Pricing externalities and moral behavior

Axel Ockenfels

Peter Werner

Ottmar Edenhofer

This version: May 2020

Abstract

To measure how moral behavior interacts with pricing regimes, we conduct highly controlled experiments in which trading creates pollution. We compare indirect pricing (a cap-and-trade mechanism in our experiment) and direct pricing (a tax) in an otherwise equivalent setting in which 'producers' are incentivized to emit CO_2. 'Judges' decide on central trading parameters that may restrict socially harmful activities. Profit maximization predicts the same producer behavior in either setting in the absence of regulation. Yet, we find a significant share of producers refraining from emitting CO_2 at all. Even though judges restrict behavior in similar ways across mechanisms, direct pricing is more effective to accommodate moral behavior than the quantity policy.

Axel Ockenfels, University of Cologne, Center for Social and Economic Behavior (C-SEB) and Department of Economics, Cologne, Germany (e-mail: ockenfels at uni-koeln.de).
Peter Werner, Maastricht University, School of Business and Economics, Department of Microeconomics and Public Economics, Maastricht, The Netherlands (e-mail: p.werner at maastrichtuniversity.nl).
Ottmar Edenhofer, Mercator Research Institute on Global Commons and Climate Change (MCC), Potsdam Institute for Climate Impact Research (PIK), and Technical University of Berlin (TU-Berlin), Germany (e-mail: edenhofer at pik-potsdam.de).

Recent research has emphasized the importance of moral concerns for decisions in markets with negative externalities[1-4]. Yet, little is known about how different market-oriented ways of curtailing pollution affect individual moral behavior. We address this question in the context of climate change mitigation, where several scholars raised fundamental moral objections against market mechanisms[5,6], and others debate about the best pricing approach to curtail greenhouse gas emissions. In particular, we study the interaction between moral behavior (in our case: individuals' willingness to restrict CO_2 emissions at a personal cost) and pricing mechanisms in highly controlled experiments capturing two standard solutions to address environmental externalities: quantity regulation in the form of cap-and-trade (where the maximum level of CO_2 emissions is fixed and the price for each unit is determined by an auction in which emitters bid for scarce emission rights) and taxing (where the price is directly set). In our experiment, "producers" have the opportunity to make profits which leads to 'real-world' CO_2 emissions, whereas "judges" set taxes or caps that may restrict producers' actions.

One specific characteristic of cap-and-trade mechanisms, and quantity regulation in general, is of particular importance for our study: Producers who decide to not buy emission allowances, in an attempt to protect the environment, may only attract other buyers that take their place. As a result, while such moral behavior may affect the market price of emission allowances, individual abatement may be ineffective in affecting total abatement. This neutralizing effect of quantity regulation is relevant not only for climate change policy, but also for a wide range of other environmental goods such as fisheries, water resources or forests and other "common pool resources"[7]. If a government fixes a maximum level of resource consumption, any voluntary reduction in the consumption by one individual makes room for the consumption by others, making it difficult or even impossible to individually and voluntarily contribute to sustain the resource. Another example for this pattern relates to the "fundamental law of road congestion"[8] suggesting that any newly created road capacity, built to relieve congestion, will soon be congested again, due to induced demand and rebound effects. This implies in turn that environmental-friendly behavior (e.g. through the use of public transport) by some road-users tends to be neutralized by other road users.

We find that in the presence of CO_2 externalities, judges are significantly more likely to tax and to cap, and producers are significantly less likely to emit CO_2, compared to otherwise identical markets without externalities – even when this comes at a cost of foregoing monetary revenues. Importantly, behavioral patterns are very similar in both mechanisms. At the same time, however, voluntary cooperation plays out very differently across mechanisms; while with indirect pricing through cap-and-trade, voluntary reduction of CO_2 emissions drives down prices and thus only makes room for polluters, the direct pricing mechanism ensures more, and more efficient, CO_2 abatement from voluntary cooperation. Hence, in our setting, individually similar 'moral' behavior aggregates into different market outcomes, depending on market institutions. In this sense, the amount of 'morality' in a market is affected by the institutional design.

Pricing mechanisms and CO_2 externalities

Most previous experimental studies on emissions markets centered on details of specific market designs[9], yet these studies do not consider the externality caused by trading in emission markets[10]. Some recent experiments have provided evidence of a willingness to pay a premium for a positive environmental impact[11] and for a motivation to remove negative externalities[12]. Finally, some studies have elicited individuals' willingness to pay to avoid real-world CO_2 emissions[13-17]. None of these studies, however, investigated trading behavior and third-party regulation in the presence of environmental externalities and their interactions with different market mechanisms.

