environmental problems is getting the prices (that is, the incentives) right*". Nevertheless, "*...such an approach is politically difficult*" (pg. 190). They are right. A recent and important example is carbon pricing. Carbon prices around the world range between less than one US dollar in Poland to 156 USD in Uruguay (World Bank, 2023). These prices are far below the most recent estimates of the social cost of carbon (Tol, 2023), as well as the levels needed to induce an abatement of greenhouse gas (GHG) emissions sufficient to avoid exceeding the 2ºC target in the Paris Agreement.

Immediately after the above statement, Thaler and Sunstein (2008) "… *suggest that along with getting the prices right (or while we are waiting for the political courage to set the prices right), we should take other nudgelike steps that can help to reduce the problem in politically more palatable ways*." (pg. 191). A similar point was made by Carlsson et al (2021): "*...if the existing tax is too low, ... a nudge could play a bigger role for policy. ... a pure green nudge could be used to complement the tax so that the combination mimics the outcome that would result from an optimal tax…. With a non-optimal tax, there is also more room for moral nudges*." Concluding that *"(w)hen taxes are not set optimally, both moral and pure nudges can be efficiency improving complements to taxes*".

These are powerful recommendations, with a huge potential impact on GHG mitigation and other important pollution policies around the world. Nevertheless, whether a specific nudge could complement a low emissions tax, and if so, by how much, remain open questions. Contributing to answering these two questions is the main motivation of this work.

Evidence regarding the relative effect of nudges and prices on negative externalities is rather limited. (See section 2 for a more detailed summary of the literature). A set of works in this literature study the effect of a nudge and a price, but not the complementarity of them. That is, they apply the nudge and the price to a different set of subjects and then compare the effect of the treatment on the outcome of interest between these two groups. They do this in different contexts, using different types of nudges and prices. These are important differences. Nevertheless, general results do emerge from this set of works. At the core, both prices and nudges seem to be effective in changing the targeted behavior, but the effect of prices is higher and more persistent in time. There are some exceptions to these general results, though.

Another set of empirical papers study the degree of complementarity between prices and nudges (when applied together, to the same population). Again, the papers differ in important features, but these differences are fewer than in the case of the first set (about a half of them study the joint effect of informing households what the level of consumption of similar households is and a price change on their level of water or electricity consumption). Nevertheless, it is harder to draw general lessons from the results of these still heterogenous works. Some of them show that a nudge could add to the effect of a price increase or the underlying prices. Others, that prices totally crowd out the effect of the nudge (or vice versa). Finally, another works found no effects.

Ours is the first paper that aims to estimate the separate and joint effect of implementing a nudge and a price on the level of an externality. More specifically, we test whether a message informing subjects what the optimal level of a negative externality is, when implemented together, adds to the effect of a tax that is insufficient to induce the optimal level of this externality.

Our experiments aim to model two different cases in which socially disconnected individuals contribute to a pollution problem. In a set of sessions, the negative externality occurs within the group of subjects in the lab. This set of experiments seeks to mimic a local public bad situation, such as garbage disposal or air pollution in a town or neighborhood. In a second set of sessions, the negative externality takes the form of a decrease of an actual donation to a real environmental NGO, without any negative externality affecting the other members of the group in the lab. In this second set of experiments, we seek to mimic the situation of a non-local public bad, in which residents of a neighborhood, city or country impose a negative externality on non-residents (outside that neighborhood, city or nation).

The paper proceeds as follows. In the second section, we present the most relevant literature. The third section presents the theoretical model used to carry out this study and the hypotheses that guide our research. Section 4 presents the experimental design, treatments and expected results, and the procedures we used to implement our experiments. In Section 5 we present the results of the experiment. Finally, in Section 6 we conclude.

2 Nudges and Prices for reducing externalities: most relevant literature

In this section, we review the most relevant literature for our work, i.e.: the empirical literature testing the degree of complementarity of nudges and prices in public goods / negative externality settings. Given the theoretical ability of nudges to reduce a negative externality but without the political and financial costs of prices, it is surprising that this literature is rather thin.

2.1 Nudges *versus* prices

Somewhat less thin is the empirical literature that compares the effect of a nudge and a price, but not the complementarity of them. That is, they applied the nudge and the price to a different set of subjects (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 important features, such as the context of the test (agricultural production, household consumption, waste separation, fishing, and other features of the framing of the lab experiment), the externality being regulated (pesticide use, waste generation, consumption of positional goods, overuse of a CPR), the type of nudges (traffic light labelling, communication, demonstration, information, graphics, normative messages, private or public disclosure of individual behavior), the type of prices (taxes, subsidies, increase in tariffs, automatic penalties, uncertain penalties), and the amount and the design of the prices being tested (neutrality of the taxes´ revenues, experimental tokens subsidies, redeemable points to be used in shopping). Beyond these differences, general results emerge; both prices and nudges seem to be effective, but the effect of prices is higher and more persistent in time.

There are some exceptions to these general results, though. For example, López et al. (2012) found that, in framed field experiments in fishing communities in Colombia, expected penalties (the prices) high enough to induce efficient levels of contributions to the public good produced lower contributions than the threat of public disclosure of individual behavior (the nudge).

Lastly, there is less consistent evidence on the relative persistence of prices versus nudges when treatments are discontinued. In some cases (Nakagawa et al., 2022; Delaney and Jacobson, 2016), nudges had more persistent effects than prices in the post treatment period. On the contrary, Ito et al. (2018) found that only the households that were treated with prices exhibit lower levels of electricity consumption than the control group in the three months after the RCT ended. Their result suggests that prices, and not nudges (moral suasion) induce habit formation (a more efficient energy use in appliances).

2.2 Nudges *plus* prices

In the literature reviewed in the previous subsection, subjects did not simultaneously face the price and the nudge. Instead, some groups of subjects faced a nudge and others faced a price. This is important because, as discussed in the introduction, we are interested in testing the degree of complementarity of prices and nudges to reduce negative externalities when prices are low, to inform policy makers about the potential role that nudges could have in complementing these low prices, in a world where high enough prices have proven to be politically difficult to implement.

As said before, empirical literature that test the degree of complementarity between prices and nudges (when applied together, to the same population) in the context of externalities is rather thin.

Panzone et al. (2018) used an experimental online supermarket to assess whether (a) being required to recall past environment-friendly behavior before shopping led consumers to purchase baskets of food products with lower carbon footprint, (b) the effect of this nudge (environmental recall) is comparable to that of an income - neutral tax on the carbon content of foods (equivalent to £70/ton of CO2eq) and (c) how they interact when imposed together. They find that the environmental recall had a similar effect to that of the tax (both decreased the carbon footprint of a food basket by around 3 kg

CO2eq) and when implemented together, the joint effect was the sum of the two separate effects.

In another shopping experiment, Hilton et al. (2014) compares the choice made by students between traveling from Toulouse to Paris by plane or train varying their prices and introducing an injunctive message (happy/unhappy face) accompanying the choice. They found that a price increase for the plane ticket coupled with a price decrease of equal value for the train ticket had more effect on choices favoring the train when these new prices were presented as a bonus/malus system due to the lower CO2 emissions of the train and the injunctive message, than when they were not. Nevertheless, the results were not statistically significant. Unluckily, the authors did not treat subjects with the injunctive message in isolation to be able to test for crowding out or in effects between prices and nudges.

Similarly, Brent and Wichman (2022) fail to find any statistically significant interaction between the underlying different blocks of tariffs and the effect of a social comparison nudge in the consumption of tap water by households. With a quasi-experimental design in which they randomly allocated "Home Water Reports" (customized reports comparing the water consumption of a household with that of similar ones, accompanied by injunctive messages), among households on one side and the other of different underlying water tariff blocks, these authors found that the effect of the social comparison (3-5%, in line with what others found) was similar between households facing different tariff blocks. At the same time, receiving a Home Water Report did not significantly change the price elasticity of water consumption. These results suggest that if prices crowd out moral values (explicit economic incentives and social preferences are substitutes) in water consumption, the effect of prices on social preferences is "categorical", not "marginal" (Bowles and Polania-Reyes, 2012). If any, it is the presence of the price, not its level, what crowds out social preferences. This may be good news. As Brent and Wichman note, this means that nudges could be a useful tool to decrease water consumption when prices are low and higher prices are not feasible.

Also, in the field of household water consumption, Hernández et al. (2023) compare the effect of sequentially applying a nudge and a tariff increase. Their results indicate that the increase in the tariff (12% for the first consumption block (1-20 m^3) and 38% for the second one (>20 m^3)) decreased water use by 11%, with respect a control untreated group. On the other hand, the nudge (a report containing a social comparison component and an informative component on the environmental effects of water use) decreased water use by 7%, compared to those households in the untreated group. However, the water consumption of the group of households that received the reports (nudge) and, six months after the nudge had ended, their tariff were increased, decreased 14% with respect to the control group of households. Because this decrease in water use is lower than the sum of the effects of the two instruments applied alone (18% =11% + 7%), the authors observed that there is a degree of complementarity between the two instruments, but this is not complete.

Another work that combines a price change and a social comparison type of nudge in an RCT, in this case with electricity household consumption, is Mizobuchi and Takeuchi (2013). These authors implemented an 8-week field experiment in Japan, in which households were either incentivized with a reward for decreasing electricity consumption (with respect to the same period of the previous year) or showed the energy consumption of other households *in addition to* the reward. Using data of 208 households (103 in the reward group, 53 in the reward + feedback and 52 in the control group), the authors found that the reward decreased electricity consumption by around 4%, while the reward + feedback did it by around 6%. This is a 50% increase in the effect. Unfortunately, they do not have a treatment group that only received the nudge (social comparison), so we do not know the isolated effect of the social comparison, to assess whether there is crowding out between this change in price (reward) and this nudge (social comparison).

A reduction in electricity consumption by households is also the outcome of interest in Sudarshan (2017), who combine, as in the two papers previously reviewed, a nudge based on social comparison and a price change. The social comparison communicated to households is the average consumption of electricity by similar households and energy saving tips. The price change is the introduction of a reward/penalty if the household´s consumption resulted below/above the average consumption of peer households. The reward/penalty rate was 2/3 of the unit price of the grid electricity and 1/3 of the unit price in the cases when the electricity was provided by diesel generators. Social comparison plus energy saving tips produced a reduction in electricity consumption of 7%, equivalent to an increase in tariffs by about 12.5%. Nevertheless, households did not change the level of electricity consumption when social comparison was accompanied by a reward/penalty scheme. This result is consistent with the monetary incentive crowding out social preferences.

