Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Fakultät Wirtschafts- und Sozialwissenschaften

# On the Interaction of Deterrence and Emotions

**Menusch Khadjavi**

**WiSo-HH Working Paper Series**
Working Paper No. 14
April 2014

WiSo-HH Working Paper Series
Working Paper No. 14
April 2014

# On the Interaction of Deterrence and Emotions

Menusch Khadjavi, University of Hamburg

ISSN 2196-8128

Font used: „TheSans UHH“ / LucasFonts

Die Working Paper Series bieten Forscherinnen und Forschern, die an Projekten in Federführung oder mit der Beteiligung der Fakultät Wirtschafts- und Sozialwissenschaften der Universität Hamburg tätig sind, die Möglichkeit zur digitalen Publikation ihrer Forschungsergebnisse. Die Reihe erscheint in unregelmäßiger Reihenfolge.

Jede Nummer erscheint in digitaler Version unter
https://www.wiso.uni-hamburg.de/de/forschung/working-paper-series/

**Kontakt:**

WiSo-Forschungslabor
Von-Melle-Park 5
20146 Hamburg
E-Mail: experiments@wiso.uni-hamburg.de
Web: http://www.wiso.uni-hamburg.de/forschung/forschungslabor/home/

# On the Interaction of Deterrence and Emotions

**Menusch Khadjavi***

University of Hamburg

February 17, 2014

Accepted for publication in the *Journal of Law, Economics and Organization*.

**Abstract**

This study analyzes deterrence schemes and their impact on stealing. The results confirm Becker's deterrence hypothesis. Moreover, crowding out of pro-social behavior occurs due to deterrence incentives: when deterrence incentives first exist and are removed later on, subsequent behavior is more selfish than without this deterrence history. This study offers evidence that (part of this) crowding out takes place via change of emotions. Without deterrence incentives in place, in a variant of the dictator game, players with pro-social emotions steal less. When players face expected punishment pro-social emotions are deactivated and do not decrease stealing; in this case self-centered emotions get activated and motivate greater stealing. This study provides support for theories on emotions in behavioral criminal law and economics and offers new insights for deterrence policy.

Keywords: Crime, Deterrence, Crowding Out, Emotions, Laboratory Experiment

JEL: C91, D63, K42

* Correspondence: Department of Economics, University of Hamburg, Von Melle Park 5, 20146 Hamburg, Germany; e-mail contact: menusch.khadjavi@philosophy-economics.de.
I am thankful for funding from the Graduate School of the Faculty of Economic and Social Sciences, University of Hamburg. Andreas Lange, Hannah Schildberg-Hörisch, Paul Smeets, Roel van Veldhuizen and participants at the 2012 ESA International Conference in New York, at the 2012 ESA North American Conference in Tucson, the University of California San Diego and the University of Hamburg provided helpful comments.

*"The advantage of the emotions is that they lead us astray."*

Oscar Wilde

## 1 Introduction

Prevention, detection and prosecution of crime involve major expenses worldwide. In the United States, the total expenditure for the legal system amounted to over 260 billion US dollars in 2010 (Kyckelhahn and Martin 2013).[1] Law and economics deterrence theories commonly build upon the seminal contribution by Gary Becker (1968) that incentives can be put to work in order to reduce crime. This so-called deterrence hypothesis has been tested and mostly supported based on evidence by empirical studies (e.g. Corman and Mocan 2000, Fisman and Miguel 2007, Levitt 1997, Witte 1980), lab experiments with students (e.g. Abbink et al. 2002, Anderson and Stafford 2003, DeAngelo and Charness 2012, Friesen 2012, Harbaugh et al. 2013, Nagin and Pogarsky 2003, Rizzolli and Stanca 2012, Schildberg-Hörisch and Strassmair 2012, Schulze and Frank 2003) as well as in an extra laboratory experiment (term coined by Charness et al. 2013, also referred to as 'artefactual field experiment' by Harrison and List 2004) with criminals (Khadjavi 2013).

However there is also a dark side to such external and extrinsic deterrence incentives. Gneezy and Rustichini (2000a), Schildberg-Hörisch and Strassmair (2012), Schulze and Frank (2003) and Khadjavi (2013) provide experimental economic evidence for detrimental effects of (deterrence) fines in laboratory, extra laboratory and field experiments.[2] Surveys by Bowles (1998), Deci et al. (1999), Frey and Jegen (2001), Gneezy et al. (2011) and Bowles and Polanía-Reyes (2012) discuss situations in which extrinsic incentives hold undesired side effects: they change the perceptions

[1] These estimates were reported on July, 1 2013 and are preliminary. Split into federal, state and local governments' expenditures, the shares of total expenditures were 19 percent, 30.5 percent and 50.5 percent respectively.

[2] Research in social psychology on the self-determination theory also provides evidence of such effects, for instance seminal contributions by Deci (1971) and Deci and Ryan (1985).

or feelings of individuals for the situation they are in. Consequently, intrinsic pro-social motivation of some individuals is crowded out and part of the desired effect is offset. Heyman and Ariely (2004) describe this process as individuals feeling to be either in a 'social market' (no extrinsic incentives) or in a 'monetary market' (when such extrinsic incentives are present). Meier (2007) labels these two situations 'moral mode' and 'exchange mode' respectively.

This work aims at investigating one possible underlying mechanism of crowding out: changes of emotions. While it is well established that crowding out occurs, the reasons why exactly individuals change their behavior remain unclear. Is it that individuals simply put on their 'moral mode hat' in some situations and replace it with an 'exchange mode hat' in others? If so, is this transition frictionless and purely rational, or is it accompanied by a change in emotions?

In this work we analyze whether subjects whose intrinsic motivation is crowded out by extrinsic incentives exhibit different emotions compared to subjects who do not face extrinsic incentives. To this end we employ a laboratory experiment and use a combined within and between subjects design. Our workhorse is the two-player stealing game (see also Schildberg-Hörisch and Strassmair 2012): player 1 is the potential criminal and makes the decision whether, and if so how much, to steal from player 2, or to abstain from stealing. Player 2 does not make a decision and only receives information on the stealing decision of player 1 and her success. There may or may not be an external deterrence institution present.[3] If this institution is present, it will detect and punish stealing by player 1 with some probability and fine in accordance with a deterrence scheme which is known to both players.

The procedure of our experiment is as follows: first, depending on the treatment, player 1 either faces extrinsic deterrence incentives or she does not. If they are present and depending on their intensity, these incentives should work to reduce player 1's stealing magnitude (or force her to

[3] If the external deterrence institution is not present, this game is similar to the dictator game in the taking domain. See Bardsley (2008) and List (2007) for dictator games with giving and taking domains.

refrain from stealing entirely). Yet these deterrence incentives may also make player 1 change from the 'moral mode' into the 'exchange mode'. Second, we elicit subjects' emotions via self-reports. Here we may find that different treatments activate different emotions. Third, subjects face a regime change (from no deterrence to deterrence or vice versa) which tests whether crowding out is present. We then test whether certain clusters of emotions, activated by extrinsic incentives, explain the stealing decision in the old regime and whether crowding out is present in the new regime.

Motivated by works such as Elster (1998), Frank (1988), Loewenstein (2000) and Rick and Loewenstein (2008), economic research increasingly engages in a deeper understanding of emotional motivation in decision making. Research on psychological games emphasize that beliefs of emotions can endogenize (the utility individuals receive from) outcomes and therefore strategies and equilibria (for instance, see Geanakoplos et al. 1989, Rabin 1993).[4] The recent experimental literature provides evidence that emotional motivation is able to significantly impact economic decision making of individuals in a set of different games (see Andrade and Ariely 2009, Bosman and van Winden 2002, Cubitt et al. 2011, Hopfensitz and Reuben 2009, Reuben and van Winden 2010).

Van Winden and Ash (2012) provide theory on behavioral determinants of criminal activity and include emotions as one component of decision making. The aim of our study is to provide evidence for their work with regard to the emotional motivation of crime. Such a motivation likely depends on the institutions that are present and how they activate and deactivate pro-social emotions like shame and guilt and negative emotions like anger and gloating.

Furthermore, Bowles and Polanía-Reyes (2012) discuss a number of factors that may influence crowding of pro-social behavior. Amongst others, they consider *moral disengagement* as one factor

---

[4] For instance, Huang and Wu (1994) use psychological games to describe how expectations and social norms may influence corruption.

of crowding out. We find this explanation likely to apply to our experiment: the focus of a criminal in our study may rest dominantly on the victim in our (risk-free) no-deterrence treatment. Conversely when deterrence incentives are present this focus shifts from the victim to the scheme. The criminal then concentrates on the optimal amount to steal, while the source of this amount, the victim's pocket, loses its focus. Consequently the moral implications are decoupled from the decision. Our hypothesis is that a criminal's emotions are informative with regard to her state of mind. If emotions differ depending on the treatment (deterrence scheme) in place, they are also likely to affect behavior.

Our results indeed suggest that emotions motivate stealing. Interestingly, criminals' emotions in our experiment depend on whether a deterrence scheme is in place and motivate behavior. In the treatment without any deterrence incentives ceteris paribus a cluster of so-called pro-social emotions *decreases* stealing. Conversely, in the face of potential punishment self-centered emotions *increase* stealing. Hence, the implemented policy does not only affect behavior directly via incentives. Rather, it also (de)activates different clusters of emotions of potential criminals which again cause pro-social behavior to crowd out. Our study therefore contributes to the interaction of deterrence incentives and emotions and to the understanding of emotional mechanisms of crowding out.

The remainder of this work is organized as follows; section 2 presents the experimental design, including the stealing game, hypotheses and procedures. The results of our study are discussed in section 3. Finally, section 4 provides a concluding discussion.

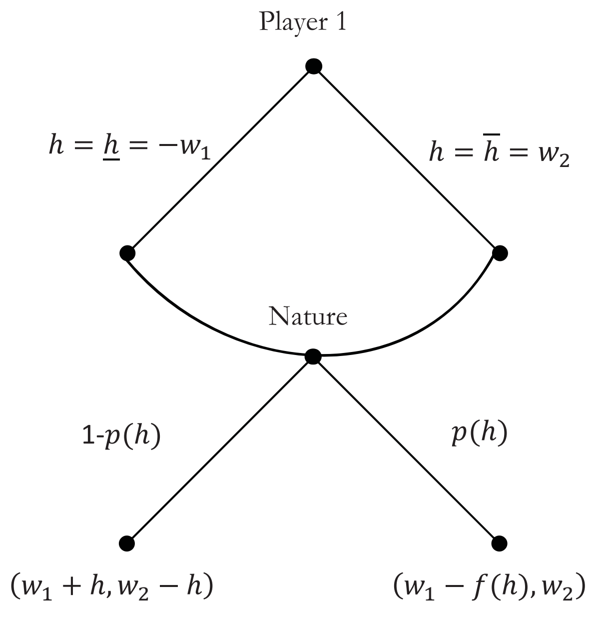
## 2 Experimental Design

In this section we first introduce the workhorse of this study: the stealing game. Next, we discuss our treatments and their aims. Based on our treatments, we provide information on the procedures of our experiment, formulate hypotheses and explain how we will test them.

### *2.1 The Stealing Game and the Treatments*

The stealing game is a two-person game with 'nature' as a stochastic player. Player 1 is the only active player who makes a decision, and player 2 solely receives information about player 1's action and her success. Hence, there is no strategic interaction between the two players. The decision problem of player 1 is to decide how much to give to or take (steal) from player 2's endowment. We denote the initial private endowments of player 1 and player 2 as $w_1$ and $w_2$, with $w_1 < w_2$ in our experiment. The action set of player 1 is the entire range from giving all of her endowment to player 2 to taking (stealing) all of player 2's endowment, i.e. for her haul we set $h_1 \in \{-w_1, \ldots, w_2\}$. This action set calibration is important, since player 1 should have the opportunity to choose no allocation change or to give to player 2. The calibration thus avoids a suggestive action set that might influence results significantly (Bardsley 2008, List 2007). This calibration is new to the experimental literature on deterrence (which usually only focuses on the taking domain) and a slight innovation of this study. Figure 1 depicts the stealing game.