In economic terms, climate change is a problem of externalities: Production and consumption decisions that cause CO_2 emissions, such as driving a car, cause harm to others who did not choose to incur that cost and who suffer from climate change (including, for example, future generations). One way to address such externalities goes back to the British economist Arthur Pigou. In *The Economics of Welfare (1920)* he proposed to increase the price of activities that create externalities, such as emitting greenhouse gases, thereby internalizing the external costs and providing a signal to promote socially desirable decision-making[18]. In a response to Pigou, Coase[19] demonstrated in *The Problem of Social Cost (1960)* that bilateral negotiations between affected parties will lead to the same social desirable outcome – irrespective of the distribution of property rights – when transaction costs, wealth or income effects are absent[20,21]. The choice of the exact mechanism to address externalities is a central question from a policy perspective and a much-discussed question in economics. Under many circumstances, cap-and-trade and direct pricing policies are equivalent regarding effectiveness, efficiency and distributional outcomes[22], because both policies generate a carbon price that internalizes the harm and the generated tax or auction revenues can be distributed accordingly[23]. However, in a situation of transaction costs, wealth or income effects, incomplete information, lack of policy credibility, myopia, or excessive discounting, or uncertainty - depending on the structure of the benefits and costs of abatement – either mechanism can be preferable[24,25]. In the context of international climate agreements, scholars have argued in favor of a price mechanism, for example because a single price can be more easily negotiated than individual quantity regulations of a large number of countries[26-29]. Moreover, quantity regulations may lead to higher price volatility for emissions and can be more prone to corruption[30]. On the contrary, it has been argued that compliance with a price regulation can be better monitored than compliance with longer-term quantity regulations[31]. Stavins and Goulder and Schein summarize the literature and the dimensions under which one option should be preferred over the other[22,32]. Our study contributes to the literature by taking into account behavioral aspects of policy choice; we show that direct pricing can be preferable in a situation in which, otherwise, both policy options are fully equivalent. The reason is that direct pricing better accommodates people's willingness to contribute to the social good despite financial incentives to do otherwise.

An experiment with CO_2 externalities

In our experiment, participants act either as "producers" or "judges." Each of the 10 producers in one market is randomly assigned an (integer) valuation from the interval [3 Euro, 12 Euro] such that each value is assigned only once. The distribution of possible valuations and one's own realized position in the distribution is known to all market participants. To monetize one's own valuation, a producer has to buy a certificate in the market; it is not possible to buy more than one certificate.

Our main treatment variation consists of the way certificates are sold (see the "Methods" Section and the Supplementary Discussion for details about the mechanisms and the decision situation). In our indirect pricing treatments, certificates are traded via an auction mechanism (cap-and-trade), specifically a multi-unit uniform price auction[33]. After the maximum number of certificates for the market is set, the uniform market price for the certificate is determined by the highest losing bid by the producers plus an increment of 0.01 Euro. In our direct pricing treatments, the price per certificate is directly fixed (tax). As described in the Supplementary Discussion, the mechanisms represent standard simple auction and price mechanisms and are fully equivalent from the perspective of profit-maximizing producers in the sense that they induce truthful bids. In particular, profit-maximizing producers should place bids equal to their valuations in all treatments, irrespective of CO_2 externalities. Moreover, the mechanisms can be presented in 'parallel' and easily understandable ways to our subjects.

Judges are not directly involved in the market, but they must determine central market parameters prior to the start of the trading. In the auction treatments, a judge decides on the maximum number of certificates that may be traded (Q, from 0 to 10 certificates). In the price treatments judges directly determine the price for a certificate in the market (P). Judges do not receive payoffs themselves. This reflects that decisions about environmental regulation are often made by committees, politicians, regulators, or judges whose monetary income and profits are often not affected by their decisions. After judges have fixed market parameters, producers make their trading decisions. Importantly, producers' choices are elicited before they are informed about the quantity or price that is relevant for their markets. With indirect pricing, this is implemented by a second-price auction mechanism, and with direct pricing, we implement the Becker-DeGroot-Marschak mechanism[34]: producers who choose to participate in the market must state their maximum acceptable price (MAP) for the certificate which is then compared to the actual price.

We conducted each price mechanism both in presence and absence of CO_2 externalities. Treatments "*Neutral Auction*" and "*Neutral Frame*" serve as control conditions for the auction and the price mechanism, respectively. In treatments with a CO_2 externality ("CO_2 *Frame Auction*," "*Price Frame*," and "*Fine Frame*") every certificate bought by the producer leads to the emission of one ton of CO_2. For every certificate not traded in the experiment, a certificate for one ton of CO_2 is bought and deleted from the EU ETS system. In this sense, abstention from trade becomes a moral decision: it can be interpreted as an attempt to contribute to a social good (any deleted certificate leads to the real abatement of one ton of CO_2) at a personal financial cost. On the contrary, in the control treatments "*Neutral Auction*," and "*Neutral Frame*", trading only affects the producers' payoffs and does *not* involve a CO_2 externality.

A final treatment variation relates to framing in the price mechanism. Under the *Price Frame*, instructions emphasized the producer as paying a price and receiving a CO_2 emission certificate in exchange. In the *Fine Frame*, the price was described as a "fine" for the environmental pollution created by the emission of one ton of CO_2. As Sandel[6] points out, the perception of a fine might differ from a price because the fine signals moral disapproval with the underlying transaction. This suggests that the *Fine Frame* might alter subjects' moral evaluation of the market transaction, and subsequently their decisions.

The experiment was conducted as a classroom experiment at different faculties of the University of Cologne at the end of 2015. Details about the experimental protocol (including the matching of judges and producers and the procedures related to CO_2 emissions) can be found in the "Methods" section.

Regulation and trade patterns

Figure 1 displays average decisions by judges separately for each experimental treatment.

