Instrument Choice and Motivation: Evidence from a Climate Change Experiment

Timo Goeschl*
University of Heidelberg

Grischa Perino
University of East Anglia

January 14, 2009

Abstract

Are prices or quantities the best regulatory instrument to align private actions with public interests in the presence of externalities? We add another dimension to this ongoing debate by experimentally analyzing the interaction between instrument choice and intrinsic motivation of regulated agents. The response of subjects facing a trade-off between real CO_2 emissions and private monetary payoffs to both a price and a quantity instrument are tested. We find evidence that taxes crowd out intrinsic motivation while emission standards are neutral. Crowding is short term persistent and not well explained by established cognitive theories of motivational crowding.

JEL codes: H23, H41, Q58, C91

Keywords: Instrument Choice, Motivation Crowding, Externalities

*Corresponding author: Alfred-Weber-Institute, University of Heidelberg, Bergheimer Str. 20, D-69115 Heidelberg, Germany, e-mail: tg203@uni-hd.de, tel.: +49 6221 54 8010, fax: +49 6221 54 8020

1 Introduction

The choice of the appropriate regulatory instruments to align private actions with public interests in the presence of externalities is one of the enduring themes of debate among public economists. In this debate, it has become customary to classify candidate instruments into one of two groups: One consists of so-called 'quantity' or 'command-and-control' instruments that specify actions in terms of outcomes. Empirically, quantity instruments dominate regulatory policy in the form of technical prescriptions and limits on activity levels. The other group comprises so-called 'price' or 'economic' instruments that specify incentives. Its most frequent empirical incarnation are tax instruments.

The debate on the relative merits of quantity and price instruments has a long and distinguished history. On basic efficiency grounds, the theoretical case for price instruments has been made forcefully over the years (Baumol and Oates 1988). At the same time, the theoretical models used for deriving the efficency of instruments have yielded more equivocal results on the relative merits of quantity and price instruments when extended to include abatement cost uncertainty (Zhao 2003), impacts on the direction of technological change (Krysiak 2008), and enforcement considerations (Montero 2002). Political economy considerations constitute another source of subtlety when considering instrument choice (Dijkstra 1999, Keohane et al. 1998, Hahn 1990). Finally, comprehensive assessments of the practical dimensions of regulatory implementation also deliver a less clear-cut verdict on prices versus quantities in regulatory policy (Harrington et al. 2004, Goulder and Parry 2008). The notion of a single 'best' regulatory instrument to align private actions with public interests would therefore appear to remain elusive.[1]

The present paper revisits the question of instrument choice by focusing on the fundamental drivers of how individuals respond to different forms of regulation. Specifically, this paper is interested in the possible interactions between instrument choice and the pro-public motivation of the regulated. The pro-public motivation in question is the willingness of economic agents to align their private actions with public interests voluntarily. This focus traces

[1] See Kerr and Newell (2003), Austin and Dinan (2005), Fischer and Newell (2008). Goulder and Parry (2008) provide an up-to-date literature review on instrument choice in environmental economics.

its origins to an empirical literature gathering evidence on the way in which explicit incentives interact with behavior that generates either positive or negative externalities, starting with Titmuss (1970). More specifically, there is a bold conjecture by Frey (1997) that in the presence of preexisting private contributions above the noncooperative level, introducing price and quantity instruments to increase contributions will have opposite effects on motivation: Explicit price incentives will lead to a crowding-out of motivation while quantity instruments behave neutrally or even crowd in motivation. Further, these crowding effects are conjectured to persist after the explicit incentives are removed. Frey (1992) applies this general conjecture to the specific case of environmental externalities that is also studied here.

What makes the question of an interaction between instrument choice and motivation newly salient are two developments: One is a recent literature of new experimental findings and theoretical foundations on which this present paper builds. This recent literature examines how explicit incentives interact with the intrinsic motives and preferences that determine people's conduct. Price incentives have been found to give rise to problematic effects (Bénabou and Tirole 2006, Frey and Oberholzer-Gee 1997). But also other types of explicit incentives can suffer from similar deleterious interactions with intrinsic motivation: Fines for missing performance targets decreased the average agent's effort in experimental contract settings (Fehr et al. 2007). Likewise, introducing fines for late pick-up of children from day-care centers led to an increase in late pick-up in a field experiment (Gneezy and Rustichini 2000a). Small changes to the contract environment appear to matter: Merely restricting the agent's choice set in a principal-agent setting can trigger more selfish behavior (Falk and Kosfeld 2006). Ellingsen and Johannesson (2008) present additional experimental evidence that control systems and pecuniary incentives both lead to what has become known as 'motivation crowding' (MC). Alongside the evidence in favor of 'motivation crowding-out', there is also evidence for *positive* interactions between explicit incentives and intrinsic motivation. Non-price incentives, in particular, are credited with giving rise to intrinsic motivation and hence leading to 'motivation crowding-in' (Frey and Jegen 2001). Against this background of new results, the question of instrument choice and motivation can therefore be asked with new methods and concepts at hand.

The second salience of the question of motivation and regulatory instruments arises from

the fact policy has set itself ambitious targets in areas such as climate change policy that will require additional environmental regulations to be imposed not only on businesses, but also on individual consumers (CCC 2008). For these economic agents intrinsic motivation is clearly relevant. Moreover, individuals' attitudes and preferences do also influence decisions of firms too a larger degree than often presumed by economists (NCF 2006).

In this paper we develop an experimental setting in which the presence and magnitude of MC effects associated with the two archetypical instruments of regulating environmental externalities, taxes and standards, can be compared directly. In an experiment where subjects face a decision between private money payoffs and contributions to mitigate climate change we are able to study how the introduction of a tax and an emission standard (enforced by a fine) affect subjects willingness to contribute to the environmental public good. To our knowledge, this is the first paper to provide a direct comparison of the behavioral impact of these common regulatory instruments in a controlled setting.

In the chosen setting, subjects take a sequence of three anonymous and independent decisions that involve a simple trade-off between providing a global public good in the form of greenhouse gas emissions reductions and collecting personal payoffs in the form of cash pay-outs. Subjects are randomly selected into one of four groups. All groups go through an initial round of choice without any explicit incentive schemes in place. Two of the four groups make their second-round decision with explicit price or quantity incentives in place, while the other two groups act as controls. In the third round, all four groups choose again without explicit incentives. We compare the choices between rounds and test whether and how experiencing an episode of explicit incentives in the second round changes subjects' behavior. This design choice makes it possible to focus not only on the instantaneous effects of regulation, but also on the intertemporal dimension of the interaction between explicit incentives and motivation that is the subject of considerable debate (Frey 1992, Bénabou and Tirole 2003, 2006). The experimental evidence supports the notion of crowding out of motivation by price instruments but not by quantity restrictions. We also find that the crowding out of the price instrument spills over from the regulated round into the round when incentives have been removed. Together, these findings can be taken as providing support for the 'Frey (1992) Conjecture'.

The final contribution of the paper is to take steps to explain the crowding effects observed. A critical review of the existing cognitive theories (Bénabou and Tirole 2003, 2006, Sliwka 2007) that explain MC reveals that due to the experimental design, their mechanisms have limited traction in explaining the present evidence. As an additional check, we test for the longer-term persistence predicted by cognitive theories by performing a follow-up session one week after the initial session. We find that the effects observed in the initial session are no longer detectable after this comparatively short period. Other mechanisms may be necessary to explain the experimental evidence, among them affective mechanisms studied in the psychological literature on social motivation (Forgas and Laham 2005).

The following section discusses the experimental design. Section 3 sets out the behavioral predictions before turning to the key results in section 4. Section 5 evaluates the results in light of theoretical considerations and related experimental findings. Section 6 concludes.

2 The Experiment

2.1 The General Set-Up

The purpose of the experimental design is to provide a platform for testing whether experiencing an episode of explicit incentives affects individuals' trade-offs between private gain and private contributions to an environmental public good. There are two sessions: The initial session 1 involves three rounds; the second session 2 - one week later - involves one round. Only session 1 is relevant for what follows in sections 2 to 4 of this paper. Treatment groups only differ in their experience of the second round in the first session in which a subsample of subjects face an episode of explicit incentives. We test the impact of a regulatory episode at two points: (i) when the explicit incentives are applied, and (ii) in a subsequent round in the same session when incentives have been removed. Subjects knew that there might be more than one round, but not how many. Payoffs and contributions were accumulated over all rounds in session 1.

The practical challenge in this experiment is to bring a public good with desirable characteristics into the laboratory to ensure that the trade-off between the public good and

private gain is real while minimizing the interaction between participants. We achieve this by using abatement of greenhouse gas emissions that contribute to global climate change as the public good. Contributions towards the public good purchase and retire carbon dioxide (CO_2) permits from phase II of the European Union Emission Trading System (EU ETS) and thus restrict the EU's regulated greenhouse gas emissions by the same amount.

Subjects choose an production level that determines both their private payoff and the contribution to the public good. Feasible production levels, corresponding payoffs and contributions are given in Table 1. The permit price at the time of the experiment was EUR 24.50 per ton of carbon emissions. During the introduction to the experiment subjects were informed about the EU ETS and the process of buying and irrevocably withdrawing permits from the aggregate quota. The feasibility of abating EU CO_2 emissions through this mechanism was demonstrated during the introduction by logging into the experimenter's carbon trading account and purchasing and retiring one permit. The basic schedule of payoffs and contributions remained constant across all periods and treatments. Production level 5 generates the maximum of the sum of payoff and contribution, implying that some amount of production may be socially desirable.

Subjects were randomly assigned IDs when entering the room and were seated according to the first letter of their ID. There was no interaction between subjects either in terms of communication, information, or payoffs at any stage of the experiment. This total anonymity was announced in advance as was the independence of decisions across periods. Horizontal interactions between participants are therefore not part of the experimental set-up. All material used in the experiment was in hard-copy, pre-printed and not individualized. Together with the random selection at the outset, subjects could therefore not reasonably deduce a relationship between their characteristics, their choices, and their set tasks. A vertical incentive structure is therefore not part of the experimental set-up.

This completes the description of the basic set-up.

2.2 Treatments

Session 1 of the experiment consisted of three rounds. In the first round, subjects in all four treatments faced an identical decision problem without explicit incentives in place. This

1	2	3	4
Production Level	Your Payoff	Climate Contribution	CO_2 Abated
0	0.00 EUR	30.00 EUR	1.22 t
1	3.00 EUR	29.50 EUR	1.20 t
2	6.00 EUR	28.50 EUR	1.16 t
3	9.00 EUR	27.00 EUR	1.10 t
4	12.00 EUR	25.00 EUR	1.02 t
5	15.00 EUR	22.50 EUR	0.92 t
6	18.00 EUR	19.00 EUR	0.78 t
7	21.00 EUR	14.00 EUR	0.57 t
8	24.00 EUR	8.00 EUR	0.33 t
9	27.00 EUR	2.00 EUR	0.08 t
10	27.00 EUR	0.00 EUR	0.00 t

Table 1: Payoffs and Climate Contributions

problem consists of the basic set-up explained above with the corresponding trade-offs given in Table 1. In the second round, the decision problems differed between the four groups. Of the four, two were regulation treatments: *TAX* and *COMMAND & CONTROL* (*C&C*), and two corresponding control treatments: *BASELINE* and *EXOGENOUS PAYOFF CHANGE* (*EX*). In the regulation treatments *TAX* and *C&C* , subjects experienced an episode of explicit incentives, in which they were told that in order to protect the environment individual emissions are metered and regulated. For each level of production a corresponding tax level was set in the *TAX* treatment. The instructions presented the amount of tax payable at each production level and the resulting net payoff while the answering form contained only the net payoff and was otherwise identical to the one in period 1. In the *C&C* treatment, a permitted threshold of 6 was stipulated and for each production level above this threshold individuals had to pay a penalty (matching the tax due at this level). Table 2 states the amounts by which payoffs were reduced. Both regulatory instruments are equivalent in the sense that both shift the payoff maximizing production level from 9 and 10 to 6. Subjects were informed that revenues from taxes and fines were neither redistributed nor spent on

1	2	3
Production Level	Tax	Penalty (*C&C* treatment)
0	0.00 EUR	0.00 EUR
1	0.50 EUR	0.00 EUR
2	0.75 EUR	0.00 EUR
3	1.50 EUR	0.00 EUR
4	3.50 EUR	0.00 EUR
5	5.50 EUR	0.00 EUR
6	8.00 EUR	0.00 EUR
7	11.50 EUR	11.50 EUR
8	15.50 EUR	15.50 EUR
9	19.50 EUR	19.50 EUR
10	21.00 EUR	21.00 EUR

Table 2: Taxes and Penalties

abating carbon dioxide.