**Figure 1.** The Stealing Game.



As is usual with giving gifts to others, there is no external institution which aims to prevent voluntary (charitable) giving. On the contrary, taking from others without their consent is the definition of the criminal act of stealing.[5] Accordingly, if player 1 decides to steal some monetary amount from player 2, she will face some fine $f$ with probability $p$. Corresponding to the description of the stealing game above, a player 1 with utility function $u_1$ faces the maximization problem of expected utility

$$\max_{h_1} EU_1 = \left(1 - p(h_1)\right)u_1(w_1 + h_1) + p(h_1)u_1\left(w_1 - f(h_1)\right).$$

In our experiment, we calibrate $w_1 = 2$ EUR and $w_2 = 10$ EUR. Note that either $f$ or $p$ or both may be a function of the amount $h_1$ player 1 tries to steal. In fact, Stigler (1970) points out that it is crucial for the punishment to fit the crime. Else, if the expected punishment was to remain constant, player 1's potential gains would increase in her amount stolen and would likely result in a corner solution (i.e. either stealing all or nothing). With this in mind, we proceed to elaborate our treatments.

The three treatments of our experiment resemble three deterrence schemes. These schemes systematically vary fine $f$ and probability $p$. In the treatment *NoDeter* there is no external institution that executes the extrinsic deterrence incentives. Hence, we implicitly set $f = 0$ and/or $p = 0$. This treatment allows us to collect information on the behavior of player 1 when she wears her 'moral mode hat'. Accordingly, player 1's maximization problem reduces to $\max_{h_1} EU_1 = u_1(w_1 + h_1)$. A rational, non-satiated and narrowly self-interested player 1 solves this problem by stealing the maximum amount, i.e. all of player 2's endowment, $h_1^* = w_2 = 10$ EUR.

[5] Merriam-Webster's online dictionary defines to steal as "to take without permission".

As noted above, especially in *NoDeter* subjects may wear their 'moral mode hats'. It may therefore be the case that player 1 holds social, other-regarding preferences or certain moral concepts, and acts on them. One prominent form of social preferences includes inequality aversion with regard to final payoff $\pi$ (Fehr and Schmidt 1999, Bolton and Ockenfels 2000), such that $u_1(\pi_1, \pi_2)$. A sufficiently inequality averse[6] player 1 will therefore not steal player 2's entire endowment, but rather prefers to achieve ex post payoff-equality.[7] Likewise, in accordance with social-welfare preferences based on the model of Charness and Rabin (2002), player 1 may prefer to increase her payoff (by stealing) as it is the lowest in the group. Another motivation to abstain from stealing (or stealing the maximum amount) is a coherent self-image (e.g. Brekke et al. 2003) motivated by moral concepts like Kantianism. If such social preferences or self-image concerns are present, we expect to find some average level of stealing which is below the standard prediction of $h_1^* = w_2$.

In the second and third treatment of our experiment, called *DeterFine* and *DeterProb*, the external deterrence institution is in place. This deterrence institution only gets activated if stealing takes place. Giving actions remain permitted. In *DeterFine* player 1 faces a probability of prosecution of

$$p_{DeterFine}(h) = \begin{cases} 0 & if \quad h \leq 0 \\ 0.5 & if \quad h > 0 \end{cases}$$

and a fine $f$ which is increasing in $h_1$. The function reads

$$f_{DeterFine}(h) = \begin{cases} 0 & if \quad h \leq 0 \\ 1.25 - 0.25h + 0.1h^2 & if \quad h > 0 \end{cases}.$$

This function is useful to calibrate treatment *DeterFine* in order to achieve an interior solution for player 1's optimal stealing. If she is risk neutral, her expected payoff is maximal at $h^*_{DeterFine} =$

6 For instance, in the standard linear Fehr-Schmidt model sufficiently inequality averse means $\beta_1 > \frac{1}{2}$.

7 Note that player 1 can achieve ex post payoff equality by stealing 4 EUR from player 2 in *NoDeter*.

$\frac{25}{4} = 6.25$ EUR. Treatment *DeterProb* was designed to offer the exact same solution $h^*_{DeterProb} = \frac{25}{4} = 6.25$ EUR and, importantly, also corresponds to *DeterFine* with regard to the expected value for all other $h$ choices. *DeterProb* meanwhile varies the probability of prosecution

$$p_{DeterProb}(h) = \begin{cases} 0 & if \quad h \leq 0 \\ 0.25 + 0.05h & if \quad h > 0 \end{cases}$$

while the fine is constant at

$$f_{DeterProb}(h) = \begin{cases} 0 & if \quad h \leq 0 \\ 2.50 & if \quad h > 0 \end{cases}.$$

**Table 1.** Treatment Overview.

| *h* | Treatments (Deterrence Schemes) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *NoDeter* | | | *DeterFine* | | | *DeterProb* | | |
| | *f* in EUR | *p* | $\pi$ in EUR | *f* in EUR | *p* | E($\pi$) in EUR | *f* in EUR | *p* | E($\pi$) in EUR |
| -2 | | | -2 | | | -2 | | | -2 |
| -1 | | | -1 | 0 | 0 | -1 | 0 | 0 | -1 |
| 0 | | | 0 | | | 0 | | | 0 |
| 1 | | | 1 | 1.10 | | -0.05 | | 0.3 | -0.05 |
| 2 | | | 2 | 1.15 | | 0.425 | | 0.35 | 0.425 |
| 3 | | | 3 | 1.40 | | 0.80 | | 0.4 | 0.80 |
| 4 | 0 | 0 | 4 | 1.85 | | 1.075 | | 0.45 | 1.075 |
| 5 | | | 5 | 2.50 | 0.5 | 1.25 | 2.50 | 0.5 | 1.25 |
| 6 | | | 6 | 3.35 | | 1.325 | | 0.55 | 1.325 |
| 7 | | | 7 | 4.40 | | 1.3 | | 0.6 | 1.3 |
| 8 | | | 8 | 5.65 | | 1.175 | | 0.65 | 1.175 |
| 9 | | | 9 | 7.10 | | 0.95 | | 0.7 | 0.95 |
| 10 | | | 10 | 8.75 | | 0.625 | | 0.75 | 0.625 |

Note again that these calibrations are in accordance with Stigler (1970), i.e. the expected punishment increases in the amount stolen. Table 1 summarizes the three treatments and their probabilities and fines for all possible stealing actions. We include both *DeterFine* and *DeterProb* in our experiment to test for differences in deterrence from increasing fines and increasing probabilities. We can therefore directly test their relative usefulness to deter stealing when expected punishment is constant across the two treatments *for every possible action*. To the best of our knowledge this is the first study with such a calibration and thereby contributes to theory on optimal deterrence (e.g. Becker 1968, Polinsky and Shavell 1979, Garoupa 1997, Garoupa 2001).

### *2.2 Procedures*

All sessions were conducted in the computer laboratory of the University of Hamburg in October and November 2011 as well as January and February 2012. Each session lasted approximately one hour. We used z-Tree (Fischbacher 2007) for programming and ORSEE (Greiner 2004) for recruitment. In total 408 subjects participated in our experiment. All subjects were students from various academic backgrounds and 50.7 percent of the participants were female. No subject participated in the experiment more than once. We conducted fourteen experimental sessions of two parts; the first part was a real-effort task to endogenously determine endowments and positions for the second part, i.e. the stealing-game part.

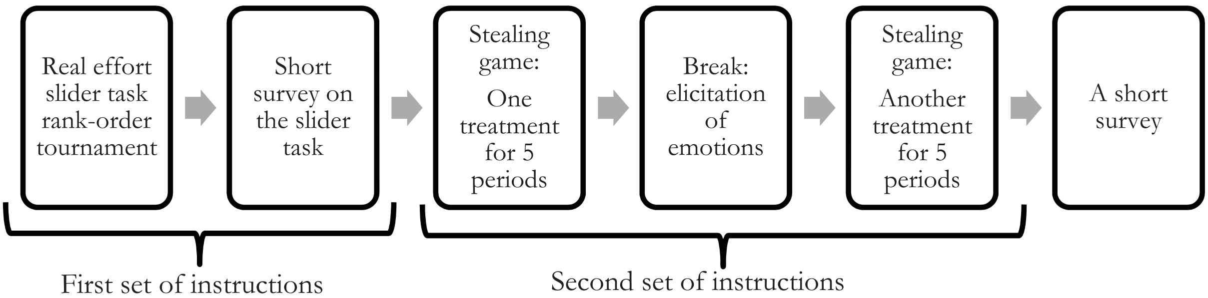
Once the participants were seated, a set of neutral instructions was handed out and read out loud by the experimenter.[8] In order to ensure that subjects understood the respective game, experimental instructions included several numerical examples and participants had to answer control questions via their computer terminals.

Let us now consider the course of the experiment more closely. Figure 2 depicts the timeline of our experiment. As outlined above, studies by Schulze and Frank (2003), Schildberg-Hörisch and Strassmair (2012) and Khadjavi (2013) find evidence for crowding out of pro-social behavior due to deterrence incentives. These three studies randomly assign subjects to be player 1 and player 2. Studies however show that the mechanism which determines a player's position in a game and her endowment may affect her decision making (see e.g. List 2007, List and Cherry 2008). Established mechanisms are the exogenous, random mechanism (by the experimental program)

8 See the instructions in appendix B for further information.

or ‘earned’ positions and endowments in a preceding endogenous mechanism. For instance, giving in a dictator game with earned endowment is usually less generous compared to giving when endowments are randomly provided (List and Cherry 2008).

**Figure 2.** Timeline of the Experiment.




For this reason, participants of our experiment engaged in a preceding real effort task before playing the stealing game. The task we use is the slider task of Gill and Prowse (2011, 2012)[9] which was part of one common rank-order tournament scheme. The more successful half of subjects received endowments of $w_2 = 10$ EUR each and later became player 2, while the less successful half of subjects received endowments of $w_1 = 2$ EUR each and later became player 1. In addition, all subjects received an initial endowment of 7 EUR each.

Subjects knew beforehand that the experiment would be split into parts and that the earned endowment from the rank-order tournament would be transferred to later parts. Also, they were informed that the final payout would depend on the events of *all* parts of the experiment. After the end of the tournament but before subjects were informed about their endowments and the

[9] The slider task is programmed in z-tree and offers a number of useful features (each effort unit is approx. equally costly, no prior knowledge is needed, etc.). Subjects see a screen of 48 sliders. All sliders have a range between 0 and 100. The default position is 0 for all sliders. Subjects then need to put the sliders exactly midway of the range, i.e. on the ‘50’ mark, to receive a point. Subjects were only able to use the computer mouse for this task and time was limited and equal for all subjects, so that no subject could collect all points. Subjects were able to conduct a test round of the task (without rank feedback).

content of the latter part of the experiment (i.e. the stealing game), subjects answered three questions regarding the task.

Hence, the slider task led to unequal endowments based on a rank-order tournament scheme: 'rich' player 2 subjects (potential victims) and 'poor' player 1 subjects (potential criminals). This design features carries two advantages. First, it increases identification with endowments and social positions and thereby potentially increases feelings of entitlement. This is important because we aim at analyzing changes in emotions and emotional involvement is crucial for this to work. Second, the competitive nature of the task may already let subjects put on their 'exchange mode hat' and crowd-out pro-social behavior *before* the actual stealing decision. Hence, this procedure works against our aim to detect and explain crowding out. If we nevertheless find crowding out to occur in our experiment, we can rule out that this pro-social behavior stems from random roles.

Subsequently, subjects received information on the latter course of the experiment that was again distributed and read out loud. Player 1 was then able to give up to 2 EUR to or steal up to 10 EUR from player 2. The stealing game was split into two halves of five periods each. We employed an absolute stranger matching. Note that due to the rank-order tournament in part 1 that included all subjects in a session, player-1 subjects could be sure that the player 2 subjects they were matched with owned higher endowments because they realized more effort in the slider task. At the end of every period player 1 and player 2 received information about the amount stolen and the payoffs of both matched subjects of this period. Hence, both subjects knew whether stealing was successful. To keep player 2 involved we offered her 1 EUR extra payoff in case she correctly guessed the decision of her matched player 1 in that period.[10]

[10] This opportunity was private information of player 2.

After five periods player-1 subjects were asked to take their time to fill in a survey on their emotions *when making their decisions*. Player-2 subjects were meanwhile asked to state their emotions *when seeing the decisions* of their matched player 1.

After the survey on emotions, the experiment continued for another five periods with some treatment change. As an alternative to a treatment change, additional sessions tried to trace out whether change in judicial procedure, i.e. simple ex-post communication between the player 1 and 2 (via an anonymous chat), is able to deter criminal activity. Such communication is a first step towards testing the effect of alternative procedure on deterrence, as suggested by proponents of restorative justice (e.g. Braithwaite 2002, Umbreit et al. 1994). Depending on the interpretation of means to ends, restorative justice can be regarded as an alternative for or improvement of retributive justice. Here the aim was to examine alternative ways of deterrence. In the results section we will briefly discuss the impact of a forced chat opportunity between player 1 and player 2.

**Table 2.** Session Overview.

| Session # | # of Subjects | Av. Age | % male | Part 1 | Part 2 1st half incentives | Part 2 1st half Chat? | Break | Part 2 2nd half incentives | Part 2 2nd half Chat? |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 28 | 24.1 | 64.3 | Real-Effort Slider Task | NoDeter | No | Self-Report of Emo-tions | DeterFine | No |
| 2 | 28 | 25.1 | 39.3 | | NoDeter | No | | DeterProb | No |
| 3 | 30 | 24.9 | 40.0 | | DeterFine | No | | DeterProb | No |
| 4 | 30 | 23.0 | 30.0 | | DeterFine | No | | DeterProb | No |
| 5 | 30 | 26.1 | 53.3 | | DeterProb | No | | DeterFine | No |
| 6 | 30 | 22.2 | 50.0 | | DeterProb | No | | DeterFine | No |
| 7 | 30 | 23.3 | 46.7 | | DeterFine | No | | NoDeter | No |
| 8 | 28 | 22.5 | 60.7 | | DeterProb | No | | NoDeter | No |
| 9 | 30 | 25.2 | 66.7 | | NoDeter | No | | NoDeter | Yes |
| 10 | 30 | 23.8 | 36.7 | | DeterFine | No | | DeterFine | Yes |
| 11 | 30 | 24.5 | 60.0 | | DeterProb | No | | DeterProb | Yes |
| 12 | 28 | 25.0 | 46.4 | | NoDeter | Yes | | NoDeter | No |
| 13 | 28 | 25.7 | 46.4 | | DeterFine | Yes | | DeterFine | No |
| 14 | 28 | 24.3 | 71.4 | | DeterProb | Yes | | DeterProb | No |
| Overall | 408 | 24.3 | 50.7 | | | | | | |

Note: The number of subjects per session is not perfectly equal over the fourteen sessions due to some registered individuals not showing up.

After all periods were played, one out of the ten periods was randomly selected for payment. Average payment over all treatments was 11.70 EUR. Table 2 summarizes the information for all 14 sessions.

### *2.3 Hypotheses*

The central aim of this study is to investigate the interaction of deterrence incentives and emotions. First, we use the stealing game and the three treatments *NoDeter*, *DeterFine* and *DeterProb* to establish the deterrence effect using a between-subject comparison. The underlying hypothesis is that stealing will be lower when an external deterrence institution is present than when it is not. This resembles Gary Becker's deterrence hypothesis. Accordingly, we formulate

**Hypothesis 1.** *Deterrence incentives cause lower average stealing, i.e.* $\bar{h}_{NoDeter} > \bar{h}_{DeterFine}$ *and* $\bar{h}_{NoDeter} > \bar{h}_{DeterProb}$.

Second, there has been a long-standing debate on the relative deterrence effectiveness and probabilities and fines (Garoupa 1997). On the one hand, Schildberg-Hörisch and Strassmair (2012) provide evidence that the both probabilities and fines work to reduce stealing and hold similar relative effectiveness and can be regarded as substitutes. On the other hand, Harbaugh et al. (2013) suggest that fines hold a stronger relative effectiveness compared to probabilities. One explanation may be risk aversion (Ehrlich 1973). To provide further evidence on this question, we include the two treatments *DeterFine* and *DeterProb* in order to investigate the relative effectiveness of increasing probabilities and fines as a function of stealing. Our design includes the novel feature that for the two treatments *expected punishment is equal for every stealing action.*

**Hypothesis 2.** *Probabilities and fines are substitutes for deterrence of stealing, i.e.* $\bar{h}_{DeterFine} = \bar{h}_{DeterProb}$.

Third, as indicated by the results of Schulze and Frank (2003), Schildberg-Hörisch and Strassmair (2012) and Khadjavi (2013), we assume that while deterrence incentives reduce average stealing, they also crowd out intrinsic motivation to act pro-socially. Note that we aim to test whether deterrence incentives in *DeterFine* and *DeterProb* make subjects put on their 'exchange mode hat'. If this is the case, then we should see that stealing in *NoDeter* is greater in the second five periods of the experiment if it is preceded by stealing in *DeterFine* or *DeterProb* beforehand. For this purpose we employ both between- and within-subject comparisons as laid out above. Hence, we formulate

**Hypothesis 3.** *Deterrence incentives crowd out intrinsic motivation for pro-social behavior. That is, stealing in NoDeter without any deterrence history is lower compared to stealing in NoDeter if preceded by DeterFine and/or DeterProb.*

Fourth, if we are able to establish that crowding out indeed occurs, then we may find evidence for the 'moral mode hat' or 'exchange mode hat' in the self-reported emotions of player-1 subjects before the regime change. We borrow the emotions we elicit in our experiment from Reuben and van Winden (2010). Amongst others, these emotions are shame and guilt. Gilbert (2003) and Haidt (2003) discuss and classify shame and guilt as moral, pro-social emotions. Hence, if deterrence incentives 'morally disengage' (Bowles and Polanía-Reyes 2012) player-1 subjects, then they should feel less shame and guilt when engaging in stealing in *DeterFine* and *DeterProb* compared to *NoDeter*. We hypothesize

**Hypothesis 4.** *Crowding out of intrinsic motivation occurs via changes in emotions when making decisions. In the act of stealing player-1 subjects in NoDeter exhibit stronger emotions such as shame and guilt compared to those in treatments DeterFine and DeterProb.*

Van Winden and Ash (2012) present theory on behavioral criminal law and economics. One of their contributions is the addition of emotions as a motivational factor of criminal activity. Our research tests their hypothesis. If incentives shape emotions of subjects, we should find differences in self-reported emotions, especially between *NoDeter* and *DeterFine/Prob*. Further, if these emotions then affect stealing, the regime changes should be able to pick up decision differences in the data.

# 3 Results

Before we formulate results concerning the hypotheses we aim at testing for the stealing game, we need to ensure comparability of sessions. For this purpose we use two Kruskal-Wallis tests. The first test analyzes the equality of populations with regard to effort levels in the real-effort slider task. The test does not reject equality of all fourteen sessions ($p = 0.4227$, for histograms see Figure A.1 in appendix A). That is, subjects in all fourteen sessions exerted similar effort in the task. The second Kruskal-Wallis test examines the perceived fairness of the slider task, which was reported on a five-point scale before subjects received information on their relative position and on the stealing game. This second test also does not reject the null hypothesis of the equality of populations across all fourteen sessions ($p = 0.4985$, for histograms see Figure A.2 in appendix A). Hence, we are able to conduct our analysis with respect to stealing in our three treatments *NoDeter*, *DeterFine* and *DeterProb*.

## *3.1 The Deterrence Hypothesis and Crowding Out of Pro-Social Behavior*

Let us next analyze the magnitudes of stealing in a between-subject comparison of deterrence schemes in the first five periods without ex-post chat opportunity. Using average stealing for Mann Whitney tests (yielding one observation per individual), we find significantly higher stealing of 7.71 EUR in *NoDeter* ($n = 43$) compared to both *DeterFine* (4.17 EUR, $n = 60$) and *DeterProb* (4.75 EUR, $n = 59$). Both deterrence treatments yield lower stealing at $p < 0.0000$. Hence, our results firmly support Becker's deterrence hypothesis (our Hypothesis 1).

Comparing *DeterFine* and *DeterProb* (with the same expected punishment for every stealing decision), we find no treatment effects (p = 0.2114). The fact that we cannot reject the null hypothesis that stealing is equally pronounced in *DeterFine* and *DeterProb* supports Hypothesis 2, i.e. substitutability of probabilities and fines.[11] Figure 3 depicts average stealing in the three treatments in the first five periods. We summarize

**Result 1.** *Our experiment supports the deterrence hypothesis and supports the general result in the literature that deterrence works. Further, we find that punishment size and probability deter stealing equally effectively.*

Disentangling the deterrence effect from time effects and demographic information does not change these results. Table 3 reports random-effects regressions for the first five periods. We include the sessions with chat opportunity and control for individual periods, availability of the chat, age, gender, and subjects being students of business and economics or of law. None of these control variables reports significant influence on stealing behavior.[12] Note that there is no chat effect in our treatments. Hence, the hypothesis that ex post communication alone deters and mitigates a part of criminal activity does not hold. Consequently we can pool the data of chat and no-chat treatments in our later analysis.

[11] Note however that our treatments only consider fines and probabilities in a medium range. An investigation on extremely high or low probabilities and fines may be an avenue for future research.

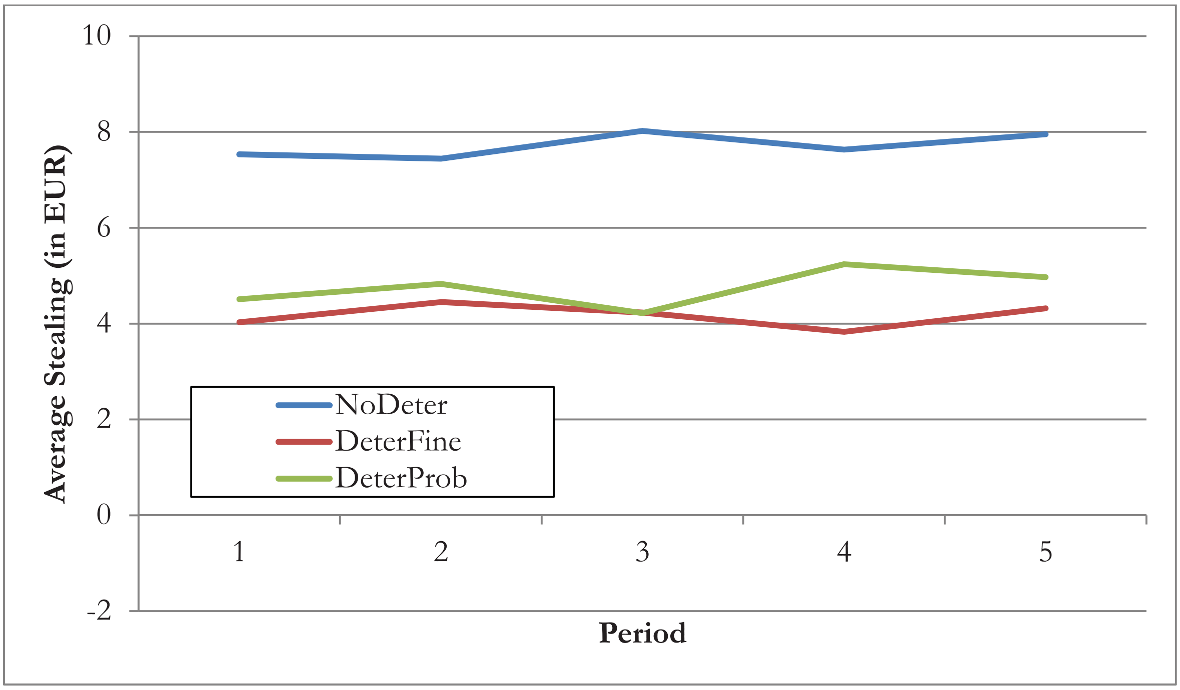
[12] There is one minor exception: specification IV reports higher stealing in period 5 (at the 10 percent level). This effect disappears when employing a time-lag of a subject's stealing in t-1.

**Table 3.** Linear Random-Effects Regressions: Testing for Treatment Effects in Criminals' Behavior, First Five Periods.

| Independent Variable | Dependent Variable: *h* (magnitude of stealing) | | | | |
|---|---|---|---|---|---|
| | I (*without chat*) | II | III | IV | V |
| $h_{i,t-1}$ (continuous, time lag) | | | | | 0.482*** (0.050) |
| Caught stealing in previous period (dummy, time lag) | | | | | 0.223 (0.255) |
| DeterFine (dummy) | -3.543*** (0.467) | -3.656*** (0.400) | -3.796*** (0.396) | -3.796*** (0.397) | -2.089*** (0.339) |
| DeterProb (dummy) | -2.964*** (0.496) | -3.038*** (0.432) | -3.195*** (0.438) | -3.195*** (0.438) | -1.684*** (0.328) |
| Chat (dummy) | | 0.289 (0.387) | 0.259 (0.386) | 0.259 (0.386) | 0.127 (0.213) |
| Age (continuous) | | | -0.027 (0.044) | -0.027 (0.044) | -0.007 (0.024) |
| Male (dummy) | | | 0.511 (0.348) | 0.511 (0.349) | 0.148 (0.201) |
| Business or economics student (dummy) | | | 0.326 (0.358) | 0.326 (0.358) | 0.145 (0.209) |
| Law student (dummy) | | | 0.685 (0.691) | 0.685 (0.692) | 0.400 (0.410) |
| Period2 (dummy) | | | | 0.272 (0.220) | |
| Period3 (dummy) | | | | 0.153 (0.192) | -0.251 (0.301) |
| Period4 (dummy) | | | | 0.124 (0.222) | -0.227 (0.215) |
| Period5 (dummy) | | | | 0.436* (0.238) | 0.097 (0.258) |
| Constant | 7.716*** (0.365) | 7.785*** (0.328) | 8.159*** (1.108) | 7.962*** (1.113) | 4.292*** (0.793) |
| Individuals | 162 | 204 | 202 | 202 | 202 |
| Observations | 810 | 1020 | 1010 | 1010 | 808 |

Note: An observation is a subject's magnitude of criminal activity in a period. Treatment *NoDeter* is the baseline. Two subjects did not enter their demographic information, so that they are excluded from analysis with demographics (regressions III to V). Robust standard errors (clustered at the individual level) in parentheses, significance: *p < 0.10, **p < 0.05, ***p < 0.01.

**Figure 3.** Average Stealing, First Five Periods without Chat.




In specification V we also include a time-lag of the magnitude of stealing and a dummy variable that indicates when subjects were caught stealing in t-1. As a by-product, specification V thus provides an opportunity to test for the so-called *gambler's fallacy* (Tversky and Kahneman 1974, Clotfelter and Cook 1993, Sundali and Croson 2006).[13] We find that subjects appear to be consistent in their choice of stealing over time; the time-lag of stealing is highly significant and subjects do not show a change in behavior when caught in the former period. We therefore formulate the two findings concerning the chat and the gambler's fallacy:

**Result 2.** *An ex-post chat opportunity of matched player 1 and player 2 does not change stealing behavior.*

**Result 3.** *Player-1 subjects do not change behavior if caught stealing in a previous period. Hence, there appears to be only minor or no biased decision making suffering from the gambler's fallacy.*

[13] Let us consider an example for the gambler's fallacy: Assume an urn that is filled with 50 black and 50 white balls. Consider drawing from this urn with replacement. The gambler's fallacy describes the bias that an individual may perceive the likelihood for a black ball to be drawn as greater than 50 percent (the objective likelihood) if black balls have not been drawn for a relatively long sequence of draws. Thus, the gambler's fallacy ignores the independence of draws.

So far, we have reported results from a between-subject-design analysis. When decision making in the first five periods was over, the experiment took a break to ask subjects for their self-reported emotions *when making the stealing decision*.[14] Before we have a closer look at these emotions we examine the effects of regime change. Here we take advantage of a within-subject design feature of our experiment. It may well be that decision making in a treatment for five periods establishes a norm that carries over to another treatment. Therefore, we are able to test two relationships: first, we test whether the deterrence hypothesis holds when there is no prior punishment. That is, we test whether subjects steal less in *DeterFine* and *DeterProb* after being in *NoDeter* for five periods. Second and conversely, we also test whether deterrence continues to work after the punishment incentives have been removed.

We find that our deterrence effects are the same whether or not player-1 subjects have a history of no deterrence for both *DeterFine* and *DeterProb* (sessions # 1 and 2 respectively, Wilcoxon signed-ranks tests, both tests yield $p < 0.0000$).[15] For the inverted case in which criminals face deterrence and these incentives are then removed, stealing increases significantly (sessions # 7 and 8, Wilcoxon signed-ranks test, $p < 0.0000$). Figures 4 and 5 provide graphs of these effects of regime change. Figure 5 also includes graphs of mean stealing in *NoDeter* in the first five periods (sessions # 1, 2, 9 and 12) and the second five periods without prior deterrence (sessions # 9 and 12). We find that stealing in the treatments with deterrence history is significantly *greater* than without such history (Mann Whitney test, $p = 0.0495$). This finding provides evidence for the crowding out of pro-social behavior by deterrence incentives and supports Hypothesis 3:

[14] Player-2 subjects (victims) were similarly asked for their emotions in the moment they received information on the stealing decisions of matched player-1 subjects. For instance, findings include that player-2 subjects exhibit greater anger, envy and irritation in *DeterFine/Prob* compared to *NoDeter*, even when controlling for the mean amount stolen from them. Such findings suggest that institutions (or procedures) influence emotional costs independent from monetary outcomes.

[15] In accordance with our data structure, we compare mean stealing of a player 1 before and after the regime change. Hence, we keep two observations (means) per individual for our analysis.

**Result 4.** *Deterrence incentives crowd out pro-social behavior. That is, stealing in NoDeter is significantly greater after facing deterrence institutions than without such a history.*

**Figure 4.** Average Stealing, Sessions *NoDeter_DeterFine* and *NoDeter_DeterProb*.

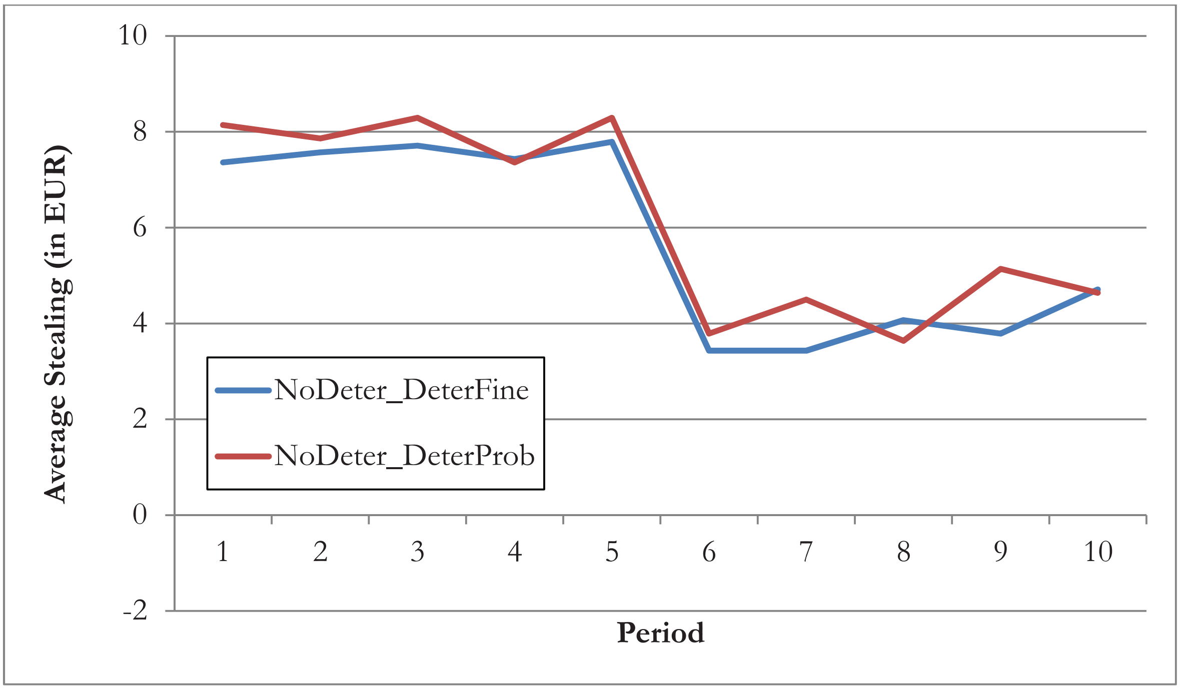
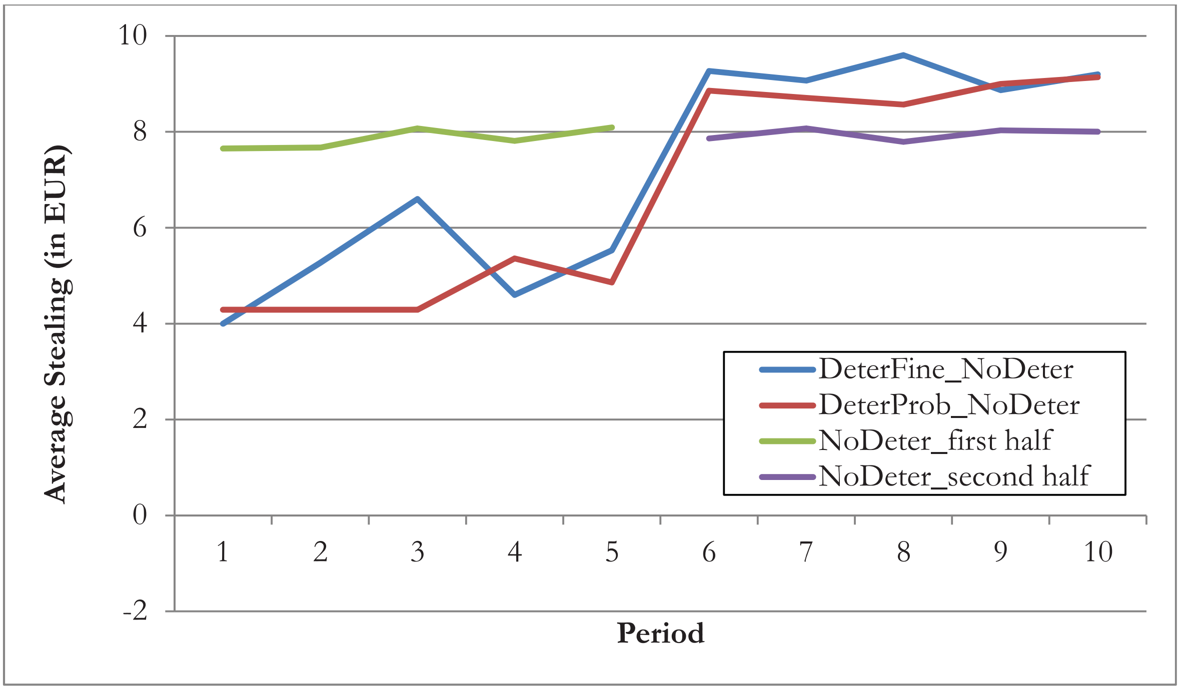



**Figure 5.** Average Stealing, Sessions *DeterFine_NoDeter* and *DeterProb_NoDeter*, Compared to *NoDeter* Stealing in other Sessions.



### *3.2 Incentives, Emotions and Behavior*

We now turn to the emotions in order to investigate their impact on behavior. Note again that after the first five periods, and before a treatment change was introduced, player-1 subjects were asked to self-report emotions they exhibited *when making the decision*. Self-reports were done on a seven-point scale.[16]

The results show that whether or not a deterrence scheme is in place (*NoDeter* vs. *DeterFine* and *DeterProb*) indeed influences emotions of criminals. That is, emotion differences mostly exist between *NoDeter* and *DeterFine* as well as *NoDeter* and *DeterProb*. Meanwhile there are few to no differences between emotions in *DeterFine* and *DeterProb*. Descriptive statistics and pairwise Mann-Whitney test for treatment differences are report in Table 4.

For instance, on average player-1 subjects in *DeterFine* and *DeterProb* feel more anger (2.63 and 2.71 points respectively) and more envy (3.28 and 3.36 points respectively) when making their decisions compared to player-1 subjects in *NoDeter* (on average 1.56 points of anger and 1.79 points of envy). On contrary, player-1 subjects in *NoDeter* on average feel more guilt (2.72 points) and shame (2.95 points), but also more gratitude (3.16 points) compared player 1 subjects in *DeterFine* (1.42 points of guilt, 1.93 points of shame, and 2.02 points of gratitude) and *DeterProb* (2.00 points of guilt, 1.97 points of shame, and 2.19 points of gratitude). All these differences are statistically significant at conventional levels based on pairwise Mann Whitney tests. More information can be found in Table 4.

[16] The list of 16 emotions corresponds directly to the list of Reuben and van Winden (2010). The aim was to elicit a balanced and neutral list of emotions, some of which are expected to be influenced (like anger, envy and guilt) and some which are not (like surprise and gloating).

**Table 4.** Descriptive Statistics and Pairwise Mann-Whitney Tests for Equality of Emotion Intensity of Player 1 Subjects.

| Emotions | Descriptive Statistics: Mean & Std. Dev. in parentheses | | | Mann-Whitney Test Results: p-values in parentheses | | |
|---|---|---|---|---|---|---|
| | *NoDeter* | *DeterFine* | *DeterProb* | NoD vs. DFi | NoD vs. DPr | DFi vs. DPr |
| Anger | 1.56 (1.22) | 2.63 (1.80) | 2.71 (1.81) | NoD < DFi (p = 0.001) | NoD < DPr (p = 0.000) | DFi = DPr (p = 0.911) |
| Shame | 2.95 (1.88) | 1.93 (1.45) | 1.97 (1.27) | NoD > DFi (p = 0.003) | NoD > DPr (p = 0.007) | DFi = DPr (p = 0.585) |
| Pride | 2.16 (1.53) | 2.37 (1.67) | 1.90 (1.49) | NoD = DFi (p = 0.523) | NoD = DPr (p = 0.309) | DFi > DPr (p = 0.071) |
| Disappoint-ment | 2.05 (1.59) | 3.97 (2.23) | 3.98 (1.95) | NoD < DFi (p = 0.000) | NoD < DPr (p = 0.000) | DFi = DPr (p = 0.891) |
| Surprise | 2.84 (2.06) | 2.73 (1.97) | 2.90 (1.93) | NoD = DFi (p = 0.986) | NoD = DPr (p = 0.752) | DFi = DPr (p = 0.646) |
| Joy | 4.14 (2.02) | 3.35 (1.89) | 3.15 (1.92) | NoD > DFi (p = 0.047) | NoD > DPr (p = 0.016) | DFi = DPr (p = 0.541) |
| Contempt | 1.51 (1.01) | 1.57 (1.16) | 1.93 (1.62) | NoD = DFi (p = 0.890) | NoD = DPr (p = 0.269) | DFi = DPr (p = 0.289) |
| Envy | 1.79 (1.47) | 3.28 (2.01) | 3.36 (2.00) | NoD < DFi (p = 0.000) | NoD < DPr (p = 0.000) | DFi = DPr (p = 0.901) |
| Irritation | 1.53 (1.10) | 3.40 (2.12) | 3.44 (2.07) | NoD < DFi (p = 0.000) | NoD < DPr (p = 0.000) | DFi = DPr (p = 0.967) |
| Guilt | 2.72 (1.91) | 1.42 (1.00) | 2.00 (1.52) | NoD > DFi (p = 0.000) | NoD > DPr (p = 0.046) | DFi < DPr (p = 0.024) |
| Regret | 2.70 (1.54) | 2.88 (1.87) | 3.15 (1.97) | NoD = DFi (p = 0.829) | NoD = DPr (p = 0.311) | DFi = DPr (p = 0.535) |
| Admiration | 1.79 (1.42) | 1.33 (0.95) | 1.69 (1.24) | NoD > DFi (p = 0.030) | NoD = DPr (p = 0.789) | DFi < DPr (p = 0.041) |
| Pity | 2.93 (1.86) | 1.43 (0.96) | 1.59 (1.07) | NoD > DFi (p = 0.000) | NoD > DPr (p = 0.000) | DFi = DPr (p = 0.360) |
| Gratitude | 3.16 (1.98) | 2.02 (1.55) | 2.19 (1.58) | NoD > DFi (p = 0.003) | NoD > DPr (p = 0.008) | DFi = DPr (p = 0.633) |
| Gloating | 1.70 (1.17) | 1.68 (1.28) | 1.73 (1.34) | NoD = DFi (p = 0.583) | NoD = DPr (p = 0.820) | DFi = DPr (p = 0.723) |
| Sadness | 1.81 (1.48) | 2.57 (1.85) | 2.44 (1.58) | NoD < DFi (p = 0.016) | NoD < DPr (p = 0.006) | DFi = DPr (p = 0.894) |
| Observations | 43 | 60 | 59 | | | |

Note: Subjects self-reported emotions on a seven-point scale. The list of emotions corresponds to Reuben and van Winden (2010).

So far, we have established a number of links in our experiment. The first link that we can support is (the classic one) that incentives influence behavior. Naturally then, behavior influences outcomes, as they are directly deterministically or stochastically related. The self-reported emotions of player-1 subjects in the three treatments demonstrate differences with respect to both pro-social emotions (like shame, guilt and gratitude) and self-centered emotions (like envy).

Note again that these subjects were asked about the emotions they exhibited *when making their decisions*. Accordingly, we are able to establish a link between the treatments and emotions. One interpretation is that there is also an intermediate link via the change from player 1 being in the 'moral mode' to being in the 'exchange mode' (Meier 2007). As we have no way to directly elicit subjects' 'modes', we postpone further explanations of this matter to future research.

In the last part of this results section, we will show that there is also a link between emotions and stealing behavior. In combination with evidence that stealing in *DeterFine* and *DeterProb* does not differ significantly, we pool subjects in *NoDeter* (n = 57) and subjects in *DeterFine* and *DeterProb* (n = 147) in two groups. Principle component analyses (PCAs) use these emotion self-reports to construct emotion clusters.[17]

Table 5 reports the emotion clusters for *NoDeter* and Table 6 reports emotion clusters for *DeterFine* and *DeterProb*. All of the analyzed components (emotion clusters) feature eigenvalues of about 1 and above; this feature commonly defines the usefulness of a component. The clusters relate to different families of emotions: 'negative emotions', 'positive emotions', 'pro-social emotions', 'self-centered emotions', and 'vicious emotions'. These families correspond to those in Reuben and van Winden (2010). The values of the different emotions in a given cluster signal how strongly the respective emotion is correlated positively or negatively with the cluster. For instance, the cluster of 'pro-social emotions' for *NoDeter* (in Table 5) is most highly and positively (proportionally) correlated with gratitude and admiration. Conversely, the cluster of 'self-centered emotions' for *DeterFine/Prob* (in Table 6) is most highly and positively (proportionally) correlated with pride and joy and most highly and negatively (anti-proportionally) correlated with guilt, shame and pity.

[17] This step follows Reuben and van Winden (2010). It is motivated by the large number of pairwise correlations of emotions as shown in tables A.1 and A.2 in appendix A. The aim is to organize these emotions in clusters.

**Table 5.** Principal Emotion Components of Player-1 Subjects, *NoDeter*.

| Emotions | Components for NoDeter | | | | |
|---|---|---|---|---|---|
| | 1st<br>Eigenvalue: 6.82<br>Proportion: 0.43 | 2nd<br>Eigenvalue: 2.15<br>Proportion: 0.13 | 3rd<br>Eigenvalue: 1.25<br>Proportion: 0.08 | 4th<br>Eigenvalue: 1.19<br>Proportion: 0.07 | 5th<br>Eigenvalue: 1.03<br>Proportion: 0.06 |
| Anger | 0.2626 | -0.2002 | 0.3569 | -0.2586 | -0.0888 |
| Shame | 0.3090 | 0.0755 | -0.0464 | -0.2269 | -0.2165 |
| Pride | 0.1209 | 0.2396 | 0.6652 | 0.1095 | 0.0243 |
| Disappointment | 0.3068 | -0.1381 | -0.1333 | 0.2277 | 0.0916 |
| Surprise | 0.2379 | 0.1324 | -0.2756 | 0.2507 | -0.3102 |
| Joy | 0.0340 | 0.5235 | 0.2698 | 0.0674 | -0.0386 |
| Contempt | 0.2979 | -0.0660 | 0.0209 | 0.1739 | 0.4604 |
| Envy | 0.2801 | -0.2288 | 0.2913 | -0.1350 | -0.0477 |
| Irritation | 0.3176 | -0.3016 | 0.0771 | 0.1482 | -0.0482 |
| Guilt | 0.2666 | 0.1643 | 0.0217 | -0.3541 | -0.1737 |
| Regret | 0.2550 | 0.1102 | -0.3209 | -0.2520 | 0.1127 |
| Admiration | 0.2955 | -0.0313 | -0.1051 | 0.3671 | 0.0836 |
| Pity | 0.1528 | 0.3610 | -0.2191 | -0.3853 | -0.0405 |
| Gratitude | 0.1583 | 0.3738 | -0.0258 | 0.4481 | -0.3300 |
| Gloating | 0.1575 | 0.3169 | -0.0369 | -0.0480 | 0.6734 |
| Sadness | 0.3231 | -0.1671 | -0.0753 | -0.0695 | -0.0897 |
| | | | | | |
| Family of emotions | negative | positive | self-centered | pro-social | vicious |
| Mean | <0.00 | <0.00 | <0.00 | <0.00 | <0.00 |
| Std. dev. | 2.61 | 1.47 | 1.12 | 1.09 | 1.01 |

Note: Principal Component Analysis of criminals' emotions in treatment *NoDeter* with and without chat. The five components account for 77.73 percent of variation.
n = 57. KMO: 0.8092.

Interestingly, similar emotion clusters are constructed by the two PCAs. We use these clusters to explain stealing in the first five periods of our experiment. Table 7 reports two OLS regressions, specification VI to explain stealing in treatment *NoDeter* and specification VII to explain stealing in treatments *DeterFine* and *DeterProb*. In these regressions we include all five clusters of emotions (while controlling for 'chat' in both specifications and '*DeterFine*' in specification VII).

**Table 6.** Principal Emotion Components of Player-1 Subjects, *DeterFine* and *DeterProb*.

| Emotions | Components for *DeterFine* and *DeterProb* | | | | |
|---|---|---|---|---|---|
| | 1st<br>Eigenvalue: 4.48<br>Proportion: 0.28 | 2nd<br>Eigenvalue: 2.30<br>Proportion: 0.14 | 3rd<br>Eigenvalue: 1.65<br>Proportion: 0.10 | 4th<br>Eigenvalue: 1.01<br>Proportion: 0.06 | 5th<br>Eigenvalue: 0.96<br>Proportion: 0.06 |
| Anger | 0.2782 | -0.2432 | 0.1805 | 0.1649 | -0.1811 |
| Shame | 0.2518 | 0.0959 | -0.3925 | -0.0084 | 0.0618 |
| Pride | 0.1399 | 0.3319 | 0.4073 | 0.0211 | 0.0242 |
| Disappointment | 0.3413 | -0.1689 | 0.1339 | -0.2749 | 0.0676 |
| Surprise | 0.2486 | 0.1526 | -0.0279 | -0.4632 | -0.2715 |
| Joy | 0.0530 | 0.4897 | 0.3199 | -0.1500 | 0.1724 |
| Contempt | 0.2867 | 0.0090 | -0.0970 | 0.2763 | -0.2970 |
| Envy | 0.2973 | -0.1242 | 0.2239 | 0.1141 | 0.0924 |
| Irritation | 0.3426 | -0.2850 | 0.2123 | -0.0134 | 0.1109 |
| Guilt | 0.2706 | 0.0581 | -0.4395 | 0.0964 | 0.1439 |
| Regret | 0.3157 | -0.0227 | -0.1592 | -0.2092 | 0.2762 |
| Admiration | 0.1976 | 0.1635 | -0.1527 | 0.1938 | -0.6600 |
| Pity | 0.1119 | 0.3040 | -0.3484 | 0.1669 | 0.3825 |
| Gratitude | 0.0920 | 0.4889 | 0.0207 | -0.2363 | -0.1674 |
| Gloating | 0.1362 | 0.2508 | 0.2247 | 0.6268 | 0.1543 |
| Sadness | 0.3451 | -0.0905 | 0.0876 | -0.0550 | 0.1226 |
| | | | | | |
| Family of emotions | negative | positive | self-centered | vicious | pro-social |
| Mean | <0.00 | <0.00 | <0.00 | <0.00 | <0.00 |
| Std. dev. | 2.12 | 1.52 | 1.29 | 1.01 | 0.98 |

Note: Principal Component Analysis of criminals' emotions in treatments *DeterFine* and *DeterProb* with and without chat. The five components account for 64.99 percent of variation.
n = 147. KMO: 0.7790.

When analyzing the impact of these clusters on stealing behavior, we find that, on average, the more pronounced the 'pro-social emotions' of player 1, the *smaller* is the amount she wants to steal in *NoDeter* (specification VI, significant at the 5 percent level). The four other emotion cluster do not impact her decision significantly. On the contrary, in *DeterFine* and *DeterProb* the more pronounced 'self-centered emotions' the *greater* is the amount player 1 wants to steal (specification VII, significant at the 5 percent level). Thus not only do our treatments influence the emotions of player-1 subjects. In addition, we find evidence that the existence of an external deterrence institution nullifies the explanatory power of pro-social emotions while it activates the explanatory power of self-centered emotions. We therefore provide evidence that emotions impact stealing behavior.[18] We summarize:

**Result 5.** *The impact of incentives on emotions is twofold. First, deterrence incentives change the intensity of player 1's emotions. Second, incentives (de)activate the emotions which are relevant for stealing behavior.*

**Table 7.** OLS Regressions: Principal Emotion Components' Impact on Stealing.

| | Dependent variable: *h* (magnitude of stealing) | |
|---|---|---|
| | VI | VII |
| Independent Variable | *NoDeter* | *DeterFine & DeterProb* |
| Negative emotions | 0.059<br>(0.119) | 0.016<br>(0.093) |
| Positive emotions | 0.146<br>(0.214) | -0.043<br>(0.130) |
| Self-centered emotions | -0.007<br>(0.280) | 0.344**<br>(0.156) |
| Pro-social emotions | -0.621**<br>(0.292) | 0.142<br>(0.204) |
| Vicious emotions | 0.363<br>(0.320) | 0.289<br>(0.196) |
| Chat | 0.110<br>(0.766) | 0.320<br>(0.514) |
| *DeterFine* | | -0.723*<br>(0.399) |
| Constant | 7.829***<br>(0.362) | 4.794***<br>(0.297) |
| Individuals | 57 | 147 |

Note: Each observation is the *average of the magnitude* of stealing of a player 1 subject over the first five periods. Standard errors in parentheses, significance: *p < 0.10, **p < 0.05, ***p < 0.01.

[18] One may argue that it is not clear whether we identify a correlation or causality. We acknowledge that this is a valid concern. The question is whether we collected information on emotions exhibited *before or during* the decision or *after* the decision. When designing our research, we very carefully formulated the question to subjects to rate to what extend they exhibited 16 emotions '*when making their decisions*'. Consequently, our data should pick up the former emotions, isolate the direction and thereby establish causality. In cited articles like Bosman and van Winden (2002), Cubitt et al. (2011) and Reuben and van Winden (2010) this is the common practice.
Nevertheless research like Rick and Loewenstein (2008) describes the interaction of emotions and behavior as a sophisticated and recursive system. While we share this point of view, we cannot think of a cleaner way to collect this data and to identify the direction than the used careful and specific questions in our controlled environment.

Our results illustrate that incentives influence behavior both directly and indirectly: directly via incentives to steal less and indirectly via emotions that turn the focus away from a pro-social, other-regarding orientation to a self-centered orientation. Or in Meier (2007)'s terms from the 'moral mode' to the 'exchange mode'. Hence, we find that one explanation for crowding out of pro-social behavior is the (de)activation of different emotions. To the best of our knowledge this study is the first to provide evidence on interaction of deterrence, crowding out and emotions.

## 4 Concluding Discussion

Beginning with Gary Becker's seminal work in 1968, the beneficial feature of deterrence incentives to reduce social welfare loss from crime has been well documented over the last fifty years. We employ experimental economic methods to provide new insights into their detrimental effect of crowding out of pro-social behavior. Further, we investigate the role of emotions for stealing decisions and the interaction of incentives with emotions. Our experiment asks whether crowding out acts through emotional motivation.

Our results support the (direct) beneficial feature of deterrence incentives. Yet our results also suggest that crowding out of pro-social behavior is an undesired by-product. Our analysis shows that incentives indeed (de)activate different clusters of emotions that then motivate behavior. Consequently, we offer support for prior findings and theories of individual decision making that integrate emotions. Van Winden and Ash (2012) provide such a theory on behavioral criminal law and economics that incorporates emotions. Emotional motivation appears to be a vehicle of crowding out of pro-social behavior. Our study therefore connects strands of literature in behavioral economics and social psychology: we regard our findings as quantitative and measureable support for Bowles and Polanía-Reyes' (2012) illustration of *moral disengagement* as a cause of crowding out. That is, deterrence incentives decrease and deactivate pro-social emotions of player-1 subjects. Likewise, our findings may be interpreted in the light of the self-determination theory in social psychology (Deci and Ryan 1985): one may interpret the creation of anger and envy as well as the activation of self-centered emotions among player-1 subjects as evidence for the violation of subjects' need for autonomy.

In the spirit of Gneezy and Rustichini (2000b)'s title "Pay enough or don't pay at all", we invert this suggestion and propose to 'punish enough or don't punish at all'. This idea finds support in results of Schildberg-Hörisch and Strassmair (2012) who find that deterrence works for high fines but not for low fines (likely due to crowding out). One caveat of our study is that we analyze a two-player game. Moral implications and emotions of an action may change behavior in the way we document it in this study. Whether the same is true for an interaction of an individual with a group or firm where no direct, single victim can be identified, pro-social emotions may be less relevant for behavior. There is clearly room for future research on this matter. We hope that this study is able to motivate further research on the interaction of emotions, extrinsic and intrinsic motivation and decision making.

## References


Abbink, K., B. Irlenbusch and E. Renner (2002). "An Experimental Bribery Game". *Journal of Law, Economics and Organization*, vol. 18, pp. 428-454.

Anderson, L.R. and S.L. Stafford (2003). "Punishment in a Regulatory Setting: Experimental Evidence from the VCM". *Journal of Regulatory Economics*, vol. 24, pp. 91-110.

Andrade, E.B. and D. Ariely (2009). "The enduring impact of transient emotions on decision making". *Organizational Behavior and Human Decision Processes*, vol. 109, pp. 1-8.

Bardsley, N. (2008). "Dictator game giving: altruism or artifact?". *Experimental Economics*, vol. 11, pp. 122-133.

Becker, G.S. (1968). "Crime and Punishment: An Economic Approach". *The Journal of Political Economy*, vol. 76, pp. 169-217.

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

Bosman, R. and F. van Winden (2002). "Emotional hazard in a power-to-take experiment". *The Economic Journal*, vol. 112, pp. 147-169.

Bowles, S. (1998). "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions". *Journal of Economic Literature*, vol. 36, pp. 75-111.

Bowles, S. and S. Polanía-Reyes (2012). "Economic Incentives and Social Preferences: Substitutes or Complements?". *Journal of Economic Literature*, vol. 50, pp. 368-425.

Braithwaite, J. (2002). *Restorative Justice and Responsive Regulation*. New York: Oxford University Press.

Brekke, K.A., S. Kverndokk and K. Nyborg (2003). "An economic model of moral motivation". *Journal of Public Economics*, vol. 87, pp. 1967-1983.

Charness, G., U. Gneezy and M.A. Kuhn (2013). "Experimental methods: Extra-laboratory experiments – extending the reach of experimental economics". *Journal of Economic Behavior & Organization*, vol. 91, pp. 93-100.

Charness, G. and M. Rabin (2002). "Understanding Social Preferences with Simple Tests". *The Quarterly Journal of Economics*, vol. 117, pp. 817-869.

Clotfelter, C.T. and P.J. Cook (1993). "The Gamber's Fallacy in Lottery Play". *Management Science*, vol. 39, pp. 1521-1525.

Corman, H. and H.N. Mocan (2000). "A Time-Series Analysis of Crime, Deterrence, and Drug Abuse in New York City". *The American Economic Review*, vol. 90, pp. 584-604.

Cubitt, R.P., M. Drouvelis and S. Gächter (2011). "Framing and free riding: emotional responses and punishment in social dilemma games". *Experimental Economics*, vol. 14, pp. 254-272.

DeAngelo, G. and G. Charness (2012). "Deterrence, expected cost, uncertainty and voting: Experimental evidence". *Journal of Risk and Uncertainty*, vol. 44, pp. 73-100.

Deci, E.L. (1971). "Effects of externally mediated rewards on intrinsic motivation." *Journal of Personality and Social Psychology*, vol. 18, pp. 105-115.

Deci, E.L. and R.M. Ryan (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Deci, E.L., R. Koestner and R.M. Ryan (1999). "A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation". *Psychological Bulletin*, vol. 125, pp. 627-668.

Ehrlich, I. (1973). "Participation in illegitimate activities: a theoretical and empirical investigation". *Journal of Political Economy*, vol. 81, pp. 521-565.

Elster, J. (1998). "Emotions and Economic Theory". *Journal of Economic Literature*, vol. 36, pp. 47-74.

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

Fischbacher, U. (2007). "Z-Tree: Zurich Toolbox for Ready-made Economic Experiments". *Experimental Economics*, vol. 10, pp. 171-178.

Fisman, R. and E. Miguel (2007). "Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets". *Journal of Political Economy*, vol. 115, pp. 1020-1048.

Frank, R.H. (1988). Passions within Reason – The Strategic Role of the Emotions. New York: Norton.

Frey, B.S. and R. Jegen (2001). "Motivation crowding theory". *Journal of Economic Surveys*, vol. 15, pp. 589-611.

Friesen, L. (2012). "Certainty of Punishment versus Severity of Punishment: An Experimental Investigation". *Southern Economic Journal*, vol. 79, pp. 399-421.

Garoupa, N. (1997). "The theory of optimal law enforcement". *Journal of Economic Surveys*, vol. 111, pp. 267-295.

Garoupa, N. (2001). "Optimal magnitude and probability of fines". *European Economic Review*, vol. 45, pp. 1765-1771.

Geanakoplos, J., D. Pearce and E. Stacchetti (1989). "Psychological Games and Sequential Rationality". *Games and Economic Behavior*, vol. 1, pp. 60-79.

Gilbert, P. (2003). "Evolution, Social Roles, and the Differences in Shame and Guilt". *Social Research*, vol. 70, pp. 1205-1230.

Gill, D. and V. Prowse (2011). *A Novel Computerized Real Effort Task Based on Sliders*. IZA DP No. 5801.

Gill, D. and V. Prowse (2012). "A Structural Analysis of Disappointment Aversion in a Real Effort Competition". *The American Economic Review*, vol. 102, pp. 469-503.

Gneezy, U., S. Meier and P. Rey-Biel (2011). "When and Why Incentives (Don't) Work to Modify Behavior". *Journal of Economic Perspectives*, vol. 25, pp. 191-210.

Gneezy, U. and A. Rustichini (2000a). "A fine is a Price". *Journal of Legal Studies*, vol. 29, pp. 1-18.

Gneezy, U. and A. Rustichini (2000b). "Pay Enough or Don't Pay at All". *Quarterly Journal of Economics*, vol. 115, pp. 791-810.

Greiner, B. (2004). "An Online Recruitment System for Economic Experiments", in: Kremer, K. and V. Macho (eds.), *Forschung und wissenschaftliches Rechnen 2003*. GWDG Bericht 63, Göttingen: Ges. für Wiss. Datenverarbeitung, pp. 79-93.

Haidt, J. (2003). "The moral emotions". In: R.J. Davidson, K.R. Scherer and H.H. Goldsmith (eds.), *Handbook of Affective Sciences*, pp. 852-870.

Harbaugh, W.T., N. Mocan and M.S. Visser (2013). "Theft and Deterrence". *Journal of Labor Research*, vol. 34, pp. 389-407.

Harrison, G.W. and J.A. List (2004). "Field Experiments". *Journal of Economic Literature*, vol. 42, pp. 1009-1055.

Heyman, J. and D. Ariely (2004). "Effort for Payment: A Tale of Two Markets". *Psychological Science*, vol. 15, pp. 787-793.