[Figure 1 here]

Judges overwhelmingly believe that pricing is an appropriate measure to induce abatement. Average regulated quantities are substantially lower than 10 units in the auction (the quantity that maximizes producer welfare). The average quantity permitted by judges in the *CO_2 Frame Auction* is 3.56 units, which is only half as high as the average quantity in the *Neutral Auction* without the externality (6.85 units), highlighting the important impact of CO_2 emissions for judges' decisions. This difference is significant as indicated by a Mann-Whitney-U (MWU) test ($p < 0.01$, $z = -8.14$, $n = 124$ versus $n = 105$). A similar picture emerges for regulated prices which are substantially higher in the case of a CO_2 externality (9.64 in *Price Frame* and 9.05 in *Fine Frame* compared to 6.13 in *Neutral Frame* where trading is not associated with CO_2 emissions). Here, too, differences between the CO_2 treatments and the control treatment are highly significant (in both cases $p < 0.01$, two-sided MWU-tests, $z = 7.76$, $n = 122$ versus $n = 130$ for the comparison *Price Frame* and *Neutral Frame* and $z = 6.83$, $n = 133$ versus $n = 130$ for the comparison *Fine Frame* and *Neutral Frame*), which excludes potentially confounding motives for interventions such as indifference, envy, or other social preferences[35,36]. Additional analyses show that a non-negligible share of judges fully restrict trade in the case of CO_2 externalities (see the Supplementary Discussion). Moreover, the *Price Frame* and the *Fine Frame* treatments do not differ from each other ($p = 0.14$, two-sided MWU-test, $z = 1.46$, $n = 122$ versus $n = 133$), suggesting that judges' interventions are rather insensitive to framing.

[Figure 2 here]

Figure 2 shows that nearly half of the producers in the *CO_2 Frame Auction* and the *Price Frame* treatment and nearly 40% of the producers in the *Fine Frame* treatment refuse to participate in trading at all, therefore foregoing all potential monetary payoffs. These shares are significantly lower than in the control conditions without CO_2 externalities (two-sample tests of proportions yield $p < 0.01$, $z = -4.53$, $n = 79$ versus $n = 84$ for the comparison *CO_2 Frame Auction* versus *Neutral Auction*, $z = -5.00$, $n = 79$ versus $n = 78$ for the comparison *Price Frame* versus *Neutral Frame* and $z = -3.88$, $n = 75$ versus $n = 78$ for the comparison *Fine Frame* versus *Neutral Frame*). As before with judges, there are no significant framing effects on participation rates in the price mechanism ($p = 0.238$, two-sample tests of proportions between *Price Frame* and *Fine Frame*, $z = -1.18$, $n = 79$ and $n = 75$).

To account for differences in valuations of those who entered the market, we calculate a producer's relative bid as the percentage share of his valuation and then compare mean relative bids (the yellow squares in Figure 2) across treatments. As some producers (4 under the auction mechanism and 16 under the price mechanism) bid more than their valuations, there is substantial variation in relative bids. Yet, once producers have entered the market, they do not seem to be significantly affected by the externality. On average, producers underbid substantially - the mean of producers' relative bids accounts for around 75% of the valuations, and differences between treatments with and without externalities are insignificant (two-sided MWU tests yield p = 0.90, z = -0.13, n = 42 versus n = 72 for the comparison *CO_2 Frame Auction* versus *Neutral Auction*, p = 0.93, z = -0.08, n = 41 versus n = 69 for the comparison *Price Frame* versus *Neutral Frame* and p = 0.17, z = 1.36, n = 46 versus n = 69 for the comparison *Fine Frame* versus *Neutral Frame*). Additional analyses (see the Supplementary Discussion) suggest that, while the valuation of a producer is positively related to the probability of participating, it has a similar impact in the treatments with and without externalities. Hence, the decision to not enter the market in the CO_2 treatments seems to be driven by a general reluctance to trade in the presence of CO_2 emissions. This result is in line with related findings[1,4] that some subjects refrain from trading when this is associated with negative real-world externalities.

Market outcomes

We apply a two-step procedure to distinguish the impact of judges' and producers' decisions for market outcomes from each other. We first calculate the theoretically induced abatement assuming that producers maximize profits and bid *according to their valuations*. Second, we consider producers' actual – 'moral' – behavior.

Figure 3 shows that induced abatement levels under profit maximization (blue bars, representing average CO_2 tons *not* emitted in the experimental markets) do not differ much between treatments, and statistical tests do not show significant differences (two-sided MWU tests yield p = 0.32, z = -1.00, n = 124 versus n = 122 for *CO_2 Frame Auction* versus *Price Frame*, p = 0.54, z = 0.61, n = 124 versus n = 133 for the comparison *CO_2 Frame*

Auction versus *Fine Frame* and $p = 0.14$, $z = 1.47$, $n = 122$ versus $n = 133$ for the comparison *Price Frame* versus *Fine Frame*). Hence, assuming profit maximization, judges' policies produce very similar abatement levels across different mechanisms, reducing the trading volume substantially and consistently across treatments by more than 60% compared to the maximum possible trading volume.

[Figure 3 here]

However, judges' and producers' behaviors may interact differently to produce market outcomes across treatments. In the second step, we calculate expected market outcomes of the judges' regulations, taking into account that many producers attempt to reduce emissions. To add actual producer choices to the analyses, we simulate 1,000 markets for each treatment with CO_2 externalities based on random draws with replacement. We pick the actual decisions of one judge and ten producers (one for each valuation level) and calculate the market outcome for this simulated market. This procedure is feasible because producers decide before knowing about the induced regulation and without interacting with other market participants. Hence, each decision of judges and producers can be treated as one statistically independent observation.