Electricity consumption, a social comparison accompanied by saving tips and a reward scheme were also the outcome of interest, the nudge, and the monetary incentive, respectively, in Dolan and Metcalfe (2013). They found that social comparison decreased consumption by 6% (0.22 standard deviations), a £100 reward for 30% reduction to be "more in line with the average consumption of other similar homes" produced a decrease in electricity consumption of 8% (0.35 standard deviations), but when implemented jointly, they have no effect in consumption. This result is consistent with monetary incentives crowding out social preferences. But, as the authors pointed out, it is also consistent with information on social norms crowding out extrinsic motivations.

Electricity saving, but in an experimental simulated environment, was also the outcome of interest for Fanghella et al (2021), who, as Dolan and Metcalfe (2013), Sudarshan (2017), Mizobuchi and Takeuchi (2013) tested the effect of monetary reward (on top of an underlying price). In their case, nevertheless, the reward was coupled with a nudge based on goal setting and feedback. None of the two instruments on their own, nor them jointly proved to decrease the simulated level of electricity consumed in this experiment, involving the operation of a washing machine.

In a laboratory experiment with a recreational fishery frame, Mackay et al. (2019) found that the communication of the average catch by others (the nudge; a descriptive social norm) increased compliance to a catch limit by 10% when the deterrence level is low (5% chance of being inspected and fined if found in noncompliance with the limit) and by an additional, not statistically significant, 3 percentage points in a higher deterrence situation (20% inspection probability). (The 20% inspection probability, without the descriptive social norm, increased compliance by 30% with respect to the low deterrence (5% inspection probability), no nudge situation). Participants in this experiment played one round per treatment, so it cannot illustrate how the treatments affect compliance in the longer run. Therefore, at least in the very short run, this result suggests that it could be a marginal crowding out of social motives by higher deterrence enforcement regimes.

Lastly, Spraggon and Oxoby (2010) tested the effect "enhanced" instructions on the effect of two incentive-based instruments in a public good game in the lab. The enhanced instructions included a detailed explanation of the marginal analysis of the decision at hand and a payoff table. The economic incentives were a tax/subsidy scheme and a tax. Providing subjects in the public good game with enhanced instructions proved to produce levels of the public good closer to its optimal level.

In summary, we identified ten papers in the literature that tested the joint effect of a nudge and a price in the context of an externality generating activity. In six of the ten papers in this literature, the nudge used is social comparison. In the rest, the nudges are environmental recall (Panzone et al. (2021), injunctive message (Hilton et al. (2014), goal setting and feedback (Fanghella et al. 2021) and "recommended play" (Spraggon and Oxoby, 2010). Eight of these ten analyzed the effect of nudges and incentives on the consumption behavior of subjects, seven of which impose a price change on top of an underlying market price. Panzone et al. (2018), Hernández et al. (2023) and Mizobuchi and Takeuchi (2013) show that a joint or sequential implementation of price *increase* and a nudge could add to the effect of the implementation of price increase alone. Comparably, Brent and Wichman (2022) found that social comparison on top of underlying prices could have an effect. On the other hand, Sudarshan (2017) and Dolan and Metcalfe (2013) found that a price change could totally crowd out the effect of the nudge. On the other end, Hilton et al. (2014) failed to find a statistically significant effect of framing a price change as a bonus/malus system and injunctive message, as compared to a plain price change. Finally, Fanghella, et al. (2021) failed to find any effect of any of these instruments, alone or jointly implemented. The result in Brent and Wichman (2022), consistent with "categorical" crowding out of social preferences by prices, not a "marginal" crowding out (Bowles and Polania-Reyes, 2012) contradicts the result obtained by Dolan and Metcalfe (2013) and Mackalay et al. (2019). The former found that a small (£10) reward for electricity conservation crowds out more of the effect of social comparison than a £100 reward. Mackalay et al. (2019) found that adding a social comparison nudge in a low deterrence enforcement regime has positive effect on compliance, while adding it in a high deterrence regime has an additional effect that is not statistically significant.

Most importantly, in terms of the contribution of our work, only Dolan and Metcalfe (2013), Panzone et al. (2021), Fanghella et al. (2021) and Hernández et al. (2023) test separately the effect of the nudge, the economic incentive and then the two instruments together, and are therefore able to estimate the actual effect of the interaction of the price and the nudge. The rest of the papers in the literature reviewed here do not test the effect of the nudge alone (Spraggon and Oxoby, 2010, Mizobuchi and Takeuchi (2013), Hilton et al. (2014), Brent and Wichman (2022), Mackay et al. (2019)) or the price alone (Sudarshan, 2017). Of the former, only Panzone et al (2018) found that the effect of the joint implementation is the sum of the two separate effects (consistent with what Bowles and Polania Reyes (2012) called the "separability assumption"). Dolan and Metcalfe (2013) and Hernández et al. (2023) found that the complementarity between a price increase and a nudge is not perfect; i.e.: there is a degree of crowding out between them. Hernández et al. (2023) is a bit different because they implement a nudge and months later an increase in tariff.

We are interested in the joint effect of a nudge and a price on an unpriced externality, not the effect of a sequential or joint implementation of a nudge and a price *change* in the level of consumption of a marketed commodity with underlying prices. In fact, all the papers in this literature test the effect of nudges and price *changes.* Therefore, to our knowledge, ours is the first paper that aims to estimate the separate and joint effect of implementing a nudge and a price on the level of an externality.

3 Theoretical Framework

In this section, we present the theoretical model from which we derive the hypotheses that we test with our experiments. We start the analysis by considering that the aggregate level of emissions from a set of sources is a public bad that affects every one of these sources in the same amount, in the form of a constant cost per unit of emissions. Later, we will refer to this case as a public bad "inside the lab", as this cost affects all the decision makers in the experiments. Then, we consider the case in which the aggregate level of emissions from a set of sources generates a negative externality that affects other agents, but not the emitters, who will only be affected if they have preferences regarding the welfare of these other agents or some sort of individual moral preferences. Later, we will refer to this case as a public bad "outside the lab", as the negative externality in our experiments will be borne by an environmental NGO and not by the decision makers in the experiment. To capture the possibility that polluters may be affected by causing a public bad on third parties, at the outset of the presentation of our model, in the next subsection, we introduce polluters with moral preferences.

3.1 Moral and amoral polluter's behavior

Suppose that a subject generates a quantity e of emissions of a given pollutant. The subject could be the owner of a firm, a citizen, or a household that produces an unspecified good and generates emissions as a byproduct, or that incorrectly disposes of its waste. In either case, generating e is beneficial for the subject, because it saves the decision maker the higher costs derived from the use of more efficient and less polluting technology and inputs, or the additional effort needed to classify, keep, and correctly dispose of waste. The net benefits of generating e are captured by the function $g(e)$, which is such that $g'(e_i) > 0$ and $g''(e_i) < 0$. Suppose that there are n sources of this pollutant, whether firms, households, or citizens. The aggregate level of emissions of the n sources is $E = \sum_{i=1}^{n} e_i$. E is a public bad. We start by assuming that E produces a negative externality (cost) of γE to *each* of the n subjects, where $\gamma > 0$ is the constant cost per unit of emissions.

Following Levitt and List (2007), we assume that each subject has not only material interests but also a desire to "do the right thing". In their words, "*decisions that an individual views as immoral, antisocial, or at odds with his or her own identity… may impose important costs on the decision-maker*" (p. 156). Levitt and List (2007) focus on three factors that may influence the level of the "moral utility": (*i*) the size of the negative externality imposed on others; the higher the level of the externality that she generates, the higher the moral cost of emitting; (*ii*) *social norms or legal rules* that govern the level of emissions in the society, and (*iii*) the extent to which the action causing the externality is or can be *scrutinized by others*. Andreoni (1989) called this type of "selfish" moral preferences "impure altruism", commonly known as "warm glow" (p. 1448-1449).

Like Levitt and List (2007), we also assume that the utility that a subject derives from emitting is additively separable in profits and a moral term $M(e_i)$ that captures this moral benefit or cost associated with the action. In sum, the Levit-List type of utility function of our subjects is given by: $U(e_i) = g(e_i) - \gamma E + M(e_i)$. Building upon Alcott and Kessler (2019), who follow Glaeser (2006, 2014) and Lowenstein y O´Donoghue (2006), we model the moral term $M(e_i)$ as a weighted average of the person´s individual moral threshold level of emissions, m_i, and the person´s perception of what the social norm about emissions is, s_i. More specifically, the moral term is $M(e_i) = \mu_i[\varphi_i(m_i - e_i) + (1-\varphi_i)(s - e_i)] = \mu_i[\varphi_i m_i + (1-\varphi_i)s - e_i]$, with $0 \leq \varphi_i \leq 1$. The parameter $\mu_i > 0$ is a moral or psychological tax/subsidy for emitting above/below the threshold level of emissions $\varphi m_i + (1-\varphi_i)s_i$, a weighted average of the individual moral threshold and the perceived social norm. Including this specification of the moral term, the utility function for each of our subjects is the following:

$$U(e_i) = g(e_i) - \gamma E + \mu_i[\varphi_i m_i + (1-\varphi_i)s - e_i]; \text{with } 0 \leq \varphi_i \leq 1 \quad (1)$$

Assuming that a subject chooses e_i to maximize (1); the first order condition, which given our assumptions is sufficient to characterize an interior optimal choice, is:

$$g'(e_i) - \gamma - \mu_i = 0 \tag{2}$$

Equation (2) implicitly defines the optimal choice of emissions of an unregulated moral subject, $e_i^{um}(\gamma, \mu_i)$, as a decreasing function of the marginal externality γ and the moral price μ_i. Calling $e_i^{ua}(\gamma)$ the choice of emissions by an unregulated amoral polluter (with μ_i = 0), it is easy to see that when marginal benefits $g'(e_i)$ are decreasing, an unregulated moral subject emits less than an unregulated amoral subject $e^{um} < e^{ua}$.

3.2 Social optimum with moral subjects

We now characterize the social optimum distribution of emissions among the group of emitters. This is given by the set $(e_1, \dots, e_n)$ that solves the following social planner problem:

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

The set of first order conditions

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

implicitly define $e_i^{mso}(\gamma, n, \mu_i)$, the socially optimum level of emissions with moral subjects. Two results are easy to show for decreasing marginal benefits $g'(e_i)$, as assumed. First, the socially optimum level of emissions with moral subjects, $e_i^{mso}(\gamma, n, \mu_i)$, is lower that the socially optimum level of emissions with amoral subjects $(\mu = 0), e_i^{aso}(\gamma, n)$. Second, first order conditions (2) and (3) imply $e_i^{mso}(\gamma, n, \mu) < e_i^{um}(\gamma, \mu)$. Note that this is true even when the moral subject fully internalizes its marginal externality $\big(\mu = \gamma(n-1)\big)$. The reason is that 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 subject ($e^{um} < e^{ua}$), but it also decreases the socially optimum level of emissions ($e_i^{mso}(\gamma, n, \mu) < e_i^{aso}(\gamma, n)$), 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 Regulation: tax, nudge, and tax + nudge

We now examine the response of moral polluters to a tax on emissions and nudges, designed to reduce the aggregate level of emissions from the unregulated level.

3.3.1 A tax on emissions

Assume the regulator sets a uniform tax *t* per unit of emissions. In this case, the utility function of the representative moral subject is given by

$$U_i = g(e_i) - \gamma E - e_i t + \mu_i[\varphi_i m_i + (1-\varphi_i)s - e_i\,]$$

The first-order condition that implicitly defines the level of emissions $e_i^{tm}(\gamma, \mu_i, t)$ that maximizes utility is

$$g_i'(e_i) - \gamma - t - \mu_i = 0 \tag{4}$$

Comparing the first order condition defining the social optimum level of emissions (in (3)) with equation (4), we can conclude that the optimal tax for moral subjects should be set as:

$$t^m = \gamma(n-1) \tag{5}$$

Note that this tax is equal to the classical Pigouvian tax in the case of amoral subjects. The "warm glow" morality of the subjects does not affect the level of the optimal tax, a result obtained by Johansson (1997). This is, again, because the "impure altruism" of the Levitt-List morality creates a new private utility/disutility, which the subjects consider when deciding how much to emit. This additional marginal moral disutility makes a moral subject to emit less than an amoral subject, it does not make her emit the socially optimal level, as the externality prevails.

3.3.2 An informative nudge

Nudges may affect the individual´s moral utility *M*. As stated before, the moral utility term is a function of the difference between the person´s action and a linear combination of her moral threshold m_i and the perceived social norm, s_i, valued at the person´s moral price μ_i.

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). In our model, s_i is either the social norm, or the subject´s perception of this social norm. We call z_s the nudge that affects s_i. Hence, $s_i = s_i(z_s)$. This nudge may take the form of communication by the regulator 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 CITAS). Likewise, we call z_m the nudge that may affect m_i. Accordingly, $m_i = m_i(z_m)$. This type of nudges may take the form of a message highlighting the relevance of a healthy environment or ecosystem and its current challenges. In fact, z_m could be any acquisition of a piece of information that may alter the person´s moral threshold regarding the externality she generates.

The effect of a nudge may differ between individuals. 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 subject. We could model this "nudgeability" of the subject using an additional parameter. For ease of exposition, we assume that subjects are equally nudgeable. In this case, the utility of a moral subject under nudges is given by:

$$U_i(e_i) = g_i(e_i) - \gamma E + \mu_i[\varphi_i m_i(z_{\mathrm{m}}) + (1 - \varphi_i)s_i(z_s) - e_i] \qquad (6)$$

Note that in equation (6), μ_i is independent of z_{m} or z_s. However, the optimal individual choice of emissions remains unchanged (i.e., the first order condition is identical to equation (4)). For nudges to affect the choice of emissions, they have to affect the moral price, μ_i, such that $\mu_i = \mu_i(z_{\mathrm{m}}, z_s)$. This is the case if, for example, receiving information that the individual should do better increases her guilt (moral price). In this case, individual utility is given by

$$U_i(e_i) = g(e_i) - \gamma E + \mu_i(z_{\mathrm{m}}, z_s)[\varphi_i m_i(z_{\mathrm{m}}) + (1 - \varphi_i)s_i(z_s) - e_i\,]$$

The optimal choice of emissions satisfies

$$g'(e_i) - \gamma - \frac{\partial \mu_i(z_{\mathrm{m}}, z_s)}{\partial Z} = 0 \qquad (7)$$

With $Z = (z_{\mathrm{m}}, z_s)$. Equation (7) implicitly defines $e_i^{zm}(\gamma, \mu_i(z_{\mathrm{m}}, z_s))$, the utility-maximizing choice of emissions by a nudgeable moral subject.

3.3.3 Tax and nudge

We now consider the possibility that a regulatory agency uses both instruments, a tax, and a nudge, in combination. Following the previous discussion, the individual utility is given by:

$$U_i(e_i) = g(e_i) - \gamma E - te_i + \mu_i(z_{\mathrm{m}}, z_s)[\varphi_i m_i(z_{\mathrm{m}}) + (1 - \varphi_i) s_i(z_s) - e_i\,]$$

An individual's optimal choice of emission in this case satisfies

$$g'(e_i) - \gamma - t - \mu_i(z_{\mathrm{m}}, z_s) = 0\,, \qquad (8)$$

which implicitly defines the optimal level of emissions when the moral individual faces a tax and a nudge, $e^{tz}((\gamma, \mu_i(z_{\mathrm{m}}, z_s), t)$.

3.4 Public bad outside the lab

As mentioned in the introduction to this section, our theoretical model considers polluters with moral preferences to capture the possibility that these may be affected in the case in which the aggregate level of emissions from a set of sources generates a public bad that affects other agents who are not decision-makers. We will refer to this case as a public bad "outside the lab", as the negative externality in our experiments will be borne by an environmental NGO and not by the decision makers in our lab. More specifically, the individual utility, in this case, is given by:

$$U_i(e_i) = g_i(e_i) + \varepsilon_i(z_{\mathrm{m}}, z_s)[D^m(z_{\mathrm{m}}, z_s) - D] \qquad (9)$$

Where D^m is the subject´s moral threshold for a public good enjoyed by third parties (say, ambient quality), D is the actual level of the public good and ε_i is the moral price of a unit of the public good. We assume that $D = D^{max} - \beta E$, with $\beta > 0$ is an impact parameter that translates the level of emissions into ambient quality. We normalize $D = 0$ when $E = \sum_{i=1}^{n} e_i = E^u$, the profit-maximizing aggregate level of emissions (the level chosen by a group of amoral ($\varepsilon_i = 0$) polluters). This is the same as assuming, arbitrarily, that $D^{max} = \beta E^u$.

The optimal choice of emissions in this setting is given by the condition $g_i{}'(e_i) - \varepsilon_i \beta = 0$. We can use this framework to analyze the effect on individual behavior from introducing regulatory measures, including an emissions tax, nudges, and a policy that combine tax and nudges. The results are qualitatively similar to those we discussed before when the public bad affects insiders and are therefore not shown here.

3.5 Hypotheses

In this the final subsection of our theoretical framework, we present the hypotheses that we evaluate with our laboratory experiments. For the case of the public bad "inside the lab", we follow the enunciation of each of the hypotheses with the corresponding proof. Because they are very similar, we do not repeat the analysis for the case of the public bad "outside the lab". Notwithstanding, as explained in section **¡Error! No se encuentra el origen de la referencia.**, we run experiments to test the same hypotheses with a public bad "outside the lab".

Hypothesis 1: *Relative to the unregulated, individual utility-maximizing level, a tax on emissions decreases the level of emissions of a moral individual.*

Proof: Comparing equation (2)) and equation (4), it is easy to see that, for decreasing marginal benefits $g'(e_i)$, as assumed, $e_i^{tm}(\gamma, n, \mu_i, t) < e_i^{um}(\gamma, \mu)$. QED.

Apart from being a building block for our main hypothesis of interest, we are interested in testing Hypothesis 1 because some authors have argued that prices may backfire in the case of moral subjects (Gneezy and Rustichini, 2000; Bowles and Polania – Reyes (2012)). To allow this possibility, Bowles and Polania-Reyes (2012) set the moral price $\mu_i(z_{\mathrm{m}}, z_s) = \mu_0(1 + \mathbf{1}(\mathrm{t} > 0)\mu_c + t\mu_t)$, where $\mu_0 \geq 0$ is a baseline moral price, μ_c measures the "categorical" effect of the tax on the moral price and μ_t measures its marginal effect. The values of μ_c and μ_t could be zero if the subject experiences "separability" between moral preferences and taxes. On the other hand, it she does not, $(\mu_c, \mu_t) \gg 0$ and the subjects experiences "moral disengagement" when confronted with a price. Our model, as noted, assumes "separability".

Hypothesis 2: *Relative to the unregulated, individual utility-maximizing level, an informative nudge decreases the level of emissions of the moral individual.*

Proof: Comparing equation (2) and equation (7), assuming $\frac{\partial \mu_i}{\partial z_m} > 0$ and $\frac{\partial \mu_i}{\partial z_s} > 0$ (nudges increase guilt), as we do, it is easy see that $e_i^z(\gamma, \mu_i(z_\mathrm{m}, z_s)) < e_i^{um}(\gamma, \mu_i)$. QED.

Hypothesis 3 (Nudges are complements to Taxes): *Individual emissions by a nudgeable moral subject under both a tax and an informative nudge are lower than under only a tax.*

*Proof***:** Comparing equation (8) to equation (4), and assuming $g_i''(e_i) < 0$, as we do, it is easy to see that $e^{(t+z)}(\gamma, \mu(z_\mathrm{m}, z_s), t) < e_i^{tm}(\gamma, n, \mu_i, t)$.