Hopfensitz, A. and E. Reuben (2009). "The importance of emotions for the effectiveness of social punishment". *The Economic Journal*, vol. 119, pp. 1534-1559.

Huang, P.H. and H.-M. Wu (1994). "More order without More Law: A Theory of Social Norms and Organizational Cultures". *Journal of Law, Economics & Organization*, vol. 10, pp. 390-406.

Khadjavi, M. (2013). *Deterrence Works for Criminals*. Working Paper.

Kyckelhahn, T. and T. Martin (2013). *Justice Expenditure and Employment Extracts*. Bureau of Justice Statistics, U.S. Department of Justice, Washington DC.

Levitt, S.D. (1997). "Using Electoral Cycles and Police Hiring to Estimate the Effect of Police on Crime". *The American Economic Review*, vol. 87, pp. 270-290.

List, J.A. (2007). "On the Interpretation of Giving in Dictator Games". *Journal of Political Economy*, vol. 115, pp. 482-493.

List, J.A. and T.L. Cherry (2008). "Examining the role of fairness in high stakes allocation decisions". *Journal of Economic Behavior & Organization*, vol. 65, pp. 1-8.

Loewenstein, G. (2000). "Emotions in Economic Theory and Economic Behavior". *American Economic Review*, vol. 90, pp. 426-432.

Meier, S. (2007). "A Survey of Economic Theories and Field Evidence on Pro-Social Behavior". In: Frey, B.S. and A. Stutzer (eds.), *Economics and Psychology – A Promising New Cross-Disciplinary Field*. Cambridge: MIT Press, pp. 51-87.

Nagin, D.S. and G. Pogarsky (2003). "An experimental investigation of deterrence: cheating, self-serving bias, and impulsivity". *Criminology*, vol. 41, pp. 167-194.

Polinsky, A.M. and S. Shavell (1979). "The Optimal Tradeoff between the Probability and Magnitude of Fines". *American Economic Review*, vol. 69, pp. 880-891.

Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics". *The American Economic Review*, vol. 83, pp. 1281-1302.

Reuben, E. and F. van Winden (2010). "Fairness perceptions and prosocial emotions in the power to take". *Journal of Economic Psychology*, vol. 31, pp. 908-922.

Rick, S. and G. Loewenstein (2008). "The Role of Emotion in Economic Behavior". In: Lewis, M., J.M. Haviland-Jones and L. Feldman Barrett (eds.), *Handbook of Emotions*, Chapter 9. New York: Guilford Press.

Rizzolli, M. and L. Stanca (2012). "Judicial Errors and Crime Deterrence: Theory and Experimental Evidence". *Journal of Law and Economics*, vol. 55, pp. 311-338.

Schildberg-Hörisch, H. and C. Strassmair (2012). "An Experimental Test of the Deterrence Hypothesis". *Journal of Law, Economics, and Organization*, vol. 28, pp. 447-459.

Schulze, G.G. and B. Frank (2003). "Deterrence versus intrinsic motivation: Experimental evidence on the determinants of corruptibility". *Economics of Governance*, vol. 4, pp. 143-160.

Stigler, G.J. (1970). "The Optimum Enfocement of Laws". *Journal of Political Economy*, vol. 78, pp. 526-536.

Sundali, J. and R. Croson (2006). "Biases in casino betting: The hot hand and the gambler's fallacy". *Judgment and Decision Making*, vol. 1, pp. 1-12.

Tversky, A. and D. Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases". *Science*, vol. 185, pp. 1124-1131.

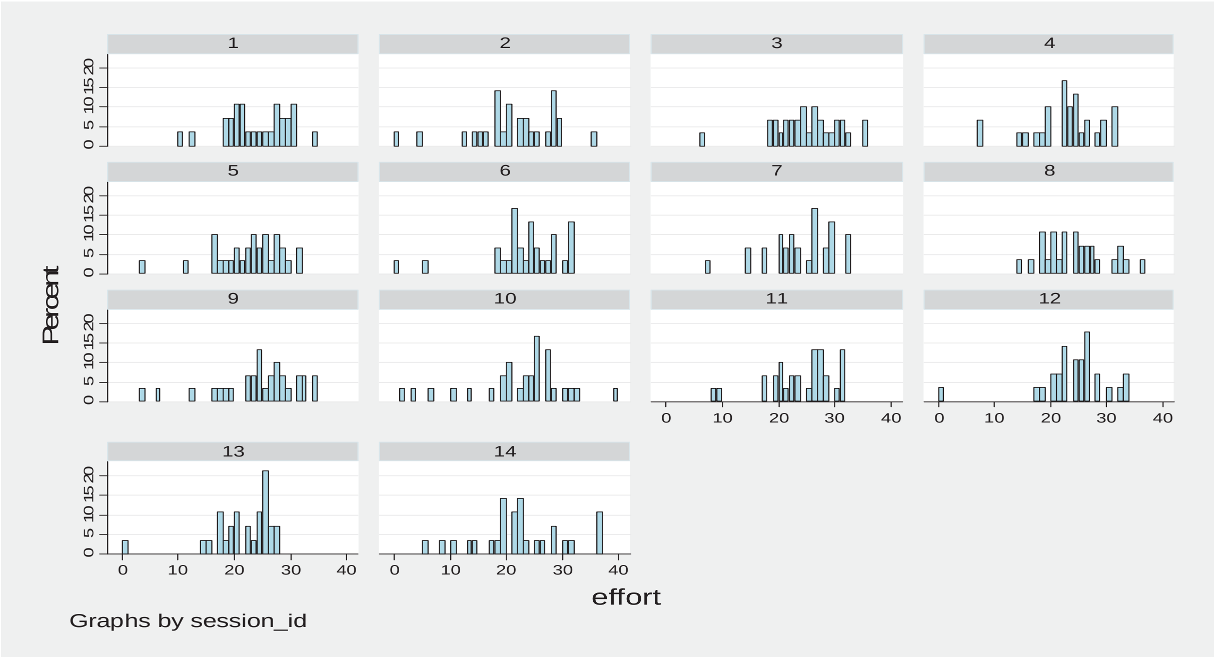
Umbreit, M.S., R.B. Coates and B. Kalanj (1994). *Victim Meets Offender: The Impact of Restorative Justice and Mediation*. Monsey: Criminal Justice Press.

van Winden, F. and E. Ash (2012). "On the Behavioral Economics of Crime". *Review of Law & Economics*, vol. 8, pp. 181-213.

Witte, A.D. (1980). "Estimating the Economic Model of Crime with Individual Data". *The Quarterly Journal of Economics*, vol. 94, pp. 57-84.

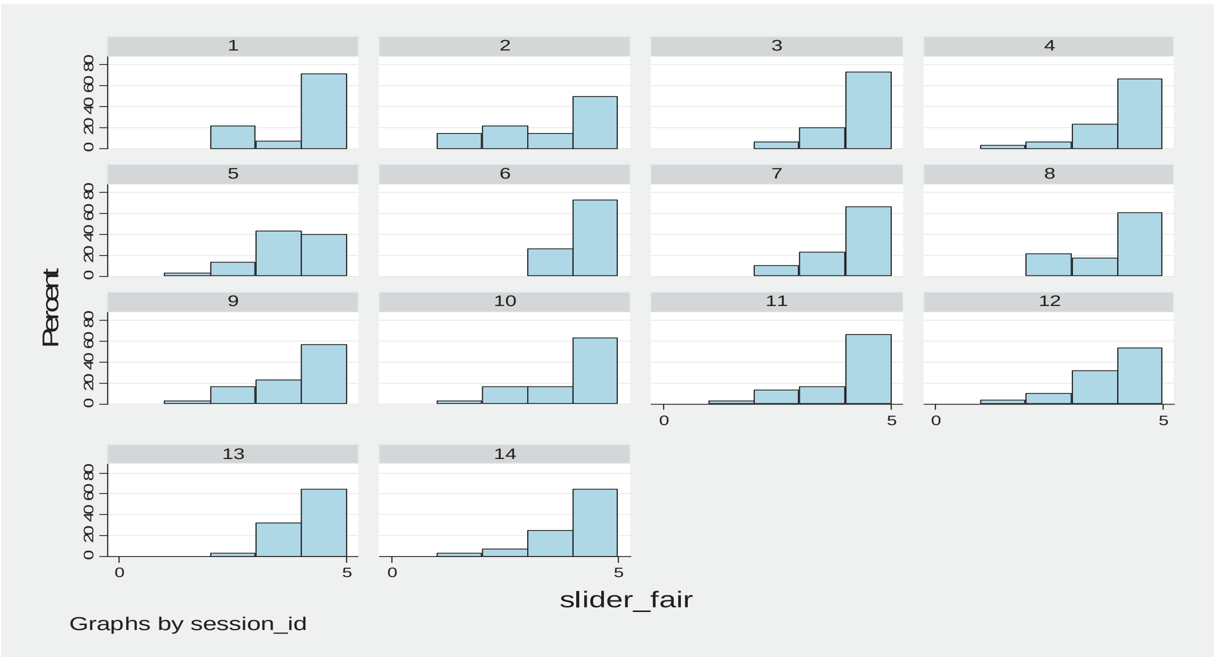
## Appendix A – Figures and Tables

**Figure A.1.** Histograms of Effort in the Real-Effort Slider-Task, per Session.




Note: A Kruskal-Wallis test cannot reject the null hypothesis of equality of populations (p = 0.4227). Mean effort over all sessions is 22.90.

**Figure A.2.** Histograms of Perceived Fairness of the Slider Task, per Session.




Note: A Kruskal-Wallis test cannot reject the null hypothesis of equality of populations (p = 0.4985). The answers were given on a five-point scale directly after the task and before subjects received information on their earnings in part 1. Mean perception of fairness over all sessions is 3.73.

**Table A.1.** Pairwise Correlation of Criminals' Emotions, Treatment *NoDeter*.

| Emotions | Anger | Shame | Pride | Disappointment | Surprise | Joy | Contempt | Envy | Irritation | Guilt | Regret | Admiration | Pity | Gratitude | Gloating |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Anger | 1.000 | | | | | | | | | | | | | | |
| Shame | **0.578** | 1.000 | | | | | | | | | | | | | |
| Pride | **0.309** | 0.187 | 1.000 | | | | | | | | | | | | |
| Disappointment | **0.444** | **0.581** | 0.110 | 1.000 | | | | | | | | | | | |
| Surprise | **0.280** | **0.478** | 0.095 | **0.431** | 1.000 | | | | | | | | | | |
| Joy | -0.066 | 0.135 | **0.369** | -0.084 | 0.107 | 1.000 | | | | | | | | | |
| Contempt | **0.492** | **0.485** | 0.258 | **0.661** | **0.390** | -0.015 | 1.000 | | | | | | | | |
| Envy | **0.717** | **0.503** | **0.266** | **0.573** | 0.243 | -0.047 | **0.512** | 1.000 | | | | | | | |
| Irritation | **0.689** | **0.587** | 0.135 | **0.743** | **0.457** | -0.166 | **0.715** | **0.775** | 1.000 | | | | | | |
| Guilt | **0.467** | **0.719** | **0.267** | **0.398** | **0.424** | 0.230 | **0.360** | **0.439** | **0.398** | 1.000 | | | | | |
| Regret | **0.279** | **0.547** | 0.053 | **0.528** | **0.385** | 0.145 | **0.474** | **0.360** | **0.382** | **0.558** | 1.000 | | | | |
| Admiration | **0.368** | **0.468** | 0.204 | **0.717** | **0.547** | 0.060 | **0.660** | **0.489** | **0.689** | **0.395** | **0.507** | 1.000 | | | |
| Pity | 0.185 | **0.416** | 0.124 | 0.134 | **0.309** | 0.218 | 0.176 | 0.175 | 0.030 | **0.381** | **0.456** | 0.154 | 1.000 | | |
| Gratitude | 0.022 | **0.366** | **0.318** | **0.331** | **0.486** | **0.401** | 0.220 | 0.085 | 0.193 | 0.247 | 0.147 | **0.384** | **0.342** | 1.000 | |
| Gloating | 0.128 | **0.287** | 0.223 | **0.265** | 0.204 | **0.289** | **0.538** | 0.113 | 0.092 | **0.319** | **0.309** | **0.277** | **0.388** | 0.213 | 1.000 |
| Sadness | **0.585** | **0.680** | 0.157 | **0.669** | **0.486** | -0.183 | **0.617** | **0.655** | **0.781** | **0.558** | **0.535** | **0.578** | **0.262** | 0.247 | 0.209 |

Note: Significant correlation between two emotions at the 5% level is noted in **bold** numbers.

