Do negative incentives encourage stealing?*

Hannah Hörisch[†‡] Christina Strassmair[§¶]
University of Munich University of Munich

April 26, 2007

Preliminary version - please do not quote

Abstract

Becker's (1968) deterrence hypothesis postulates that crime rates (weakly) decrease in "negative incentives", i.e. the severity of punishment and the detection probability. In sharp contrast, a growing empirical literature documents that small incentives often backfire by crowding out intrinsic motivation to behave in a socially desired way. We conduct a neutrally framed laboratory experiment to test whether negative incentives work. In our experiment, subjects can steal from other participants' payoffs. Different treatments vary by the severity of punishment and the detection probability. Our aggregate results clearly reject the deterrence hypothesis: except for very high levels of incentives, on average subjects steal more the higher the negative incentives. However, our results are well in line with crowding out of fairness concerns.

Keywords: Law and economics, Incentives, Experiments

JEL classification: K42, C91, D63

*We thank seminar participants at the University of Munich and the IMEBE conference in Malaga for helpful comments on earlier versions of this paper, Klaus Schmidt for his valuable support during the preparation of this paper, and Wolfgang Hering and Mara Scheib for their support during the implementation of the experiment.

[†]*Affiliation:* Seminar for Economic Theory, University of Munich, Ludwigstraße 28 (Rgb.), 80539 Munich, Germany, email: hannah.hoerisch[at]lrz.uni-muenchen.de.

[‡]Financial support from the SFB Transregio 15 is highy appreciated.

[§]*Corresponding author's address:* Seminar for Economic Theory, University of Munich, Ludwigstraße 28 (Rgb.), 80539 Munich, Germany, email: christina.strassmair[at]lrz.uni-muenchen.de.

[¶]Financial support from the ENABLE *Marie Curie Research Training Network*, funded under the $6th$ Framework Programme of the European Union, is highy appreciated.

1 Introduction

Economists firmly believe in the power of incentives. Since incentives adjust costs or benefits of an action we can induce rational, payoff maximizing individuals to exert a desired behavior by setting incentives properly. Becker's (1968) deterrence hypothesis is one specific application of that general mechanism. The deterrence hypothesis postulates that crime rates fall with the severity and the probability of punishment.

Becker's seminal paper has inspired numerous theoretical extensions such as introducing limited liability, imprisonment, accidental harms, repeat offenders, errors in law enforcement.[1] Still empirical evidence concerning the deterrence hypothesis is mixed: both severity and probability of punishment are usually, but not always found to have a significant negative impact on crime.[2] Conclusions drawn from field data are often not very reliable as they are confounded by methodological problems: due to data availability the deterrence hypothesis, originally a theory of individual behavior, is tested with aggregate data which causes simultaneity bias and necessarily suffers from omitted variables. Measurement error is widespread since not all crime is reported. Furthermore, field data usually just report the behavior of offenders and not that of the general population. In addition they only contain information on the choices that offenders actually made, but not on all options available to them. All these problems do not exist in the laboratory.

Our lab experiment directly tests Becker's deterrence hypothesis in a controlled environment that permits to exogenously vary negative incentives, i.e. detection probability and punishment. We ask a very basic but important question, namely: do negative incentives work?

By investigating the effect of negative incentives our paper contributes to the growing literature on backfiring of small incentives. For example, Fehr and Falk (2002) discuss the interaction between economic incentives and the desire for social approval. They conlcude that "giving norm violators the opportunity to free themselves from following a social norm by making them pay for the norm violation may backfire".[3] Frey and Jegen (2001) survey the mainly empirical literature on crowding out of intrinsic motivation due to the introduction of incentives. They

[1] Polinsky and Shavell (2000a) and Garoupa (1997) provide comprehensive overviews on the economic theory of optimal law enforcement.

[2] Compare Eide (2000) whose survey article focuses on empirical tests of the deterrence hypothesis.

[3] Fehr and Falk (2002), p.711.

stress that introducing incentives has two countervailing effects: besides the standard relative price effect, incentives may undermine intrinsic motivation. With small incentives the relative price effect is small and the latter, counterproductive effect may dominate. These findings on backfiring incentives are mostly based on "positive incentives", i.e. incentives that are designed to encourage socially desirable behavior. In contrast, our experiment investigates the effects of negative incentives.[4] Furthermore, to the best of our knowledge we are the first to test the existence of crowding out of intrinsic motivation in the context of criminal behavior such as stealing. Thus, our findings will hopefully add new insights to the ongoing discussion under which circumstances incentives work as predicted by standard theory or backfire.

To test the deterrence hypothesis in the lab we have chosen one of the simplest possible designs: two players, A and B, are randomly matched. Player A is a passive player. Player B can decide how much to take away (steal) from player A's initial endowment. With probability $1 - p$, player B's theft is undetected and his chosen amount is transferred from player A to player B. With probability p, however, player B's theft is detected and player B receives his initial endowment minus a fix fine f.[5]

We conduct six different treatments in which we vary p and f. Our benchmark treatment T1 sets $p = f = 0$. Treatments T2, T3, and T4 investigate the range of small negative incentives, i.e. levels of incentives such that the expected payoff of stealing everything is larger than the one of stealing nothing. Treatment T5 is characterized by a combination of p and f such that stealing everything generates about the same expected payoff as stealing nothing. Finally, in treatment T6 the expected payoff of stealing everything is remarkably smaller than the one of stealing nothing. Each subject participates in two different treatments sequentially. This design allows us to study stealing behavior from different perspectives. The static perspective analyses subjects' behavior across different treatments which are played first. The dynamic perspective analyses how individual behavior evolves from the treatment played first to the treatment played second. Both perspectives deliver valuable insights: the static perspective compares different regimes, and the dynamic perspective studies the effect of a regime change. Furthermore, we use data of a questionnaire filled out at the end of the experiment in order to control for individual characteristics.

[4] Other lab or field experiments that focus on the effect of negative incentives are Gneezy (2003), Gneezy and Rustichini (2000), and Bohnet, Frey, and Huck (2001).

[5] In case player B does not intend to take away anything, he never has to pay the fixed fine f.

Both our static as well as our dynamic results clearly reject the prediction derived from the deterrence hypothesis that the transfer amount should be monotonically (weakly) decreasing in p and f. In contrast, we find that small incentives backfire: on average subjects transfer more in the treatments with small negative incentives than in the absence of negative incentives. Only very large incentives lead to substantial levels of deterrence. We show that our data are compatible with a model of two types: selfish types who do react to negative incentives as predicted by the deterrence hypothesis and fair types whose fairness concerns are gradually crowded out by negative incentives. Our results from the benchmark treatment T1 show that about 50 % of subjects are selfish, while the other 50 % have fairness concerns.

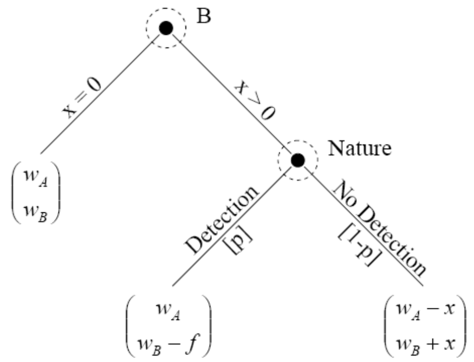
Laboratory experiments on criminal behavior are scarce. Falk and Fischbacher (2002) explore the influence of social interaction phenomena on commiting a crime. Bohnet and Cooter (2005), Tyran and Feld (2006), and Galbiati and Vertrova (2005) investigate whether law can act as "expressive law", i.e. prevent crime by activating norms that prohibit commiting a crime although it pays off. Whether law can serve as a coordination device for equilibrium selection is a further topic adressed in Bohnet and Cooter (2005). Tyran and Feld (2006) also compare the effects of exogenously imposed and endogenously chosen small negative incentives. Except for Tyran and Feld (2006) none of the existing experiments focuses on deterrence - it is simply assumed that deterrent incentives work. Our experiment aims at closing that gap.

The paper proceeds as follows. Section 2 presents the model, section 3 the experimental design and the hypotheses to be tested. The static and dynamic experimental results are summarized and discussed in section 4. In section 5 we then check whether different framing of actions has a significant impact on our results. Section 5 conlcudes.

2 The Model

We consider the simplest possible model of stealing with two agents. The set of agents corresponds to $N = \{A, B\}$. Agent A is intially endowed with w_A, and agent B is initially endowed with w_B, $w_A > w_B$. While agent A is a passive player, agent B can steal any amount $x \in [0, w_A]$ from agent A's endowment. If $x = 0$, i.e. nothing is stolen, agents A and B both receive their initial endowments w_A and w_B. If $x > 0$, with probability $(1 - p) \in [0, 1]$ B's theft is not detected and x is transferred from A to B; with probability p, however, B's theft is detected, x is not transferred from A to B and, on top of that, agent B has to pay a fixed fine f that is independent of the amount $x > 0$. The structure of the model is summarized in Figure 1.

Figure 1. Structure of the model of stealing



Player B's optimal decision will depend on the specific form of his utility function:

Assumption A: Player B has standard preferences.

Player B is completely selfish and, therefore, either steals as much as possible, i.e. w_A, or nothing. This depends on the relative size of p, f and w_A as well as on his risk attitude. For a risk neutral player the optimal stolen amount x^* is

$$x^* = \begin{cases} 0 & if \quad p > \frac{w_A}{w_A+f} \\ \in [0, w_A] & if \quad p = \frac{w_A}{w_A+f} \\ w_A & if \quad p < \frac{w_A}{w_A+f} \end{cases}.$$

The higher p or the higher f, the less attractive is the option to steal everything compared to $x = 0$. For sufficiently high values of p and f, a risk neutral player B does not steal anything. Consequently, the stolen amount is weakly decreasing in p and in f. The presence of risk aversion reduces the attractiveness of $x = w_A$ for $p > 0$ and $f > 0$. Hence, the set of p, f, w_A combinations for which stealing everything is optimal reduces. In contrast, the presence of risk affection enlarges this set. Independent of player B's risk attitude, the deterrence hypothesis holds, namely the stolen amount is monotonically (weakly) decreasing in p and in f.

Assumption B: Player B has social preferences.

Any outcome based social preferences model would deliver results similar to those presented below. We choose the model with inequity averse agents by Fehr and Schmidt (1999) due to its widespread application. If player B is inequity averse, he might steal an amount in the interior of his strategy set. This depends crucially on his advantageous inequity aversion, captured by parameter β. If $\beta > \frac{1}{2}$, he either steals $\frac{w_A - w_B}{2}$ or nothing:

$$x^* = \begin{cases} 0 & if \quad p > \frac{(w_A - w_B)*\left(\frac{1}{2}+\alpha\right)}{(w_A - w_B)*\left(\frac{1}{2}+\alpha\right)+f*(1+\alpha)} \\ \in [0, \frac{w_A - w_B}{2}] & if \quad p = \frac{(w_A - w_B)*\left(\frac{1}{2}+\alpha\right)}{(w_A - w_B)*\left(\frac{1}{2}+\alpha\right)+f*(1+\alpha)} \\ \frac{w_A - w_B}{2} & if \quad p < \frac{(w_A - w_B)*\left(\frac{1}{2}+\alpha\right)}{(w_A - w_B)*\left(\frac{1}{2}+\alpha\right)+f*(1+\alpha)} \end{cases},$$

with α capturing disadvantageous inequity aversion. If $0 < \beta < \frac{1}{2}$, player B either steals as much as possible, i.e. w_A, or nothing:

$$x^* = \begin{cases} 0 & if \quad p > \frac{w_A*(1-\beta)-w_B*\beta+\alpha*(w_A-w_B)}{w_A*(1-\beta)-w_B*\beta+\alpha*(w_A-w_B)+f*(1+\alpha)} \\ \in [0, w_A] & if \quad p = \frac{w_A*(1-\beta)-w_B*\beta+\alpha*(w_A-w_B)}{w_A*(1-\beta)-w_B*\beta+\alpha*(w_A-w_B)+f*(1+\alpha)} \\ w_A & if \quad p < \frac{w_A*(1-\beta)-w_B*\beta+\alpha*(w_A-w_B)}{w_A*(1-\beta)-w_B*\beta+\alpha*(w_A-w_B)+f*(1+\alpha)} \end{cases}.$$ [6]

Again, risk aversion diminishes the set of p, f, w_A combinations for which stealing a strictly positive amount is optimal, whereas risk affection enlarges this set. As with assumption A, p and f reduce the attractiveness of stealing and hence the deterrence hypothesis holds.

Assumption C: Player B has social preferences that are crowded out by the presence of incentives.

There exists a vast literature on crowding out of intrinsic motivation through the presence of incentives. In this literature, the term intrinsic motivation is defined in a relatively broad way and may well apply to fairness concerns. Up to now, crowding out of intrinsic motivation has not been formalised in a rigorous way.[7]

[6] Interestingly, the set of p, f, w_A combinations for which stealing a strictly positive amount is optimal is strictly smaller under assumption B than assumption A.This is true, since

$$\max\left\{\frac{(w_A - w_B)*\left(\frac{1}{2}+\alpha\right)}{(w_A - w_B)*\left(\frac{1}{2}+\alpha\right)+f*(1+\alpha)}, \frac{w_A*(1-\beta)-w_B*\beta+\alpha*(w_A-w_B)}{w_A*(1-\beta)-w_B*\beta+\alpha*(w_A-w_B)+f*(1+\alpha)}\right\} < \frac{w_A}{w_A+f},$$

for $\beta > 0$.

[7] We will not do that here either.

Verbally, the presence of incentives induces individuals to be more selfish, and to be less intrinsically motivated to behave fairly. We will capture this process in the following way:

$$U = \lambda(p, f) * s + [1 - \lambda(p, f)] * g,$$

where s is the individual payoff, and g is the utility of an inequity averse agent. The individual utility function, U, is simply a linear combination of s and g. The core of the crowding out assumption is that $\lambda(p, f)$, the weight of s, (weakly) increases in p and in f.

$\lambda(p, f)$ may evolve in quite different ways: the empirical results of Gneezy and Rustichini (2000a) and Gneezy (2003) suggest that the introduction of incentives causes a discontinous jump in behavior as captured by

$$\lambda(p, f) = \begin{cases} 0 & if \quad p = 0 \\ 1 & if \quad p > 0 \;\&\; f > 0 \end{cases}.$$

According to Frey and Oberholzer-Gee (1997), however, $\lambda(p, f)$ may also evolve smoothly. The effect of crowding out also depends on the functional form of g. This can be nicely shown by comparing two different examples. First, assume g as described in Fehr and Schmidt (1999). Here, if $[1 - \lambda(p, f)] * \beta > \frac{1}{2}$ (e.g. $\beta > \frac{1}{2}$ and $\lambda(0, 0) = 0$), an agent sets $x = \frac{w_A - w_B}{2}$. However, if $[1 - \lambda(p, f)] * \beta < \frac{1}{2}$ (e.g. $\beta > \frac{1}{2}$ and $\lambda(p > 0, f > 0)$ sufficiently positive), an agent will steal everything. Hence, an agent with constant α and β may discontinously increase the stolen amount x in the intensity of negative incentives.[8] Second, let us assume g to be quadratic in the payoff difference between the two agents. In this case, an agent may smoothly increase his stolen amount in the intensity of negative incentives.[9]

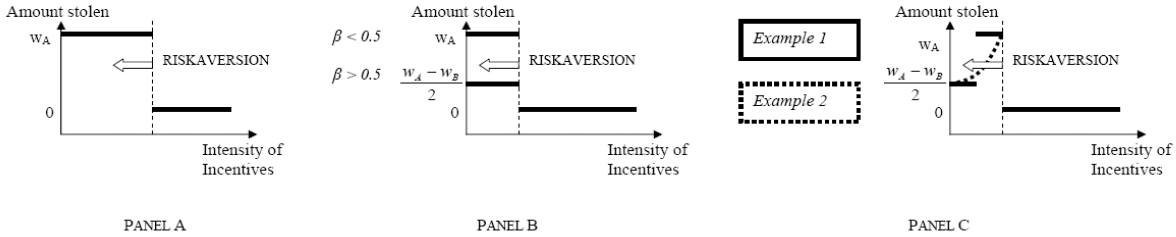
Independent of the specific functional form of $\lambda(p, f)$ and g, for individuals with U as described above we would predict that player B's stolen amount may be increasing in the detection probability p and the fine f. This stands in sharp contrast to the deterrence hypothesis. However, for very high values of p and f, it is optimal to set $x = 0$, irrespective of whether the agent is selfish or fair-minded. For this range of values, negative incentives deter people from stealing for any $\lambda(p, f)$. The range of these values depends on the specific risk attitude. Risk aversion enlarges this range, whereas risk affection reduces it.

The possible relations between the amount stolen and the intensity of negative incentives are summarised in Figure 2.

[8] This holds, even if $\lambda(p, f)$ evolves smoothly.

[9] This only holds for a smooth evolution of $\lambda(p, f)$.

Figure 2. Relations between the stolen amount and the intensity of incentives



3 Experimental design and hypotheses

3.1 Experimental procedure and treatments

Across the different treatments we vary the detection probability p and the fixed fine f, while holding w_A and w_B constant at levels 90 and 50, respectively. Table 1 presents the treatments.

Table 1: Treatments

Treatment	p	f	Expected profit of x=90
T1	0.0	0	140
T2	0.6	6	82.4
T3	0.5	25	82.5
T4	0.6	20	74
T5	0.7	40	49
T6	0.8	40	36

In the benchmark treatment T1 no negative incentives are implemented. Hence, it is just the mirror image of a dictator game.

In treatments T2 and T3, stealing everything yields virtually the same expected payoffs. Hence, the intensity of negative incentives in these two treatments is rather the same, though achieved by different levels of p and f. This offers us the possibility to test whether p and f are interchangeable instruments. For treatments T4 to T6, we constantly increase the intensity of negative incentives, i.e. the expected payoff of stealing everything decreases.[10]

[10]Moreover, one can compare the expected payoffs from stealing for different levels of $x > 0$: The expected payoff from stealing is strictly lower in treatment T6 than in treatment T5 and in

Each experimental session consisted of three independent parts: first a stealing decision in one treatment, second a stealing decision in another treatment, and third a dictator game. After these three parts, participants had to fill out a questionnaire eliciting data on their age, sex, subject of studies and risk aversion. In order to get an approximation of subjects' risk aversion, the questionnaire included a Holt and Laury (2002) table[11] that was paid.

At the beginning of each session, participants were told that the session will consist of three independent parts, out of which only one randomly chosen part would be paid out for all participants. After each single part, only the instructions for the following part were handed out. Subjects were not given any feedback before the end of the experiment. In part 3 the dictator could donate any amount of his initial endowment of 90 to a randomly matched passive player with an initial endowment of 50. The behavior of the dictator indicates his advantageous inequity aversion parameter β. The conducted sessions are presented in Table 2.

Table 2: Session Plan

Treatment	Part 1	Part 2	Part 3	Questionnaire	Number of participants
T1T3	T1	T3	DG	Yes	38
T3T1	T3	T1	DG	Yes	38
T2T3	T2	T3	DG	Yes	18
T3T2	T3	T2	DG	Yes	20
T2T4	T2	T4	DG	Yes	38
T4T2	T4	T2	DG	Yes	36
T5T6	T5	T6	DG	Yes	32
T6T5	T6	T5	DG	Yes	38

DG: dictator game

The matching of the participants was perfect stranger.[12] Players B of part 1 remained players B in part 2 and were players A (the dictators) in part 3. Therefore, the passive players, named player A in part 1 and part 2 and player B in part 3, remained passive throughout all the three parts of the session.

treatment T5 than in treatment T4, for any $x > 0$. For any $x > 5$, the expected payoff from stealing is strictly lower in treatment T4 than in treatment T3. The expected payoff from stealing is strictly lower in treatment T3 than in treatment T2, for any $x < 89$.

[11] The translated table can be found in the appendix.

[12] A matched couple is never matched again in the following parts.

Organizing the sessions in this way enables us to analyze the stealing behavior in the following two ways: First, we can compare stealing behavior in part 1 across the different treatments. This is the cleanest comparison, since individual's behavior in part 1 is not influenced by any preplay. Second, we can analyse how individuals adapt their behavior to the parameter change from part 1 to part 2. Since the structure of the stealing model is relatively simple and straight forward, we assume that a change in behavior from part 1 to part 2 is stimulated by the change in parameters rather than learning.

Our experimental sessions were run in November 2006 and March 2007 at the experimental laboratory of the SFB 504 in Mannheim. 258 students of the Universities of Mannheim and Heidelberg participated in the experiment. Subjects were randomly assigned to sessions and could only take part once. The sessions were framed neutrally[13] and lasted on average about 40 minutes. Subjects were not paid any show-up fee[14] and earned 12.34 € on average.

3.2 Hypotheses

Becker's (1968) deterrence hypothesis is supported by the assumption of standard preferences (assumption A) as well as by the assumption of Fehr and Schmidt (1999) preferences (assumption B). Therefore, our analysis focuses on the following hypothesis.

Hypothesis 1:

The stolen amount is monotonically (weakly) decreasing in the detection probability p and the fixed fine f.

More specifically, risk neutral subjects with standard preferences steal everything in treatments T1 to T4 and nothing in treatments T5 and T6. Risk neutral subjects with sufficiently strong Fehr and Schmidt (1999) preferences, i.e. $\beta > \frac{1}{2}$ and $\alpha > 1$, steal 20 in treatments T1 to T4 and nothing in treatments T5 and T6.

In contrast, assumption C postulates the following hypothesis.

[13] Translated instructions for player B can be found in the appendix.

[14] In case a subject did not earn anything in the randomly selected part and in the Holt and Laury (2002) table, nothing was paid out. This happened in six cases.

Hypothesis 2:

Negative incentives crowd out fairness concerns. Therefore, the stolen amount is (weakly) increasing in the detection probability p and the fixed fine f if p and f are in the range of rather small values. This range is larger, the less risk averse an agent is.

What all different assumptions have in common is the following.

Hypothesis 3:

Very high values of the detection probability p and the fixed fine f deter individuals from stealing. The range of these values is larger, the stronger individual risk aversion.

More specifically, risk neutral subjects of type A, B or C do not steal anything in treatments T5 and T6. Slightly risk loving subjects may steal a strictly positive amount in treatment T5.

4 Results

4.1 Static results - Behavior in part 1

4.1.1 Summary Statistics

Benchmark treatment

Our experimental data in treatment T1 show how much people steal in the absence of negative incentives. Figure 4.'s first graph summarizes the distribution of x in the benchmark treatment.

As already noted previously, treatment T1 is the mirror image of a dictator game. Therefore, we can compare the behavior observed in T1 (i) with standard results of dictator games and (ii) with behavior in part 3. For step (i) we use experimental data from Forsythe et al. (1994). In line with their paper, we can identify two types of players: selfish players and fair-minded players. In their benchmark treatment[15], about 45 % of the participants are "pure gamesmen" who donate nothing, and the rest donates a strictly positive amount.[16] These types of players correspond remarkably well to the 47 % (52.5 %) selfish participants in treatment T1 who steal

[15] Here, we use their results of the paid dictator game conducted in April with a pie of 5 $.

[16] Forsythe et al. (1994) split the group of fair-minded people into those, who (more or less) equate payoffs, and those who prefer to have a premium compared to the other player. The shares

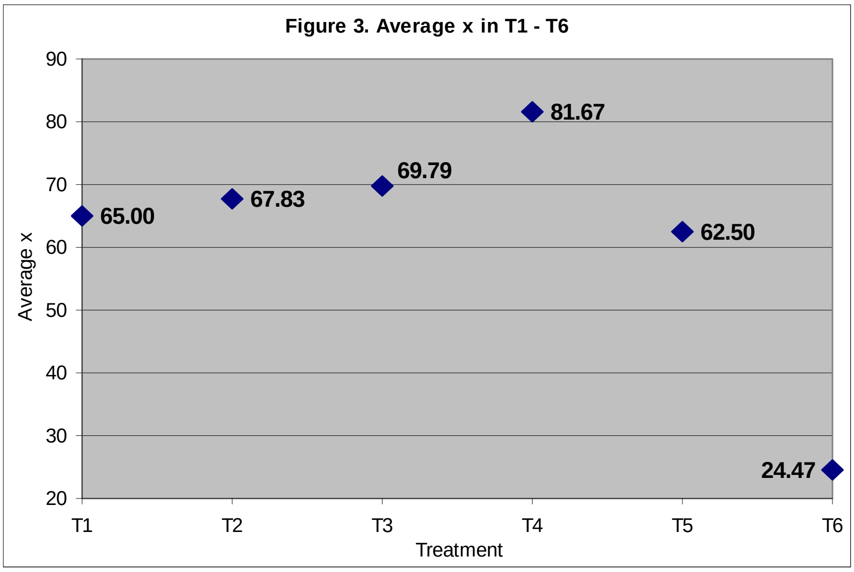
everything (between 80 and 90 units) while the others steal a strictly positive amount below 90 (80). Hence, the experimental results of our benchmark treatment are well in line with standard results from previously conducted studies.

For step (ii) we calculate the correlation coefficient between the stolen amount in T1 (in part 1) and the donated amount in the dictator game of part 3. As treatment T1 is the mirror image of part 3 we expect a highly negative correlation coefficient. This is indeed the case: Pearsons' correlation coefficient is -0.747 which is significantly different from 0 at conventional significance levels (p-value < 0.01). Hence, subjects who played T1 in part 1 behave consistently in part 1 and in part 3.

To summarize, we have found that slightly less than 50% of subjects have standard preferences as in assumption A, while a bit more than 50% have social preferences. To be able to differentiate whether or not fairness concerns are crowded out by the introduction or an increase of negative incentives, i.e. whether assumption B or C characterizes the subjects with social preferences, we must have a closer look at the other treatments.

Treatments T2 to T6

Figure 3. summarizes the average stolen amount per treatment.



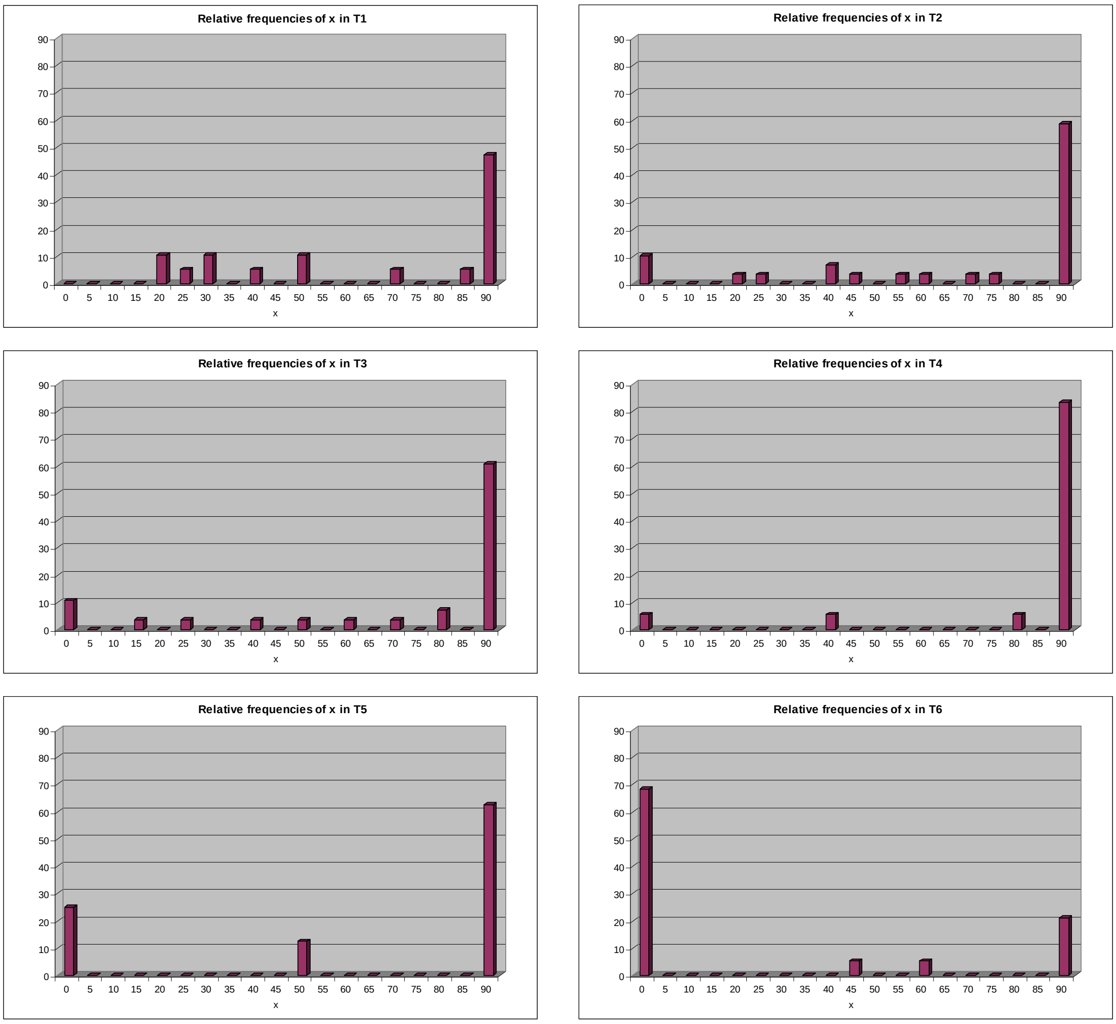
The bold numbers next to the data points represent the exact average amounts stolen. Treatments are arranged by the intensity of negative incentives. The average

of these two groups in Forsythe et al. (1994) fit remarkable well to the fraction of players who steal around 20 points in T1, and the fraction of players who steal something between 30 and 85 points in T1, respectively.

stolen amount increases in the range of small incentives (from T1 to T4), while it decreases in the range of relatively high incentives (T5 and T6). The relationship between the average stolen amount and the intensity of negatives incentives is rather inverted-U shaped than monotonically decreasing as assumption A and B predict.

Figure 4. provides an overview on how the distributions of the stolen amount (displayed in five-steps invervals) vary by treatment.

Figure 4. Distributions of the stolen amount per treatment



The fraction of individuals stealing everything increases by treatment from T1 to T4. In treatment T4, this fraction peaks at more than 80 % which is considerably higher than the corresponding 47 % in the absence of incentives in treatment T1. From treatment T5 onwards, this fraction decreases. The share of subjects who do not steal anything increases over all treatments. This share is moderate in treatments T2, T3 and T4 ($\leq$ 10 %), quite substantial in treatment T5 (about 25 %), and largest in treatment T6. Nearly 70 % of individuals are deterred in treatment

T6. Interestingly, there are always subjects stealing intermediate levels, most so in the benchmark treatment. However, the share of subjects stealing intermediate amounts decreases in the intensity of incentives. Moreover, while these subjects steal amounts centred around lower levels in treatment T1, they tend to steal higher intermediate amounts in the presence of higher negative incentives. This is evidence in favor of gradual as opposed to discontinous crowding out.[17] [18]

In sum, higher negative incentives shift mass to the borders of the support. For relatively small levels of the detection probability p and the fixed fine f (T1 to T4), mass predominately moves towards the upper border. With relatively high negative incentives, mass shifts towards the lower border. The latter effect can be explained by the deterrence hypothesis. The first effect, however, stands in sharp contrast to that.

4.1.2 Analysis of hypotheses

To ensure that the variation of the detection probability p and the fixed fine f are sufficiently high to stimulate significant incentive effects in behavior we first test whether the behavior in any treatment differs significantly from the behavior in any other treatment. The Kruskal-Wallis test confirms that behavior differs significantly across treatments ($p < 0.01$).

As a next step, we make pairwise treatment comparisons, i.e. we test which treatments are significantly different from each other. Table 3 displays the two sided p-values of Mann-Whitney-U tests.

Treatment T6 is significantly different ($p < 0.01$) from any other other treatment since a lot of people are deterred and steal nothing. Furthermore, behavior in treament T4 is different from the behavior in the other treatments. The (one sided) hypothesis that people steal less in T4 than in T1, T2, and T3 can be rejected ($p = 0.015$, 0.041, and 0.058, respectively).

In order to control for individual characteristics when comparing the different treatments, we run the Tobit regression presented in Table 4.

[17] This aggregate pattern would be also compatible with individuals whose behavior is driven by fairness concerns and by the notion that "stealing has to pay". Such individual would then steal $x = 9$ in T2, $x = 25$ in T3, and $x = 30$ in T4, given that they are risk neutral. However, we do not observe any risk neutral subjects stealing such amounts.

[18] Inequity aversion on averages, rather than on outcomes, would also predict an increase of the average stolen amount from treatment T1 to treatment T4. Individuals equating average payoffs would steal $x = 54.5$ in T2, $x = 52.5$ in T3, and $x = 65$ in T4. However, we observe only very few individuals stealing such amounts.

Table 3: Pairwise comparisons of all treatments (Mann-Whitney-U tests)

	T2	T3	T4	T5	T6
T1	0.573	0.468	0.030	0.800	0.001
T2	-	0.815	0.081	0.893	0.0004
T3	-	-	0.115	0.780	0.003
T4	-	-	-	0.141	0.0001
T5	-	-	-	-	0.009

Table 4: Parametric comparison of all treatments (Tobit regression)

Stolen amount	Coefficient	P-value
Intercept	70.65	0.436
Sex (1 if male, 0 else)	29.24	0.193
Age	-1.37	0.693
Economist (1 if economist, 0 else)	28.20	0.217
Risk attitude (0 if risk averse, 1 else)	57.42	0.014
DG (donated amount in part 3)	-0.52	0.513
T2	32.81	0.320
T3	27.63	0.396
T4	90.11	0.039
T5	-21.69	0.562
T6	-135.70	0.001

Number of observations: 129
Tx: 1 in treatment Tx, 0 else

The significantly positive coefficient of treatment T4 confirms that on average subjects with same characteristics steal more in treatment T4 than in treatment T1. The strong deterrence effect in treatment T6 is reflected by the significant negative coefficient. Moreover, on average risk averse subjects steal significantly less.[19] Table 4 at hand, we can reject the hypothesis that the coefficient of T4 is smaller or equal than 0 (p-value = 0.019). We cannot reject the hypotheses that the (positive) coefficients of T2 and T3 are smaller or equal to 0 (p = 0.160 and 0.198, respectively).

[19] Results change little when interaction terms of risk attitude with treatment dummies are inserted in the tobit specification. Coefficients and p-values of treatments T4 and T6 remain qualitatively the same.

So far, our findings can be summarized in the following results:

Result 1:

Based on the results presented in Figure 3, Table 3 and Table 4, we reject hypothesis 1, the deterrence hypothesis, that predicts the average stolen amount to be monotonically decreasing in the intensity of negative incentives. Non-parametric comparisons as well as the tobit regression coefficients reveal that the average stolen amount increases significantly in the range of small values of the detection probability p and the fixed fine f (T1 to T4).

Result 2:

Figure 4. indicates that in the range of relatively small detection probabilities p and fines f (T1 to T4), implementing or incresing negative incentives shifts mass from interior values of x to rather high values of x. Hence, the increase of the average amount stolen in the intensity of negative incentives is triggered by the subjects stealing intermediate amounts when no negative incentives are present. These facts are well in line with hypothesis 2, namely that negative incentives crowd out fairness concerns.

Result 3:

Within the range of rather high negative incentives (T5 and T6), individuals are deterred. Still, Figure 4. shows that a non-negligible fraction of subjects steals a strictly positive amount. These subjects, however, are significantly (Mann-Whitney-U test, $p < 0.05$) less risk averse than their counterparts in the same treatment. For these subjects the intensity of incentives is too low to deter them from stealing. Furthermore, we know from Table 4 that risk neutral or loving subjects steal on average significantly more than risk averse ones. These facts are well in line with hypothesis 3.

From results 1 to 3, we conclude that crowding out of fairness concerns is the driving force behind our results in the treatments with small negative incentives. In contrast to Becker's (1968) deterrence hypothesis, small incentives seem to increase the average stolen amount.

4.2 Dynamic results - Behavior in part 1 and part 2

Our static results show that *across different treatments* negative incentives may not deter subjects but rather "provoke" them to steal more within the range of rather small negative incentives. This analysis, however, does not shed light on the question, how *the same individuals* react to a change of negative incentives. Our experimental design enables us to have a closer look on this question as well. When crowding out of fairness concerns is an issue, as documented by our static results, we should find this phenomenon in our dynamic analysis as well.

The existing literature on crowding out of intrinsic motivation has mainly focused on the effects of introducing or increasing incentives.[20] Thus, it is an open question whether the removal of incentives gives rise to crowding in of intrinsic motivation, or whether intrinsic motivation remains crowded out once crowding out has taken place. The following empirical analysis may shed light on that issue.

If fairness concerns were crowded out on a long term basis fairness concerns that have been crowded out in part 1 remain crowded out in the second part independent of the intensity of negative incentives in part 2.[21] Thus, we should observe sequence effects.

4.2.1 Identification of sequence effects

Table 5 displays the results of Mann-Whitney-U tests that test whether there is a significance difference in behavior when a treatment is played in part 2 instead of part 1.[22]

At a significance level of 10 %, we identify significant sequence effects in treatments T1, T2, T4 and T5. While the presence of sequence effects could be caused by long lasting crowding out, it could also be due to anchoring. In order to have a more precise idea what exactly drives the dynamics we have a closer look at within subjects adaptations.

4.2.2 Evolution of individual behavior from part 1 to part 2

Introduction or increase of incentives from part 1 to part 2

[20] An exception is Gneezy and Rustichini (2002) who first introduce and, after some period of time, removed incentives again. Their findings are compatible with long lasting crowding out of fairness concerns.

[21] Crowding out may even influence the behavior in the dictator game in part 3. In contrast, the behavior in the Holt and Laury (2002) table cannot be influenced by long term crowding out of fairness concerns, since these decisions only influence the own payoff.

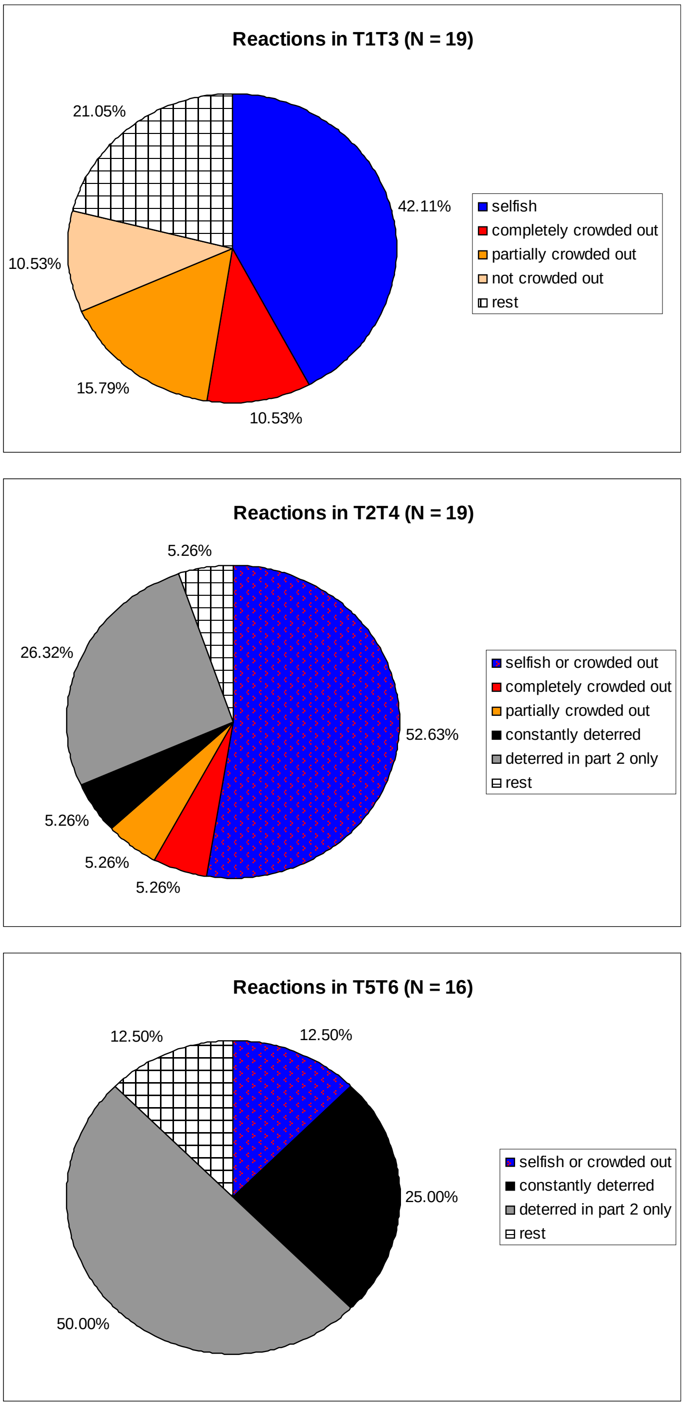
[22] Treatment T2 and T3 are played second in two different sessions. The observations from these treatments when played secondly can be pooled according to Mann-Whitney-U tests (p=0.71 and p=0.34 respectively).

Table 5: Non-parametric comparisons of different sequences (Mann-Whitney-U test)

Treatment	played first in	played second in	p-value (two sided)
T1	T1T3	T3T1	0.082
T2	T2T3 T2T4	T3T2 T4T2	0.099
T3	T3T1 T3T2	T1T3 T2T3	0.676
T4	T4T2	T2T4	0.061
T5	T5T6	T6T5	0.014
T6	T6T5	T5T6	0.617

In three different sessions we increased the intensity of incentives: first, from no incentives (T1) to small incentives (T3), second, from small incentives (T2) to intermediate incentives (T4), and third, from large incentives (T5) to even larger incentives (T6). Figure 5. summarizes how individuals react to these three parameter changes.

Figure 5. Reactions to an increase of incentives



In session T1T3, we can perfectly distinguish whether an individual is selfish or fair since treatment T1 is played in part 1. Again, we identify about 45 % selfish people who steal everything in both treatments. In part 1, about 35 % are somehow fair-minded and steal intermediate amounts. In part 2, these fairness concerns are completely crowded out in a third of these cases, i.e. the same subjects steal everything, and are partially crowded out in another third of these cases, i.e.

the same subjects moderately increase their stolen amounts from part 1 to part 2. For the last third of fair-minded subjects, the change in the intensity of negative incentives may be too small to trigger a reaction in behaviour. As in the following sequence comparisons, the checkered slices stand for the individuals whose behavior can neither be explained by crowding out of fairness concerns nor the deterrence hypothesis: they decrease their stolen amount to a level higher than zero when confronted with an increase in incentives.[23]

In part 1 of the session T2T4, some fairness concerns may have already been crowded out. That could be the reason why the fraction of subjects stealing constantly everything is larger than in session T1T3.[24] Still, some subjects exhibit fairness concerns in part 1. These are partially or completely crowded out in part 2. Interestingly, no fair-minded subject keeps his decision constant. Since the intensity of negative incentives already amounts to an intermediate level in part 2, we can observe some deterrence. These individuals are especially risk averse.

Hypothesis 3 postulates that we should see a high fraction of deterrence in session T5T6. Moreover, the fraction of subjects behaving fairly should be rather small, first, due to crowding out, and second, due to the high risk. The third graph of Figure 5. shows that this is indeed the case. The 12.5 % of subjects stealing in at least one of the two parts are significantly less risk averse.

As in the static results, we can identify crowding out of fairness concerns. In our dynamic setting, a substantial number of fair-minded individuals steals more when negative incentives are introduced or increased in part 2. The larger incentives are in part 1 the smaller this fraction is. This is very plausible since the fraction of selfish or already crowded out people is larger in part 1 if incentives are higher.

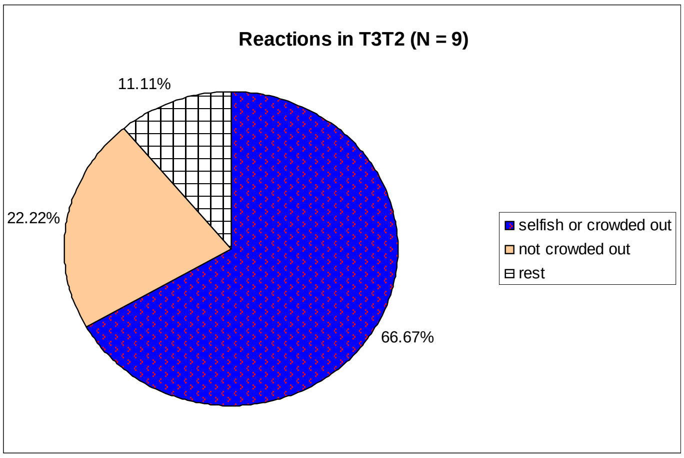
[23]From treatment T1 to treatment T2, two individuals (contained in the rest) switch to stealing nothing. This cannot be explained by deterrence, since their estimated risk aversion is relatively small.

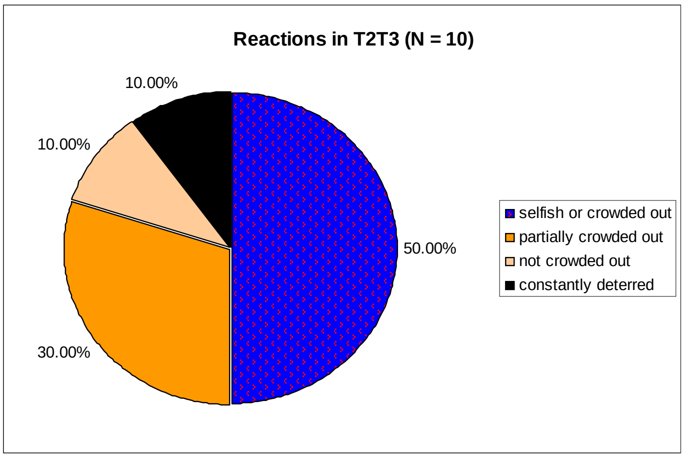
[24]More precisly, it is slightly less than the sum of the fraction of the selfish from session T1T3 plus the fraction of the completely crowded out from session T1T3.

Constant incentives with different detection probability and fine in part 1 and 2

The intensity of negative incentives in treatment T2 and T3 is very similar. Therefore, hypothesis 1 as well as hypothesis 2 predict constant behavior. Figure 6. summarizes how behavior actually changed from part 1 to part 2 in sessions T2T3 and T3T2.

Figure 6. Reactions to constant incentives





In session T3T2, about 90 % of subjects do not change their behavior: about 65 % keep stealing everything and about 20 % keep stealing the same intermediate amount.

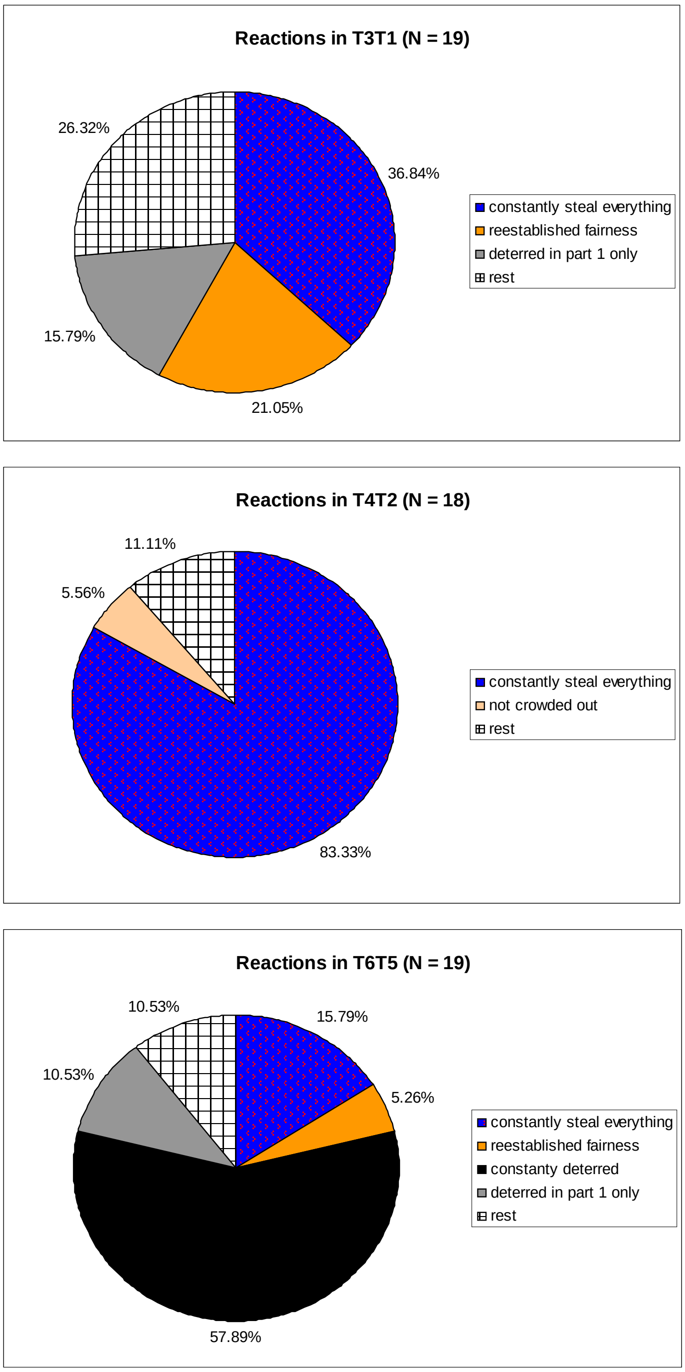
In session T2T3, the majority of subjects does not change their behaviour: 50 % keep stealing everything, 10 % keep stealing nothing, and another 10 % keep stealing the same intermediate amount. Nearly a third, however, increase their stolen amount. These subjects steal such low amounts in part 1 that a parameter change from treatment T2 to T3 corresponds to an increase of the intensity of negative incentives. Therefore, we classify these subjects as "crowded out".

Due to the small number of observations (9 in sequence T3T2, and 10 in sequence in T2T3) our results for constant incentives should be interpreted very carefully. Still, it is remarkable that in both sessions the majority of subjects does not alter their behaviour. Furthermore, the 30 % of subjects increasing their stolen amount in session T2T3 can be explained by crowding out of fairness concerns.

Removal or decrease of incentives from part 1 to part 2

As mentioned before, it is not clear whether fairness concerns - when crowded out in part 1 - can be reestablished through a removal or decrease of incentives in part 2. In three different sessions, we decrease the intensity of incentives: first, from small incentives (T3) to no incentives (T1), second, from intermediate incentives (T4) to small incentives (T2), and third, from largest incentives (T6) to large incentives (T5). Figure 7. summarizes the reactions of individuals in sessions T3T1, T4T2, and T6T5.

Figure 7. Reactions to a decrease of incentives



In session T3T1, the fraction of people constantly stealing everything is a little bit smaller than in session T1T3, but still within the same dimension. About 20 % of subjects decrease their stolen amount from part 2 to part 1 which can interpreted as reestablished fair behavior. Thus, after a period of small incentives fairness concerns seem to be crowded out lastingly. Further 15 % of subjects flip from stealing nothing to stealing a strictly positive amount. Since these subjects have a relatively high degree of risk aversion, they were probably deterred in part 1.

In session T4T2, the picture is quite different. More than 80 % keep stealing the same amount, all except one steal everything. This may indicate that fairness concerns remain crowded out, at least when switching from intermediate incentives to small incentives.[25]

In session T6T5, the majority of subjects do not steal anything as the deterrence hypothesis and crowding out of fairness concerns suggest. Similar to session T5T6, about 15 % of subjects constantly steal everything.[26] One individual decreases his positive amount stolen from part 1 to part 2. We interprete this behavior as returned fairness. Two subjects were deterred in part 1 and steal everything in part 2.

The presence of relatively small incentives does not seem to crowd out fairness concerns lastingly when incentives are completely removed in the next part. However, when intermediate incentives were set and are only reduced to a smaller level, fairness concerns seem to be crowded out lastingly. In session T6T5, we observe a lot of deterred, in particular constantly deterred, subjects.

The following results summarize our findings on within subject behavior:

Result 4:

When the intensity of negative incentives is increased but remains in the range of a relatively small detection probabilty p and fine f, a substantial fraction of fair-minded individuals behaves more selfishly. This is well in line with hypothesis 2 and reinforces our static results.

Result 5:

When the intensity of negative incentives is decreased within the range of a relatively small detection probabilty p and fine f it depends on the initial level of incentives and the intensity of the decrease whether crowded out fairness concerns are reestablished.

Result 6:

Within the range of relatively large incentives, a lot of subjects are deterred as hypothesis 3 suggests.

Our results on dynamic behaviour are well in line with crowding out of fairness concerns for fair types which strengthens our static results. Of course, some single

[25]Furthermoe, this could explain why we observe a significant sequence effect for treatment T2 which is either played after harsher or constant incentives, but not for treatment T3 which is played after T2 or T1.

[26]In treatment T6, negative incentives are that intense that the majority of subjects does not steal anything. This is why we do not observe any significant sequence effect for treatment T6.

subjects whose individual behavior from part 1 to part 2 cannot be explained are found in every session.

5 Extension - Framing of stealing

Up to now, we have analyzed data obtained in neutrally framed experiments to focus on the pure and isolated incentive effect. Policy makers often have an additional tool to setting negative incentives, namely framing of behavior. Labelling a transfer decision with $x > 0$ as "stealing" and talking about the fixed fine f as a "penalty" instead of minus points may affect the decision to steal.[27] Furthermore, framing of behavior may not only affect the decision to steal in the absence of incentives (in treatment T1), but might also interact with negative incentives. Thus we run another two sessions with framing, one without and one with negative incentives (see Table 6).

Table 6: Framed sessions

Treatment	Part 1 (framed)	Part 2 (framed)	Part 3 (not framed)	Questionnaire (not framed)	Number of participants
T1T4f	T1	T4	DG	Yes	38
T4T1f	T4	T1	DG	Yes	32

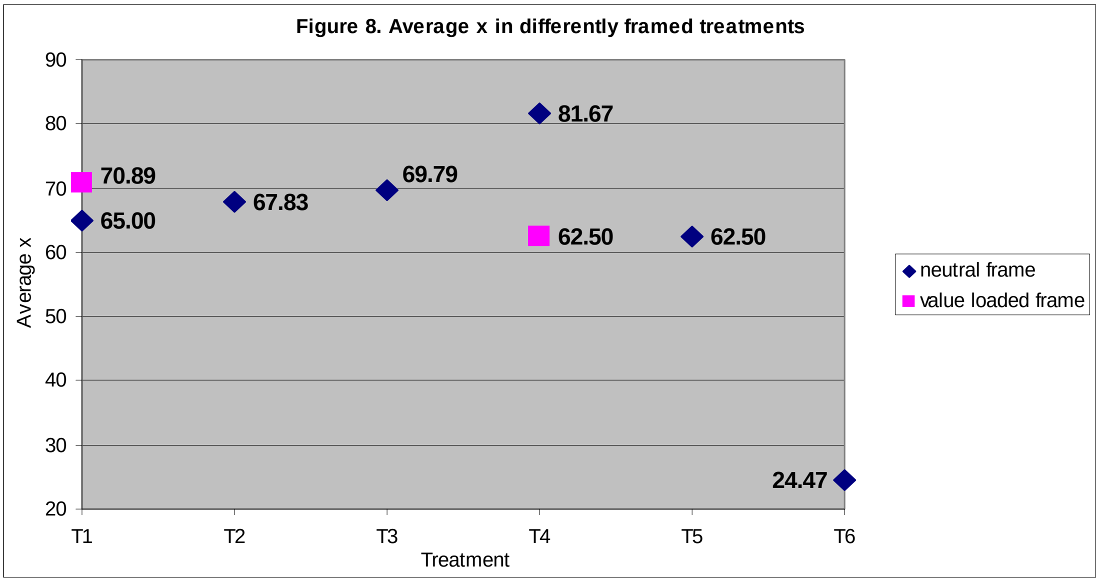
DG: dictator game

Among the treatments with negative incentives we chose treatment T4, since the intensity of negative incentives is (i) not as high that the majority of individuals is already deterred as in the neutrally framed treatments T5 and T6, and (ii) high enough to observe a significant incentive effect in the static comparison of the neutrally framed treatment T1 with T4.

5.1 Static Results - Behavior in part 1

Figure 8. summarizes the average stolen amounts in the neutrally framed as well as the framed treatments.

[27] Except for the framing of these two terms, the neutral and framed treatments are identical.

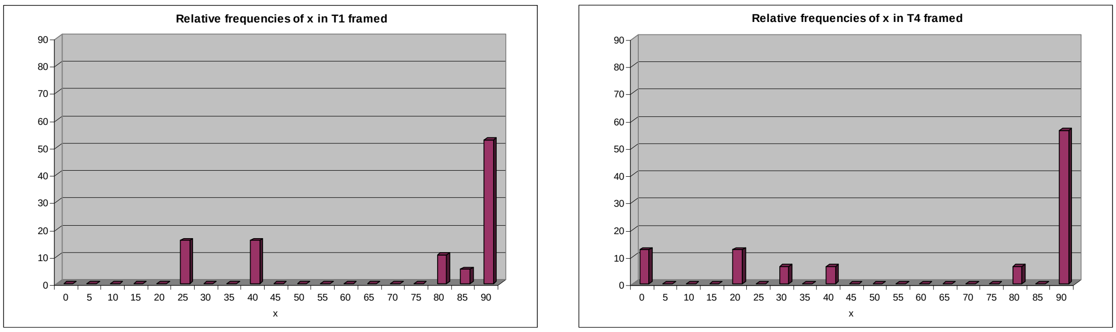


The average stolen amount in the framed treatment with no incentives is not significantly different from the one in the neutrally framed benchmark treatment (Mann-Whitney-U test, $p > 0.5$). In addition, the distribution of x in the framed T1 is very similar[28] to the one in the benchmark treatment depicted in the first graph of Figure 9.. In the absence of incentives, framing does not seem to affect the individual decision to steal: selfish subjects remain selfish, fair-minded individuals remain fair-minded.

In treatment T4, however, we identify a significant framing effect at a 10 % significance level (Mann-Whitney-U test, $p = 0.075$). The distribution of the stolen amount seems to change with the implementation of the frame as indicated in the second graph of Figure 9..

[28] Actually, more similar than the neutrally framed treatments T2 and T3, since intermediate levels are still centered on a relatively low level.

Figure 9. Distributions of the stolen amount in the framed treatments

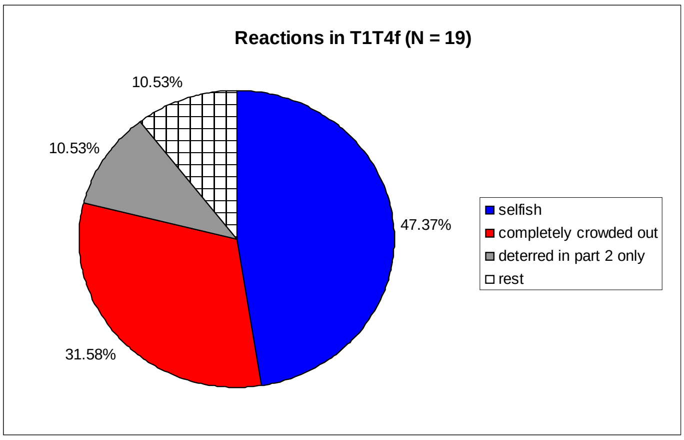


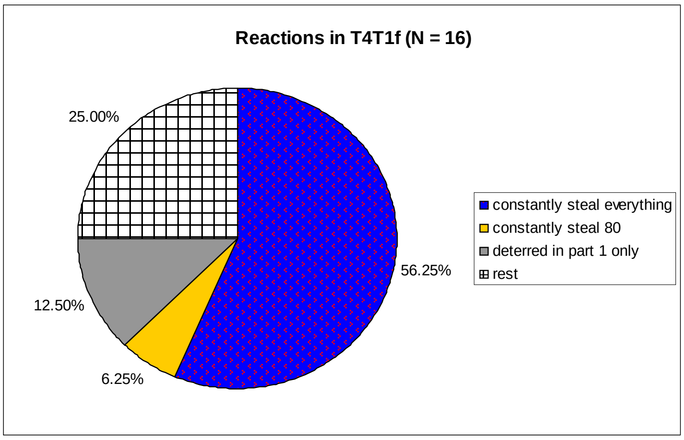
Therefore, we might conclude that framing affects stealing behavior when incentives are present. One possible explanation that is in line with our data is that selfish subjects are unaffected by the frame but with a frame fairness concerns are not crowded out by the presence of incentives. We now check whether the conclusions from our static analysis also hold in the dynamic setting.

5.2 Dynamic Results - Behavior from part 1 to part 2

In the framed treatments, sequence effects are not significant at conventional significance levels (Mann-Whitney U test, p = 0.3877 for T1 and p = 0.3689 for T4). Nevertheless, p-values are too low to safely pool the data of the two different parts. Figure 10. summarizes individual reactions to parameter changes.

Figure 10. Reactions to parameter changes in the framed sessions





In session T1T4f, incentives are increased from the level of no incentives (T1) to the level of intermediate incentives (T4). As without any frame, about 47 % of subjects are selfish and steal constantly everything regardless of the intensity of incentives. Due to the relatively harsh parameter change, more than 30 % of individuals flip from intermediate stolen amounts to stealing everything. This observation of crowding out fits nicely to the dynamic results of the neutrally framed sessions T1T3 and T2T4. As the intensity of negative incentives in T4 is at an intermediate level, we observe some deterrence.

In session T4T1f, intermediate incentives are implemented in part 1 (T4) and are completely removed in part 2 (T1). About 55 % of subjects constantly steal everything.[29] One individual remains stealing 80 points and about 12 % were deterred in the first part and switch to strictly positive amounts in the second part.

[29] These subjects may be selfish or crowded out lastingly. However, we concluded from Figure 8. and the second graph of Figure 9. that not so much crowding out happened in the first part. That may well be the reason, why we do not see any behaviour which we would classify as "returned fairness".

In contrast to our framed static results in T4, we could conclude from the dynamic observations of session T1T4f that crowding out is also present with framing, namely when one significantly increases incentives while holding the frame constant. Since we start from intermediate incentives and only minor crowding out in session T4T1f, crowding out does not seem to play a major role when negative incentives are completely removed.

5.3 Framing effects

In sum, differences between neutral and framed treatments are small: (i) we do not observe any significant framing effect in treatment T1, and (ii) the dynamic results document crowding out of fairness concerns through negative incentives. It still remains puzzling why subjects confronted with intermediate incentives plus a frame steal significantly less than when they are confronted with the same incentives and no frame, or when they have already experienced one period without incentives and the same frame. Within the scope of this paper we are not able to solve this puzzle. Still, our extension shows that even when frames are present crowding out of incentives is a non-negligible phenomenon.

6 Conclusion

We started with the question "Do negative incentives work?": More precisely, we presented an experimental test whether crime rates are indeed monotonically (weakly) decreasing in the severity of punishment and the detection probability as predicted by Becker's (1968) deterrence hypothesis. Our experimental results clearly reject the deterrence hypothesis: for small negative incentives, the average stolen amount is *increasing* in the severity of punishment and the detection probability. Only very high incentives deter the majority of subjects.

A close look at individual behavior reveals an explanation for these findings. Our data reflect the behavior of two different types of subjects: about 45 % are selfish and behave as predicted by the deterrence hypothesis. But the aggregate results are driven by the second type, namely the about 50 % fair minded subjects. Their fairness concerns are crowded out by the presence of negative incentives. Fair subjects steal only low amounts (roughly amounts needed to equate payoffs) in the absence of incentives. However, their fairness concerns are reduced whenever incentives are introduced or increased which stimulates them to steal larger amounts. Our finding that small incentives may backfire is well in line with the literature on crowding out of intrinsic motivation through positive incentives. We document that crowding out can also (i) apply to fairness concerns and (ii) be caused by negative incentives (compare footnote 4).

Our design has focused on the domain of small incentives, i.e. levels of incentives that for risk-neutral individuals substantially reduce the expected gain from commiting a crime, but do not completely neutralize it. Small incentives are especially relevant in real life: in Germany, the clearance rate for thefts with (without) aggravating circumstances was 14 % (44 %) in 2005.[30] Polinsky and Shavell (2000b) point out that also the severity of punishment is often quite low in relation to what potential offender are capable to pay.

What are the take aways from our results? Let us start by saying that any policy implications should be treated very carefully: for example, testing criminal behavior in the lab might abstract from social norms that could be a driving force reducing criminal behavior. Even if this were the case our results would still provide a strong test of Becker's deterrence hypothesis that purely relies on incentive effects. Furthermore, our framed treatments can be considered as a robustness check of our neutrally framed results: Talking about "stealing" and "penalties" certainly introduces a very strong moral connotation. Still, the results from the framed treatments document that small incentives might backfire, though to a slightly smaller extent. Put provokingly, our results then suggest to punish criminal activities hard or not to punish them at all to avoid mild punishment to be counterproductive.