Hypothesis 4 (Taxes are complements to Nudges). *Individual emissions by a nudgeable moral subject under both a tax and an informative nudge are lower than under only a nudge.*

*Proof***:** Comparing equation (8) to equation (7), and assuming $g_i''(e_i) < 0$, $\frac{\partial \mu_i}{\partial z_m} > 0$ and $\frac{\partial \mu_i}{\partial z_s} > 0$ (nudges increase guilt), as we do, it is easy to see that $e^{(t+z)}(\gamma, \mu(z_\mathrm{m}, z_s), t) < e_i^z\big(\gamma, \mu_i(z_\mathrm{m}, z_s)\big)$.

As stated above, we are going to test these four hypotheses for the case of a local public good affecting the decision makers ("inside the lab") and for the case of another type of public bad, one that affects third parties, different from the decision makers, who are the sources of the externality. Although we formally test the four hypotheses enunciated, our hypothesis of interest is Hypothesis 3. This is our main instrument to answer whether a specific nudge could complement a low emissions tax or not, and if so, by how much. Moreover, we are also interested in providing evidence on whether the results of the test for Hypothesis 3 differ between the situation represented by a public bad inside the lab and that of the public bad outside the lab.

4 Experimental Design

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

The objective of our research is to study the relative effectiveness of alternative instruments to control negative externalities. Starting from a baseline situation without regulation, we study the effectiveness of three policy interventions: a uniform tax on production, an informative message (nudge), and a combination of the tax and the nudge. To do this, we conducted a series of lab experiments with university students in Montevideo, Uruguay. In these experiments, there is a group of individual producers. Each group consisted of five subjects. We framed the experiment as a neutral production decision of an unspecified good. Every subject had a production capacity of 10 units (whole numbers). The production of these units generates economic benefits for their producer. The schedule of marginal benefits (benefit per unit produced) is presented in Table 1 and is the same for each producer and throughout the experiments. Each individual decides how many units of the unspecified good to produce.

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

Our design considers a situation under which production activities cause a negative externality between producers (insiders) (public bad “inside the lab”). We also study the effectiveness of these instruments in a situation under which the externality causes damage to a group of individuals who are not the decisions makers in the experiment (outsiders) (public bad “outside the lab”). We start by presenting the design of the former (experiments with a public bad inside the lab) in the next subsection.

Subsequently, in section 4.2, we do the proper with the design of the experiments in which the public bad occurred outside the laboratory.

4.1 Externality on insiders experiments (public bad inside the lab)

Apart from generating an economic benefit to its producer, each unit of production in these experiments generates a cost to each member of the group; that is, production generates a public bad. Therefore, the economic benefits of each individual in the group are given by the total production benefits minus the value of damages caused by the aggregated group production. More specifically, in our experiment, the net economic benefits for an individual producer are

$$\pi_i(e_i) = g(e_i) - \gamma E$$

Where π_i refers to profits from production, e_i is the production level of individual i, $g(e_i)$ refers to the profits obtained when producing e_i units of the good (according to the marginal benefits schedule of Table 1), $\gamma = 2$ and $E = \sum_{i=1}^{5} e_i$.

4.1.1 Treatments and theoretical benchmarks

To study the complementarity of a price and a nudge to reduce a negative externality that a group of subjects impose on each other, we implemented the following treatments for the public bad inside the lab context:

T1. Baseline: In this treatment, subjects decide freely and uncoordinatedly the number of units each produce in each round. With the chosen parameterization, producing 10 units is a dominant strategy for those interested in maximizing profits (amoral, zero moral price subjects). Of course, given the public bad, if all end up producing 10 units, the individual profit is $90. Instead, if the 5 subjects in the group produce 5 units each, every subject earns $185, the maximum possible. Therefore, while 10 units is the Nash equilibrium, 5 units is the social optimum.

T2. Low tax: The second treatment considers a uniform tax on production (T2). We set the level of the tax to UY$ 5 per unit. At this level, an amoral profit maximizer individual 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 welfare of the group (5 units per individual). Our choice of the level of the tax is consistent with our motivation to study the complementarity of nudges and prices when taxes are low due to political-economy reasons.

T3. Nudge: Our nudge consists of a message informing the level of individual production that maximizes group benefits to the subjects (5 units). Of course, the actual level of individual production when facing a nudge will depend on how responsive individuals are to the intervention. Since we do not observe the moral term of the utility function of subjects, we are unable to provide a theoretical benchmark.

More specifically, 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."

Providing information is one of the most important means of nudging people (Sunstein, 2014). Information provision can take several forms. One is to inform what other people are doing. This gives information of what the social norm is with respect to the behavior being studied. Other types of information can be the disclosure of the environmental impacts of consumption choices, or environmental information. The majority of the messages in the most relevant literature have these features. Our message is a simple, informative nudge. It mimics the situation in which citizens receive a message with tips on how to behave in order to avoid negatively impacting the environment in a significant and welfare decreasing way.

Our message does not have an explicit moral appeal. Nevertheless, it does implicitly call for a subject interested behaving consistently with the group´s greater good to produce 5 units. In this sense, our message classifies also as a suggestion. More precisely, our message is a message of suggested play with an implicit moral suasion based on utilitarianism.

Similar to our nudge, Antinyan et al. (2020) informs participants of the joint welfare maximizing consumption bundle. Moreover, they find that the main channel by which this type of nudge operates is by increasing the moral price and psychological cost of the externality generating subject. Messages of "suggested play" have proved to be effective in the presence of heterogenous 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).

T4. Low tax + nudge: Lastly, as we are interested in testing the complementarity of a tax on emissions that is insufficient to induce the optimal level of emissions and a nudge, we include a treatment in which we implement both instruments at the same time. 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 is slightly difference and adds a reference to the tax revenues. This modification is necessary, given that a tax decreases profits but not welfare. You could think of our experiments with a public bad inside the lab as a local pollution problem that affects the members of the community, who are taxed by a national regulator and the revenues collected from this tax go to the general treasury.

A full rebate of the tax revenues among the 5 five subjects, according to some rule, would have made the modification of the message unnecessary. Two reasons are behind our decision not to implement a rebate. First, actual rebates are not full rebates (at least, revenues have to finance the implementation of the pollution control program, including tax collection and administration costs). Implementing any other rebate different than a full one in our experiments would have needed a similar message. Second, apart from financing the implementation of the emissions control program, revenues frequently finance environmental education campaigns, restoration of habitats, or defense measures (adaptation in the case of climate change). These are all important features that could foster the support of actual emissions tax programs. Nevertheless, reflecting these features in the simple context of our experiments with the public bad inside the lab would have complicated the setting without adding value to the test of our main hypothesis.

4.2 Externality on outsiders' experiments (public bad outside the lab)

These experiments are the same as those described in the context of an externality affecting the group of individual producers, with the exception that in this case the externality does not affect the profits of the producers in the lab, but it reduces the amount to be transferred to another group, which is passively affected by the negative externality generated by the group participating in the experiment. Specifically, in these experiments, participants had the opportunity to decide how much to produce of a good from which they could not only earn money for themselves, but they could also contribute to donate to an NGO. The NGO chosen in this case was SOCOBIOMA (www.socobioma.org), which operates in Uruguay, rescuing injured animals and rewilding them after they recover. SOCOBIOMA also promotes education and research based on rescued animals.

In these experiments, the net economic benefits for an individual producer are $\pi_i(e_i) = g(e_i)$, where, as before, $g(e_i)$ is the total profits function consistent to the marginal benefits schedule of Table 1. Note the absence of the term γQ, consistent with the absence of a public bad inside the lab. Instead, subjects were told that each unit produced reduces the donation to SOCOBIOMA by U$10, which is the sum of the "internality" of $2 ($\gamma$) that a producer imposed upon itself and the externality of $8 ($2*4) that a producer imposed on the other 4 members of the group, when deciding to produce a unit of the good, in the experiments with the public bad inside the lab. Likewise, to make the experiments with the public bad outside the lab more comparable to those with the externality inside the lab, the initial donation on the table at the beginning of each round was $500. This is the donation the NGO receives if the group production is zero and is equal to the maximum level of the externality inside the lab (when the 5 subjects decide to produce 10 units each).

We ran four treatments in this context. These treatments parallel the treatments considered in the context of an externality affecting only the members of the group of participants (insiders); that is, a baseline treatment-no regulation (T5), tax treatment (T6), nudge treatment (T7), and a tax and nudge treatment (T8).

If individuals do not care about the outsiders (NGO), individual production is expected to be 10 in the baseline. Likewise, it is expected to be 10 units in the nudge treatment, under the assumption of an amoral profit maximizer producer. The actual level of individual production when facing a nudge (T6) depends on how much they care about the work of the NGO and how responsive they are to the intervention. Under the combination of tax and nudge we expect the level of production of a profit maximizer subject to be no higher than the one under the tax. In this second set of experiments, subjects faced a tax of UY$ 7 instead of a tax of UY$ 5. The reason is that without the internality of UY$ 2, subjects would have produced more than the 6-7 units that a tax of UY$ 5 induces in the case of the public bad inside the lab. Therefore, we add UY$ 2 to the tax so that the expected level of production from a profit maximizer individual was, again, 6-7.

In sum, we constructed a total of eight treatments for these experiments. The parameters per treatment and theoretical benchmarks of the experiments are presented in Table 2.

Table 2. Parameters by Treatments and Theoretical Benchmarks

	Treatment	Tax per unit	Theoretical Benchmarks	
			Amoral subjects	Moral subjects
Public bad inside the lab	T1: Baseline	-	10	≤ 10
	T2: Low tax	$ 5	6-7	≤ 6-7
	T3: Nudge	-	10	≤ 10
	T4: Low tax + Nudge	$ 5	6-7	≤ 6-7
Public bad outside the lab	T5: Baseline	-	10	≤ 10
	T6: Low Tax	$ 7	6-7	≤ 6-7
	T7: Nudge	-	10	≤ 10
	T8: Low tax + Nudge	$ 7	6-7	≤ 6-7

4.3 Procedures

We conducted computer-based experiments in the Experimental Economics Laboratory of the University of Montevideo (UM), which was conditioned to prevent the participants from communicating with each other, or from seeing the monitor or another participant. The sessions were planned so that there was a maximum of 40 subjects per session since that was the capacity of the laboratory.

