



Do Negative Emotions Explain Punishment in Power-to-Take Game Experiments?

Fabio Galeotti

► **To cite this version:**

Fabio Galeotti. Do Negative Emotions Explain Punishment in Power-to-Take Game Experiments?. Working paper GATE 2015-04. 2015. <halshs-01128873>

HAL Id: halshs-01128873

<https://halshs.archives-ouvertes.fr/halshs-01128873>

Submitted on 10 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

WP 1504 – March 2015

Do Negative Emotions Explain Punishment in Power-to-Take Game Experiments?

Fabio Galeotti

Abstract:

An important branch of economic research on emotions has used power-to-take game experiments to study the impact of negative emotions, such as anger, irritation and contempt, on the decision to punish. We investigate experimentally the role that the specific punishment technology adopted plays in this context, and test to what extent punishing behavior can be truly attributed to negative emotions. We find that a large part (around 70%) of the punishment behavior observed in previous PTTG studies is explained by the technology of punishment adopted instead of negative emotions. Once this effect is removed, negative emotions do still play an important role, but the efficiency costs associated to them are much smaller.

Keywords:

Emotions, punishment, power-to-take, experiment

JEL codes:

A12, C72, C91

Do Negative Emotions Explain Punishment in Power-to-Take Game Experiments?

Fabio Galeotti*

Université de Lyon, Lyon, F-69007, France

CNRS, GATE Lyon Saint-Etienne, Ecully, F-69130, France

* GATE Lyon St. Etienne, 93 Chemin des Mouilles, 69130 Ecully, France

E-mail: galeotti@gate.cnrs.fr; Tel.: + 33 (0) 4 72 86 60 38

Version: March 2015

Abstract

An important branch of economic research on emotions has used power-to-take game experiments to study the impact of negative emotions, such as anger, irritation and contempt, on the decision to punish. We investigate experimentally the role that the specific punishment technology adopted plays in this context, and test to what extent punishing behavior can be truly attributed to negative emotions. We find that a large part (around 70%) of the punishment behavior observed in previous PTTG studies is explained by the technology of punishment adopted instead of negative emotions. Once this effect is removed, negative emotions do still play an important role, but the efficiency costs associated to them are much smaller.

Keywords: Emotions, punishment, power-to-take, experiment.

JEL classification: A12, C72, C91

1. Introduction

This study investigates the role played by the punishment technology in driving the behavior of responders in power-to-take game (PTTG) experiments, and tests to what extent it can be truly attributed to negative emotions, such as anger, irritation and contempt. The concept of emotion is an elusive one (Frijda, 2007). In psychology and neuroscience, it is used to describe a range of discrete, synchronized, and time-limited responses (including subjective experience, expression, bodily responses and action tendencies) of an individual to internal or external events (stimuli) which are relevant or significant to that individual (Phelps, 2009). A person experiences positive emotions, such happiness and joy, when his or her interests are satisfied or facilitated, and negative emotions, such as sadness and contempt, when his or her interests are prevented or opposed. Psychologists and neuroscientists have accumulated considerable empirical evidence suggesting that emotions play a very important role in the decision-making process (see, e.g., Frijda, 1986; Damasio, 1994; LeDoux, 1996; Picard, 1997). In particular, one important aspect of emotions is that they motivate the individual towards particular actions depending on the context, type and intensity of the emotion experienced (Frijda, 1986). For instance, a fearful individual will have a tendency to run away from the stimulus responsible for his or her fear, while an envious person will tend to seek to possess the valued thing or person (Lazarus, 1991).

Over the last few decades, economists have also started to pay attention to the complexity of emotions on economic scenarios and have been trying to capture the range of possible roles that emotions play in the economic decision-making process. For instance, emotions have been proposed as an explanation for important economic phenomena such as cooperation (Frank, 1988; Fehr and Gächter, 2002) and decision-making under risk (Loewenstein *et al.*, 2001), and are seen to have important consequences for many other economic phenomena, such as inter-temporal choices (Rick and Loewenstein, 2008), competition (Kräkel, 2008), bidding behavior (Bosman and Riedl, 2004) and bargaining behavior (Pillutla and Murnighan, 1996). The advent of neuroeconomics has then further pushed forward the interest of economists on the role played by emotions in the economic decision-making process (for a review on emotions and neuroeconomics, see Phelps, 2009). More recently, emotions have also been associated with the automatic and impulsive decision system of dual-process theories, according to which human behavior is the result of the interaction between this system and a controlled, reflective and rational one (Alos-Ferrer and Strack, 2014; Brocas and Carrillo, 2014; Schulz *et al.*, 2014).