Subjects in the *BASELINE* treatment faced exactly the same decision as in period 1. In the *EX* treatment private payoffs were adjusted to match those in the *TAX* treatment, but without giving any reason for the change. The *EX* treatment allows us to disentangle the effects of the change in the relative price of carbon abatement from any effect due to the source of this change - here the imposition of explicit incentives in the form of a tax in the *TAX* treatment. The answer sheets that subjects completed in the *TAX* and *EX* treatments were exactly the same.

In round 3, all subjects faced the same decision problem as in round 1. This continuation in case of the *BASELINE* treatment and the reversion to round 1 in the *EX* treatment was implemented without further explanation. In the *TAX* and *C&C* treatments, subjects were informed that emissions will neither be measured nor sanctioned in this period and that the payoff matrix was therefore exactly the same as in period 1. They also knew - as before - that their action would not affect future payoffs.

A total of 172 students participated in session 1 spread over two sub-sessions each lasting

about one hour. Both sub-sessions included all four treatments. For sub-session 1A we recruited 43 subjects by an open call to students of different departments (e.g. economics, political science, physics, humanities etc.). The 129 subjects of sub-session 1B were recruited from a single lecture. We found no evidence that behavior of subjects in the two sub-sessions differed significantly. 10 participants had to be excluded from the analysis since we could not rule out that they participated in more than one sub-session. After the third round some participants (5 in sub-session 1A and 10 in sub-session 1B) were picked at random and both their private payoffs and contribution to abate CO_2 materialized. This was also announced in advance. Individual payoffs ranged from EUR 0.- to EUR 81.- and 30 permits for a total of 30 tons of CO_2 were bought and retired. The experiment was conductedbetween May and July, 2008.

3 Behavioral Predictions

Subjects in the experiment differ only with respect to events in round 2. There, subjects in treatments *TAX* and *C&C* experience an episode of exposure to explicit incentive schemes, i.e. a change in incentives explicitly framed as interventions aiming to change behavior, that would induce a cash pay-out maximizing subject to reduce its production level to 6. Subjects in treatment *EX* experience a tax-equivalent reduction in cash pay-outs without an explicit intervening character. Subject in the *BASELINE* treatment experience no change in incentives. In rounds 1 and 3 and in the follow-up session one week later, all subjects face an identical choice situation. The focus here is on predictions how differences in regulatory experiences change the decision of subjects in round 2 and later rounds compared to their round 1 behavior and the *BASELINE* treatment.

Hypothesis 1 (round 2 behavior) *: Production levels in round 2 are determined by the relative price effect only.*

The experiment offers subjects an implicit price schedule for CO_2 emissions abatement. Upper and lower bounds of subjects' implicit marginal willingness to pay for abatement services can therefore be inferred on the basis of subjects' choice of production level. From

round 1 to round 2, all treatments except *BASELINE* involve a change in the price schedule. These changes are identical for *TAX* and *EX* for all production levels and for *C&C* for production levels above 6. Up to 6, *C&C* retains the price schedule of the *BASELINE*. Conventional economic theory would assert *invariance with respect to instruments*: The circumstances of the price shifts observed by the subjects are immaterial for their marginal willingness to pay. The presence of explicit incentive schemes will not interact with intrinsic motivation. The subject's marginal willingness to pay is revealed in round 1 and determines production levels in round 2.

Note that hypothesis 1 does not make a prediction on levels in round 1. Levels will depend on individual's preferences for climate change abatement given their interpretation of the public goods situation. Even subjects with preferences for the public good can follow the 'large economy' logic (Andreoni 1988) of the climate change game and arrive at a zero willingness to pay on account of the literally billions of players involved in the public good provision. Alternatively, subjects can conclude that the experiment offers an offset opportunity for their consumption of private goods that have an adverse effect on the public good (Vicary 2000, Kotchen 2009). In this case, even with a very large number of players, individuals will rationally choose production levels below 9 (Kotchen 2009). Changes from round 1 to round 2 are restricted to simple changes in relative prices. The fundamental characteristics of the public goods situation remain therefore unaffected.

Hypothesis 1 will fail if there is a sufficiently strong interaction between the regulated individual's willingness to pay and the explicit incentives provided by the regulatory instrument. The presence of such an interaction is at the heart of experimental and theoretical contributions to understanding MC. A prerequisite is that the experimental public goods situation is interpreted as an opportunity for prosocial or moral behavior such that individuals can act upon their pro-public motivation. If present, this has two implications. One is that subjects will rationally choose production levels below 9 in round 1 in line with the strength of their pro-social motivation. The second is that since this motivation relates to subtle inter-personal relationships, altering these relationships by introducing explicit incentives can give rise to an unintended reduction in cooperative behavior. Both in vertical (Bénabou and Tirole 2003, Sliwka 2007) and horizontal relationships (Bénabou and Tirole 2006), the

introduction of externally provided incentives interacts with preexisting information imperfections to generate reductions in cooperative effort by individuals. Jointly, these theoretical perspectives on MC give rise to cognitive explanations for motivation crowding and could explain why hypothesis 1 fails. We discuss in section 5 how the experimental set-up relates to the mechanisms required for a cognitive interaction between treatments and the relevant motivational preferences.

Hypothesis 2 (round 3 behavior) *: Production levels in round 3 are the same as in round 1, adjusted for any baseline effects.*

Round three consists of the same answer sheet as round 1 and presents the original price schedule. Under invariance with respect to instruments, only factors common to all treatments would be admissible for explaining changes in production levels and hence in line with conventional economic theory. These common factors are taken into account by comparing changes in production levels across treatments and by adjusting for the behavior of subjects in the *BASELINE* treatment.

Note that hypothesis 2 can hold even if hypothesis 1 fails: If preference shifts occur in round 2 despite the behavioral predictions of hypothesis 1, they may only be effective within the round. The relationship between the two hypotheses is therefore a question of short-term persistence of incentive-induced changes in preferences.

The presence of such persistence - even in the long run - has been postulated by theoretical models of MC (Bénabou and Tirole 2003, 2006). The reason is that the cognitive processes modeled there are based on Bayesian learning. Even if regulation has been removed, subjects still take any information previously revealed into account. Cognitive motivation crowding (CMC) is therefore a candidate explanation if hypothesis 2 fails. Evidence from field experiments provides grounds for predicting such a failure: Short-term spillovers from incentive episodes into later non-incentive episodes have been observed in other examples of public goods contribution (Meier 2007).

We provide a more detailed discussion of the contribution of CMC to explaining the experimental evidence in section 5 and now turn to the experimental evidence used to test these hypotheses.

4 Results

We start with some basic information on the behavior of the four groups in the three rounds before testing the two key behavioral hypotheses set up in the preceding section.

Figure 1 reports on average production levels of each treatment group in each round, with 95% confidence intervals added. Round 1 offers information on subjects' willingness to pay for abatement before explicit incentive schemes are introduced. While mean production levels are different, these first-round differences are not statistically significant as the confidence intervals indicate: p-values of pairwise Mann-Whitney tests range from 0.2281 to 0.9038. Average production levels are substantially below the pure private payoff maximum of 9: The median production level is 6 in all groups. While not *per se* conclusive evidence for the presence of intrinsic pro-social motivation (rather than selfish offset motives) among subjects, this provides a pool of subjects for whom crowding out could conceivably be observed when explicit incentives are introduced in round 2. There is also a high degree of variation between subjects with individual choices spanning the entire range of feasible production levels.

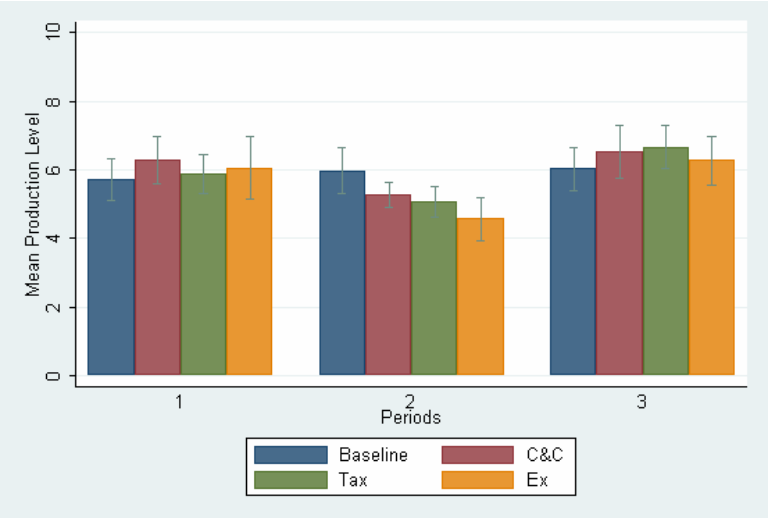

Figure 1: Production Levels

4.1 Instantaneous impacts of explicit incentives

The analysis of the within-round impact of the explicit incentives introduced in round 2 requires comparing choices between round 1 and round 2 and differences between treatment groups. As a first simple test, we check, along the lines of Falk and Kosfeld (2006), whether the introduction of relative price changes does shift behavior between rounds and across treatments. To make rounds 1 and 2 comparable despite the fact that in treatments *EX*, *TAX*, and *C&C* production levels above 6 are effectively ruled out, round 1 choices for these treatments are adjusted such that all production levels above six are set equal to six. Since the median production level in that round is 6 for all treatments, the medians are not affected by the Falk-Kosfeld adjustment.

Result 1 *Production levels in rounds 1 and 2 are the same for all treatments on an adjusted basis (where required), except where the price shift is not induced by explicit incentives (EX treatment).*

While containing non-adjusted data, figure 1 already hints at this result, with production levels in the *EX* treatment exhibiting the greatest reduction among the treatment groups. As far as the *BASELINE* treatment is concerned, result 1 is unsurprising. Since subjects experience no change in the choice situation, there is no reason for WTP to change. The absence of a statistically significant change in behavior in treatments *TAX* ($p = 0.3932$) and *C&C* ($p = 0.1754$) is less obvious. For *C&C*, result 1 implies that despite the exogenous introduction of an upper limit on production levels that is enforced with a penalty, there is no *prima facie* evidence of a "hidden cost of control" (Falk and Kosfeld 2006) in this experimental setting. This evidence confirms the finding in Falk and Kosfeld (2006) that exogenously introduced controls do not lead to motivation crowding. At the same time, it contradicts the notion that the stipulated pollution threshold signals something like a social norm (or experimenter's expectations) that subjects should choose a production level of six. At least those with lower production levels do not respond accordingly.

For *TAX*, the interpretation of result 1 requires more subtlety. Two countervailing effects are plausibly present at the same instant. The first is a relative price effect that shifts production levels down as abatement becomes relatively cheaper. The operation of the

relative price effect is evident in the significant drop of production levels for the subjects in the *EX* treatment ($p = 0.0119$ Wilcoxon signed-rank test), the only treatment in which production levels change significantly. The second possible mechanism is an interaction effect known as the 'cost of price incentives' (Frey and Oberholzer-Gee 1997) that shifts production levels up and is a well-established phenomenon in the empirical literature (Frey and Oberholzer-Gee 1997, Gneezy and Rustichini 2000b,a, Meier 2007).

The gist of result 1 is borne out by more detailed comparisons between treatments. The median production level in period two is significantly lower in the *EX* than in the *TAX* treatment ($p = 0.083$, median test when values equal to the median are assigned to the above group). While there is no significant difference in adjustment from period one to two between the *TAX* and the *BASELINE* treatment (all production levels truncated at six) ($p = 0.3368$) there is a difference between the same adjustment between the *EX* and the *BASELINE* treatments ($p = 0.0107$). There are no significant differences between the *C&C* and the *BASELINE* treatment when the latter is adjusted such that all production levels above six are set to six.

Result 2 *Relative price changes are insufficient for explaining the behavioral changes induced by explicit price incentives, but explain behavioral changes induced by explicit quantity incentives.*

Result 2 relates to the specific predictions contained in hypothesis 1 on how behavior will change between rounds and between treatments. Comparing rounds 1 and 2, Hypothesis 1 predicts that those subjects experiencing a change in relative prices will adjust their choice of production level in line with their original willingness to pay, which should be independent of changes in relative prices. Such changes in relative prices occur for all subjects in treatments *TAX* and *EX* and for those subjects in *C&C* that chose an production level above 6 in round 1. Comparing between treatments, hypothesis 1 maintains that the context of the price change is immaterial to subjects' decisions. Observed changes should therefore not differ between treatments *TAX* and *EX* that are payoff equivalent and differ only in terms of an explicit incentive schemes being present in the case of *TAX*.

We implement hypothesis 1 by determining a minimum willingness to pay for each subject on the basis of their first period choice. This minimum WTP is then used to determine the predicted production level in round 2. This predicted production level is compared with the actual choice. Note that for the *C&C* treatment the predicted choice coincides with round 1 production levels restricted to values not larger than six. Hence, for the *C&C* treatment we already tested for hypothesis 1 by comparing adjusted round 1 choices with those in round 2. There is no significant difference and hypothesis 1 is confirmed.

Testing whether there are significant deviations from the behavior predicted by hypothesis 1 in the *TAX* and*EX* treatments, the findings are that hypothesis 1 is rejected for the *TAX* treatment ($p = 0.00$), but also for its control, the *EX* treatment ($p = 0.00$). This means that in both cases, factors other than relative price effects must play a role. In the *TAX* treatment, on average subjects' adjustment falls short of the predicted adjustment by 1.91, whereas the difference is 1.42 in the *EX* treatment. In both treatments no one adjusts by more than hypothesis 1 predicts.

Result 3 *Explicit price incentives lead to a reduction in pro-social behavior relative to a situation without explicit incentives.*

Result 3 is supported by tests of the *TAX* and the *EX* treatments for statistically significant differences. Examining the average deviation from full adjustment to the price change, subjects in the *TAX* treatment adjust substantially less from round 1 to round 2 than those in the *EX* treatment ($p = 0.0740$ Mann-Whitney test). Since the two treatments are payoff equivalent, we attribute this statistically significant difference to an interaction between preferences (WTP) and the explicit incentives. Their presence shifts preferences regarding voluntary abatement effort and diminishes subjects' willingness to fully adjust to the new price schedule. In other words, result 3 provides evidence for the presence of MC of explicit price incentives in an experimental setting that disentangles price effects from motivational effects.

4.2 Short-term persistence of crowding effects

We now turn to the question of how an episode of explicit incentives influences subsequent behavior once the incentives have been removed. Results 1 to 3 establish the presence of MC vis-à-vis explicit price incentives in round 2. If the MC detected in round 2 are manifestations of a (Bayesian-) learning process, that suggests intertemporal persistence of the change in preferences of those subjects that have experienced an emissions tax. These behavioral changes may therefore spill over into the subsequent period, even though the incentives are no longer present.

Comparing production levels in rounds 1 and 3 provides evidence on short term persistence of behavioral changes triggered by explicit incentives applied in round 2. In round 3, subjects face again an identical decision to round 1. The only difference lies in their regulation experience, i.e. their treatment in round 2. As figure 1 shows, average production levels in the regulated treatments readjust overall - as expected - in the direction of levels chosen in round 1. Here we carefully examine this re-adjustment treatment by treatment.

Hypothesis 2 predicts that the readjustment process is complete. Table 3 shows the average and median production levels in round 1 and 3. Readjustment takes place in all treatments, with averages and medians in round 3 returning to their round 1 level. In the case of the *TAX* treatment, however, subjects choose production levels that are significantly higher ($p = 0.0006$) than in round 1. The median level of production increases from 6 to 7, while it remains at the original level for all other treatments. Testing for hypothesis 2 shows that average production levels significantly increased in the *TAX* treatment, but not in the *C&C* ($p = 0.5194$) and the *EX* ($p = 0.8865$) treatments, leading to result 4.

Result 4 *Once explicit incentive schemes have been removed, production levels overshoot only after a regulatory episode involving price incentives.*

The average adjustment from rounds 1 to 3 in the *TAX* treatment is significantly larger than that in the *EX* ($p = 0.0129$) and the *C&C* ($p = 0.0453$) treatments. Since there is a small, but significant increase in the *BASELINE* ($p = 0.0718$), the increase in the *C&C* treatment can be discounted somewhat, but despite being smaller than in the *BASELINE*, the difference is not statistically significant ($p = 0.3427$). Compared to the *BASELINE*

		Treatment			
		TAX	*EX*	*C&C*	*BASELINE*
Round 1	Average	5.89	6.06	6.29	5.72
	Median	6	6	6	6
Round 3	Average	6.67	6.27	6.53	6.02
	Median	7	6	6	6

Table 3: Mean and Median Production Levels in Rounds 1 & 3

treatment there is a small decrease in production levels for *C&C* which points more towards crowding in than crowding out of intrinsic motivation to contribute to the public good. Hypothesis 2 can therefore not be rejected for the *C&C* and *EX* treatments: There are no crowding effects that persist from round 2 into round 3. The overshooting result in the case of the *TAX* treatment, on the other hand, contradicts hypothesis 2 for explicit price incentives. This leads to result 5.

Result 5 *Crowding-out of explicit price incentives persists in the short run.*

Seen in the context of the MC observed in round 2, results 4 and 5 provide evidence for the short-term persistence of changes in preferences or motivation induced by the application of explicit price incentives in round 2. After removing any change in relative prices only subjects in the treatment that generated MC in round 2 increase their production level from round 1 to round 3 at a statistically significant rate, raising the median production from 6 to 7. This provides evidence that the type of MC detected in round 2 is not only incentive-specific, but also short-term persistent. The use of explicit price incentives continues to reduce intrinsic motivation to contribute to the environmental good even once the incentives are removed. This short-term persistence is at odds with the idea that a social-norm or the experimenter's expectation explain the crowding effect observed in round 2. Production levels do not move towards six, but quite the opposite. Round 3 in the *TAX* treatment is the only incidence across all treatments in rounds 1 & 3 where production levels are significantly different from (i.e. above) six ($p = 0.0262$). We do not find a similar 'hidden cost of control' effect for command-and-control regulation. The quantity restriction has no statistically significant impact on production levels in the subsequent period.

5 Discussion

5.1 Evaluation of the experimental evidence

Summarizing the results of the experiment, we find that explicit price incentives trigger crowding-out of contributions to a global public good, even after taking into account imperfect adjustments to changes in relative prices in a neutrally framed context. By contrast, an episode of explicit quantity controls enforced in a way that - in economic terms - is broadly equivalent to the price incentive does not lead to crowding-out. Testing for short-run persistence, the crowding effect of the price incentive survives into a subsequent round in which the incentives are no longer operational. For the quantity restriction, there is weak evidence for crowding in.

It is important to put these results into the broader context of the recent literature on motivation crowding. Empirically the presence of crowding effects has been established for a wide range of situations and instruments (Gneezy and Rustichini 2000a,b, Falk and Kosfeld 2006, Meier 2007). However, both field and laboratory experiments focus on small scale principal-agent relationships (with the exception of Meier (2007)) more suitable e.g. for workplace settings than for public policy. Moreover, none of these studies compares quantity and price based interventions. The theoretical literature explaining crowding phenomena relies on preexisting informational imperfections. The introduction of incentive instruments provides new information in vertical relationships (Bénabou and Tirole 2003, Sliwka 2007) or changes the signal value of contributions to the public good in horizontal relationships (Bénabou and Tirole 2006). This literature models the reduction in motivation as a rational response to a change in the informational context triggered by the introduction of an incentive scheme. The interaction between incentives and motivation is therefore purely cognitive.

The theoretical literature provides both concepts and terminology that appear to fit our experimental findings. We proceed by testing whether the appearance survives scrutiny.

5.2 Can Cognitive Theories Explain the Evidence?

CMC theories demonstrate how the interaction of explicit incentives and informational imperfections can give rise to MC. Their explanatory power with respect to the experimental evidence presented here is determined by two aspects. One is the relationship between the mechanisms required for CMC and the experimental design. The other is the relationship between the predictions of CMC and the experimental evidence. We address these two aspects in the following.

CMC requires horizontal or vertical interaction channels for the interaction between information and incentives to play out. Through a number of experimental design features, however, these channels are shut down in the present experimental setting. One feature is the common knowledge at the outset of the experiment on (a) the randomness of group assignment, (b) the independence of actions between rounds, and (c) on the independence of the type and sequence of incentives provided (if any) from individual or group behavior. On the basis of this common knowledge, participants have no reason for expecting that the incentives are in some way conditioned on their individual or aggregate behavior (Bénabou and Tirole 2003). Also, the randomness of assignment and independence of choices across rounds rules out the conditional behavior studied in Sliwka (2007) as an explanation as subjects cannot learn something about the environment from observing the explicit incentives provided. The second feature is the anonymity of decisions among participants. This anonymity shuts down reputation effects along the lines of Bénabou and Tirole (2006). The anonymity combined with the independence of choices also rules out horizontal interactions between subjects that would be a precondition for peer pressure effects (Kandel and Lazear 1992, Fischer and Huddart 2008). In short, there are no obvious cognitive reasons to our knowledge for expecting MC to be present in this setting. This contrasts with the settings in Andreoni (1993), Cardenas et al. (2000), Gneezy and Rustichini (2000a) and Gneezy and Rustichini (2000b).

One exception to the above are any effects subjects' self-image might have on their behavior (Bénabou and Tirole 2006). In general, self-image effects would seem almost impossible to avoid in experiments of this type. However, if self-image is the driving force behind any observed crowding effect, we should be able to distinguish it from other drivers if the

predicted direction and persistence of effects differ. Note that the experimental set-up also should not give rise to a vertical relationship between subjects and experimenter as it is common knowledge in the experiment that the experimenter is completely precommitted, demonstrates that he has no outcome-specific interest, and that the experimenter cannot condition on observed individual or group behavior in any way. However, subjects decisions are not perfectly anonymous to the experimenter who could conceivably identify the winning individuals during the pay-off procedure. This might interfere with subjects' choices if they have beliefs about the experimenter's expectations.[2] This potential interference is of relevance only as far as the effects of regulatory interventions are concerned, whereas its potential impact on the absolute level of contributions is immaterial to the subsequent analysis. If subjects believe that the explicit incentive schemes in round 2 signal experimenter's expectations or a social norm, this would result in a shift of contributions towards six in both the *TAX* and *C&C* treatments for rounds 2 - 3. This possibility was examined in section 4 and found not be supported by the evidence. Moreover, it fails to explain the difference between the command-and-control and the tax intervention.

As a control, subjects were given an exit questionnaire that tested for attitudes vis-à-vis emissions trading and other relevant factors that might interact with subject's observed behavior. For example, subjects may object to the idea that polluters can continue polluting as long as they purchase permits rather than being obliged to abate. The questionnaire asked subjects about their views on the idea of trading pollution rights. There is no detectable relationship between these views and behavior in the laboratory. Alternatively, subjects may find that the aggregate emissions quota is set too high or low, potentially biasing their decisions. The only finding in this context is the predictable results that subjects with a stronger environmental orientation did provide more contributions to climate change abatement than those with a low level. While contributing to our understanding of the level of contributions, though, environmental orientation did not interact with changes in behavior in any detectable way. The experimental set-up was also careful to introduce explicit prices regarding the environmental good for all treatments at the same time, namely

[2]There is mixed evidence on the relevance of experimenter blindness (Hoffman et al. 1994, Bolton et al. 1998).

at the start of the experiment. This set-up ensures that the results pick up only incentive-specific effects rather than a possible opposition of subjects to the general idea of putting a price on environmental goods. These attitudes can therefore not explain observed effects and to not correlate with behavior.

With the obvious channels of most cognitive mechanisms of MC shut down and evidence inconsistent with communication of experimenter's expectation or social norms via interventions only self-image concerns remain of the established drivers of MC. To check whether the Bénabou and Tirole (2006) type of self-image concern can explain the crowding out of price instruments in this setting we perform a 'long-run' persistence test.

5.3 Testing for long-term persistence

To test for the long-term persistence of the observed crowding effects participants of subsession 1B were invited to participate in a follow-up session exactly one week later. This was feasible without a previous announcement since they were recruited from within a weekly lecture course. They again faced the same payoff matrix as in period 1. All treatments had the same instructions. 70 students participated in the follow-up with 62 being included in the analysis.

Hypothesis 3 (round 4 behavior) *: Production levels in round 4 are the same as in round 1, adjusted for any baseline effects.*

Using standard theory to predict round 4 behavior the same arguments apply as for hypothesis 2. The difference is that round 4 tests for evidence of long-term persistence. Rejecting hypothesis 3 implies that the preference changes induced by round 2 persist for a non-trivial amount of time and have therefore altered something more fundamental in subjects' valuation of climate change abatement contributions in either direction. Self-image concerns based on cognitive mechanisms would predict a long-run persistence of any difference between instruments observed in the short-run. This is a result of the Bayesian updating process that takes into account all information previously revealed.