In each session, participants signed consent forms and were then randomly assigned to groups of five. A maximum of six groups of five participated in a particular session. At the beginning of each session, the experimenter read the instructions aloud with PowerPoint slides highlighting the main points and illustrating the instructions. Practice rounds were conducted. Control questions were asked about the procedures to determine whether the subjects were ready to participate in the experiments. Each session started with the baseline treatment, where no incentive was applied, and then it was followed by a second treatment consisting in a policy intervention to control the externality. The intervention was either a uniform tax on production, an informative message, or a combination of both. The total number of rounds per session was 10, which was equally divided between the two treatments we run in that particular session.

After the experimental rounds ended, subjects answered a questionnaire about socio-economic characteristics and environmental attitudes, including attitudes towards SOCOBIOMA, the environmental NGO considered in our experiments.

In addition to the earnings from the exercises, participants were paid UY$ 150 for showing up on time for the experiment. The payment procedure preserved the confidentiality of their decisions. We did this in the following way. First, we stored the personal earnings data in a different file from the one where we stored the decisions in the activities and the answers to the questionnaire. The experimenter did not see the subjects' personal information at any time. The payment procedure was as follows. When the activity finished, the experimenter left the room, and the assistant extracted the information of the participants´ final earnings from the server. With this information, the assistant prepared the payments and a receipt for each participant. Once this task was completed, the assistant proceeded with the payments. Each participant received their earnings 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 made in the experiment.

Participants in the experiments with the public bad outside the lab were also informed about the procedure to pay the NGO, through the Accounting Office of the University of Montevideo, who was in charge of transferring the money to SOCOBIOMA. Once the amount was transferred, we uploaded the donation receipt to the UM Experimental Economics Laboratory website (https://econlab.um.edu.uy/). This was also informed to participants.

To recruit the participants, we used ORSEE.

Table 3. Subjects per treatment

Public bad inside the lab		
Treatment	Groups	Subjects
T1: Baseline	37	185
T2: Tax	14	70

T3: Nudge	13	65
T4: Tax + Baseline	10	50
Sub-Total	37	185
Public bad outside the lab		
Treatment	Groups	Subjects
T5: Baseline	39	195
T6: Tax	13	65
T7: Nudge	13	65
T8: Tax + Baseline	13	65
Sub-Total	39	195
Total	76	380

4.4 Participants characteristics

A total of 800 participants registered on the website, but only 580 of them were eligible as undergraduate college students. A total of 415 subjects from different universities signed up and presented themselves to the experiments. In total, we conducted 34 experimental sessions, recruiting 380 subjects (76 groups of 5 subjects). Eighteen of the 34 sessions, with a total of 185 students, in 37 groups, corresponded to public bad inside the lab sessions. The rest, (16 sessions, 195 subjects, 39 groups of 5) were public bad outside the lab sessions.

Most of the participants were students from the University of the Republic (224), followed by students from the University of Montevideo (10), Catholic University (10), and Universidad ORT (4). The rest of the students were enrolled in other universities in the country.

In the public bad outside the lab experiments most subjects were students between 18 and 23 years old, 54% of whom were women. Most were from the University of Montevideo (49.7%), while 42.6% are from the University of the Republic, 3.6% are from the Catholic University, 1.0% are from ORT University, 0.5% are from Universidad de la Empresa, and 2.6% are from another universities. As for majors, 23.1% study economics, 12.3% study media science, 6.7% study international Business, 8.2% study business Administration, 7.7% study public accountant, 7.2% study law, 1.0% study data science, 1.0% study finance, 4.1 % study engineering, 0.5% study statistics and the other 28.2% is distributed in other majors.

In the public bad in the lab experiment most subjects were between 18 and 21 years old. There were 50% female subjects. Most were from the University of the

Republic (76.2 %). 17.8% were from the University of Montevideo, 1.62 % from the Catholic University, 1.1% from ORT University, and 3.2% from another universities. With regard to subjects' major, 57.3% study economics, 6.5% study business Administration, 65% study engineering, 6% study to be a public accountant, 3.2% study medicine, 2.2% study media science, 1.6% study law, 1.6% study statistics, 1.6% study psychology, 1.1% study data science, 1.1% study finance, 1.1% study architecture and the other 10.3% is distributed in other majors. Appendix A includes descriptive statistics for responses related to environmental questions.

Table 4. Descriptive statistics of the subjects

Variable		Subjects	
		Public bad inside the lab	Public bad outside the lab
Gender	M	46%	50%
	F	54%	50%
Age	17-18	8%	21%
	19	24%	22%
	20	21%	17%
	21	14%	13%
	22	7%	7%
	23-30	19%	16%
	30-49	7%	4%
Home Income	Less than $38000	12%	19%
	Between $38000 and $53600	16%	11%
	Between $53600 and $72000	17%	14%
	Between $72000 and $99000	16%	13%
	Between $99000 and $163000	18%	19%
	More than $163000	21%	24%
University	UDELAR	76%	42%
	Universidad de Montevideo	18%	50%
	Universidad Católica	2%	4%
	ORT	1%	1%
	Other	3%	3%
University Degree	Economics	57%	23%
	Business Administration	6%	8%
	Engineering	6%	4%
	Public Accountant	6%	8%
	Medicine	3%	3%
	Media Science	2%	12%
	Law	2%	7%
	Statistics	2%	
	Psychology	2%	3%
	Data Science	1%	1%
	Finance	1%	1%
	International Business		7%
	other	12%	23%
Political Preferences	Right	25%	34%
	Left	22%	17%
	Middle	53%	49%

5 Results

In this section we present the results of our work. First, we present a descriptive analysis of the overall results and non-parametric tests. Then, we present the results based on regression analysis (parametric tests).

5.1 Overall results: descriptive analysis and non-parametric tests

We present descriptive statistics for the individual level of production for the public bad inside and outside the lab experiments in Table 5 and Figure 1. In both scenarios, we observe that the mean and median individual production levels in the Tax-only treatment, the Message-only treatment, and the Message + Tax treatment are lower than in the corresponding Baseline treatments. Particularly, in the experiments with the public bad inside the lab, the mean production level is 0.47 units lower in the Message treatment (T3), 1.03 units less in the Tax treatment (T2) and 1.07 units less in the Message + Tax treatment (T4), than in the Baseline treatment, respectively. In experiments with the public bad outside the lab context, the average production level in the Message treatment (T7) is 0.7 units less than in the Baseline treatment, while it is 1.24 units lower in the Tax treatment (T6) and 1.41 units lower in the Message + Tax treatment (T8) than in the Baseline treatment. All these differences are statistically significant. (Stata outputs of these tests are included in Appendix B).

Although the different interventions do not have different effects in terms of the median production, another general observation is that the mean individual production level in the Tax and (Tax + Message) treatments are lower than in the Message-only treatment. In the experiments inside the lab, the difference between the mean production level under the Tax-only and the Nudge-only treatment is statistically significant, according to the *t*-test. Consistently, we reject that the data come from the same distribution, according to the ranksum test. Therefore, we can conclude that the mean effect of the Tax-only intervention is higher than that of the Nudge-only intervention. At the same time, according to the *t*-test, we can reject the hypothesis that the mean production levels under the (Tax + Message) treatment and that of the Nudge-only treatment are the same, in favor of the alternative hypothesis that the former is lower than the latter. Consistently, we reject the hypothesis that the data originate from populations with the same distribution or median, as indicated by the ranksum and median tests. On the other hand, we cannot reject that the medians, the distribution from which the observations come from, and the mean production of the (Tax + Message) and the Tax-only treatments are the same.

With respect to the experiments with the public bad outside the lab, based on the *t*-test results, we can conclude that the differences in average production under the Nudge-only treatment are statistically significant compared to both the Tax treatment and the (Tax+Message) treatment. However, as in the case of the experiments with the public bad inside the lab, we cannot reject the hypothesis that the average production level in the Tax treatment is equal to that of the (Tax+Message) treatment, suggesting that both have a similar effect. Regarding the outcomes of the non-parametric tests, the ranksum test indicates that the samples from each treatment do not come from populations with the same distribution. Nevertheless, we reject the hypothesis that median production level of the two interventions is the same.

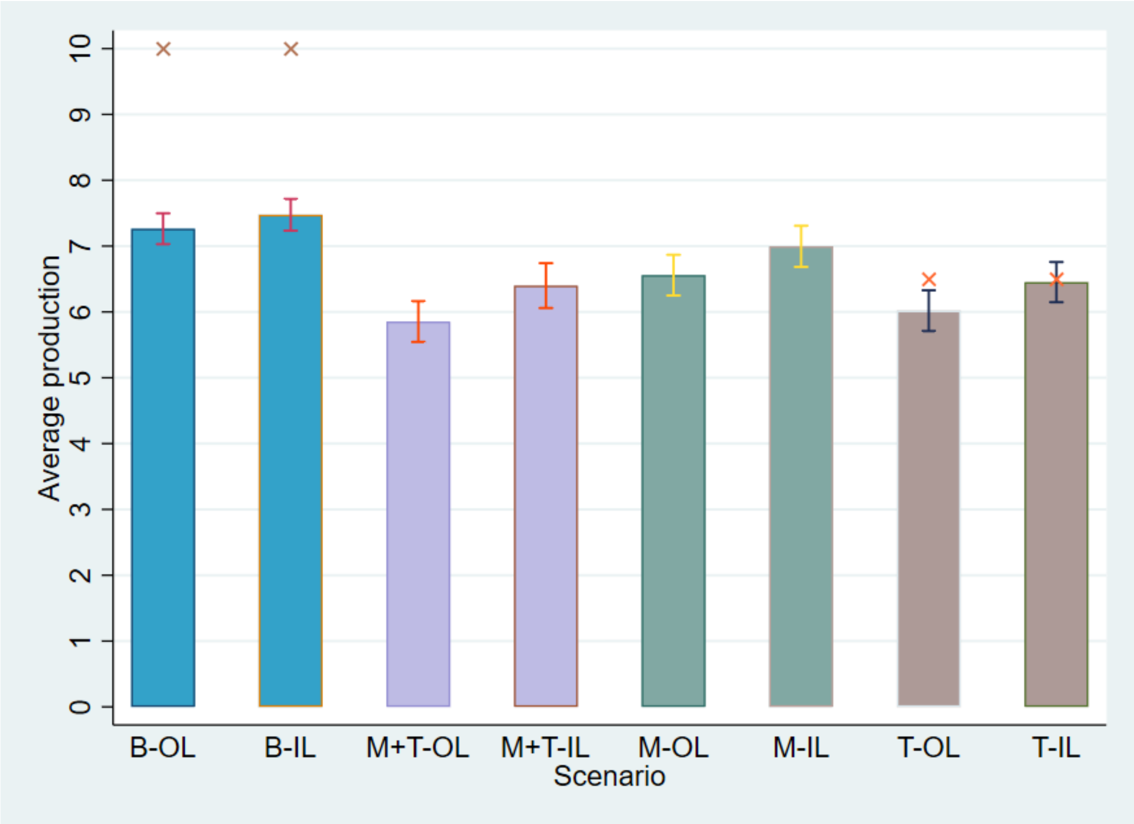
In summary, according to the tests performed, we can conclude that the Tax-only treatment is more effective than the Nudge-only treatment, which combining both instruments (Message + Tax) enhances the effect beyond applying the message in isolation but is not more effective than implementing the Tax alone.

Finally, comparing the results of the experiments inside versus outside the lab, the Tax, and the Tax + Message treatments seem to be more effective in reducing production with respect to the statistically equal baselines average levels of production Baseline in the public bad outside the lab experiments than in the public bad inside the lab experiments. We do not observe such a difference when comparing the effect of the Message.

Table 5. Descriptive Statistics for Public Bad Inside the Lab and Public Bad Outside the Lab Experiments

Public Bad Inside the Lab		T1: Baseline	T2: Tax	T3: Message	T4: Tax + Message
Individual Quantity produced per round (q)	Mean	7.47	6.44	7.00	6.40
	Median	8	6	6	6
	Std. Deviation	2.26	2.11	2.23	1.95
	Theory	≤ 10	≤ 6-7	≤ 10	≤ 6-7
	Social optimum	5	5	5	5
Public Bad Outside the Lab		T5: Baseline	T6: Tax	T7: Message	T8: Tax + Message
Individual Quantity produced per round (q)	Mean	7.26	6.02	6.56	5.85
	Median	8	6	6	6
	Std. Deviation	2.65	1.56	2.39	1.54
	Theory	≤ 10	≤ 6-7	≤ 10	≤ 6-7
	Social optimum	5	5	5	5
Individual average donation per round		27.36	40.74	31	43.2

Figure 1: Mean individual production by treatment and scenario



Note: The symbol "x" benchmarks the level of production of a profit maximizer subject. "IL" stands for experiment with the public bad inside the lab. "OL" stands for experiment with the donation to the NGO, the public bad outside the lab. "B" stands for baseline treatment, "M" message treatment and "T" for tax treatment. The treatment indicator for each scenario presented in the graph is: B-CD=T5; B-SD=T1; M+T-CD=T8; M+T-SD=T4; M-CD=T7; M-SD=T3; T-CD=T6; T-SD=T2.

5.2 Regressions

To complement the results of the nonparametric tests previously presented, we carried out an econometric analysis. Our outcome in this analysis is the level of production of subject *i* in round *t* (q_{it}). Models (1) and (3) in Table 6 present the results of random-effects linear panel regressions without and with controls for the public bad inside the lab experiments. The corresponding results for the public bad outside the lab experiments are presented in columns (2) and (4). Finally, we also performed pooled a pooled regression analysis whose results we present in column 5. The covariates that we include in specifications (3) and (4) are the university major of the subjects, their declared household income, their political ideology, their level of trust in others and how much they care about the NGO. This information was collected in the post-experiments survey.

The results suggest that in both contexts (public bad inside the lab and public bad outside the lab) the three interventions are effective in decreasing the average individual production level. Nevertheless, the different interventions do not have the same effect. In both settings, we find that:

(a) the effect of the tax is higher than that of the message ($|\hat{\beta}_{Tax}| > |\hat{\beta}_{Message}|$). The value of the *t*-statistic in the context of public bad outside the lab is 0.53 and in the setting of the public bad inside the laboratory is 0.56. The associated p-values for each test are 0.002 and 0.001, respectively.

(b) the effect of the Tax + Message is higher than the effect of the Message ($|\hat{\beta}_{Tax+Message}| > |\hat{\beta}_{Message}|$). The value of the *t*-statistic in the public bad outside the lab setting is 0.69 and in the public bad inside the laboratory setting 0.60. The associated p-values for each test are 0.000 and 0.002, respectively.

(c) the effect of combination of the two instruments (Tax + Message) is not different from the effect of the Tax ($|\hat{\beta}_{Tax+Message}| \approx |\hat{\beta}_{Tax}|$). The value of the *t*-statistic in the public bad outside the lab setting is 0.16 and in the public bad inside the laboratory setting 0.04. The associated p-values for each test are 0.35 and 0.82, respectively.

Finally, based on the estimates presented in column (5), the pooled model, and the corresponding *t*-tests (not shown), we did not find any significant differences in the behavior of the average subject in the corresponding treatments between the experiments inside and outside the lab, with one exception. We reject (at the 10% level) that the effect of the Message + Tax is the same between the experiments with the public bad inside the lab and the public bad outside the lab ($|\hat{\beta}^{IL}_{M+T}| = |\hat{\beta}^{OL}_{M+T}|$), in favor of the alternative hypothesis that the effect of the Message + Tax inside the lab is lower than when the public bads occurs outside the lab $|\hat{\beta}^{IL}_{M+T}| < |\hat{\beta}^{OL}_{M+T}|$.

Table 6. Linear Random Effect Models-Level of Production

	(1) Public bad Outside the Lab	(2) Public Bad inside the Lab	(3) Public bad Outside the Lab	(4) Public Bad inside the Lab	(5) Pooled
Message (M)	-0.72***	-0.47***	-0.73***	-0.48***	-0.48***
	(0.12)	(0.13)	(0.12)	(0.13)	(0.13)
Tax (T)	-1.24***	-1.03***	-1.25***	-1.03***	-1.02***
	(0.12)	(0.13)	(0.12)	(0.13)	(0.12)
Message+Tax (M+T)	-1.40***	-1.07***	-1.38***	-1.06***	-1.07***
	(0.12)	(0.15)	(0.12)	(0.15)	(0.14)
Outside the lab (OL)					-0.21
					(0.17)
OL*M					-0.23
					(0.18)
OL*T					-0.22
					(0.18)
OL*M+T					-0.34*
					(0.19)
Constant	7.26***	7.47***	8.39***	5.45***	7.47***
	(0.13)	(0.11)	(0.57)	(1.08)	(0.12)
Chi-squared	246.00	124.22	311.26	137.43	368.84
N	1950.00	1850.00	1950.00	1850.00	3800.00
Controls	No	No	Yes	Yes	No
MvsT	0.53	0.56	0.53	0.56	0.55
	(.174)	(.176)	(.172)	(.177)	(.177)
TvsM+T	0.16	0.04	0.13	0.03	0.17
	(.174)	(.189)	(.172)	(.19)	(.177)
MvsM+T	0.69	0.60	0.66	0.58	0.71
	(.174)	(.192)	(.172)	(.193)	(.177)

6 Conclusions

We find that, for both settings (a public bad inside and outside the lab), compared to the Baseline treatment (no tax, no nudge), both the Tax (Hypothesis 1) and Nudge (Hypothesis 2) were effective in decreasing the average level of a negative externality. We find also that the average level of the negative externality under the joint implementation of the Tax and Nudge intervention is lower than under only a Nudge, consistent with our Hypothesis 4 (Taxes are complements to Nudges). Nevertheless, we did not find evidence in favor of the hypothesis that Nudges are complements to Taxes (Hypothesis 3). We could not reject that the average individual level of negative externality in the under the joint implementation of the Tax and the Nudge is equal to that in the average level of this externality when only a Tax is implemented.

To conclude, our experiments provide evidence consistent with the effect of a tax on a negative externality being higher than those of a nudge in the form of a message informing players what the optimal level of the externality is, implicitly suggesting what to play to a utilitarian player. We obtain this result even though the implemented tax was only half of the tax needed to induce the socially optimum level of the externality. Moreover, such a low tax may add to the effect obtained by the implementation of the nudge alone. The reverse is not true, nevertheless, suggesting that taxes may complement nudges but not the other way around. This evidence is inconsistent with the policy recommendation of implementing nudges to complement low taxes, while we wait for the political will to increment taxes to develop.

7 References

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

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Ricardian equivalence. *Journal of Political Economy*, 97(6), 1447-1458.

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.

Barmettler, Franziska., Fehr, Ernst, and Zehnder, Christian. (2012). "Big experimenter is watching you! Anonymity and prosocial behavior in the laboratory." *Games and Economic Behavior* 75(1): 17-34.

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

Bicchieri, C., Dimant, E., Gächter, S, and Nosenso, D. (2019). "Social Proximity and the Evolution of Norm Compliance." *Games and Economic Behavior* 132: 59-71.

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

Brent, Daniel and Wichman, Casey J., Do Behavioral Nudges Interact with Prevailing Economic Incentives? Pairing Experimental and Quasi-Experimental Evidence from Water Consumption. Available at SSRN: https://ssrn.com/abstract=4162438

Buchholz, 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.

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.

Carlsson, Frederik and Johansson-Stenman, Olof (2019). “Optimal Prosocial Nudging.” Working Paper in Economics No 757, Department of Economics, University of Gothenburg.

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

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.

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).

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.

Glaeser, Edward L., 2006. "Paternalism and Psychology," Working Paper Series rwp06-006, Harvard University, John F. Kennedy School of Government.

Glaeser, Edward L, 2014. "The Supply of Environmentalism: Psychological Interventions and Economics," Review of Environmental Economics and Policy, Association of Environmental and Resource Economists, vol. 8(2), pages 208-229.

Hernández, F., Jaime, M. M., & Vasquez, F. Are Price Changes More Effective than Nudging to Reduce Water Use? *Available at SSRN 4441756*.

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.

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

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

López, Maria Claudia and Murphy, James J. and Spraggon, John M. and Stranlund, John. (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.

Loewenstein, George and O'Donoghue, Ted (2006) "We Can Do This the Easy Way or the Hard Way": Negative Emotions, Self-Regulation, and the Law," *University of Chicago Law Review*: Vol. 73: Iss. 1, Article 10.