An important branch of economic research on emotions utilized experiments to study the impact of negative emotions, particularly anger, irritation and contempt, on the decision to punish (Bosman and van Winden, 2002; Bosman *et al.*, 2005; Ben-Shakhar *et al.*, 2007; Hopfensitz and Reuben, 2009; Joffily *et al.*, 2013).¹ This stream of research started with the seminal work of Bosman and van Winden (2002) on the PTTG. In the PTTG, there are two players, the ‘take authority’ (with income Y_{take}), and the ‘responder’ (with income Y_{resp}). The game is divided into two stages. In the first stage, the take authority selects a take rate $t \in [0,1]$, which is the proportion of the responder’s income that will be transferred to the take authority at the end of the game. In the second stage, the responder chooses a destroy rate $d \in [0,1]$, which is the proportion of Y_{resp} that will be destroyed. Therefore, the payoffs of the game are $(1-t)(1-d)Y_{\text{resp}}$ for the responder, and $Y_{\text{take}} + t(1-d) Y_{\text{resp}}$ for the take authority.

If the subjects are rational profit-maximizing agents, the responder should not destroy if the take rate is less than 1, and should be indifferent between all possible destroy rates if the take rate is 1. Hence, from backward induction, the take authority should select $t = 1 - \varepsilon$, where ε is an infinitesimal positive number. The PTTG can be interpreted as an ultimatum game with continuous opportunities to punish and can describe many economic situations where an agent can take away any part of the endowment of another agent (e.g. taxation, monopolistic pricing and principal-agent relationships) (Bosman and van Winden, 2002).

Most of the literature on the PTTG investigates the role played by negative emotions on responders’ behavior through physiological (Ben-Shakhar *et al.*, 2007) and self-report measures (Bosman and van Winden, 2002; Bosman *et al.*, 2005). Both measures were found to be related to destruction decisions. In particular, participants who experienced intense anger, irritation and contempt punished their counterparts more often and more severely. This result seems to identify these negative emotions as the main driving force for punishing behavior, and the principal source of efficiency costs in this context.²

¹ An individual experiences anger, irritation and/or contempt when he or she disapproves of someone else’s action and, specifically in the case of anger and irritation, is also displeased about the undesirable event related to that action (Ortony *et al.*, 1988). According to the psychological literature on emotions (see, e.g., Lazarus, 1991), the action tendency that results from these negative emotions is an attack (which can be, for instance, physical, verbal or symbolic) on the blameworthy agent. In economic environments, this may be translated into the destruction of the agent’s resources, as it is hypothesized in the aforementioned economic studies on emotions and punishing behavior. Note that this impulse to attack does not necessarily result in actual behavior, and is often inhibited for personal and social reasons (Averill, 1983; Lazarus, 1991). If we look at the other discrete emotions considered in our study, they are not usually associated with aggressive tendencies (Lazarus, 1991). According to Ortony *et al.* (1988), some of these emotions (such as happiness and sadness) may not even involve an action tendency.

² Other well-established findings from the experimental literature on the PTTG show that people appropriate almost 60% of responders’ income, while only 20% of the responders destroy income and usually all of it

This literature on the PTTG adopts a technology of punishment characterized by a *non-constant* ‘fine-to-fee’ ratio. This parameter identifies “the income reduction for the targeted subject relative to the cost for the subject who requested the punishment” (Casari, 2005:107). The advantage of using this technology of punishment is that it reflects many economic and everyday situations modeled by the PTTG to a fairly accurate extent. Take, for instance, all the cases where a principal can set a less favorable incentive scheme (e.g. lower piece rate) for the agent in order to appropriate a larger share of the profit. The agent can retaliate by exerting a lower effort or going on strike in order to reduce the profit, even if this goes against his or her own material incentives (Bosman and van Winden, 2002). If the principal is particularly unfair, the agent can even quit the job and give up his or her entire salary to prevent the principal from benefitting from the agent’s work. This will have little cost for the agent but a large one for the principal. On the contrary, if the principal is only slightly unfair, the agent will bear a much higher cost from punishing the principal. Another example is when, in a bipartisan coalition government, one political party (the leading one) takes the majority of the places in the cabinet or dictates most of the coalition’s political agenda. The members of the other party may decide to reduce their support and even bring the government to a collapse, even though this will cause them to step down from power. The cost they will have to bear to punish the other party’s members is small (large) if the leading party has been very (only marginally) unfair. In both examples, the ‘fine-to-fee’ ratio is not constant and increases with the initial offence of the principal, in one case, and the leading party, in the other.