**Table A.2.** Pairwise Correlation of Criminals' Emotions, Treatments *DeterFine* and *DeterProb*.

| Emotions | Anger | Shame | Pride | Disappointment | Surprise | Joy | Contempt | Envy | Irritation | Guilt | Regret | Admiration | Pity | Gratitude | Gloating |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Anger | 1.000 | | | | | | | | | | | | | | |
| Shame | **0.206** | 1.000 | | | | | | | | | | | | | |
| Pride | 0.124 | 0.058 | 1.000 | | | | | | | | | | | | |
| Disappointment | **0.489** | **0.270** | 0.109 | 1.000 | | | | | | | | | | | |
| Surprise | **0.194** | **0.285** | 0.160 | **0.331** | 1.000 | | | | | | | | | | |
| Joy | -0.147 | -0.024 | **0.518** | 0.029 | **0.209** | 1.000 | | | | | | | | | |
| Contempt | **0.285** | **0.258** | 0.121 | **0.319** | **0.362** | -0.041 | 1.000 | | | | | | | | |
| Envy | **0.417** | **0.184** | **0.187** | **0.417** | **0.184** | 0.056 | **0.284** | 1.000 | | | | | | | |
| Irritation | **0.631** | **0.178** | 0.112 | **0.682** | **0.241** | -0.084 | **0.312** | **0.537** | 1.000 | | | | | | |
| Guilt | 0.158 | **0.528** | -0.028 | **0.281** | **0.213** | -0.045 | **0.341** | **0.216** | **0.264** | 1.000 | | | | | |
| Regret | **0.201** | **0.308** | 0.111 | **0.441** | **0.333** | 0.005 | **0.293** | **0.373** | **0.408** | **0.495** | 1.000 | | | | |
| Admiration | **0.197** | **0.258** | 0.151 | 0.130 | **0.205** | 0.064 | **0.314** | **0.174** | 0.105 | **0.297** | 0.204 | 1.000 | | | |
| Pity | -0.033 | **0.333** | 0.071 | -0.022 | 0.122 | **0.210** | **0.167** | -0.031 | -0.062 | **0.321** | 0.199 | 0.126 | 1.000 | | |
| Gratitude | -0.098 | **0.171** | **0.315** | 0.060 | **0.288** | **0.488** | 0.045 | -0.009 | -0.147 | 0.138 | 0.075 | **0.270** | **0.243** | 1.000 | |
| Gloating | 0.120 | 0.067 | **0.286** | 0.055 | 0.063 | **0.311** | **0.239** | **0.189** | 0.129 | 0.117 | 0.074 | 0.114 | 0.153 | **0.207** | |
| Sadness | **0.346** | **0.302** | **0.205** | **0.529** | **0.284** | 0.021 | **0.419** | **0.450** | **0.576** | **0.255** | 0.481 | **0.203** | 0.102 | 0.036 | 0.142 |

Note: Significant correlation between two emotions at the 5% level is noted in **bold** numbers.

## Appendix B – English Translations of the Instructions

*[Sample instructions for the treatment with* DeterFine *in the first five periods and* NoDeter *in the second five periods. The spacing is slightly adjusted to make Appendix B more compact.]*

General Instructions for Participants

Welcome to the Experiment Laboratory!

You are now taking part in an economic experiment. You will be able to earn a considerable amount of money, depending on your decisions and the decisions of others. It is therefore important that you read these instructions carefully.

The instructions which we have distributed to you are solely for your private information. **It is prohibited to communicate with other participants during the experiment.** Should you have any questions please raise your hand and an experimenter will come to answer them. If you violate this rule, we will have to exclude you from the experiment and from all payments.

During the experiment you will make decisions **anonymously**. Only the experimenter knows your identity while your personal information is confidential and your decisions will not be traceable to your identity.

For your participation in this experiment you receive an initial endowment of **7 Euros**. The additional calculation of your payment depends on your decisions and the decisions of other participants. At the end of the experiment the payment will be made to you in cash.

The experiment consists of multiple, interrelated parts. All parts are payment relevant.

Information for Part 1

All participants work on the same task. The task is to earn points by locating sliders on the computer screen **exactly at '50'**. The sliders are adjustable between 0 and 100, i.e. 50 is the exact middle. Initially all sliders are located at 0. For each slider that you locate at 50, you will earn **one point**.

After all participants worked on the task, the sum of points per individual will be ranked by size. All participants that belong to the more successful half receive **10 Euros**, while all participants who belong to the less successful half receive **2 Euros**. These incomes will be transferred to the next parts of the experiment. Your final payment depends on all parts of the experiment.

*An example:*
Let us assume that four participants A, B, C and D earn the following points: A – 10 Points, B – 13 Points, C – 7 Points and D – 17 Points. Then the participants will be ranked accordingly:

1. Participant D (17 Points)
2. Participant B (13 Points)
3. Participant A (10 Points)
4. Participant C (7 Points)

Below you see the **slider screen:**


You can first try out the task in a **practice round**. This practice round is not payment relevant. Subsequently the **payment relevant Task** begins. While the practice round lasts 200 seconds, you will get **140 seconds** for the payment relevant task.

Please leave the keyboards turned around. The violation of this rule will lead to exclusion of the experiment. We will also ask you to answer three questions after the task.

Thereafter you will receive information for part 2 of the experiment.

Information for Part 2 *[Circulated after Part 1 was over]*

Please turn around your keyboard so that you can use it in part 2.

The participants will be split into **two groups**: **persons 1** and **persons 2**. If you belong to the more successful half of the participants in part 1 and therefore earned 10 Euros, you are person 1. If you belong to the less successful half of the participants and therefore earned 2 Euros, you are person 2. **In part 2 you will always stay in the same group, that in accordance with the result of part 1 you are always either person 1 or person 2.**

Part 2 consists of **10 periods** in which person 2 always faces the same decision making. In each period a person 1 and a person 2 are **randomly matched**. After a period is over, you will **never** again be matched with the same person in later periods.

Person 1 does **not** make an active decision and is only able to **guess** the decision of person 2.

**The Decision Making of Person 2**

Each person 2 faces the same decision making. Your task is to decide about a transfer between your account and the account of person 1. At the beginning of each period each person 2 has **2 Euros in her/his account from part 1**, while there are **10 Euros in the account of the randomly matched person 1**. The income of part 1 was therefore transmitted to part 2 for all participants.

Each person 2 needs to decide how many Euros she/he wants to transfer between her/his account and the account of person 1. The transfer may be between -2 and 10 Euros (only whole numbers). A negative transfer means that **person 2** would like to transfer money from her/his account to the account of **person 1**. A positive transfer means that he/she wants to transfer money from the account of person 1 to her own account.

*Two examples:*
Assume that you are person 2 and that you want to transfer 2 Euros to your account from the account of person 1. Then your transfer should be “2”.
If you would like to transfer 1 Euro from your account to the account of person 1, then your transfer should be “-1”.

Negative and neutral transfers have a **probability of success of 100 percent**, while positive transfers may **fail** with a certain probability. Failing means that person 2 does not receive the

transfer. If a transfer fails, then person 2 will also need to pay a fee. This **fee** will be deducted from the account of person 2 and **not** transferred to the account of person 1.
The probabilities of success and the fees are dependent on the size of the positive transfer. The table below shows how high the probabilities and the fees are for different transfers:

| **Transfer** | **Probability of Success** | **Fee in Case of No Success** |
|---|---|---|
| -2 Euro | 100 % | - |
| -1 Euro | 100 % | - |
| 0 Euro | 100 % | - |
| 1 Euro | 50 % | 1.10 Euro |
| 2 Euro | 50 % | 1.15 Euro |
| 3 Euro | 50 % | 1.40 Euro |
| 4 Euro | 50 % | 1.85 Euro |
| 5 Euro | 50 % | 2.50 Euro |
| 6 Euro | 50 % | 3.35 Euro |
| 7 Euro | 50 % | 4.40 Euro |
| 8 Euro | 50 % | 5.65 Euro |
| 9 Euro | 50 % | 7.10 Euro |
| 10 Euro | 50 % | 8.75 Euro |

The income of person 2 is the sum of 7 Euros for participation, 2 Euros for the less successful slider task performance in part 1 and the consequences of her/his transfer in part 2:

**Income of Person 2** =
7 Euros + 2 Euros + Transfer (if successful) - fee (if not successful)

The income of person 1 analogously is the sum of 7 Euros for participation, 10 Euros for the more successful slider task performance in part 1 and the consequences of the person 2's transfer:

**Income of Person 1** =
7 Euros + 10 Euros – Consequences of the Transfer

After 5 periods we will ask you some questions. We ask you to take your time for the questions and think hard about your answers. Thereafter the experiment will continue with a similar decision making for person 2.
After the 10 periods of part 2 are over, the experiment will end. At the end of the experiment you will receive the payment of **one of the ten periods** of part 2 in cash. That period is chosen **randomly**. Therefore each period is potentially payment relevant. All payments are done in private and other participants will not see what you have earned.

Information on the Course of Events of the Experiment

In the beginning of each period you will see an **input screen**. This input screen is the same for all persons 2:

| Überweisung | Erfolgswahrscheinlichkeit | Gebühr bei keinem Erfolg |
|---|---|---|
| -2 Euro | 100 % | - |
| -1 Euro | 100 % | - |
| 0 Euro | 100 % | - |
| 1 Euro | 50 % | 1.10 Euro |
| 2 Euro | 50 % | 1.15 Euro |
| 3 Euro | 50 % | 1.40 Euro |
| 4 Euro | 50 % | 1.85 Euro |
| 5 Euro | 50 % | 2.50 Euro |
| 6 Euro | 50 % | 3.35 Euro |
| 7 Euro | 50 % | 4.40 Euro |
| 8 Euro | 50 % | 5.65 Euro |
| 9 Euro | 50 % | 7.10 Euro |
| 10 Euro | 50 % | 8.75 Euro |

Sie sind Person 2. Sie können zwischen Ihrem Konto und dem Konto einer Ihnen zufällig zugeordneten Person 1 Geld transferieren. Person 1 kann keinen Transfer durchführen.

Ihre Eingabe kann zwischen -2 und 10 Euro liegen.

Eine negative Zahl gibt an, dass Sie Geld von Ihrem Konto auf das Konto von Person 1 transferieren.

Eine positive Zahl gibt an, dass Sie Geld von dem Konto von Person 1 auf Ihr Konto transferieren.

Wieviele Euro möchten Sie zwischen sich und einer Ihnen zufällig zugeordneten Person 1 transferieren?

As described, the account of person 2 holds 2 Euros in each period, while the account of person 1 holds 10 Euros in each period. Person 2 makes a decision on the transfer between the two accounts by entering a whole number between -2 and 10 into the input window. You can click on this window with the mouse. When person 2 has made his/her decision, she/he needs to press the **OK-button**. When you press the OK-button you cannot change your decision in this period.

After all persons 2 have made their decisions, these decisions will be displayed to persons 1. Persons 1 will know how many Euros person 2 wanted to transfer, whether person 2 was successful and the incomes of both persons are at the end of the period.
Persons 2 will also know whether the transfers were successful and the incomes of both persons at the end of the period.

As described the **income of person 2** is

**7 Euros + 2 Euros + Transfer (if successful) - Fee (if not successful)**

Before the experiment will continue all participants need to answer control questions on the screen. These questions aim to familiarize you with the rules of the experiment.
Do you still have question concerning the experiment? In this case please raise your hand.

## Additional Information for Part 2

*[Circulated after the first five periods and the emotions questionnaire were past]*

The **course** of part 2 **remains the same**. The decision making of part 2 however **slightly changes** for the next 5 periods: The **probability of success** of transfer from now on is always 100 %. This means that the **fees** for positive transfer are **abolished**. This also means that the **table** of the first 5 periods is not valid any longer and the decisions of person 2 will be **always** executed directly.

In case you have any questions concerning the changes and the next 5 periods, please raise your hand.