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.

Miao, Haoran, Trandafir, Simona and Uchida, Emi. (2017). "The effect of informational nudges to promote voluntary behavior to reduce nonpoint source pollution: A randomized controlled trial in the field." Unpublished working paper.

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, Luca, Ulph, Alistair, Zizzo, Daniel John, Hilton, D., Clear, Adrian. (2021). "The impact of environmental recall and carbon taxation on the carbon footprint of supermarket shopping." *Journal of Environmental Economics and Management* 102137.

Pellerano, J. A., Price, M. K., Puller, S. L., & Sánchez, G. E. (2017). Do extrinsic incentives undermine social norms? Evidence from a field experiment in energy conservation. *Environmental and resource Economics*, *67*, 413-428.

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

Sunstein, Cass and Thaler, Richard. (2009). *NUDGE: Improving Decisions About Health, Wealth, and Happiness.* Penguin Books.

Sunstein, Cass. (2017). "Nudges that fail." *Behavioral Public Policy* 1(1): 4-25.

Vélez, María Alejandra, Murphy, James, and Stranlund, John. (2010). "Centralized and Decentralized Management of Local Common Pool Resources in the Developing World: Experimental Evidence from Fishing Communities in Colombia." *Economic Inquiry* 48(2): 254-265.

World Bank Group. 2023. State and Trends of Carbon Pricing 2019. Washington, DC: World Bank. © World Bank.

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.

8 Appendix A

Table A.1: Descriptive statistics for environmental questions

Variable		Subjects	
		Public bad inside the lab	Public bad outside the lab
"It is necessary to apply taxes to polluters"	Disagree (0,1 or 2)	6%	7%
	Neutral (3)	9%	7%
	Agree (4, 5 or 6)	85%	86%
"Indicate how often do you Reuse your shopping bags"	Never	14%	22%
	Seldom	53%	45%
	Sometimes	16%	16%
	Often	12%	10%
	Always	5%	7%
"Indicate how often do you bring empty bottles to the recycling bins"	Never	29%	21%
	Seldom	26%	23%
	Sometimes	16%	18%
	Often	18%	23%
	Always	11%	15%
"Do you approve of the use of messages by a government to inform the population about the environmental effects of our behaviors? "	Yes	97%	98%
	No	3%	2%

9 Appendix B

9.1 Non-Parametric Tests

9.1.1 Public bad inside the laboratory

Table B.1.1: Baseline vs Message

Median test: H0: the median for the baseline (B) game and for the M game are the same).

```
Median test

   Greater
  than the          Tratamiento
    median        Base    Mensaje        Total

        no         439        187          626
       yes         486        138          624

     Total         925        325        1,250

          Pearson chi2(1) =    9.7726   Pr = 0.002

   Continuity corrected:
          Pearson chi2(1) =    9.3736   Pr = 0.002
```

Ranksum test: H0: Both samples (B and M) come from populations with equal distribution

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

 Tratamiento       obs     rank sum     expected

        Base       925     594311.5     578587.5
     Mensaje       325     187563.5     203287.5

    combined      1250       781875       781875

unadjusted variance     31340156
adjustment for ties   -1154237.3
                      ----------
adjusted variance       30185919

Ho: q(Tratam~o==Base) = q(Tratam~o==Mensaje)
             z =   2.862
    Prob > |z| =   0.0042
```

Table B.1.2: Baseline vs Tax

Median test: H0: the median for the B game and for the T game are the same

```
Median test

   Greater |
  than the |      Tratamiento
    median |      Base   Impuesto |     Total
-----------+----------------------+----------
        no |       439        231 |       670
       yes |       486         94 |       580
-----------+----------------------+----------
     Total |       925        325 |     1,250

          Pearson chi2(1) =  53.9385   Pr = 0.000

   Continuity corrected:
          Pearson chi2(1) =  52.9931   Pr = 0.000
```

Ranksum test - H0: Both samples (B and M) come from populations with equal distribution.

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

 Tratamiento |      obs    rank sum    expected
-------------+---------------------------------
        Base |      925    619471.5    578587.5
    Impuesto |      325    162403.5    203287.5
-------------+---------------------------------
    combined |     1250      781875      781875

unadjusted variance    31340156
adjustment for ties  -813600.14
                     ----------
adjusted variance      30526556

Ho: q(Tratam~o==Base) = q(Tratam~o==Impuesto)
             z =   7.400
    Prob > |z| =   0.0000
```

Table B.1.3: Baseline vs (Message + Tax)

Median test: H0: the median for the B game and for the (M+T) game are the same.

Median test

Greater than the median	Tratamiento Base	Mensaje+I	Total
no	**439**	**177**	**616**
yes	**486**	**73**	**559**
Total	**925**	**250**	**1,175**

```
         Pearson chi2(1) =   42.9882   Pr = 0.000

   Continuity corrected:
         Pearson chi2(1) =   42.0574   Pr = 0.000
```

Ranksum test - H0: Both samples (B and (M+T)) come from populations with equal distribution.

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

Tratamiento	obs	rank sum	expected
Base	**925**	**578225**	**543900**
Mensaje+Impu	**250**	**112675**	**147000**
combined	**1175**	**690900**	**690900**

```
unadjusted variance    22662500
adjustment for ties  -648204.67
                     ----------
adjusted variance      22014295

Ho: q(Tratam~o==Base) = q(Tratam~o==Mensaje+Impuesto)
             z =   7.316
    Prob > |z| =   0.0000
```

Table B.1.4: Message vs Tax

Median test: H0: the median for the M game and for the T game are the same.

Median test

Greater than the median	Tratamiento Mensaje	Impuesto	Total
no	159	187	346
yes	166	163	329
Total	325	350	675

```
         Pearson chi2(1) =   1.3692   Pr = 0.242

   Continuity corrected:
         Pearson chi2(1) =   1.1948   Pr = 0.274
```

Ranksum test - H0: Both samples (M and T) come from populations with equal distribution.

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

Tratamiento	obs	rank sum	expected
Mensaje	325	119416.5	109850
Impuesto	350	108733.5	118300
combined	675	228150	228150

```
unadjusted variance   6407916.67
adjustment for ties   -168752.92
                      ----------
adjusted variance     6239163.75

Ho: q(Tratam~o==Mensaje) = q(Tratam~o==Impuesto)
             z =   3.830
    Prob > |z| =   0.0001
```

Table B.1.5: Message vs (Message + Tax)

Median test: H0: the median for the M game and for the (M + T) game are the same.

Median test

Greater than the median	Tratamiento Mensaje	Mensaje+I	Total
no	159	147	306
yes	166	103	269
Total	325	250	575

```
         Pearson chi2(1) =   5.5368   Pr = 0.019

   Continuity corrected:
         Pearson chi2(1) =   5.1472   Pr = 0.023
```

Ranksum test – Ho: Both samples (M and (M + T)) come from populations with equal distribution.

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

Tratamiento	obs	rank sum	expected
Mensaje	325	100252	93600
Mensaje+Impu	250	65348	72000
combined	575	165600	165600

```
unadjusted variance   3900000.00
adjustment for ties   -196505.73
                      ----------
adjusted variance     3703494.27

Ho: q(Tratam~o==Mensaje) = q(Tratam~o==Mensaje+Impuesto)
             z =   3.457
    Prob > |z| =   0.0005
```

Table B.1.6: Tax vs (Message + Tax)

Median test: Ho: the median for the T game and for the (M + T) game are the same.

Median test

Greater than the median	Tratamiento Mensaje+I	Impuesto	Total
no	147	187	334
yes	103	163	266
Total	250	350	600

```
         Pearson chi2(1) =   1.7049   Pr = 0.192

   Continuity corrected:
         Pearson chi2(1) =   1.4942   Pr = 0.222
```

Ranksum test – Ho: Both samples (T and (M + T)) come from populations with equal distribution.

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

Tratamiento	obs	rank sum	expected
Mensaje+Impu	250	75130	75125
Impuesto	350	105170	105175
combined	600	180300	180300

```
unadjusted variance   4382291.67
adjustment for ties   -106215.24
                      ----------
adjusted variance     4276076.43

Ho: q(Tratam~o==Mensaje+Impuesto) = q(Tratam~o==Impuesto)
             z =   0.002
    Prob > |z| =   0.9981
```

9.1.2 Public bad outside the laboratory

Table B.2.1: Baseline vs Message

Median test: Ho: the median for the B game and for the M game are the same.

Median test

Greater than the median	Tratamiento Base	Mensaje	Total
no	439	187	626
yes	486	138	624
Total	925	325	1,250

```
         Pearson chi2(1) =    9.7726   Pr = 0.002

   Continuity corrected:
         Pearson chi2(1) =    9.3736   Pr = 0.002
```

Ranksum test – Ho: Both samples (B and M) come from populations with equal distribution.

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

Tratamiento	obs	rank sum	expected
Base	925	594311.5	578587.5
Mensaje	325	187563.5	203287.5
combined	1250	781875	781875

```
unadjusted variance    31340156
adjustment for ties  -1154237.3
                     ----------
adjusted variance      30185919

Ho: q(Tratam~o==Base) = q(Tratam~o==Mensaje)
             z =   2.862
    Prob > |z| =   0.0042
```

Table B.2.2: Baseline vs Tax

Median test: Ho: the median for the B game and for the **T** game are the same.

Median test

Greater than the median	Tratamiento Base	Impuesto	Total
no	439	255	694
yes	486	95	581
Total	925	350	1,275

```
          Pearson chi2(1) =  66.0348   Pr = 0.000

   Continuity corrected:
          Pearson chi2(1) =  65.0148   Pr = 0.000
```

Ranksum test: – Ho: Both samples (B and T) come from populations with equal distribution.