Although the results of the standard PTTG can be easily generalized to many real-world situations, they may lack internal validity, especially in their attempt to link punishment behavior with negative emotions. In particular, the non-constant ‘fine-to-fee’ ratio adopted in previous studies of the PTTG may explain all or part of the relationship between punishing behavior and negative emotions observed in this context. In the standard PTTG, the income reduction for the authority is ΔY_{resp} , whereas the cost for the responder to

(Bosman and van Winden, 2002; Reuben and van Winden, 2010). Small differences were observed between an effort treatment (where endowments were earned by performing a preliminary individual real effort task) and a no-effort treatment (where endowments were exogenously given by the experimenter) (Bosman *et al.*, 2005). A group version of the PTTG, where decisions were made by groups, produced the same results qualitatively as the no-group experiment (Bosman *et al.*, 2006). Furthermore, in a three-player version of the PTTG with one take authority and two responders, Reuben and van Winden (2008) show that responders who knew each other from outside the laboratory punish and coordinate more than strangers. The PTTG has also been used to study the influence of participation (Albert and Mertins, 2008), gender pairing (Sutter *et al.*, 2009) and waiting time (Galeotti, 2013) on economic decision-making.

punish is $d(1 - t)Y_{resp}$. Therefore, the ‘fine-to-fee’ ratio is $tdY_{resp}/d(1 - t)Y_{resp} = t/1 - t$, where t (the proportion of the responder’s income that will be transferred to the take authority) is an endogenous and non-constant parameter. This implies that the ‘demand’ for punishment is higher when t is higher (i.e. the ‘fine-to-fee’ ratio is higher). In other words, for high take rates the responder has a higher incentive to punish his or her counterpart, whereas for low take rates the incentive is lower. This means that when the offence is severe, subjects may punish because punishing is cheaper and not, or not only, because they experience anger, irritation or contempt. It is important to test whether or not this is the case in order to provide a correct measure of the efficiency costs associated with negative emotions (what Bosman and van Widen (2002) refer to as *emotional hazard*).

In order to investigate to what extent the punishment behavior observed in previous studies on PTTGs is explained by the specific punishment technology adopted rather than negative emotions, we conducted a laboratory experiment. We varied the extent to which the punishment technology embedded a variable or constant ‘fine-to-fee’ ratio.³ Emotions were assessed through self-report measures, as in previous studies.⁴ As mentioned earlier, a PTTG which employs a variable ‘fine-to-fee’ ratio is a more natural environment, and has a broader external validity than a PTTG with a constant ‘fine-to-fee’ ratio. The latter, on the other hand, provides more internally valid results. This suggests that an experimental approach of the kind we propose, where both PTTGs with constant and variable ‘fine-to-fee’ ratios are considered, would be extremely useful to complement previous studies on PTTGs, and could contribute to a better understanding of the role played by negative emotions in this context.

To give a very brief overview of our results, we found that a large part (around 70%) of the punishment behavior observed in previous PTTG studies is explained by the technology of punishment adopted rather than negative emotions. In a PTTG with a constant ‘fine-to-fee’ ratio, negative emotions still play an important role, but much smaller. This paper is organized as follows: Section 2 presents the experimental design, Section 3 describes

³ Real-world situations captured by a PTTG with a constant ‘fine-to-fee’ ratio are less common than those modeled by a PTTG with a variable ‘fine-to-fee’ ratio, but still present. An example is the punitive destruction of own resources which have been (or are about to be) taken by an enemy in a conflict. Although these resources are gone, a party can decide to destroy them at a certain constant marginal cost (e.g. X amount of explosive for each unit of resources to be destroyed) in order to punish the counterpart. In war conflicts, this could be done not only for punitive but also strategic or operational reasons, and it is usually referred to as a *scorched earth strategy*.

⁴ For a discussion on the reliability of self-reports in measuring emotions, see Bosman and van Winden (2002) and Hopfensitz and Reuben (2009).

the theoretical background and the behavioral hypotheses, Section 4 reports the results, and Section 5 concludes.

2. Experimental Design

The experiment was conducted between March and September 2012 at the University of East Anglia, with 282 students participating over many sessions.⁵ Each session lasted on average 50 minutes. No subject was allowed to participate in more than one session. Subjects received a show-up fee of £5 and earned on average £9.41 (around 15 US dollars). In order to ensure the greatest comparability of our experiment with previous literature, we tried to replicate, as close as possible, the experimental procedures adopted in previous PTTG experiments. In particular, we (a) conducted a paper-and-pencil experiment, (b) employed the same instructions, exercises, examples and procedures as previous PTTG studies,⁶ (c) avoided any particular or suggestive terminology during the sessions, such as ‘take authority’ or ‘take rate’, (d) adopted the same payment procedure as Bosman and van Winden (2002), and (e) assessed emotions on a 7-point Likert scale via self-reports after each subject learned about the decision of their counterpart.⁷ More details about the experimental procedure are provided below.⁸

Upon arrival, each subject was randomly assigned the role of participant A (take authority) or participant B (responder) by drawing a letter from an urn, then randomly allocated to a computer workstation which was isolated from other workstations via partitions. The instructions were then distributed and read aloud to provide common

⁵ Details of the socioeconomic characteristics of the subjects and experimental instructions are contained within the on-line appendix.

⁶ Minor adjustments to the original instructions were made to fit them to our laboratory routines, monetary payments and comparability of our treatments.

⁷ In our and previous work on the PTTG, emotions were assessed via single items. There is an ongoing debate on the reliability of multi-item versus single-item measures of psychological constructs (see, e.g., Gogol *et al.*, 2014). Single-item measures are generally less reliable than the corresponding multi-item scales. Hence, a potential limitation of our and previous experiments is that different emotions might be associated with different effects on behavior because of a different degree of reliability between the measures of emotions. Single-item measures have, on the other hand, the advantage of being easy and fast to fill out. This was quite desirable in our experiment since we assessed many emotions. For example, if the questionnaire was too long and tedious, subjects might lose concentration or experience other emotions, such as irritation, annoyance, frustration or resentment (which could, in turn, affect their answers to the questionnaire). Future studies could be devoted to evaluate whether, and to what extent, the use of single-item measures of emotions is a concern in PTTG experiments.

⁸ We also ran separate sessions for UK students and non-UK students. This is because we wanted to re-create, at least for half of the sessions (i.e. those with UK subjects), a laboratory environment analogous to the one used in previous PTTG studies where most of the participants shared a similar cultural background; and to control, in a systematic way, that negative emotions were not driven by the subjects’ perceptions of cultural dishomogeneity in a given session. However, the data analysis did not detect any differences in emotions and punishment behavior between UK and non-UK students.

information to the subjects. Two individual computerized exercises followed in order to check the subjects' understanding of these instructions. Clarifications were individually provided to subjects with incorrect answers. After completion of these exercises, each participant A was randomly matched with a participant B by asking each participant A to randomly choose a coded envelope which was linked to a certain participant B. Each participant A was then asked to fill in the take rate, that is the proportion of participant B's endowment that would be transferred to participant A at the end of the experiment, on the form that was placed inside the envelope. Afterwards, the envelopes containing the forms were collected and given to all participants B who were asked to complete the form with the destroy rate, that is the proportion of their endowment that will be destroyed. The envelopes with the forms inside were then given back to all participants A, who could take note of the decision of their corresponding matches. Subsequently, each subject was asked to fill in a questionnaire concerning emotions, expectations about the decision of their counterpart,⁹ and personal information. Meanwhile, the envelopes were collected and handed to the cashier who was outside the laboratory and, hence, not present during the experiment. Subjects were then privately paid one-by-one outside the laboratory by the cashier.

As in Bosman and van Winden (2002), we assessed a list of eleven emotions. To not direct the attention of the subjects to specific emotions, the list included both negative emotions that previous studies had found relevant for explaining punishing behavior observed in the PTTG (i.e. anger, irritation, contempt), other less influential negative emotions (i.e. envy, jealousy, sadness, shame, fear), positive emotions (i.e. joy, happiness) and neutral emotions (i.e. surprise). Subjects were asked to state how much they felt each emotion on a 7-point Likert scale when they learned about the decision of their counterparts. The scale ranged from "no emotion at all" to "high intensity of the emotion" (Bosman and van Winden, 2002).

We ran two different treatments where we varied the nature of the 'fine-to-fee' ratio embedded in the punishment technology (constant or variable).¹⁰ In one treatment, the 'fine-to-fee' ratio was increasing in the take rate and ranged from 0 to infinite. This treatment was an exact replication of previous PTTG experiments (i.e. Bosman and van Winden, 2002;

⁹ Each participant B was asked to indicate which percentage of his or her endowment he or she expected participant A would decide to transfer to himself or herself; each participant A was asked to indicate which percentage of the transfer he or she expected participant B to destroy.

¹⁰ In total, we collected 70 independent observations under a variable 'fine-to-fee' ratio, and 71 under a constant 'fine-to-fee' ratio. An independent observation is a pair consisting of a responder and a take authority. The variation in the number of independent observations across the two treatments is due to different rates of attendance across sessions.

Bosman *et al.*, 2005). In the other treatment, we employed a constant ‘fine-to-fee’ ratio, which was therefore independent of the take rate. Most of the literature on punishment behavior in economics usually employs ‘fine-to-fee’ ratios ranging from 0 to 4 (Nikiforakis and Normann, 2008; Carpenter, 2007). We chose a constant ‘fine-to-fee’ ratio equal to 2.¹¹ Note that the purpose of our study was not to show how people react to changes in the price of punishing (as in Nikiforakis and Normann, 2008; Carpenter, 2007), but to provide a correct measure of the efficiency costs attributed to negative emotions. To this end, a constant ‘fine-to-fee’ ratio equal to 2 is optimal as it maximizes comparability with the other treatment where the ‘fine-to-fee’ ratio is variable, and with previous PTTG experiments. Note in fact that a ‘fine-to-fee’ ratio equal to 2 corresponds, under a variable ‘fine-to-fee’ ratio, to a take rate of $2/3$, which is roughly equivalent to the mean and median take rate observed in previous PTTG experiments, and obtained in our treatment with a variable ‘fine-to-fee’ ratio. Hence, when we compare the two treatments, we have approximately half the observations where the incentive to punish is lower in the “constant fine-to-fee ratio” treatment, and the other half where the incentive is higher. This allowed us to measure the bias, if there was any, which may have occurred in the standard PTTG for both sides of the distribution of the take rates: when $t < 2/3$ (i.e. when the variable ‘fine-to-fee’ was lower than the constant ‘fine-to-fee’ ratio), and when $t > 2/3$ (i.e. when the variable ‘fine-to-fee’ was higher than the constant ‘fine-to-fee’ ratio). A constant ‘fine-to-fee’ ratio equal to 2 has also been extensively used in previous economic experiments to study punishment behavior (e.g. de Quervain *et al.*, 2004; Cubitt *et al.*, 2011) and, therefore, it enables comparisons with other studies which do not use the PTTG as a vehicle of research.

Figure 1 displays how the ‘fine-to-fee’ ratio evolved over different values of the take rate in the two treatments. The graph clearly shows that under a variable ‘fine-to-fee’ ratio the effectiveness of the punishment increases exponentially as the take rate increases. As a consequence, subjects might punish simply because it is more ‘convenient’ to do so and not (or not only) because they experience negative emotions (which is to be expected for increasing take rates). In other words, in this set-up, the idiosyncratic features of the punishment technology might induce an effect on behavior which can be erroneously

¹¹ A value of 2 means that the cost of punishing is half of how much the punishment reduces participant A’s endowment. In order to employ a constant ‘fine-to-fee’ ratio equal to 2, we simply stated in the instructions that “for each 1% of his or her endowment that participant B decides to destroy, 10 pence of the transfer to participant A will be destroyed as well”. In addition, we allowed subjects to deduct the cost of punishing from their show-up fee, if needed. For this reason, the show-up fee was set at the level of £5 to ensure that, at worst, subjects (in particular, participants B) could leave the experiment with £2.50 in their pockets. This ensured that participants B could punish participants A for any possible value of the take rate without incurring losses.

attributed to negative emotions. This possibility was instead ruled out in the constant ‘fine-to-fee’ ratio treatment.

[Figure 1 about here]

3. Theoretical background

In this section, we briefly discuss the theoretical implications of having a constant versus variable ‘fine-to-fee’ ratio, embedded in the punishment technology, for the behavior and emotions of the responder in the PTTG, and present the hypotheses that can be tested. We build on the model that Loewenstein (2000) proposed to describe the impact of visceral factors in the utility function of an agent. Let $U_i(p_i, s_i)$ be the utility function of a generic responder i , in which p_i is the consumption activity, in our case the size of the punishment, and s_i the visceral state, in our case the experience of anger, irritation or contempt. If we call p_i^* the optimal punishment, this may rise if the intensity of the negative emotion (s_i) rises as well ($\partial p_i^* / \partial s_i \geq 0$). In the context of the PTTG, the responder may experience emotional distress when, for instance, the decision of his or her counterpart is perceived as unkind or when the distribution of the income becomes unequal, which may occur every time that t_i increases. To mitigate this distress, the responder may decide to punish.

Anger, irritation or contempt may not be the only explanation for the decision to punish. The responder may, for instance, punish because he or she feels the need to comply with what he or she believes is the appropriate behavior in the laboratory (Zizzo, 2010) or because he or she misunderstands the instructions or the incentives in the experiment.¹² In addition to these alternative explanations of punishing behavior, the decision to punish may also be sensitive to more traditional economic incentives, such as the cost per unit of punishment that the responder needs to pay in order to damage the counterpart, which is a measure of the effectiveness of the punishment (Carpenter, 2007; Nikiforakis and Normann, 2008). To formalize all this, let us assume that optimal punishment (p_i^*) depends on the intensity of the negative emotion (anger, irritation or contempt) which, in turn, depends on the take rate experienced by the individual (t_i), the effectiveness of the punishment, which is equivalent to the ‘fine-to-fee’ ratio (σ_i), and a generic parameter θ_i which captures everything

¹² A significant part of the behavior observed in experiments may be attributed to confusion or mistakes, as, for instance, Andreoni (1995) found in the context of public good games. Confusion and mistakes may explain why a responder punishes in the PTTG, since the only possible direction in which a responder can make a mistake is towards punishment.

else, including confusion, mistakes and experimenter demand effects.¹³ We can write the optimal punishment as:

$$p_i^* = p_i^*(s_i(t_i), \sigma_i, \theta_i)$$

The optimal level of punishment weakly increases in the take rate ($\partial p_i / \partial t_i \geq 0$). In addition, it weakly increases in the fine-to-fee ratio ($\partial p_i / \partial \sigma_i \geq 0$).¹⁴ Under a variable ‘fine-to-fee’ ratio, σ_i is an increasing function of t_i . Remember that $\sigma_i = t_i / (1 - t_i)$. This means that the punishment may be even higher if t_i increases (i.e. t_i may have a *multiplier effect* on the decision to punish by means of the non-constant ‘fine-to-fee’ ratio). This multiplier effect is given by $\partial [t_i / (1 - t_i)] / \partial t_i = 1 / (1 - t_i)^2$, and is exponentially increasing in t_i . In contrast, if the ‘fine-to-fee’ ratio is a constant ($\sigma_i = \bar{\sigma}$), its impact on the decision to punish should be the same across different levels of t_i (i.e. there is no multiplier effect). By comparing a situation where the punishment technology embeds a constant ‘fine-to-fee’ ratio with a situation where the punishment technology is characterized by a variable ‘fine-to-fee’ ratio, we can not only measure how much of the punishment is actually attributable to the multiplier effect rather than emotional distress,¹⁵ but also whether negative emotions matter at all. Note in fact that it is possible that $\partial s_i / \partial t_i \geq 0$ and $\partial p_i / \partial s_i = 0$. If the latter is the case, punishment is better explained by the parameter θ_i rather than anger, irritation or contempt. Based on this simple model that describes the motivations of the responder to punish, we can formulate the following three hypotheses.

Hypothesis 1. In the standard PTTG, the variable ‘fine-to-fee’ ratio produces a multiplier effect on the decision to punish, that is $(\partial p_i / \partial \sigma_i) (\partial \sigma_i / \partial t_i) > 0$, and thus the role of other factors, including negative emotions, is inflated.

We can test this hypothesis by studying how the difference in punishment evolves between a constant and a variable ‘fine-to-fee’ ratio when the take rate increases. The null hypothesis is that there is no such multiplier effect, that is $(\partial p_i / \partial \sigma_i) (\partial \sigma_i / \partial t_i) = 0$.

¹³ For simplicity, we also assume perfect separability between θ_i and s_i , that is θ_i does not affect s_i and *vice versa*.

¹⁴ If the responder displays a rational and self-interested behavior, he or she should never destroy if $t_i < 1$, and be indifferent between any levels of punishment if $t_i = 1$. This is irrespective of the size of the ‘fine-to-fee’ ratio. In other words, $\partial p_i / \partial t_i \geq 0$ and $\partial p_i / \partial \sigma = 0$. If his or her behavior is driven by reciprocity (Rabin, 1993; Duwfenberg and Kirchsteiger, 2004), the punishment should be more likely when t_i increases (i.e. $\partial p_i / \partial t_i \geq 0$), *caeteris paribus*, and less likely when σ increases (i.e. $\partial p_i / \partial \sigma \geq 0$), *caeteris paribus*. If the responder cares about equality (Fehr and Schmidt, 1999), the punishment should be more likely when t_i increases (i.e. $\partial p_i / \partial t_i \geq 0$), *caeteris paribus*. For any given level of t_i , the punishment should also be weakly increasing in σ (i.e. $\partial p_i / \partial \sigma \geq 0$). The proofs are contained in the on-line appendix.

¹⁵ Note that it is not possible to separate the multiplier effect, caused by the variable ‘fine-to-fee’ ratio, from the arousal effect, caused by the negative emotion, using statistical analysis on the data collected from previous PTTG studies. This is because of the very high correlation between negative emotions and the take rate.

Hypothesis 2. Once the multiplier effect is removed, the intensity of anger, irritation or contempt does still explain the punishing behavior of the responder, that is $\partial p_i / \partial s_i > 0$.

We can test this hypothesis against the null hypothesis that punishing is due to something else (e.g. confusion, experimenter demand effects) by looking at whether these negative emotions do still explain punishing behavior when the ‘fine-to-fee’ ratio is constant.

Hypothesis 3. Due to the multiplier effect, anger, irritation and contempt erroneously predict much more punishing behavior under a variable ‘fine-to-fee’ ratio compared to a constant ‘fine-to-fee’ ratio, that is $\partial p_i^V / \partial s_i > \partial p_i^C / \partial s_i$, where p_i^V is the level of punishment with a variable ‘fine-to-fee’ ratio, and p_i^C is the level of punishment with a constant ‘fine-to-fee’ ratio.

Hypothesis 3 can be tested by comparing how much of the punishing behavior is explained by these negative emotions under a variable versus a constant ‘fine-to-fee’ ratio.

4. Results

In this section, we first check whether there was any difference in the behavior of the take authorities across the treatments. Then we move to the main focus of this study, that is the punishing behavior of the responder, to test whether there exists any difference in the standard PTTG compared to the modified version of the PTTG where the punishment technology embeds a constant ‘fine-to-fee’ ratio. Next, we look at emotions and the extent to which anger, irritation and contempt explain the punishing behavior of the responders. Finally, we briefly consider the expectations of the responders, and whether they help to understand behavior and emotions.

Behavior of the take authorities. Table 1 displays the take rates for the two treatments. The values are in line with previous PTTG experiments. If we compare the behavior of the subjects under constant and variable ‘fine-to-fee’ ratios, there was no significant difference between the take rates (Mann-Whitney test, $p = 0.191$).¹⁶

[Table 1 about here]

¹⁶ A Tobit regression (reported in the on-line appendix), where we control for the demographic characteristics of the subjects, confirms this result. In this regression, we also find that non-UK subjects display significantly higher take rates than UK subjects. Different conjectures (discussed in the on-line appendix) might explain why non-UK take authorities appropriated more resources than UK take authorities. In the on-line appendix, we also analyze the emotions of the take authorities.

Behavior of the responders. We define the *punishment rate* as the proportion of the amount taken by the take authority that was destroyed by the responder.¹⁷ As discussed in Section 3, the punishment rate may depend on the amount taken by the take authority because of several psychological reasons, such as inequality aversion and reciprocity. Hence, we need to control, in the analysis of the punishing behavior, for the impact of the take rate. Figure 2 displays the scatterplots and the locally weighted smoothed regressions of the punishment rate as a function of the take rate, for each of the treatments. When the ‘fine-to-fee’ ratio was variable, the punishment rate rose exponentially as the take rate increased. When the ‘fine-to-fee’ ratio was instead constant, subjects increased their punishment less in response to higher take rates. This is preliminary evidence in favor of Hypothesis 1. In particular, it seems that, under a variable ‘fine-to-fee’ ratio, there was a multiplier effect at work which induced more punishment from the responders when the take rate increased. We tested this more formally with non-parametric techniques. In particular, we grouped the take rates into four classes based on their distribution: the very low take rates (the bottom 25%), the low take rates (between 25% and 50%), the high take rates (between 50% and 75%), and the very high take rates (the top 25%).¹⁸ This distinction enabled us to investigate the punishing behavior controlling for different levels of the take rate, and also discriminated between take rates for which the constant ‘fine-to-fee’ ratio was lower than the variable ‘fine-to-fee’ ratio, and *vice versa*. Note in fact that the constant ‘fine-to-fee’ ratio crosses the variable ‘fine-to-fee’ ratio at $t_i = 2/3$ (see Figure 1), which is slightly above the median of the distribution of the take rates. Hence, for the very low and low take rates, the incentive to punish was higher under a constant ‘fine-to-fee’ ratio. In contrast, for the high and very high take rates, the opposite is true, that is the incentive to punish was higher under the variable ‘fine-to-fee’ ratio.¹⁹

[Figure 2 about here]

Table 2 displays the average punishment rates for the two treatments. If we compare the punishment rate of sessions characterized by a constant ‘fine-to-fee’ ratio with sessions where the ‘fine-to-fee’ ratio was variable, we find that, when the take rates were very high,

¹⁷ The punishment rate coincides with the destroy rate under a variable ‘fine-to-fee’ ratio, and is equal to $2d/t$ under a constant ‘fine-to-fee’ ratio, that is the amount destroyed ($10 \text{ pence} \times d$) over the amount taken ($5 \text{ pounds} \times t$).

¹⁸ The 25th quartile corresponds to a take rate of 0.5, the 50th quartile (median) to a take rate of 0.6, and the 75th quartile to a take rate of 0.8. Hence, t_i is classified as very low take rate if $t_i \leq 0.5$, low take rate if $0.5 < t_i \leq 0.6$, high take rate if $0.6 < t_i \leq 0.8$, and very high take rate if $t_i > 0.8$.

¹⁹ To be precise, among the high take rates, there were 3 out of 38 observations at $t_i = 0.65$, which were slightly below $t_i = 2/3$ and where, therefore, the incentive to punish can be considered as identical between the constant and variable ‘fine-to-fee’ ratio.

punishment was significantly more severe under a variable ‘fine-to-fee’ ratio (Mann-Whitney test, $p = 0.017$). In contrast, when the take rates were very low, subjects seemed to punish more under a constant ‘fine-to-fee’ ratio. However, the difference was not statistically significant ($p = 0.286$). For low and high take rates, the punishment rates were very similar across the two treatments, and the difference was not statistically significant (Mann-Whitney test, $p > 0.1$).

[Table 2 about here]

We tested the robustness of these findings using Tobit regression analysis (see Table 3).²⁰ The dependent variable was the punishment rate (p_i). The independent variables were, in Regression 1, the take rate received from the take authority (t_i), a dummy variable which took value 1 when a constant ‘fine-to-fee’ ratio was employed, the nationality of the subjects (non-UK = 1 for non-UK), and two interaction terms of the dummy used to identify the constant ‘fine-to-fee’ ratio with nationality and the take rate respectively. In Regression 2, we also controlled for the experience of the subjects in previous experiments,²¹ their gender (Male = 1 for male subjects), and age. The coefficient of the variable ‘take rate’ was positive and statistically significant, meaning that the take rate from the take authority negatively affected the punishment behavior of the responder when the ