Marcelo Caffera* and Carlos Chávez

The Regulatory Choice of Noncompliance in the Lab: Effect on Quantities, Prices, and Implications for the Design of a Cost-Effective Policy

DOI 10.1515/bejeap-2014-0202
Published online February 5, 2016

Abstract: Recent theoretical developments show the conditions under which it is cost-effective for the regulator to induce perfect compliance in cap-and-trade programs. These conditions are based on the ability that a regulator with perfect information has to induce the firms to emit any desired level with different combinations of the number of permits supplied to the market and the monitoring probability, assuming that firms are expected profit maximizers. In this paper, we test this hypothesis with a series of laboratory experiments. Our results suggest that firms may behave significantly different from what these models predict precisely when the different combinations of the supply of permits and the monitoring probability induce compliance versus noncompliance. More specifically, by allowing noncompliance in a manner consistent with theory, the regulator could produce a decrease in emissions and an increase in the market price of tradable permits that is not predicted by the theoretical models. The implications for the cost-effective design of environmental policy are discussed.

Keywords: emission standards, emissions trading, enforcement, environmental policy, laboratory experiments
JEL Classification: C91, L51, Q58, K42

1 Introduction

Cost-effectiveness is an important criterion for policy design in the environmental economics literature. Notwithstanding, this literature has only recently began

***Corresponding author: Marcelo Caffera,** Departamento de Economía, Universidad de Montevideo, Prudencio de Pena 2544, Montevideo 11600, Uruguay,
E-mail: marcaffera@um.edu.uy
Carlos Chávez, Departamento de Economía, Research Nucleus in Environmental and Natural Resource Economics, and Interdisciplinary Center for Aquaculture Research (INCAR), Universidad de Concepción, Concepción, Chile

to study the cost-effective design of environmental policy when not only abatement but also enforcement costs are included in the calculations. When considering enforcement costs, a relevant issue for the regulator is whether or not it is cost-effective to induce perfect compliance. More precisely, the question is the following: If a regulator wants to cap the aggregate level of emissions from a set of n firms at a certain level Q, what would be cheaper: to set the legal cap at Q and perfectly enforce the program, or to set the legal cap at a lower level, but allow violations such that the aggregate level of emissions is Q? Stranlund (2007) was the first to address this question for the case of a cap-and-trade program. Because in both alternatives each individual firm emits the same amount, the difference in costs between the two alternatives is given by the difference in monitoring and sanctioning costs borne by the regulator. Stranlund (2007) shows that (a) this difference depends on the fine structure and (b) inducing perfect compliance with a marginal penalty tied to the equilibrium price of permits minimizes the total expected costs of the pollution control program. Arguedas (2008) proved that these conclusions are extendable for the case of a single firm confronted to an emission standard. Finally, Caffera and Chávez (2011) proved that these conclusions are also valid for the case in which a regulator caps the emissions of n firms with emission standards and when not only abatement but also monitoring and sanctioning costs differ between firms. They also compare the total cost of a cap-and-trade program with that of an expected cost-minimizing allocation of abatement responsibilities (emission standards) and monitoring probabilities. They do it for the case in which it is cost-effective for the regulator to induce perfect compliance, and for the case in which it is not. Their results show that: (a) when it is cost-effective to induce perfect compliance, a cap-and-trade program minimizes the total cost of capping aggregate emissions only if monitoring costs are the same across firms or the marginal penalty for violations is constant, and (b) when it is not cost-effective to induce perfect compliance, the conditions under which a cap-and-trade program minimizes total costs are implausible.

A fundamental aspect for the comparison of the cost of inducing perfect compliance versus the cost of allowing violations in these works is the possibility that a perfectly informed regulator has to induce each individual firm to emit a given desired level of pollution with different combinations of the aggregate supply of emission permits (or the level of emission standards) *and* the monitoring probability. In this paper, we present the results of a series of laboratory experiments designed to test this possibility.

The theoretical complementarity of inspections and emission permits as leverages for capping emissions has not been previously examined in the experimental literature. Among the aspects of the enforcement of cap-and-trade systems and emission standards that have been previously examined in the experimental literature are: the existence and extent of a direct and indirect (through the permit price) effect of enforcement on emissions trading programs (Murphy and Stranlund 2006), the possibility of targeting enforcement in emissions trading programs and emissions standards programs (Murphy and Stranlund 2007), the effect of environmental framing (Cason and Raymond 2010), the perception of policy fairness as driver of the subjects' truthfulness in emission reports and compliance behavior (Raymond and Cason 2011), and the level of violations, emissions and prices of permits in the context of dynamic enforcement, banking and random emission shocks (Cason and Gangadharan 2006).

The set of the different combinations of the aggregate supply of permits (or emission standards) and the monitoring probability to which the firms respond with the same level of emissions are derived from the standard models on which the economic theory of enforcing pollution control policy is based. In these, the firms are assumed to be expected profit maximizers. In our experiments, we compare the behavior of firms when the regulator induces perfect compliance and when it induces the same level of emissions as in the perfect compliance case, but allowing a certain level of violations. We do this for two different regulatory programs, cap-and-trade and emission standards. In both programs, the desired levels of emissions are induced cost-effectively, i. e.: setting the monitoring probability arbitrarily close to the minimum possible level that induces the desired level of emissions.

Contrary to what the standard theoretical models predict, the results of our experiments suggest that firms may behave significantly different when confronted to a combination of supply of permits and monitoring probability that induces them to comply, as compared to the case in which they are confronted to a pair of supply of permits and monitoring probability that induces expected profit maximizers to emit the same of emissions, but violating their permits holding. More specifically, by inducing noncompliance in a manner consistent with theoretical recommendations, the regulator could produce a nonsought significant decrease in emissions and increase in the market price of tradable permits.

In the case of tradable permits, on average, the aggregate level of emissions is 10.5 % higher than the cap when the regulator induces compliance, and it is 12.5 % lower than the cap when it induces violations. In the case of emission standards, the average aggregate level of emissions is 13.5 % higher than the cap when it induces compliance, but about the same level predicted by the model based on rational, risk-neutral agents when the regulator induces violations (1.5 % lower than the cap).

According to these results, given a penalty function, a regulator needs a higher monitoring probability than that predicted by the conventional theory to actually induce perfect compliance in a cap-and-trade scheme or an emission standards scheme. Interestingly, at the same time, for a given decrease in the supply of permits, a regulator could actually decrease the monitoring probability more than what is suggested by the conventional theory if the objective is to induce the same level of emissions as in the perfect compliance case, but allowing noncompliance. In terms of cost-effectiveness, this means that the regulator could save more monitoring costs than those predicted by the conventional theoretical models by allowing noncompliance in the case of tradable permits. In other words, allowing noncompliance in tradable permits could be more convenient in terms of the cost savings in the real world than in the theoretical world of profit-maximizing, risk-neutral agents.