The green bars in Figure 3 represent average realized abatement levels in number of CO_2 tons not emitted across treatments resulting from these simulations. These abatement levels represent the market outcomes combining actual behavior of judges and producers. The figure shows that, despite the similarity in induced abatement by judges, actual abatement differs substantially between mechanisms. In the *CO_2 Frame Auction*, the mean number of abated tons accounts for 7.04 and is thus close to the induced level by judges under the assumption of profit-maximizing producers. Under the price mechanism, however, actual abatement is substantially higher than under the auction mechanism (by about 29% in *Price Frame* and by about 19% in *Fine Frame*). The differences in actual abatement levels between the auction mechanism and the price mechanisms are significant (two-sided MWU tests yield $p < 0.01$, $z = -23.06$, $n = 1{,}000$ versus $n = 1{,}000$ for *CO_2 Frame Auction* versus *Price Frame*, and $p < 0.01$, $z = -15.27$, $n = 1{,}000$ versus $n = 1{,}000$ for *CO_2 Frame Auction* versus *Fine Frame*). The reason for the difference is that in the price mechanism, voluntary abatement by producers (i.e. the decision to refrain from trading) directly results in lower total emissions. In the auction mechanism, however, this is not necessarily the case as voluntary abatement increases the supply of certificates and thus makes room for other producers to buy certificates and emit CO_2. Thus, under the price mechanism, as a combination of judges' and producers' choices, the average number of abated CO_2 tons is close to the maximum number of certificates (= 10 tons).

[Table 1 here]

Table 1 lists further simulation results. In markets where regulation was not prohibitive ($Q > 0$ or $P \leq 12$), average prices in the *CO_2 Frame Auction* are much lower (2.59 Euro) than under the *Price Frame* (8.46 Euro) and the *Fine Frame* (7.85 Euro), as is corroborated by two-sided MWU tests: $p < 0.01$, $z = -28.24$, $n = 826$ versus $n = 718$ for *CO_2 Frame Auction* compared to *Price Frame*, and $p < 0.01$, $z = -26.76$, $n = 826$ versus $n = 807$ for *CO_2 Frame Auction* compared to *Fine Frame*). When producers voluntarily stay out of the market, they decrease the scarcity of certificates and drive down prices under the auction mechanism. In 38.7% of these simulated *CO_2 Frame Auction* markets, the market price even accounts for zero. Yet, since the price is fixed in the price treatments through judges' decisions, direct pricing is immune against such neutralizing effects.

Voluntary abatement also affects the efficiency of abatements: for a given market regulation, producers with a high (low) valuation are those who should trade (abate) to maximize social welfare. We calculate the efficiency of the abatement for markets in which regulation was not prohibitive but at the same time induced at least an abatement of one CO_2 ton. Our efficiency measure is the share of inefficient producers who, given the regulation and profit-maximizing bidding by producers, should have abated and in fact did so. For example, if the regulation (quantity or price) would induce the emission of five certificates under profit-maximizing bidding, only the producers with the five highest valuations should produce. If one of the five less efficient producers obtains a certificate, our efficiency measure for this market would account for 80%. For the *CO_2 Frame Auction*, the average efficiency measure accounts for 71.7%. Thus, in nearly 29% of the cases, an inefficient producer receives a certificate. Average shares in the price treatments are with

98.0% (*Price Frame*) and 91.2% (*Fine Frame*) substantially higher (and also significantly so, as two-sided MWU tests indicate: $p < 0.01$, $z = -30.47$, $n = 750$ versus $n = 666$ for *CO_2 Frame Auction* compared to *Price Frame*, and $p < 0.01$, $z = -21.78$, $n = 750$ versus $n = 696$ for *CO_2 Frame Auction* compared to *Fine Frame*). Here, an inefficient producer receives a certificate only if he places a bid higher than his valuation. In this sense, direct pricing is more efficient concerning the achievement of the *ex-ante* abatement goal.

Finally, our measure for producer welfare is the percentage share of obtained profits relative to the maximum achievable profits without regulation (i.e., Q =10 or P = 0). Due to restrictive regulations and the high likelihood to stay out of the market, producer welfare is generally low. The endogeneity of prices in the auction market increases profits of the remaining producers so that producer welfare is substantially higher here (25.1%) than under direct pricing where realized profits account for only a small fraction of totally achievable payoffs (5.1% in *Price Frame* and 6.8% in *Fine Frame,* two-sided MWU yield $p < 0.01$, $z = 22.57$, $n = 1{,}000$ versus $n = 1{,}000$ for *CO_2 Frame Auction* compared to *Price Frame*, and $p < 0.01$, $z = 21.73$, $n = 1{,}000$ versus $n = 1{,}000$ for *CO_2 Frame Auction* compared to *Fine Frame*).

Discussion

Our experiment provides evidence for substantial moral behavior in the sense that, irrespective of the pricing mechanism, we observe a remarkable willingness to voluntarily exert costly efforts to reduce CO_2 emissions. Yet, the two market mechanisms lead to markedly different outcomes. In the quantity regime, moral behavior may be neutralized because refraining from trading increases the supply of emission allowances, and this makes room for more CO_2 emissions by other producers, counteracting the voluntary abatement. This suggests that quantity regulation might not be the most effective way to induce, and make use of, moral behavior in environments characterized by incentives for free-riding, such as common pool resources. On the other hand, direct pricing is associated with more, and more effective, moral behavior, as offsetting individual efforts to reduce CO_2 emissions is not possible here. This advantage of direct pricing applies for a wide range of challenges in governing environmental externalities, and it adds a behavioral dimension to the quantity-versus-price debate discussed above. More generally, our results indicate that the choice of market policy may have a strong impact on the effectiveness of moral behavior for aggregate outcomes.

Participants from non-student samples might differ in attitudes towards climate change and market mechanisms for governing CO_2 abatements. Therefore, it would be interesting to investigate whether responses to market mechanisms in the presence of externalities would be similar for more representative population samples. Our data allow us to obtain some indications on how variations in participants' attitudes towards environmental issues may correlate with decisions. In the Supplementary Discussion, we report results of regressions testing for links between a decision-maker's environmental awareness and moral behavior. Indeed, we find that a proxy measure for environmental awareness is significantly correlated with both judges' and producers' choices: the higher this measure, the higher is the likelihood of judges to prevent trade and the likelihood of producers to refrain from market participation. Moreover, producers who perceive the danger from climate change to be very big are significantly less likely to enter the market. Taken together, these results indicate that variations in environmental concerns may have explanatory power for behavioral variations in our experiment.