With little theory to provide guidance on the time scales of behavioral changes, round 4 is spaced one week from the session containing rounds 1 through 3. The smaller sample (62

		Treatment			
		TAX	*EX*	*C&C*	*BASELINE*
Round 1	Average	6.00	6.29	6.67	5.80
	Median	6	7	7	7
Round 4	Average	7.00	6.93	7.27	6.27
	Median	9	7.5	8	7

Table 4: Mean and Median Production Levels in Rounds 1 & 4

Treatment	# subjects
BASELINE	15
C&C	15
TAX	17
EX	14

Table 5: Number of Participants in Round 4

subjects) of returned subjects is statistically not distinguishable from the larger population that participated in rounds 1 to 3. Inspection of the average and median production levels in table 4 seems at first to suggest significant effects. The median production level has increased in all treatments, most notably in the *TAX* treatment. Comparisons of average production levels in period one and four confirm this observation of increased production levels only for the the *TAX* ($p = 0.0668$) treatment and for the *BASELINE* ($p = 0.0149$). However, variances within treatment groups are considerable. Comparing effects across treatments fails to pick up longer-term treatment effects between the initial and final round of the experiment one week later.

Result 6 *There are no significant differences in the adjustments from round 1 to round 4 between treatments.*

Hypothesis 3 can therefore not be rejected. When testing for longer-run persistence of these effects by repeating the last round one week later, crowding effects are no longer present. This makes it unlikely that subjects' concern about their self-image (Bénabou and

Tirole 2006) drive the crowding effects observed in this experiment. This renders the last of the established drivers of MC implausible.

5.4 Other candidate explanations

The limited traction of established cognitive drivers of MC raises the question of what alternative mechanism is likely to be activated by price incentives rather than quantity incentives, but would fail to lead to a longer term change in behavior. One plausible candidate for an alternative mechanism that combines an instrument-specific response with a short-, but not long-term persistence is affect-based behavior, leading to affective motivation crowding (AMC) as opposed to CMC.

Affect-based behavior requires a 'valence' basis, i.e. an association of positive or negative emotions with an event (Forgas and Laham 2005). A plausible valence basis for the effects observed in the experiment derives from the fact that its choice situation can be readily regarded as a moral decision in which dichotomous categories of 'good' and 'bad' apply. If this moral dichotomy is at the root of emotional attitudes vis-á-vis regulation, different forms of regulation can differ by how well they map into those categories. If they fail to map well along subjects' moral views, this can generate feelings of being treated unfairly and therefore lead to a behavioral reversal (De Dreu and Steinel 2006, Pillutla and Murnighan 1996).

A quantity control instrument of the type applied in round 2 creates two distinct regimes and thus maps well into a dichotomous view of 'good' and 'bad' behavior. It creates one regime for those within the threshold and another for those outside. In rounds 1 and 3 we asked subjects what they regard as the socially optimal production level.[3] The mean across all subjects was 4.98 in round 1 and 5.19 in round 3, the median in both rounds was 5. Since the threshold level of 6 is close to what subjects regard as the 'social optimum', it is likely to divide the set of subjects close to where subjects themselves would conjecture that a threshold between good and bad would lie.

The price instrument, on the other hand, is only marginally selective rather than dividing behavior into dichotomous moral categories. If subjects regard themselves as incurring

[3]The question was on the same answering sheet subjects used to choose their production level (see the instructions in the appendix).

sacrifices in order to 'do the right thing', the introduction of a tax in round 2 can appear to condemn every choice in round 1, however prosocial, with the exception of a level of zero. The implicit message of a tax instrument that no behavior is good enough not be penalized will then plausibly conflict with subjects' attitudes regarding their behavior because it fails to acknowledge the moral categories that may motivate individuals. Requiring a decision on production levels while such sentiments of being treated unfairly do linger in subjects' minds will plausibly lead to an reduction in the willingness to contribute.

To test for this we compare the mean distance between a subject's choice and what he or she regards as socially optimal. Note that choice and self-stated 'social optimum' are significantly different for all treatments ($p = 0.0000$ in round 1 and $p < 0.001$ in round 3 for all treatments except the *BASELINE* where $p = 0.0796$). The only treatment where this distance significantly increases ($p = 0.0157$) from round 1 to 3 is the *TAX* treatment. The p-values for the other treatments are between 0.7039 and 0.9667. Note that there is no significant change in what subjects in the *TAX* treatment consider to be socially optimal ($p = 0.7232$). Hence, exposure to an explicit price incentive scheme does not crowd out moral standards themselves but (temporarily) the motivation to adhere to them in a public goods context. The stability of moral standards fits well with the short term persistence of crowding effects observed. Once the affective response to the tax has ebbed away, subjects revert to their previous relation between moral sentiment and actual behavior.

Both the observation of price-induced MC at the same time as the absence of control-induced MC in this public goods choice situation and the line of reasoning on the valence basis of this difference are very close to the original discussions by Frey (1992, 1997) that first brought the notion of motivation crowding into the economic literature. 'Frey motivation crowding' (FMC) consists of deleterious price-based incentives for environmental public goods alongside rather benign effects of command and control on intrinsic motivation. These effects are also observed in the experiment presented here. At the same time, FMC is explicitly related back to a psychological literature that emphasizes not only cognitive, but in particular also affective drivers of human behavior (DeCharmes 1968). AMC is therefore a plausible alternative as well as complementary explanation for FMC alongside CMC.

6 Implications

The focus of our present experiment is - in our mind - novel for two reasons: The first is that it directly compares two archetypical policy instruments and finds that - contrary to the behavioral prediction - instrument choice and motivation appear to interact in a way predicted by Frey (1997). The second is that these effects are generated by an experimental design that provides only limited traction for the cognitive mechanisms developed in the theoretical literature. Affective mechanisms offer a plausible alternative explanation of the evidence.

What is the significance of these findings? We believe that there are two areas of significance. One area is the development of a comprehensive theory of MC. Given the extensive empirical literature, the need to develop a precise understanding of its nature is imperative. Cognitive theories of MC are obviously important steps towards developing such an understanding. At the same time, MC is not sufficiently well understood to be attributed to cognitive mechanisms alone: MC of the type described by Frey is found in settings in which CMC would not predict its presence. Our experimental results show that FMC may be attributable to either CMC, AMC, or both. The second area is policy. Given the incompleteness of current MC theories, it may be too early to formulate policy recommendations to tackle MC: Behavior under regulation may systematically differ from its predictions. Alternative remedies might be passed over at significant cost.

The link between instrument choice and motivation established in this paper adds an important new dimension to the problem of 'prices versus quantities'. Further research should shed more light on the nature and underlying drivers of this type of interaction.

References

Andreoni, J.: 1988, Privately provided public goods in a large economy: The limits of altruism, *Journal of Public Economics* **35**, 57–73.

Andreoni, J.: 1993, An experimental test of the public-goods crowding-out hypothesis, *American Economic Review* **83**(5), 1317–1327.