The paper is organized as follows. In Section 2, we present the main hypotheses we want to evaluate with our laboratory experiments. Section 3 contains a description of the experimental design and procedures. Section 4 presents the results. Finally, in Section 5, we put forward concluding remarks from our work.

2 Hypotheses

In this section, we present the main hypotheses that we evaluate with our laboratory experiments. The conceptual models from which our hypotheses are derived are presented in Section 1 of Online Appendix.[1] These follow directly from the works of Stranlund (2007), Arguedas (2008) and Caffera and Chávez (2011).

Assume a cost-minimizing regulator with perfect information about the marginal benefits that each of the regulated firms derive from pollution. This regulator is interested in capping the aggregate emissions of these firms at a level E. To enforce this cap, the regulator inspects firms with a predetermined probability, π, which is known by the firms. If an inspection discovers a violation, the firm is fined with a monetary fine f, which is an increasing function of the units of emissions in excess of the legal level. We consider two policy instruments: tradable pollution permits and emission standards. Our first hypothesis concerns the case in which the regulator chooses to implement a system of tradable pollution permits.

Hypothesis 1: *Assume that, given a penalty function f, a regulator induces perfect compliance in a system of tradable pollution permits* $(L = E)$ *by auditing each of*

1 The Online Appendix is available at http://www2.um.edu.uy/marcaffera/investigacion/OnlineAppendix_Paper_1_ BEJEAP.pdf.

the polluting firms with a pre-determined probability π^c. Then, this regulator can induce the same level of individual emissions that induce in this case of perfect enforcement by varying the aggregate supply of permits and the monitoring probability according to $\frac{d\pi}{dL} = \frac{\pi f''}{nf'} > 0$, where f' and f'' are the first- and the second-order derivatives of the penalty function and n is the number of regulated firms.

Our second hypothesis concerns a system of emission standards. In this case, the leverages by which the regulator can induce compliance or allow violations such that the level of emissions of each individual firm remains constant are the emission standard set for firm i (s_i) and the probability that the firm i faces of receiving an inspection (π_i).

Hypothesis 2: *Assume that, given a penalty function f, a regulator induces perfect compliance in a system of emission standards $\left(\sum_{i=1}^{n} s_i = E\right)$ by auditing each of the polluting firms with a pre-determined probability π_i^c. Then, this regulator can induce the same level of individual emissions that induces in this case of perfect enforcement by varying the individual emission standards and the monitoring probability according to $\frac{d\pi_i}{ds_i} = \frac{\pi_i f''}{f'} > 0$.*

3 Experimental Design and Procedures

3.1 Experimental Design

We framed the experiments as a neutral production decision of an unspecified fictitious good q, from which the subjects obtained benefits. Every subject had a production capacity of 10 units (whole numbers), but the production of these units did not derive the same benefits for every subject. Each subject had one of four possible marginal benefits (obtained from Cason and Gangadharan (2006)), which gave place to four “types” of subjects. The four marginal benefits are depicted in Table A.1 of the Online Appendix. These schedules of marginal benefits were the same through all the experiments and were randomly assigned between subjects.

3.1.1 Tradable Permits

In the permit experiments, subjects had to possess a permit in order to be legally able to produce one unit of the good. Consequently, subjects had to decide how much to produce of the fictitious good and how many permits to buy or sell. In order to buy or sell permits, subjects participated in a double-auction market, one permit at a time. A market was formed by eight subjects, two of each type. After their decision, at the end of each period, the subjects were audited with a known homogeneous predetermined and exogenous probability π. If audited, the number of units produced by the subject i in that period (q_i) was compared with the number of permits possessed by the subject i (l_i) at the end of the period. If the level of production chosen was higher than the number of permits possessed, the subject was automatically fined according to a penalty function of the form $f(v) = \varphi v + (\gamma/2)v^2$, with $\varphi > 0$ and $\gamma > 0$. We use this fine structure, which implies an increasing marginal fine $f'(v) = \varphi + \gamma v$, because it is never optimal to induce or allow violations when the marginal fine is constant, as we know from the previously cited literature. Arguedas (2008) calls the parameter φ the linear component of the fine and the parameter γ the gravity component. The subjects had the information on the probability of inspection that they faced and on the marginal fine for every level of violation in their screens at every moment before making their decisions.

We constructed two treatments for the case of markets for permits (see Table 1). In treatment M1, the total number of tradable permits supplied to each group of eight subjects was 40. The initial allocation was four permits for subjects of types 1 and 2, the prospective buyers, and six permits for subjects of types 3 and 4, the prospective sellers. We chose this initial allocation of permits as opposed to a homogeneous allocation (five each) as a way to foster the market activity (the number of expected trades is 10).[2] The enforcement parameters took the values $\varphi = 100$, $\gamma = 66,67$ and $\pi = 0.6$. This probability is sufficient to induce all types of firms to comply with their permit holdings under the assumption of risk neutrality.[3] The resulting perfect compliance equilibrium price of the market is expected to be between 74 experimental pesos (E$) and E$ 80. In contrast, treatment M2 induces violations of the permit holdings. This is done by decreasing the total number of permits supplied to 20 (initial allocations and expected number of trades halved) and by decreasing the monitoring probability from 0.6 to 0.30. With this parameterization, under the assumption

2 The theoretical predictions for our experiments are included in Section 2 of the Online Appendix.

3 It also induces compliance for the risk-averse individuals (Stranlund 2008). We deal with risk aversion in Section 4.

Table 1: Treatments.

Treatment	Regulation	Monitoring probability by firm's type				Fine parameter values		Policy induces	Number of tradable permits supplied/ aggregate standard	Equilibrium price/ emission standards	Expected aggregate level of emissions
		Type 1	Type 2	Type 3	Type 4	Phi (φ)	Gamma (γ)				
M1	Tradable	0.60	0.60	0.60	0.60	100	66.66	Compliance	40	80–74	40
M2	permits	0.30	0.30	0.30	0.30			Violations	20		
S1	Standards	0.60	0.65	0.63	0.66			Compliance	40	Type 1 = 7	
										Type 2 = 6	
										Type 3 = 4	
										Type 4 = 3	
S2		0.24	0.26	0.32	0.32			Violations	20	Type 1 = 4	
										Type 2 = 3	
										Type 3 = 2	
										Type 4 = 1	

of expected profit maximization, the treatment M2 induces the same equilibrium price of permits and individual level of emissions as the treatment M1 does.[4] Hence, the expected level of aggregate emissions remains in 40 units. This is a unique feature of our design. Another unique feature of our design is that each subject participates in both the M1 and M2 treatments.