Methods

In line with the behavioral economics literature, the term "moral" in our setting is used to describe behavior that attempts to protect a social value (in our setting: protecting the environment through CO_2 abatement) at a personal financial cost. Defined in this way, moral behavior is not necessarily socially efficient, that is, equating marginal social benefits and marginal social costs to obtain the greatest possible total social benefit. (Indeed, climate policies are sometimes criticized for producing highly inefficient patterns of CO_2 pricing[37]). While we acknowledge that one might reasonably argue that only efficient CO_2 abatement can be described as 'moral', we refrain from this high standard in our study because this would require knowledge about the 'true' social cost of CO_2 emissions, or about our subjects' beliefs about this cost, which we do not have. Moreover, we believe that our simpler notion captures a common perception of what constitutes moral behavior in our setting.

Given the characteristics of the decision situation in our experiment, the null hypothesis for producer behavior is simple: irrespective of the market mechanism and the existence of the CO_2 externality, strictly profit-maximizing producers should place bids equal to their valuations (see the Supplementary Information for details). If producers care about the externality created by trading, this should lead to less aggressive trading behavior relative to the control treatments without the negative CO_2 externality.

For judges, several hypotheses are conceivable. If judges neither care about producers' payoffs nor about the (small) climate effect of trading, and since they themselves do not receive any decision-dependent payment in our experiment, they should be indifferent between all potential quantities and prices and thus could simply randomize their choices. If, however, judges care about their peers' experiment payoffs earned in the role of producers, they should leave prices at zero. Yet, if judges care about externalities and social efficiency, they might want to set the price P and the quantity Q in a way that the (resulting) price P^* from both mechanisms equals their beliefs about the marginal social costs of the emission of one ton of CO_2.

Our study was conducted in compliance with the common ethical standards in experimental economics. We collected the data as a classroom experiment in November and December 2015 in several lectures at the University of Cologne. Participation in the experiment was voluntary. Subjects were informed that all decisions and payments would be anonymous and confidential. Moreover, it was explained to the subjects that all decisions and associated payoffs would be implemented in exactly the same way as described in the experimental instructions. In the first classroom sessions, we elicited the decisions of the judges for all treatments. In the second step, one week after the judges' sessions, we collected the data for producer decisions. We conducted our experiment in three different faculties of the University of Cologne (Human Sciences, Management, Economics, and Social Sciences, Mathematics and Natural Sciences). This was done to make sure that our results would not be influenced by variations in general attitudes towards environmental protection and market solutions in the context of CO_2 emissions that may exist among student groups from specific faculties. Moreover, with the inclusion of lectures from different faculties, one advantage of our classroom study is that we can address a much broader student population than what would have been possible when we would have asked students to register for our experiments and then come to our laboratory. In fact, even though we strive to attract students from all faculties to our laboratory, students from business and economics are well-overrepresented among our registered subjects. That is, the social distance of subjects in different sessions would almost surely be much smaller with a laboratory study.

Moreover, with classroom experiments we can avoid a selection problem that arises in laboratory experiments: Laboratory participants have actively decided to be part of the laboratory subject pool. Therefore, in principle, members of laboratory subject pools may have different (unobservable) characteristics than other students; e.g., they might be more driven to make money. This potential selection problem is not present in our classroom setting that includes all students of a particular lecture.

In every lecture in which we conducted our experiment, we implemented all five treatments. In addition, we elicited self-reported data on demographic characteristics and general attitudes towards environmental protection and markets in a post-experimental questionnaire.

The experiment sessions were conducted with the help of several research assistants. Prior to each session, experiment instructions for each treatment were put into closed envelopes. Because we did not know the exact number of participants in a classroom in advance, to be on the safe side, we prepared more envelops for each session than we actually needed *ex-post*. These envelopes were then shuffled and randomly handed out by the assistance to participants in each lecture until all participating students received one envelope. This procedure ensured that there was no systematic *ex-ante* bias in the sense that certain treatments were run more often with subjects from one particular lecture or faculty. At the same time, the actual numbers of observations for each treatment and role (judge or producer) may differ somewhat as the result of the random assignment of instructions to participants and because of the noise in the actual number of participants in each lecture.

As described above, all decisions were anonymous. While we cannot completely rule out that students communicated about the experiment between the weeks in which we collected decisions of judges and producers, we took measures to keep the probability of this happening low. In particular, we selected lectures for the experiment that minimized overlap of student groups: The data for judges' and producers' sessions were collected from the whole 50,000 student pool at the University of Cologne, in lectures that were unrelated to each other (e.g., lectures from different study programs and different faculties). Indeed, a typical laboratory experiment would recruit subjects from a much narrower pool of subjects, typically all studying at the same faculty, likely yielding a higher risk of between-subject communication between the sessions.

To additionally control for the possibility of previous knowledge of the decision situation, we elicited data about participation in the lectures related to our experiment among all producers. As a result of this, we excluded the data from altogether 33 producers who at the time of the particular session stated that they had participated in a lecture in which we already conducted our experiment. In addition, 3 producers did not enter a decision. All in all, we were able to collect usable responses for altogether 614 judges and 395 producers, totaling 1,009 subjects.