Austin, D. and Dinan, T.: 2005, Clearing the air: The costs and consequences of higher cafe standards and increased gasoline taxes, *Journal of Environmental Economics and Management* **50**, 562–582.

Baumol, W. J. and Oates, W. E.: 1988, *The theory of environmental policy*, Cambridge University Press, Cambridge, UK.

Bénabou, R. and Tirole, J.: 2003, Intrinsic and extrinsic motivation, *Review of Economic Studies* **70**, 489–520.

Bénabou, R. and Tirole, J.: 2006, Incentives and prosocial behavior, *American Economic Review* **96**(5), 1652–1678.

Bolton, G. E., Katok, E. and Zwick, R.: 1998, Dictator game giving: Rules of fairness versus acts of kindness, *International Journal of Game Theory* **27**, 269–299.

Cardenas, J. C., Stranlund, J. and Willis, C.: 2000, Local environmental control and institutional crowding-out, *World Development* **28**(10), 1719–1733.

CCC: 2008, Building a low-carbon economy - the UK's contribution to tackling climate change, *Report*, Committee on Climate Change, London.

De Dreu, C. and Steinel, W.: 2006, Social decision making in fuzzy situations: Motivated decision making and strategic choice, *in* D. De Cremer, M. Zeelenberg and J. K. Murnighan (eds), *Social Psychology and Economics*, Routledge, London, pp. 55–78.

DeCharmes, R.: 1968, *Personal Causation: The Internal Affective Determinants of Behavior*, Academic Press, New York.

Dijkstra, B. R.: 1999, *The Political Economy of Environmental Policy: A Public Choice Approach to Market Instruments*, Edward Elgar, Cheltenham, UK.

Ellingsen, T. and Johannesson, M.: 2008, Pride and prejudice: The human side of incentive theory, *American Economic Review* **98**(3), 990–1008.

Falk, A. and Kosfeld, M.: 2006, The hidden costs of control, *American Economic Review* **96**(5), 1611–1630.

Fehr, E., Klein, A. and Schmidt, K. M.: 2007, Fairness and contract design, *Econometrica* **75**(1), 121–154.

Fischer, C. and Newell, R. G.: 2008, Environmental and technology policies for climate mitigation, *Journal of Environmental Economics and Management* **55**(2), 142–162.

Fischer, P. and Huddart, S.: 2008, Optimal contracting with endogenous social norms, *American Economic Review* **98**(4), 1459–1475.

Forgas, J. P. and Laham, S. M.: 2005, The interaction between affect and motivation in social judgments and behavior, *in* J. P. Forgas, K. D. Williams and S. M. Laham (eds), *Social Motivations. Conscious and Unconscious Processes*, Cambridge University Press, Cambridge, pp. 168–193.

Frey, B. S.: 1992, Pricing and regulating affect environmental ethics, *Environmental and Resource Economics* **2**, 399–414.

Frey, B. S.: 1997, *Not Just for the Money: An Economic Theory of Personal Motivation*, Edward Elgar Publishing, Cheltenham, UK.

Frey, B. S. and Jegen, R.: 2001, Motivation crowding theory, *Journal of Economic Surveys* **15**(5), 589–611.

Frey, B. S. and Oberholzer-Gee, F.: 1997, The cost of price incentives: An empirical analysis of motivational crowding-out, *American Economic Review* **87**(4), 746–755.

Gneezy, U. and Rustichini, A.: 2000a, A fine is a price, *Journal of Legal Studies* **29**(1), 1–17.

Gneezy, U. and Rustichini, A.: 2000b, Pay enough or don't pay at all, *Quarterly Journal of Economics* **115**(2), 791–810.

Goulder, L. H. and Parry, I. W.: 2008, Instrument choice in environmental policy, *Review of Environmental Economics and Policy* **2**(2), 152–174.

Hahn, R. W.: 1990, The political economy of environmental regulation: Towards a unifying framework, *Public Choice* **65**(1), 21–47.

Harrington, W., Morgenstern, R. D. and Sterner (eds.), T.: 2004, *Choosing Environmental Policy. Comparing Instruments and Outcomes In the United States and Europe*, Resources for the Future, Washington.

Hoffman, E., McCabe, K., Shachat, K. and Smith, V.: 1994, Preferences, property rights, and anonymity in bargaining games, *Games and Economic Behavior* **7**, 346–380.

Kandel, E. and Lazear, E. P.: 1992, Peer pressure and partnership, *Journal of Political Economy* **100**(4), 801–817.

Keohane, N., Stavins, R. N. and Revesz, R. L.: 1998, The positive political economy of instrument choice in environmental policy, *in* P. Portney and R. Schwab (eds), *Environmental Economics and Public Policy: Essays in Honor of Wallace E. Oates*, Edward Elgar, Cheltenham, UK.

Kerr, S. and Newell, R. G.: 2003, Policy-induced technology adoption: Evidence from the U.S. lead phasedown, *Journal of Industrial Economics* **51**(3), 317–343.

Kotchen, M. J.: 2009, Voluntary provision of public goods for bads: A theory of environmental offsets, *Economic Journal* **forthcoming**.

Krysiak, F. C.: 2008, Prices vs. quantities: The effects on technology choice, *Journal of Public Economics* **92**(5-6), 1275–1287.

Meier, S.: 2007, Do subsidies increase charitable giving in the long run? matching donations in a field experiment, *Journal of the European Economic Association* **5**(6), 1203–1222.

Montero, J.-P.: 2002, Prices versus quantities with incomplete enforcement, *Journal of Public Economics* **85**, 435–454.

NCF: 2006, Phase I EU ETS trading decisions, *Report*, New Carbon Finance, London.

Pillutla, M. M. and Murnighan, J. K.: 1996, Unfairness, anger, and spite: Emotional rejections of ultimatum offers, *Organizational Behavior and Human Decision Processes* **68**(3), 208–224.

Sliwka, D.: 2007, Trust as a signal of a social norm and the hidden costs of incentive schemes, *American Economic Review* **97**(3), 999–1012.

Titmuss, R.: 1970, *The Gift Relationship: From Human Blood to Social Policy*, Allen and Unwin, London.

Vicary, S.: 2000, Donations to a public good in a large economy, *European Economic Review* **44**, 609–618.

Zhao, J.: 2003, Irreversible abatement investment under cost uncertainties: tradable emission permits and emission charges, *Journal of Public Economics* **87**(12), 2765–2789.