3.1.2 Standards

In the standards experiments, subjects faced a maximum allowable level of production (the standard) and had to decide how much to produce. The auditing procedure was exactly the same as in the case of tradable permits; except that in the case of standards a violation is defined as $q_i - s_i > 0$, where s_i is the legal maximum level of production (the standard) set for type i firms. Similar to the case of tradable permits, we constructed two treatments for the case of emission standards. These are labeled S1 and S2 in Table 1. In treatment S1, the emission standards are 7, 6, 4 and 3 for firm types 1–4, respectively. The monitoring probabilities are 0.6, 0.65, 0.63 and 0.66 (violations are fined with the same penalty function; $\varphi = 100$ and $\gamma = 66, 67$). This policy induces compliance for expected profit-maximizing subjects, so the expected aggregate level of production is 40 units in a group of eight subjects. In treatment S2, the standards are decreased for every type of subject, so that the aggregate cap of emissions is 20, but monitoring probabilities are decreased so as to keep the predicted level of emissions at 40 units, the same level as in treatment S1. Therefore, Treatment S2 induces violations.

3.2 Experimental Procedures

The experiments were programmed and conducted with the software z-Tree (Fischbacher 2007) in a computer laboratory at the University of Montevideo. Participants were recruited from the undergrad student population of the University of Montevideo, the University of the Republic, the Catholic University and ORT University; all in the city of Montevideo, Uruguay. We organized the experiments in tradable permits sessions and standards sessions.

4 We call "emissions" the output chosen by the subjects although, as we have already mentioned, we framed the experiment as a neutral production decision.

A tradable permits experimental session consisted of 20 rounds, during which the subjects played 10 rounds of each of the two market treatments (M1 and M2). A standards session consisted of 20 rounds during which the subjects played 10 rounds of each of the two standards treatments (S1 and S2). The order of the M-treatments as well as of the S-treatments was varied, such that half the subjects played the compliance treatment first, and half played the violation treatment first.

Subjects participating in a permit or standards session were randomly assigned into groups of eight individuals.[5] Each eight-subject group comprised a permits market or a group of firms regulated by the same set of standards.

Before the beginning of the experiments, instructions were handed out to subjects. The instructions (see Online Appendix, Section 3) were read aloud and questions were answered. Prior to the first round of the first treatment, subjects played two trial rounds of the first treatment in the standards sessions, and three trial rounds of the first treatment in the permit sessions. In the standards sessions each period lasted 2 min. In the permit sessions, each period lasted 5 min, to give subjects time to make their bids, asks and to decide how many units to produce and how many permits to buy or sell.

After all subjects in the group had made their decision, the computer program automatically produced a random number between 0 and 1 for each subject. If this number was below the informed probability of being monitored, the subject was inspected, as explained in the instructions. Subjects were informed in their screen whether they had been selected for inspection or not, and the result of the inspection (violation level, total fine and net profits after inspection). After this, subjects were informed in their screen the history of their decisions in the game, the history of inspections and the history of profits, up to the last period just played. After 20 seconds in this screen, the next period began automatically.

At the end of each session we conducted a questionnaire to gather some basic socio-economic and attitudinal characteristics of the subjects. One of these characteristics was the subjects' level of risk aversion. To elicit this, as part of the questionnaire we conducted a Holt and Laury (Holt and Laury 2002) type of test. In this test, the subjects were confronted to 10 choices between a certain amount of money (labeled Option A, equal to US$ 40, and fixed across the 10

5 The number of subjects showing up for a session was not always multiple of eight. This was not a problem in the standards experiments because in these experiments the subjects do not relate with each other in any form. In the case of market treatments we selected the subjects by order of arrival, completing groups of eight subjects. Excess subjects were paid the show-up fee.

choices) and a lottery (labeled Option B). In the lottery, subjects could earn either US$ 15 or US$ 65. The probability of winning the higher prize varied from 0.1 to 1 between choices 1 and 10. Our measure of risk aversion is the number of the choice in which the subject switches to Option B. It then varies between 1 and 10, with 10 being the highest value of risk aversion. (In the tenth choice the higher prize of the lottery, higher than the certain amount in option A, has a probability equal to 1, so every subject should choose the lottery in the 10th choice). A risk-neutral subject should switch from option A (the certain amount) to option B (the lottery) in the fifth or sixth choice. We informed the subjects that after completing the questionnaire, one subject was going to be chosen from the pool of subjects in the room and that she was going to be paid according to her decisions in the Holt and Laury choices by drawing a number between 1 and 10 from an urn. If the subject selected the lottery in the drawn choice, the lottery was conducted with the corresponding probabilities in the form of colored balls in an urn.

Overall, sessions lasted between 2 and 2.5 hours. Subjects were paid the equivalent to US$ 7 for showing up on time and earned more money from their participation in the experiment.[6] The exchange rate between the experimental and Uruguayan pesos was set in order to produce an average expected payment for the participation in the experiment that was similar to what an advanced student could earn in the market for 2 h of work. Total payments ranged between US$ 30 and US$ 5 in the standards sessions, with a mean value of US$ 19, a median of US$ 18 and a standard deviation of US$ 4. For the case of the tradable permit sessions, payments ranged between US$ 23 and US$ 7, with a mean value of US$ 20, a median of US$ 20 and a standard deviation of US$ 2.

A total of 96 subjects participated in five permits sessions, comprising a total number of 12 eight-subject groups. Each of these groups played one M1 and one M2 treatment. We therefore ran a total of 12 M1 treatments and 12 M2 treatments. A total of 61 subjects participated in two standards sessions, enough number to complete 7 eight-subject groups. Each of these groups played one S1 and one S2 treatment. We therefore ran a total number of seven S1 treatments and seven S2 treatments. We allowed participation of a reduced number of subjects in multiple sessions because we had a thin pool. In the market experiments, 12 subjects participated two times and 2 subjects three times. In the case of standards, one subject participated two times. If we do not count repeating

6 In the first session we paid a show-up fee of US $5.

subjects, therefore, the number of different subjects that participated in the market experiments is 80; and 60 in the standards experiments.[7]

4 Results

In this section, we present the results of our work. We present the outcomes of the permits experiments first and then those of the standards experiments.

4.1 Overall Results of the Market Experiments

Table 2 presents the summary statistics of key variables in the market experiments with the values that theory predicts for the case of cost-minimizing, risk-neutral agents. It can be seen first that the average price of permits traded in the treatment M1 (E$ 80.3) is approximately within the predicted range of [E$ 74, E$ 80]. On the contrary, the average price of permits traded for the treatment M2 (the violation treatment) is E$ 104.3, above the predicted range.