The matching of judges and producers to markets for the determination of experimental payoffs was implemented in the following way: after producers' decisions were collected, one producer with a valuation equal to each of the 10 values in the interval of [3 Euro; 12 Euro] was randomly picked and assigned to one market. Due to the ex-ante uncertain number of experimental producers, we were not always able to form complete markets with exactly 10 producers. In cases of the auction treatment where producers with one or more valuations were missing in the market, we randomly selected producers from other markets with the missing valuations and included their bids as dummy players. In the auction market, it was necessary to have 10 producers, one for each possible valuation, because the instructions told producers that final prices and allocations will be the result of the competition of such 10 traders. The random matching of these 10 producers that form a market was done only *ex-post*, in line with our instructions to the subjects. In those cases in which we did not have enough producers to form the required market (see below on how the sample sizes came about), we filled up the corresponding market with randomly chosen other producers. This procedure allowed us to *ex-post* compute prices and allocations for all participating producers, including in those markets that initially lack some producers. (Each producer was paid the profit from only one market – the market he or she was first assigned to –, of course.) Moreover, this procedure ensured that all incentives for all subjects were exactly as stated in the instructions, and fully in line with our theory.

In the price treatments, rematching of producers is not necessary because each stated price by a producer only has an impact on his/her own payoff.

One judge was randomly chosen whose decision (maximum quantity or price) was binding for the particular market. Based on the imposed regulation by the judge and the decisions of the producers, market prices, allocations of CO_2 certificates and profits for the producers (calculated as the difference between private valuations and certificate prices) were realized. Participants in the role of producers received feedback about the realized market outcome in the lectures one week after the experiment together with their payments. These payments were distributed in closed envelopes to maintain anonymity and privacy of the payoffs realized in the experiment. Average payoffs for producers accounted for 2.03 Euro (standard deviation: 3.44 Euro). If producers had realized losses, these losses were set to zero ex-post.

The implementation of the CO_2 externality was done in collaboration with the organization "TheCompensators*." For every ton of CO_2 that was not realized in our experimental markets (either because the number of traded certificates was restricted by the judge or because producers placed too low or no bids at all), we bought one EU ETS emission certificate through the organization which was then deleted from the system. Altogether 194 CO_2 certificates were bought and deleted in the course of the experiment by the "TheCompensators*."

Data availability

All data reported in this paper is available on the *nature.com* website as supplementary information files.

Code availability

The data was analyzed with the software Stata. A Stata do-file that allows for the replication of the analyses reported in our paper is available for download together with our data set.

References

1 Falk, A. & Szech, N. Morals and Markets. *Science* **340**, 707-711 (2013).

2 Bartling, B., Weber, R. A. & Yao, L. Do Markets Erode Social Responsibility? . *The Quarterly Journal of Economics* **130**, 219-266 (2015).

3 Kirchler, M., Huber, J., Stefan, M. & Sutter, M. Market Design and Moral Behavior. *Management Science* **62**, 2615-2625 (2015).

4 Sutter, M., Huber, J., Kirchler, M., Stefan, M. & Walzl, M. Where to look for the morals in markets? *Experimental Economics* **23**, 30-52 (2020).

5 Page, E. A. Cashing in on climate change: political theory and global emissions trading. *Critical Review of International Social and Political Philosophy* **14**, 259-279 (2011).

6 Sandel, M. J. *What money can't buy - The moral limits of markets*. (Farrar, Strauss and Giroux, 2013).

7 Ostrom, E. Coping with tragedies of the commons. *Annual Review of Political Science* **2**, 493-535 (1999).

8 Duranton, G. & Turner, M. A. The Fundamental Law of Road Congestion: Evidence from US Cities. *American Economic Review* **101**, 2616-2652 (2011).

9 Friesen, L. & Gangadharan, L. Environmental Markets: What Do We Learn From The Lab? *Journal of Economic Surveys* **27**, 515-535 (2013).

10 Plott, C. R. Externalities and Corrective Policies in Experimental Markets. *The Economic Journal* **93**, 106-127 (1983).

11 Sapci, O., Wood, A. D., Shogren, J. F. & Green, J. F. Can verifiable information cut through the noise about climate protection? An experimental auction test. *Climatic Change* **134**, 87-99 (2016).

12 Jakob, M., Kübler, D., Steckel, J. C. & van Veldhuizen, R. Clean up your own mess: An experimental study of moral responsibility and efficiency. *Journal of Public Economics* **155**, 138-146 (2017).

13 Löschel, A., Sturm, B. & Vogt, C. The demand for climate protection—Empirical evidence from Germany. *Economics Letters* **118**, 415-418 (2013).

14 Diederich, J. & Goeschl, T. Willingness to Pay for Voluntary Climate Action and Its Determinants: Field-Experimental Evidence. *Environmental and Resource Economics* **57**, 405-429 (2014).

15 Diederich, J. & Goeschl, T. To mitigate or not to mitigate: The price elasticity of pro-environmental behavior. *Journal of Environmental Economics and Management* **84**, 209-222 (2017).

16 Löschel, A., Sturm, B. & Uehleke, R. Revealed preferences for voluntary climate change mitigation when the purely individual perspective is relaxed – evidence from a framed field experiment. *Journal of Behavioral and Experimental Economics* **67**, 149-160 (2017).

17 Lohse, J., Goeschl, T. & Diederich, J. H. Giving is a Question of Time: Response Times and Contributions to an Environmental Public Good. *Environmental and Resource Economics* **67**, 455-477 (2017).

18 Pigou, A. C. *The economics of welfare*. (London : Macmillan and co., ltd., 1920).

19 Coase, R. H. The Problem of Social Cost. *Journal of Law and Economics* **3**, 1-44 (1960).

20 Stavins, R. N. in *The New Palgrave Dictionary of Economics* (eds N. Durlauf Steven & Lawrence Blume) (Palgrave Macmillan, 2008).

21 Klenert, D. *et al.* Making carbon pricing work for citizens. *Nature Climate Change* **8**, 669-677 (2018).

22 Goulder, L. H. & Schein, A. R. Carbon taxes versus cap and trade: a critical review. *Climate Change Economics* **04**, 1350010 (1350011-1350028) (2013).

23 Chichilnisky, G. & Heal, G. Who should abate carbon emissions?: An international viewpoint. *Economics Letters* **44**, 443-449 (1994).

24 Weitzman, M. L. Prices vs. quantities. *Review of Economic Studies* **41**, 477-491 (1974).

25 Fuss, S. *et al.* A Framework for Assessing the Performance of Cap-and-Trade Systems: Insights from the European Union Emissions Trading System. *Review of Environmental Economics and Policy* **12**, 220-241 (2018).

26 MacKay, D., Cramton, P., Ockenfels, A., Stoft, S. Price carbon - I will if you will. *Nature* **526** 315-316 (2015).

27 Weitzman, M. L. in *Global Carbon Pricing - The Path to Climate Cooperation* (eds Peter Cramton, David JC MacKay, Axel Ockenfels, & Steven Stoft) 125 - 148 (MIT Press, 2016).

28 Cramton, P., Ockenfels, A. & Tirole, J. Policy Brief—Translating the Collective Climate Goal Into a Common Climate Commitment. *Review of Environmental Economics and Policy* **11**, 165-171 (2017).

29 Kornek, U. & Edenhofer, O. The strategic dimension of financing global public goods. *European Economic Review* **forthcoming** (2020).

30 Nordhaus, W. D. To Tax or Not to Tax: Alternative Approaches to Slowing Global Warming. *Review of Environmental Economics and Policy* **1**, 26-44 (2007).

31 Cramton, P., Ockenfels, A. & Stoft, S. in *Global Carbon Pricing - The Path to Climate Cooperation* (eds Peter Cramton, David JC MacKay, Axel Ockenfels, & Steven Stoft) 221-241 (MIT Press, 2016).

32 Stavins, R. N. in *Environmental and Energy Policy and the Economy* (eds Matthew J. Kotchen, James H. Stock, & Catherine D. Wolfram) 8 - 64 (National Bureau of Economic Research, 2020).

33 Krishna, V. *Auction Theory*. (Academic Press, 2002).

34 Becker, G. M., Degroot, M. H. & Marschak, J. Measuring utility by a single-response sequential method. *Behavioral Science* **9**, 226-232 (1964).

35 Fehr, E. & Schmidt, K. M. A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics* **114**, 817-868 (1999).

36 Bolton, G. E. & Ockenfels, A. ERC: A Theory of Equity, Reciprocity, and Competition. *The American Economic Review* **90**, 166-193 (2000).

37 Gollier, C. & Tirole, J. Negotiating Effective Institutions Against Climate Change. *Economics of Energy & Environmental Policy* **4**, 5-28 (2015).

Correspondence and requests for materials should be addressed to O.E.

Acknowledgments

We thank the editor and two reviewers of our paper, conference and seminar participants in Berlin, Cologne, Mannheim and Paderborn for valuable comments and suggestions. We thank Leona Jung for excellent research assistance. Financial support of the German Research Foundation for the experiment through the research unit "Design & Behavior" (FOR 1371) is gratefully acknowledged. This project also received support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 741409; the results reflect the authors' view, the ERC is not responsible for any use that may be made of the information it contains) and from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/1– 390838866.

Author contributions

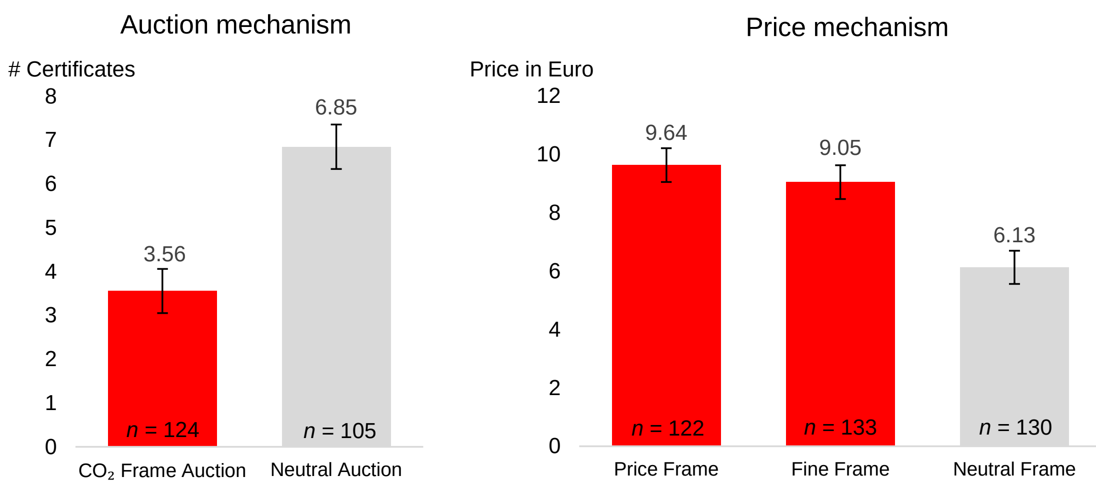
A.O. devised the project, A.O. and P.W. designed the experiment with support by O.E., P.W. carried out the experiment, performed the numerical simulations and conducted the statistical tests with support by A.O., and all authors wrote the manuscript.

Competing Interests Statements

We declare that none of the authors have competing interests.

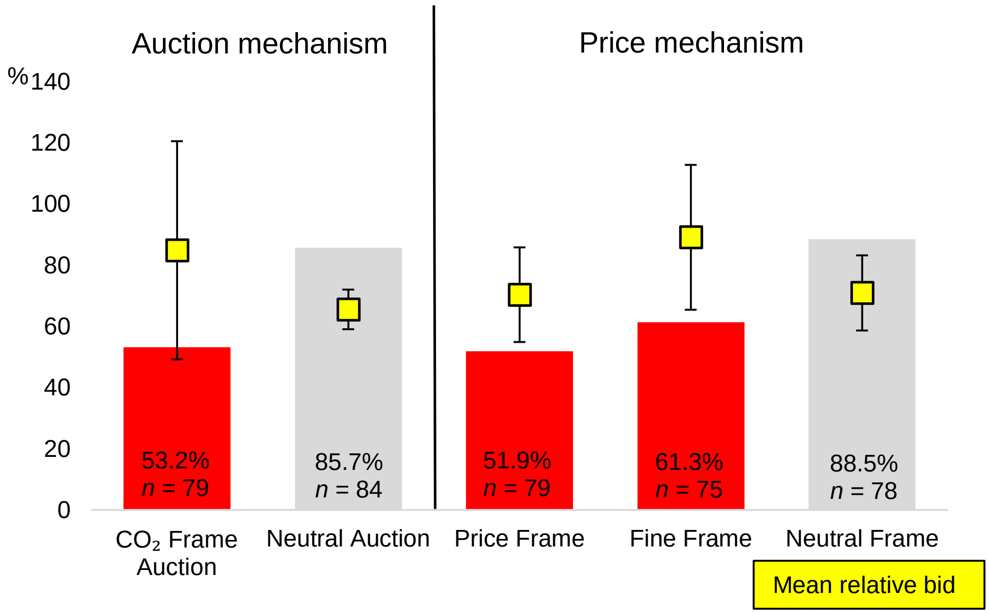
Figures

Title Figure 1: Judges' decisions – Descriptive statistics



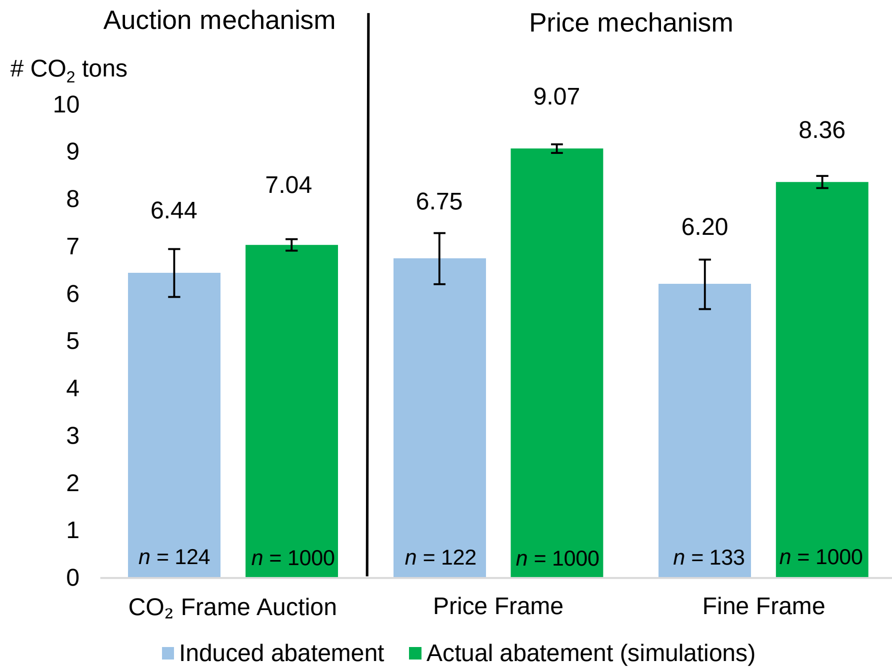
Legend Figure 1: Bars indicate mean values for judges' decisions in each experimental treatment. Error bars refer to 95%-confidence intervals.

Title Figure 2: Producer decisions – Descriptive statistics



Legend Figure 2: Bars show the percentage share of producers who participate in trading. The number of observations per treatment (n) are displayed in each bar. Yellow squares indicate the mean relative bid of producers who entered the market in each treatment (i.e. the bid the producer placed in % of his valuation). Errors bars refer to 95%-confidence intervals for mean relative bids.

Title Figure 3: Induced and actual abatement (in number of CO_2 tons not emitted)



Legend Figure 3: Blue bars indicate mean values for the induced abatement (in tons of CO_2 not emitted) by the judges' decisions under the assumption of profit-maximizing producers. Green bars indicate mean values for the actual abatement in 1000 simulated markets based on observed judges' and producers' decisions. The number of observations per treatment (*n*) are displayed in each bar. Errors bars refer to 95%-confidence intervals.

Tables

Title Table 1: Average outcomes of market simulations per treatment

	Market price in Euro (std. dev)	Efficiency of ex-ante abatement – Share of inefficient producers that abated (std. dev.)	Producer welfare – Share of maximum achievable payoffs (std. dev.)
CO_2 Frame			
Auction	2.59 (2.74)	0.717 (0.169)	0.251 (0.223)
Price Frame	8.46 (2.79)	0.979 (0.104)	0.051 (0.119)
Fine Frame	7.85 (3.13)	0.912 (0.139)	0.068 (0.176)