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

Tratamiento	obs	rank sum	expected
Base	925	638559	590150
Impuesto	350	174891	223300
combined	1275	813450	813450

```
unadjusted variance    34425417
adjustment for ties  -864757.81
                     ----------
adjusted variance      33560659

Ho: q(Tratam~o==Base) = q(Tratam~o==Impuesto)
             z =   8.356
    Prob > |z| =   0.0000
```

Table B.2.3: Baseline vs (Message + Tax)

Median test: - Ho: the median for the **B** game and for the **(M+T)** game are the same.

```
Median test

   Greater |
  than the |        Tratamiento
    median |      Base   Mensaje+I |     Total
-----------+-----------------------+----------
        no |       439         177 |       616
       yes |       486          73 |       559
-----------+-----------------------+----------
     Total |       925         250 |     1,175

          Pearson chi2(1) =  42.9882   Pr = 0.000

   Continuity corrected:
          Pearson chi2(1) =  42.0574   Pr = 0.000
```

Ranksum test: – Ho: Both samples (B and (M+T)) come from populations with equal distribution.

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

 Tratamiento |      obs    rank sum    expected
-------------+---------------------------------
        Base |      925      578225      543900
Mensaje+Impu |      250      112675      147000
-------------+---------------------------------
    combined |     1175      690900      690900

unadjusted variance    22662500
adjustment for ties  -648204.67
                     ----------
adjusted variance      22014295

Ho: q(Tratam~o==Base) = q(Tratam~o==Mensaje+Impuesto)
             z =   7.316
    Prob > |z| =   0.0000
```

Table B.2.4: Message vs Tax

Median test: - Ho: the median for the **M** game and for the **T** game are the same.

Median test

Greater than the median	Tratamiento Mensaje	Impuesto	Total
no	159	187	346
yes	166	163	329
Total	325	350	675

```
         Pearson chi2(1) =   1.3692   Pr = 0.242

   Continuity corrected:
         Pearson chi2(1) =   1.1948   Pr = 0.274
```

Ranksum test: - Ho: Both samples **(M and T)** come from populations with equal distribution.

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

Tratamiento	obs	rank sum	expected
Mensaje	325	119416.5	109850
Impuesto	350	108733.5	118300
combined	675	228150	228150

```
unadjusted variance   6407916.67
adjustment for ties   -168752.92
                      ----------
adjusted variance     6239163.75

Ho: q(Tratam~o==Mensaje) = q(Tratam~o==Impuesto)
             z =   3.830
    Prob > |z| =   0.0001
```

Table B.2.5: Message vs (Message + Tax)

Median Test: - Ho: the median for the **M** game and for the **(M+T)** game are the same.

Median test

Greater than the median	Tratamiento Mensaje	Mensaje+I	Total
no	159	147	306
yes	166	103	269
Total	325	250	575

```
         Pearson chi2(1) =   5.5368   Pr = 0.019

   Continuity corrected:
         Pearson chi2(1) =   5.1472   Pr = 0.023
```

Ranksum test: - Ho: Both samples **(M and (M+T))** come from populations with equal distribution.

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

Tratamiento	obs	rank sum	expected
Mensaje	325	100252	93600
Mensaje+Impu	250	65348	72000
combined	575	165600	165600

```
unadjusted variance   3900000.00
adjustment for ties   -196505.73
                      ----------
adjusted variance     3703494.27

Ho: q(Tratam~o==Mensaje) = q(Tratam~o==Mensaje+Impuesto)
             z =   3.457
    Prob > |z| =   0.0005
```

Table B.2.6: Tax vs (Message + Tax)

Median Test: - Ho: the median for the **T** game and for the **(M+T)** game are the same.

Median test

Greater than the median	Tratamiento Mensaje+I	Impuesto	Total
no	147	187	334
yes	103	163	266
Total	250	350	600

```
         Pearson chi2(1) =    1.7049   Pr = 0.192

   Continuity corrected:
         Pearson chi2(1) =    1.4942   Pr = 0.222
```

Ranksum test: - Ho: Both samples **(T and (M+T))** come from populations with equal distribution.

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

Tratamiento	obs	rank sum	expected
Mensaje+Impu	250	75130	75125
Impuesto	350	105170	105175
combined	600	180300	180300

```
unadjusted variance   4382291.67
adjustment for ties   -106215.24
                      ----------
adjusted variance     4276076.43

Ho: q(Tratam~o==Mensaje+Impuesto) = q(Tratam~o==Impuesto)
             z =   0.002
    Prob > |z| =   0.9981
```

9.1.3 Pooled sample (Inside and outside the lab)

Table B.3.1: Baseline vs Message

Median Test: - Ho: the median for the **B** game and for the **M** game are the same.

Median test

Greater than the median	Tratamiento Base	Mensaje	Total
no	904	391	1,295
yes	996	259	1,255
Total	1,900	650	2,550

```
          Pearson chi2(1) =  30.6410   Pr = 0.000

   Continuity corrected:
          Pearson chi2(1) =  30.1399   Pr = 0.000
```

Ranksum test: - Ho: Both samples **(B and M)** come from populations with equal distribution.

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

Tratamiento	obs	rank sum	expected
Base	1900	2491636	2423450
Mensaje	650	760889	829075
combined	2550	3252525	3252525

```
unadjusted variance    2.625e+08
adjustment for ties   -9817909.7
                      ----------
adjusted variance      2.527e+08

Ho: q(Tratam~o==Base) = q(Tratam~o==Mensaje)
             z =   4.289
    Prob > |z| =   0.0000
```

Table B.3.2: Baseline vs Tax

Median test: - Ho: the median for the **B** game and for the **T** game are the same.

```
Median test

   Greater
  than the          Tratamiento
    median        Base   Impuesto      Total

        no         904        527      1,431
       yes         996        123      1,119

     Total       1,900        650      2,550

          Pearson chi2(1) = 220.6859   Pr = 0.000

   Continuity corrected:
          Pearson chi2(1) = 219.3277   Pr = 0.000
```

Ranksum test: - Ho: Both samples **(B and T)** come from populations with equal distribution.

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

 Tratamiento        obs    rank sum    expected

        Base       1900     2618116     2423450
    Impuesto        650      634409      829075

    combined       2550     3252525     3252525

unadjusted variance   2.625e+08
adjustment for ties  -6800309.4

adjusted variance     2.557e+08

Ho: q(Tratam~o==Base) = q(Tratam~o==Impuesto)
             z =  12.173
    Prob > |z| =   0.0000
```

Table B.3.3: Baseline vs (Message + Tax)

Median test: - Ho: the median for the **B** game and for the **(M+T)** game are the same.

Median test

Greater than the median	Tratamiento Base	Mensaje+I	Total
no	**904**	**475**	**1,379**
yes	**996**	**100**	**1,096**
Total	**1,900**	**575**	**2,475**

Pearson chi2(**1**) = **219.5310** Pr = **0.000**

Continuity corrected:
Pearson chi2(**1**) = **218.1135** Pr = **0.000**

Ranksum test: - Ho: Both samples **(B and (M+T))** come from populations with equal distribution.

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

Tratamiento	obs	rank sum	expected
Base	**1900**	**2551668**	**2352200**
Mensaje+Impu	**575**	**512382**	**711850**
combined	**2475**	**3064050**	**3064050**

unadjusted variance	**2.254e+08**
adjustment for ties	**-6160254.8**
adjusted variance	**2.193e+08**

Ho: q(Tratam~o==Base) = q(Tratam~o==Mensaje+Impuesto)
z = **13.471**
Prob > |z| = **0.0000**

Table B.3.4: Message vs Tax

Median test: - Ho: the median for the **M** game and for the **T** game are the same.

```
Median test

   Greater |
  than the |        Tratamiento
    median |   Mensaje   Impuesto |     Total
-----------+----------------------+----------
        no |       337        389 |       726
       yes |       313        261 |       574
-----------+----------------------+----------
     Total |       650        650 |     1,300

          Pearson chi2(1) =   8.4353   Pr = 0.004

   Continuity corrected:
          Pearson chi2(1) =   8.1140   Pr = 0.004
```

Ranksum test: - Ho: Both samples **(M and T)** come from populations with equal distribution.

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

 Tratamiento |      obs    rank sum    expected
-------------+---------------------------------
     Mensaje |      650    451727.5      422825
    Impuesto |      650    393922.5      422825
-------------+---------------------------------
    combined |     1300      845650      845650

unadjusted variance    45806042
adjustment for ties  -1445545.1
                     ----------
adjusted variance      44360497

Ho: q(Tratam~o==Mensaje) = q(Tratam~o==Impuesto)
             z =   4.339
    Prob > |z| =   0.0000
```

Table B.3.5: Message vs (Message + Tax)

Median test: - Ho: the median for the **M** game and for the **(M+T)** game are the same.

Median test

Greater than the median	Tratamiento Mensaje	Mensaje+I	Total
no	337	397	734
yes	313	178	491
Total	650	575	1,225

Pearson chi2(1) = 37.5718 Pr = 0.000

Continuity corrected:
Pearson chi2(1) = 36.8591 Pr = 0.000

Ranksum test: - Ho: Both samples **(M and (M+T))** come from populations with equal distribution.

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

Tratamiento	obs	rank sum	expected
Mensaje	650	438499.5	398450
Mensaje+Impu	575	312425.5	352475
combined	1225	750925	750925

```
unadjusted variance     38184792
adjustment for ties   -1946704.1
                      ----------
adjusted variance       36238088

Ho: q(Tratam~o==Mensaje) = q(Tratam~o==Mensaje+Impuesto)
             z =   6.653
    Prob > |z| =   0.0000
```

Table B.3.6: Tax vs Message + Tax

Median test: - Ho: the median for the **T** game and for the **(M+T)** game are the same.

Median test

Greater than the median	Tratamiento Mensaje+I	Impuesto	Total
no	**397**	**389**	**786**
yes	**178**	**261**	**439**
Total	**575**	**650**	**1,225**

```
         Pearson chi2(1) =  11.2241   Pr = 0.001

   Continuity corrected:
         Pearson chi2(1) =  10.8277   Pr = 0.001
```

Ranksum test: - Ho: Both samples **(T and (M+T))** come from populations with equal distribution.

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

Tratamiento	obs	rank sum	expected
Mensaje+Impu	**575**	**331746**	**352475**
Impuesto	**650**	**419179**	**398450**
combined	**1225**	**750925**	**750925**

```
unadjusted variance     38184792
adjustment for ties   -1497739.6
                      ----------
adjusted variance       36687052

Ho: q(Tratam~o==Mensaje+Impuesto) = q(Tratam~o==Impuesto)
             z =  -3.422
    Prob > |z| =   0.0006
```