A second result that can be observed in Table 2 is that the mean levels of violations are within 0 and 1 unit for the four types of firms in the compliance treatment. This result is in line with the literature. Other authors have obtained positive average violations in treatments that induce compliance in theory before. (See, for example, Murphy and Stranlund 2006, 2007; Stranlund, Murphy, and Spraggon 2011, 2013. Cason and Gangadharan (2006) do not present descriptive statistics for levels of violations.) However, although the mode of violations is 0 for the four types of firms as in the other studies, the overall compliance rate (66 %) is lower than in the case of Stranlund, Murphy, and Spraggon (2011) (we find no other paper that reported the compliance rate). The difference may be explained by differences in design. While Stranlund et al. (2011) decided to "over-enforce" in their perfect compliance treatment, we do not. On the contrary, we set the expected marginal penalty arbitrarily small enough to induce compliance because we are interested in comparing the cost-effective designs of the pollution control programs. Notwithstanding this

7 In the case of standards, each of these 60 subjects provides an independent set of observations, because subjects in the standards experiments do not interact with each other. In the case of the tradable permits experiments, because the individuals in the market interact with each other buying or selling permits, the unit of observation is the market. Nevertheless, because of our thin sample (12 markets), and because the feature of repeating subjects, we also test our hypothesis by treating the sample as an unbalanced panel of 80 subjects.

Table 2: Comparison of predicted results with summary statistics for permit experiments.

Market treatment 1 – compliance treatment		Mean price per period	Number of transactions per period	Type 1 ($l_0 = 4$)			Type 2 ($l_0 = 4$)			Type 3 ($l_0 = 6$)			Type 4 ($l_0 = 6$)		
				q	*l*	*v*	*q*	*l*	*v*	*q*	*l*	*v*	*q*	*l*	*v*
Theory		**74–80**	**10**	**7**	**7**	**0**	**6**	**6**	**0**	**4**	**4**	**0**	**3**	**3**	**0**
Experiments	Mean	80.3	8.5	6.5	5.8	0.8	6.5	5.9	0.7	4.8	4.4	0.4	4.3	3.9	0.3
	Std. dev.	14.1	2.6	1.3	1.7	1.4	1.4	1.5	1.5	1.2	1.0	0.6	1.4	1.2	0.6
	Conf. int./mode	[77.7, 82.7]	(8.0, 9.0)	6/7	7	0	6	6	0	5	4	0	3	3	0
	Median	79.6	8	6	6	0	6	6	0	5	4	0	4	4	0
	# Obs.	118	118	236	236	236	236	236	236	236	236	236	236	236	236
Market treatment 2 – violation treatment		**Mean price per period**	**Number of transactions per period**	**Type 1 ($l_0 = 2$)**			**Type 2 ($l_0 = 2$)**			**Type 3 ($l_0 = 3$)**			**Type 4 ($l_0 = 3$)**		
				q	*l*	*v*	*q*	*l*	*v*	*q*	*l*	*v*	*q*	*l*	*v*
Theory		**74–80**	**6**	**7**	**4**	**3**	**6**	**3**	**3**	**4**	**2**	**2**	**3**	**1**	**2**
Experiments	Mean	104.3	6.1	5.5	3.1	2.4	4.8	2.7	2.1	3.9	2.2	1.7	3.3	2.0	1.3
	Std. dev.	11.0	3.1	1.7	1.5	2.1	1.9	1.2	2.0	1.8	1.2	1.6	1.4	0.9	1.2
	Conf. int./mode	[101.9, 106.6]	[5.4, 6.8]	5	4	1	4	3	2	3	3	2	3	2	1
	Median	105.8	5	5	3	2	4	3	2	3	2	2	3	2	1
	# Obs.	86	86	172	172	172	172	172	172	172	172	172	172	172	172

difference, 90 % of violations are 0 or 1 unit violations in our case. Moreover, 37.5 % of the subjects complied in every round, 81.3 % of the subjects had an average violation level equal to 1 or less and 98 % had an average violation of 1.7 or less.[8]

In the treatment M1, we can note further that although, on average, emissions are lower than predicted for firms of type 1 and are higher than predicted for firms of types 2, 3 and 4, the modal behavior is consistent with the theory. As for permit holdings, these are higher than predicted for the firms of types 3 and 4 (the net sellers) and are lower than predicted for firms of types 1 and 2 (the net buyers), but again the modal behavior is consistent with theory.

We cannot say the same thing for the case of the treatment M2, though. In this case, both the average and modal violations are below their predicted values. As it can be seen in Tables A.4.1 and A.4.2, these observations are valid independently of the order in which the treatments were played, although there are some observable differences in the mean price and mean number of trades in both treatments, depending on whether they were played first or second. In general, the treatment played first is closer to the theoretical levels. This observation is suggestive of an order effect in the level of the average price and the number of trades. We explore the issue of an order effect in more detail below.

Lower than predicted violations, on average, are also reported by the previously cited papers. Cason and Raymond (2010) found that subjects underreport "well below" the level of what would be predicted for a risk-neutral or even risk-averse subject. It is difficult to draw a clean picture of the reasons behind this result by looking at Table 2. One noticeable result is that type 4 firms

8 Six subjects went bankrupt in the market experiments; five were net expected buyers of permits, five were emitting the maximum level of emissions (10 units) when they were caught and went bankrupt, five went bankrupt in the treatment that induced violations, and three went bankrupt in the first treatment played (these were allowed to play the second treatment starting with the initial endowment). By design, bankruptcy was not possible in the permits experiments if every subject behaved as an expected profit maximizer. Moreover, subjects received an initial endowment at the beginning of each treatment. This endowment was conceived as a source of funding for the subjects in the case they experience losses in one or more periods (as a result of a behavior not consistent with an expected profit maximizer or a risk-averse subject). The amount of the endowment was calculated considering a quite unlikely scenario for subjects with the lowest marginal benefits (type 4 subjects). It was therefore sufficient to cover a considerable amount of losses spanning through more than one period. In order not to lose the total number of observations of the markets in which a subject went bankrupt, we use the observations of these markets in the periods during which the eight subjects of the market were active. The results that we show below do not change qualitatively if we use only the groups in which no subject went bankrupt.

sold one permit less than what was expected, on average, and type 1 firm bought one permit less than expected, on average. Types 3 and 4 firms, on the other hand, demanded a number of permits closer to their expected levels. Nevertheless, probably because the higher than expected price, firms of types 1 and 2 emitted less than expected. We can say also that on average the expected sellers tended not to sell as many permits as predicted. At the same time, the average number of transactions per period was what theory predicted. Both results together may suggest that traded permits may have changed hands more than once in M2.

4.2 Hypothesis 1 Tests

It can be seen in Table 2 that, although both treatments were designed to induce the same level of individual emissions, the mean, the median and the modal level of emissions (q) were higher in the treatment M1 than in the treatment M2 for all four types of firms, except for modal emissions of type 4 firms, which were equal.

4.2.1 Nonparametric Tests

To evaluate Hypothesis 1, we first perform a set of nonparametric, Wilcoxon signed-rank tests. Because each experimental market provides only one independent observation (Davis and Holt 1993), we perform the comparison of emissions between treatments at the market level (aggregated across the eight subjects and averaged across periods). Moreover, we do this by matching the pair of aggregate market emissions in the violation treatment with that of the compliance treatment for the same market. This is to exploit the fact that the observations samples are "related" (i. e. the same group of subjects was exposed to both treatments). More formally, the null hypothesis of this test is $Ho: Q_j^{M1} = Q_j^{M2}$, where j refers to a specific eight-subject market and Q refers to the aggregate level of emissions. The result of this test leads us to reject the null hypothesis, in favor of the alternative that the aggregate market level of emissions is higher in the compliance treatment than in the violation treatment ($p = 0.0022$, $z = 3.059$, $n = 12$). This result is robust to the order of treatments.[9]

9 The aggregate level of emissions is higher in the M1 treatment than in the M2 treatment when M1 is played first ($p = 0.0277$, $z = 2.201$, $n = 6$). The same result is obtained when M2 is played first. Moreover, in spite of the violation of the independence of observations, we also conducted

Although the p-value of the test is remarkably low, we believe that the fact that it includes the observations corresponding to 14 subjects that participated two times in the experiments, and two subjects that participated three times, may weaken its result. A natural solution would be to perform the same test using only sessions (markets) in which no subject was repeating. But we cannot do this because we have a low number of observations of such groups. Instead, we address this issue by conducting an econometric analysis controlling for observations belonging to a subject that is repeating participation.[10]

4.2.2 Regressions

In Table 3, we present the results of a random-effects regression analysis with the individual decisions as the unit of observation, in which we add the corresponding assumption regarding the distribution of the error terms. To be more precise, the specification of the estimated equation is the following:

$$e_{it} = f(\text{VIOLATION TREATMENT}_{it}, \text{ FIRMTYPE}_{i}, \text{ PERIOD}_{t}, \text{ REPEAT}_{it}, \text{ OTHER CONTROLS}) \quad [1]$$

where e_{it} is the level of emissions of subject i in round t, VIOLATION TREATMENT is a dummy variable equal to 1 if the treatment is M2, and equal to 0 if the treatment is M1; FIRM TYPE is a set of three dummy variables to control for firms' type, PERIOD_t is another indicator variable for each of the 10 periods used in these regressions; and REPEAT_{it} is an indicator variable equal to 1 for the observations corresponding to a subject that is participating for her second or third time. The other controls employed in the regressions depend on the specification. In specification 1, we included a dummy variable to control for the order of the treatments (*First* equals 1 if the treatment to which the observation belongs was played first). We also included an interaction of this variable with "*Violation Treatment*", as a way to disentangle any possible difference in the order effect between the violation and compliance treatments. In specification 2, we included a set of dummy variables to control for possible market

the same test matching the pairs of the average level of emissions in each treatment for each individual. We also reject the null hypothesis in this case ($p = 0.0000$; $z = 5.624$; $n = 80$). We obtain the same result if we match the average emissions by type of firm ($p = 0.0000$, $z = 4.933$, $n = 46$).

10 It is worth mentioning that we also reject the null hypothesis if we conduct the test of type-of-firm emissions using only markets in which no subject was repeating ($n = 20$; $z = 2.931$; $p = 0.0034$).

Table 3: Linear random-effect models for Hypothesis 1 (market treatments).

Dependent variable:	**(1)**	**(2)**	**(3)**	**(4)**
Level of emissions	**Coefficient (std. error)**	**Coefficient (std. error)**	**Coefficient (std. error)**	**Coefficient (std. error)**
Violation treatment (M2)	−1.374***	−1.311***	−1.136***	−1.026***
	(0.249)	(0.293)	(0.161)	(0.193)
First	−0.263	−0.217		
	(0.231)	(0.243)		
Violation treatment × first	0.460	0.539		
	(0.406)	(0.459)		
Type 2	−0.0372	−0.00664	0.00673	0.0745
	(0.199)	(0.206)	(0.208)	(0.226)
Type 3	−1.356***	−1.962***	−1.346***	−1.785***
	(0.274)	(0.215)	(0.233)	(0.293)
Type 4	−2.571***	−2.596***	−2.638***	−2.398***
	(0.189)	(0.269)	(0.194)	(0.262)
Repeat	−0.767***	−0.934***	−0.834***	−0.625*
	(0.172)	(0.272)	(0.207)	(0.363)
Risk aversion		−0.00798		0.0263
		(0.0878)		(0.0998)
Period dummies	Yes	Yes	Yes	Yes
Market dummies	No	No	Yes	Yes
Constant	6.645***	6.627***	6.578***	6.348***
	(0.199)	(0.598)	(0.332)	(0.683)
N	1,631	1,107	1,631	1,107
*N*_clust	80	61	80	61

Note: ***p < 0.01, **p < 0.05, *p < 0.1.

(group of eight subjects) effects.[11] In both cases, we included random individual effects, and we clustered errors by subjects. The results that we present do not change in any significant way if we cluster errors by group (market) instead of subjects. Finally, we ran both specifications with and without controlling for risk aversion. The level of risk aversion for each subject is a discrete variable that

11 We do not include the price as an explanatory variable because this is a fundamental channel by which the enforcement regime affects emissions and violation choices in a cap-and-trade system. Moreover, according to theory (Malik 1990) and experimental evidence (Murphy and Stranlund 2006) an increase in the enforcement effort by the regulator does not directly affect the level of emissions of the firms, but only through the permit's price. See also Section 4.3.

takes the values 1–10, using the results of the Holt and Laury test as explained in Section 3.2. [12]

It can be seen in Table 3 that the level of individual emissions is between 1 and 1.3 units lower in the violation treatment than in the compliance treatment, depending on the specification used. This result leads us to reject Hypothesis 1, as we did with the nonparametric tests in the previous section. Interestingly, the rejection of Hypothesis 1 does not depend on whether or not we control for risk aversion. We also do not observe an order effect statistically different from zero in the level of emissions.[13] The coefficients of the variables indicating type are as expected. Not shown, the group of subjects and period dummies show no statistically significant effect, except for one group and one period. The sign of the indicator variable "Repeat" is negative. This result is consistent with learning. In fact, subjects who participated in more than one session earned more when they were participating for their second and third time than what they earned in their first time. Nevertheless, the result may have other possible interpretations. Beyond these, its inclusion is for purposes of consistency of our variable of interest, M2, and the results show that the rejection of Hypothesis 1 is robust to repetition.

Apart from these specifications, we did some robustness checks. In one of these, we re-estimated the models splitting the periods in two subperiods: rounds 1–5, and rounds 6–10. The results do not change in any significant way. The coefficient of interest does not change in magnitude or statistical

12 When we include the risk-aversion dummies we lose the observations of 18 subjects that made inconsistent choices in the Holt and Laury test. We also dropped from the sample the observations of an additional individual that revealed an extreme preference for risk in the test (opted for the lottery in the 10 choices, risk aversion = 1) but did not behave consistently with this choice, biasing the estimation of some of the other risk-aversion dummies. We keep only the first participation of the subjects who revealed different levels of risk aversion between participations. Among the remaining 61 subjects, the mean level of risk aversion is 7.2, the median 7.0, the minimum 4 and the maximum 10.

13 We did observe an order effect in the level of violations. In estimations not shown in this paper (available upon request) we found that the order of the treatment had a statistically significant effect on the observed *level* of the violations of the compliance treatment (M1). Subjects tended to violate more in M1 if the violation treatment (M2) was played first, as compared to what they violate in M1 if M1 was played first. On the other hand, the level of violations of the firms in the M2 treatment did not depend on whether this treatment was played before or after the treatment M1. The order effect in the compliance treatment may be seen as an "anchoring effect" (Tversky and Kahnema, 1974; Ariely, Loewenstein, and Prelec 2003) in enforcement regimes. In terms of policy implication, it may suggest that a regulator that previously allowed violations in a cap-and-trade program needs a relatively more stringent enforcement strategy and/or more time to induce perfect compliance, than a regulator that induced perfect compliance in the first place.

significance. In another specification, we use risk-aversion dummies instead of a discrete linear variable. Again, the results do not change in any significant way. These results are available upon request.[14]

4.3 The Effect of Inducing Noncompliance on the Price of Permits

We have provided experimental evidence against the hypothesis that a regulator has the possibility of inducing each of the regulated firms to emit at a given level by manipulating the supply of permits and the monitoring probability in the manner suggested by the theoretical models of enforcement in cap-and-trade schemes. Nevertheless, we have not explored the channel by which this effect takes place. One obvious channel is the price of the pollution permits. As it is well known, the price of the pollution permits determines the level of emissions of the firms that participate in a cap-and-trade scheme, together with its individual characteristics (particularly, their marginal pollution benefits) (Stranlund and Dhanda 1999). To explore this channel, we ran two additional regressions whose results are presented in Table 4. The panel variable in these regressions is the group of eight subjects that comprise the market, and the time variable is the round. In the first of these regressions, whose results we show in Table 4, the dependent variable is the average price at which the permits were traded. We can see that the average price of the permits traded was higher in the treatment that induces violations than in the treatment that induces compliance, although both treatments were designed to produce the same equilibrium price. Moreover, the size of this difference depends on the order in which the treatments were played. The average price was around E$ 13.9 higher in the M2 treatment than in the M1 treatment when the M2 was played after the M1 treatment, but it was E$ 24.3 (= 13.93 − 9.669 + 19.99) higher in the M2 treatment if this was played before M1. Moreover, the order effect has a negative sign for the compliance treatment (−9.669) and a positive sign for the violation treatment (−9.669 + 19.99). This result suggests that there may be a partial anchoring effect of the price in the first treatment over the price of the second treatment, irrespective of which treatment was played first.

This result is consistent with the finding that firms emit less than expected in this same treatment. Moreover, as it can be seen in Table 2, on average, type 4

14 Our result of interest is robust to using fixed effects instead of random effects, except in the case in which we control for the order effect and we discard all the subjects showing risk-aversion inconsistencies, although with fixed effects we cannot use risk-aversion controls.

Table 4: Linear random-effect models for prices (market treatments).

	Dependent variable:
	Average price of permits traded in the period
Control variables:	**Coefficient (std. error)**
Violation treatment (M2)	13.93***
	(3.881)
First	−9.669***
	(2.838)
Violation treatment × first	19.99***
	(4.534)
Number of subjects that repeat in the market	−1.134 (0.990)
Period dummies	Yes
Constant	83.18***
	(2.803)
N	204
*N*_clust	12

Note: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

firms keep one permit more than predicted and type 1 firms buy one permit less than predicted. We are unable to provide a complete explanation for the reason behind this result. A possible explanation is an endowment effect: type 4 firms received three permits and are expected to demand one permit, while type 1 firms received two permits at the beginning of the experiment and are expected to demand four permits. Because of this initial allocation, we believe that type 4 firms may be affected by an endowment effect. Another possible effect is risk aversion. The expected levels of violations are derived assuming risk neutrality. A risk-averse, prospective seller firm may not be willing to sell as much permits as predicted at the expected equilibrium price. Similarly, a risk-averse, prospective buyer may be willing to pay a premium over the expected price to avoid being out of compliance for as many units as a risk neutral would be.

4.4 Overall Results of the Standards Experiments

In this section, we present the results for the standards experiments. Recall that in the standards experiments the regulator sets the expected cost-minimizing allocation of emissions by fiat, in the form of individual standards, instead of trusting a market for pollution permits for this task.[15]

Table 5 shows the summary statistics of key variables in the emission standards experiments and the value of those variables that theory predicts for the case of expected profit-maximizing agents.[16] The results indicate that

Table 5: Comparison of predicted results with summary statistics for emission standards experiments.

Treatment S1 (compliance treatment)		**Type 1 ($s = 7$)**		**Type 2 ($s = 6$)**		**Type 3 ($s = 4$)**		**Type 4 ($s = 3$)**	
		q	***v***	***q***	***v***	***q***	***v***	***q***	***v***
Theory		**7**	**0**	**6**	**0**	**4**	**0**	**3**	**0**
Experiments	Mean	7.7	0.7	6.5	0.5	4.8	0.8	3.6	0.6
	Std. dev.	1.0	1.0	1.2	1.2	0.8	0.8	0.9	0.9
	Mode	7	0	6	0	4	0	3	0
	Median	7	0	6	0	5	1	3	0
	# Obs.	130	130	140	140	120	120	160	160
Treatment S2 (violation treatment)		**Type 1 ($s = 4$)**		**Type 2 ($s = 3$)**		**Type 3 ($s = 2$)**		**Type 4 ($s = 1$)**	
		q	***v***	***q***	***v***	***q***	***v***	***q***	***v***
Theory		**7**	**3**	**6**	**3**	**4**	**2**	**3**	**2**
Experiments	Mean	6.5	2.5	5.9	2.9	3.6	1.6	3.2	2.2
	Std. dev.	1.3	1.3	2.3	2.3	1.0	1.0	1.8	1.8
	Mode	7	3	5	2	4	2	2	1
	Median	7	3	6	3	4	2	3	2
	# Obs.	130	130	140	140	120	120	160	160

15 It is not an issue of analysis here, but for this to be possible, the regulator should perfectly observe the marginal benefits of the firms. The relative advantage of a market for pollution permits to perform the task of allocating emission responsibilities in a cost-minimizing way is based precisely on the fact that regulators do not have perfect information on the firms' marginal benefits of pollution.

16 We discard the observations of six subjects that went bankrupt, unexpectedly. (Expected profits were positive for all subjects if behaved as expected profit maximizers. Moreover, subjects were given an initial allocation of E$ 525. This endowment allowed firms of type 4, for example, to cover 35 periods of losses if they behaved as expected profit maximizers and were audited and fined in every period of the violation treatment.) Bankrupted subjects were three type 4 subjects, two type 3 subjects and one type 1 subject.

subjects violated the standards in the treatment S1 (the compliance treatment), on average, although they were not supposed to. Nevertheless, the modal behavior of subjects is consistent with this model. The median behavior is also consistent, except for type 3 firms. At the same time, in the violation treatment (S2), the average levels of violations turned out to be somewhat lower than those predicted for type 1, 2 and 3 firms, and somewhat higher than predicted for firms of type 4. If we look to the modes, type 1 and 3 subjects performed as the theory predicts, while types 2 and 4 emitted one unit less than predicted. Overall, it seems that the expected profit maximizer model does better in predicting the average behavior of firms in the violation treatment for the case of emission standards than for the case of tradable pollution permits, while it does basically the same job in the compliance treatment.

4.5 Hypothesis 2 Tests

We now address the test of Hypothesis 2. In the case of the standards experiments we have only one subject that participated twice in these experiments. This individual was assigned type 3 in one participation and type 4 in another. To assure independence of observations, we only included the type 3 observations corresponding to this individual in the tests.[17]

4.5.1 Nonparametric Tests

If we look at the average levels of emissions of subjects in the treatment S1 versus treatment S2 in Table 5, the comparison is not as clean as in the case of tradable permits. For the cases of subjects of types 1 and 3, the average level of emissions is lower in the violation treatment than in the compliance treatment. However, this difference is much closer to the theoretical prediction for subject types 2 and 4 (0.6 and 0.4 units, respectively).

To perform a formal test for Hypothesis 2, we perform a Wilcoxon signed-ranks test matching the pair of average (across periods) levels of emissions for each individual in each treatment. The null hypothesis is that the average level of emissions of the individual in the treatment S1 is equal to the average level of emissions of the same individual in treatment S2 (or, more formally, that the

17 We decided to discard type 4 and keep his type 3 observations due to the fact that we had more observations for type 4 firms. Anyway, the results do not change in any significant way if we exclude the type 3 observations, or if we exclude the individual as a whole.

observations are drawn from the same distribution). This test clearly rejects the null hypothesis ($z = 4.267$, $p = 0.0000$, $n = 54$). This result is robust to the order of treatments. The individual level of emissions is higher in the S1 treatment than in the S2 treatment when the S1 treatment is played first ($p = 0.03$, $z = 2.97$, $n = 20$). The same result is obtained when S2 is played first ($p = 0.0016$, $z = 3.155$, $n = 34$).

4.5.2 Regressions

In this section, we estimate a linear random-effects model to explore associations between exogenous covariates and the levels of emissions and as an additional test of Hypothesis 2.[18] The specification of our econometric model for the case of standards mimics the specification 1 of our econometric model for the case of tradable permits. We do not run a specification 2 in this case since there is no reason to control for group effects in the standards experiments. Table 6 presents the results with and without controlling for risk aversion.[19]

Our econometric analysis shows that, everything else equal, the level of emissions is around 0.7 units lower in the violation treatment than in the compliance treatment. This result is consistent with the result of the Wilcoxon test performed in the previous section, by which we rejected Hypothesis 2. The result is the same one we obtained in the case of tradable permits. Therefore, we can conclude that regulators may not be able to induce the same level of emissions pooling the leverages of the emission standards and the monitoring probabilities in the quantities suggested by the standard theoretical models of enforcement.

Apart from the main result previously commented it is interesting to note that, as we do in the case of tradable permits, we do not observe an order or anchoring effect in the case of standards. Finally, by design, the type of firm affects the level of emissions; firms with lower abatement costs tend to emit less as compared with the highest marginal abatement cost firm type.

18 Results do not change qualitatively if we use fixed effects instead of random effects.

19 Similarly to what we did in the case of tradable permits, in the case of standards we discard the observations of 11 subjects that made inconsistent choices in the Holt and Laury test of risk aversion, and one more subject that revealed an extreme preference for risk in the test (risk aversion = 1) but did not behave consistently with this choice. Among the remaining 42 subjects, the mean level of risk aversion is 7.1, the mean 7.0, the minimum 4 and the maximum 10.

Table 6: Linear random-effect model (standards treatments).

Dependent var.:	**Linear model**	**Linear model**
Level of emissions	**Coeff. (std. err.)**	**Coeff. (std. err.)**
Violation treatment	−0.723**	−0.709**
	(0.340)	(0.327)
First	0.322	0.141
	(0.200)	(0.217)
Violation treatment × first	−0.317	0.0675
	(0.497)	(0.467)
Type 2	−0.852**	−0.970**
	(0.372)	(0.377)
Type 3	−2.945***	−2.692***
	(0.262)	(0.428)
Type 4	−3.679***	−3.893***
	(0.260)	(0.273)
Risk aversion		−0.237*
		(0.127)
Period dummies	Yes	Yes
Constant	7.156***	8.895***
	(0.248)	(0.885)
N	1,080	840
*N*_clust	54	42

Note: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

5 Conclusions

We study the compliance behavior of firms under different enforcement regimes in a laboratory setting. We do this for two policy instruments: a system of transferable permits and a system of standards. We evaluate whether a regulator can induce a given level of emissions on individual firms by using different combinations of the aggregate supply of tradable permits (or the emission standards) and the monitoring probability, as suggested by the conventional theoretical models of compliance, such that in one combination the regulator induces perfect compliance while in the other one it allows violations.

Our results provide experimental evidence against this possibility. We find that when a regulator chooses to attain the desired cap of emissions allowing violations, emissions are lower than expected, and lower than in the compliance treatment. This happens both in the case of tradable permits and standards, although it could be said that this result is somewhat less robust for the case of standards.

Another noticeable result is that when the regulator induces violations in the case of tradable permits it increases the average price at which permits are traded, independently of the fact that the number of trades are less or equal to what the theory predicts.

Even though a complete explanation of the observed behavior is beyond the scope of this work, there are several possible behavioral reasons related to our main results. Considering its relevance to our study, we mention just two of them. First, for the case of a transferable emissions permit system, it may be possible that an endowment effect is present (Kahneman, Knetsch, and Thaler 1990, Shogren and Taylor 2008). Our initial allocation of permits was devoted to foster market activity, somewhat skewed to prospective sellers. However, these low-benefit firms tended to hold more permits than expected, and high-benefit firms tended to hold less permits than expected. Consistent with the observe increase in the price of permits we observe a lower than the expected level of emissions. Another possible explanation is that compliance behavior of subjects is driven by individual moral motivations beyond the monetary incentives (Tyler 1990). In this case, even though we have a neutral framing, compliance behavior could be determined by the individual moral status, which might be an additional deterrent to the expected penalty.

Our results have direct implications for the design of a cost-effective environmental policy. According to our results, the aggregate level of emissions in the experimental market is about 25 % lower in the violation treatment than in the compliance treatment, when they should be equal in theory. Moreover, the aggregate level of emissions in the violation treatment is 12 % lower than the predicted level. Both results indicate that a cap-and-trade program that, following theoretical recommendations, is designed to allow a certain level of violations (because of the relative costs of sanctioning versus monitoring perfect compliance), would have higher than expected abatement costs because firms would not violate as much as predicted. On the other hand, because firms violate less, the regulator would spend less on imposing fines. If, as it might be argued, the aggregate costs of abating those extra unpredicted units of emissions are higher than the saved costs of imposing fines, a cap-and-trade program that is designed to allow a certain level of violations would be more costly than expected. Consequently, even when the marginal costs of monitoring perfect compliance are higher than the marginal costs of sanctioning, it may not be cost-effective to design a cap-and-trade program that allows violations, contrary to what is established by the theoretical literature. More research is needed to better calibrate the compliance behavior of firms under a cap-and-trade program before a more definite policy recommendation could be given regarding the relative cost-effectiveness of inducing compliance versus noncompliance. This calibration would have to be centered at the relative monitoring effort needed to induce the desired level of emissions under a perfectly enforced cap-and-trade program and an imperfectly enforced one.

The previous observation is not as conclusive in the case of emission standards. In this case, the aggregate level of emissions is around 13 % higher than the predicted level in the compliance treatment but is practically the predicted in the case of the violation treatment. Although we reject Hypothesis 2, it appears that it might be easier for a regulator to decrease the level of the emission standards and the monitoring effort to keep the aggregate level of emissions constant at the desired cap while allowing violations to reduce enforcement costs. Nevertheless, this would depend on the relative costs and benefits of the reallocation of emissions and violations that occur among the different types of firms.

Finally, albeit less robust, another result that we observe in our experimental settings is a type of anchoring effect when switching enforcement regimes in tradable permits. Firms tend to violate more in the perfect compliance treatment if this is played after a violation treatment (as opposed to be played before). Puzzlingly, we do not observe the same effect for emission standards. The anchoring effect on violations, although deserves to be studied further, may suggest that a regulator that previously allowed violations in a cap-and-trade program may need a relatively more stringent enforcement strategy or more time to induce perfect compliance, than a regulator that induced perfect compliance in the first place.

Acknowledgments: The authors are grateful to two anonymous referees and to the Editor in Chief of the journal for their helpful comments and the authors gratefully acknowledge the financial support provided by the Agencia Nacional de Investigación e Innovación (ANII) – Fondo Clemente Estable – Uruguay, under project no. FCE_2009_1_2801. Marcelo Caffera also acknowledges support from the NEXUS Fulbright Program. Chávez acknowledges the additional partial funding for this research received from CONICYT-FONDAP under project no. 15110027. The authors thank Eduardo Cancela, Silvana Acosta, Juanita Bloomfield, Nicolás González and Facundo Danza, who provided invaluable help in programming and assistance.

Funding: Agencia Nacional de Investigación e Innovación – Uruguay (Grant/Award Number: FCE_2009_1_2801); NEXUS Fulbright Program (Grant/Award Number: Exchange Visitor Program No. G-1-00005); and CONICYT-FONDAP (Grant/Award Number: 15110027).

References

Arguedas, C. 2008. “To Comply or Not to Comply? Pollution Standard Setting under Costly Monitoring and Sanctioning.” *Environmental and Resource Economics* 41:155–68.

Ariely, D., G. Loewenstein, and D. Prelec. 2003. “Coherent Arbitrariness: Stable Demand Curves without Stable Preferences.” *Quarterly Journal of Economics* 118:73–105.

Caffera, M., and C. Chávez. 2011. “The Cost-Effective Choice of Policy Instruments to Cap Aggregate Emissions with Costly Enforcement.” *Environmental and Resource Economics* 50 (4): 531–57.

Cason, T. N., and L. Gangadharan. 2006. “Emissions Variability in Tradable Permits Market with Imperfect Enforcement and Banking.” *Journal of Economic Behavior and Organization* 61:199–216.

Cason, T. N., and L. Raymond. 2010. “Framing Effects in an Emissions Trading Experiment with Voluntary Compliance.” *Research in Experimental Economics* 14:77–114.

Davis, D. D., and C. A. Holt. 1993. *Experimental Economics*. Princeton University Press.

Fischbacher, U. 2007. “Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments.” *Experimental Economics* 10 (2):171–8.

Holt, C. A., and S. K. Laury. 2002. “Risk Aversion and Incentive Effects.” *The American Economic Review* 92 (5):1644–55.

Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1990. “Experimental Tests of the Endowment Effect and the Coase Theorem.” *Journal of Political Economy* 98 (6): 1325–48.

Malik, A. S. 1990. “Markets for Pollution Control When Firms Are Noncompliant.” *Journal of Environmental Economics and Management* 18:97–106.

Murphy, J. J., and J. K. Stranlund. 2006. “Direct and Markets Effects of Enforcing Emissions Trading Programs: An Experimental Analysis.” *Journal of Economic Behavior and Organization* 61:217–33.

Murphy, J. J., and J. K. Stranlund. 2007. “A Laboratory Investigation of Compliance Behavior under Tradable Emissions Rights: Implications for Targeted Enforcement.” *Journal of Environmental Economics and Management* 53:196–212.

Raymond, L., and T. N. Cason. 2011. “Can Affirmative Motivations Improve Compliance in Emissions Trading Programs?” *Policy Studies Journal* 39 (4):659–78.

Shogren, J., and L. Tylor. 2008. “On Behavioral-Environmental Economics.” *Review of Environmental Economics and Policy* 2 (1):26–44.

Stranlund, J. K. 2008. “Risk Aversion and Compliance in Markets for Pollution Control.” *Journal of Environmental Management* 88:203–10.

Stranlund, J. K. 2007. “The Regulatory Choice of Noncompliance in Emissions Trading Programs.” *Environmental and Resource Economics* 38:99–117.

Stranlund, J. K., and K. K. Dhanda. 1999. “Endogenous Monitoring and Enforcement of a Transferable Emissions Permit System.” *Journal of Environmental Economics and Management* 38:267–82.

Stranlund, J. K., J. J. Murphy, and J. M. Spraggon. 2011. “An Experimental Analysis of Compliance in Dynamic Emissions Markets.” *Journal of Environmental Economics and Management* 62 (3):414–29.

Stranlund, J. K., J. J. Murphy, and J. M. Spraggon. 2013. “Imperfect Enforcement of Emissions Trading and Industry Welfare: A Laboratory Investigation.” In *Handbook on Experimental*

Economics and the Environment, edited by J. List and M. Price, 265–88. Cheltenham: Edward Elgar Publishing.
Tversky, A., and D. Kahneman. 1974. "Judgement under Uncertainty: Heuristics and Biases." *Science* 185 (4157):1124–31.
Tyler, T. 1990. *Why People Obey the Law*. New Haven and London: Yale University Press.