7 Appendix

7.1 Experimental sessions and instructions

The order of events during each experimental session was the following: Subjects were welcomed and randomly assigned a cubicle in the laboratory where they took their decisions in complete anonymity from the other participants. The random allocation to a cubicle also determined a subject's role in all three parts. Subjects were handed out the general instructions for the experiment as well as the instructions for part 1. After all subjects had read both instructions carefully, and all remaining

[30] Polizeiliche Kriminalstatistik (2005), Table 23.

questions were answered we proceeded to the decision stage of the first part. Part 2 and 3 were conducted in an analogous way. We finished each experimental session by letting subjects answer a questionnaire that asked for demographic characteristics and included a Holt and Laury (2002) table. In the questionnaire, this table was explained in detail and it was highlighted that one randomly drawn decision was paid out in addition to the earnings in the previous parts.

Instructions, the program, and the questionnaire were originally written in German. The translated general instructions, the translated instructions for part 1 of the neutrally framed treatment T4, and the translated Holt and and Laury (2002) table can be found in the following section. Instructions for part 2 and part 3 as similar to part 1 as possible. For the framed treatments, we used the expression "steal any integer amount between 0 and 90 from participant A" instead of "choose any integer amount between 0 and 90 that shall be transferred from participant A to you", and the term "minus a penalty of x points" instead of "minus an amount of x points".

7.1.1 Translated general instructions

General explanations concerning the experiment

Welcome to this experiment. You and the other participants are asked to make decisions. Your decisions as well as the decisions of the other participants will determine the result of the experiment. At the end of the experiment you will be paid **in cash** according to the **actual** result. So please read the instruction thoroughly and think about your decision carefully.

During the experiment you are not allowed to talk to the other participants, to use cell phones or to start any other programs on the computer. The neglect of these rules will immediately lead to the exclusion from the experiment and all payments. If you have any questions, please raise your hand. An experimenter will then come to your seat to answer your questions.

During the experiment we talk about points instead of Euros. Your total income will therefore be calculated in points first. At the end of the experiment the total amount of points will be converted into Euros according to the following exchange rate:

1 point = 15 Cents.

The experiment consists of three **independent** parts in which you can accumulate points. Before each part only the instruction of this part will be handed out.

During the experiment neither you nor the other participants will receive any information on the course of the experiment (e.g. decisions of other participants or results of a particular part).

The results of each single part will be calculated only after all three parts will be finished. **Then, one of these three parts will be chosen randomly. At the end of the whole experiment only this part will be paid out in cash according to your decisions.**

1

7.1.2 Translated instructions for part 1 of the neutrally framed treatment T4

Part 1

In this part, there are **participants in role A and participants in role B**. **In this part, you were randomly assigned to role B**. You will be randomly and anonymously matched to another participant in role A. This random matching lasts only for this part! The matched participant will not be matched to you in the following two parts again. Neither before nor after the experiment, will you receive any information about the identity of your matched participant. Likewise, your matched participant will not receive any information about your identity.

You - as participant B - have an initial endowment of 50 points. Participant A has an initial endowment of 90 points.

As a participant in role B you can choose any **integer amount** between 0 and 90 points (including 0 and 90) **which shall be transferred from participant A to you**. Participant A does not make any decision. In order to make your decision, please enter your chosen amount on the corresponding computer screen and push the OK button.

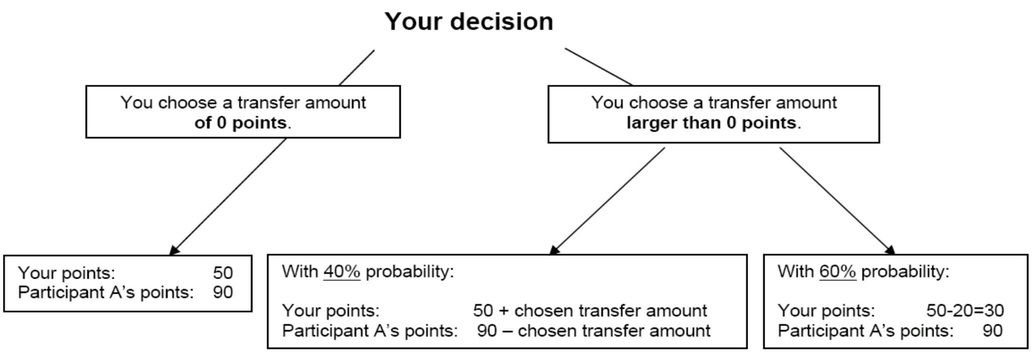
- If you choose a transfer amount of 0 points, you will receive your initial endowment of 50 points, and participant A will receive his initial endowment of 90 points.
- If you choose a transfer amount larger than 0 points,
    - **with <u>40%</u> probability you will receive your initial endowment of 50 points plus your chosen transfer amount and participant A will receive his initial endowment of 90 points minus your chosen transfer amount.**
    - **with <u>60%</u> probability you will receive your initial endowment of 50 points minus an amount of 20 points, i.e. 30 points, and participant A will receive his initial endowment of 90 points.**

<u>Example 1:</u> You choose a transfer amount of 22 points. With <u>40%</u> probability you will receive 50 + 22 points = 72 points, and with <u>60%</u> probability you will receive 50 – 20 points = 30 points. Participant A will receive 90 – 22 points = 68 points with 40% probability, and his initial endowment of 90 points with <u>60%</u> probability.

<u>Example 2:</u> You choose a transfer amount of 0 points. You will receive 50 points. Participant A will receive 90 points.

2

The course of action of part 1 is illustrated by the following graph:



If you have any questions, please raise your hand. An experimenter will then come to your seat in to answer your questions.

7.1.3 Translated Holt and Laury (2002) table

Decision	Option A	Option B
Decision 1	10 points	25 points with a probability of 10% 0 points with a probability of 90%
Decision 2	10 points	25 points with a probability of 20% 0 points with a probability of 80%
Decision 3	10 points	25 points with a probability of 30% 0 points with a probability of 70%
Decision 4	10 points	25 points with a probability of 40% 0 points with a probability of 60%
Decision 5	10 points	25 points with a probability of 50% 0 points with a probability of 50%
Decision 6	10 points	25 points with a probability of 60% 0 points with a probability of 40%
Decision 7	10 points	25 points with a probability of 70% 0 points with a probability of 30%
Decision 8	10 points	25 points with a probability of 80% 0 points with a probability of 20%
Decision 9	10 points	25 points with a probability of 90% 0 points with a probability of 10%
Decision 10	10 points	25 points with a probability of 100% 0 points with a probability of 0%

Participants had to decide ten times, whether they prefer option A to option B. One decision was randomly chosen and paid out at the end of the experiment.[31] All decisions had an equal probability to be chosen in the end.

References

[1] Becker, G. S., 1968. Crime and punishment, an economic approach. Journal of Political Economy 76 (2), 169-217.

[2] Bohnet, I., Frey, B. S., Huck, S., 2001. More order with less law, on contract enforcement, trust and crowding. American Political Science Review 95 (1), 131-144.

[31] All decisions had an equal probability to be chosen in the end.

[3] Bohnet, I., Cooter, R. D., 2005. Expressive law, framing or equilibrium selection?. Working paper.

[4] Bundeskriminalamt (Ed.), 2005. Polizeiliche Kriminalstatistik Bundesrepublik Deutschland, Berichtsjahr 2005, Wiesbaden, http://www.bka.de/pks/pks2005/download/pks-jb_2005_bka.pdf.

[5] Eide, E., 2000. Economics of criminal behavior. In: Bouckaert, B., De Geest, G. (Eds.), Encyclopedia of Law and Economics, Vol. V, Cheltenham, Edward Elgar, 345-89.

[6] Falk, A., Fischbacher, U., 2002. Crime in the lab, detecting social interaction. European Economic Review 46 (4-5), 859-869.

[7] Fehr, E., Falk, A., 2002. Psychological foundations of incentives. European Economic Review 46 (4-5), 687-724

[8] Fehr, E., Schmidt, K., 1999. A theory of fairness, competition, and cooperation. Quarterly Journal of Economics 114 (3), 817-868.

[9] Forsythe, R., Horowitz, J. L., Savin, N. E., Sefton, M., 1994. Fairness in simple bargaining experiments. Games and Economic Behavior 6 (3), 347-369.

[10] Frey, B. S., Jegen, R., 2001. Motivation crowding theory. Journal of Economic Surveys 15 (5), 589-611.

[11] Frey, B. S., Oberholzer-Gee, F., 1997. The costs of price incentives, an empirical analysis of motivation crowding-out. American Economic Review 87 (4), 746-755.

[12] Galbiati, R., Vertova, P., 2005. Law and behaviours in social dilemmas, testing the effect of obligations on cooperation. Working Paper.

[13] Garoupa, N., 1997. The theory of optimal law enforcement. Journal of Economic Surveys 11 (3), 267-295.

[14] Gneezy, U., 2003. The w effect of incentives. Working paper.

[15] Gneezy, U., Rustichini, A., 2000a. A fine is a price. Journal of Legal Studies 29 (1), 1-17.

[16] Gneezy, U., Rustichini, A., 2000b. Pay enough or don't pay at all. Quarterly Journal of Economics 115 (3), 791-810.

[17] Holt, C. A., Laury, S. K., 2002. Risk aversion and incentive effects. American Economic Review 92 (5), 1644-1655.

[18] Polinsky, A. M., Shavell, S., 2000a. The economic theory of public enforcement of law. Journal of Economic Literature 38 (1), 45-76.

[19] Polinsky, A. M., Shavell, S., 2000b. The fairness of sanctions, some implications for optimal enforcement theory. American Law and Economics Review 2 (2), 223-237.

[20] Tyran, J.-R., Feld, L. P., 2006. Achieving compliance when legal sanctions are non-deterrent. Scandinavian Journal of Economcis 108 (1), 135-156.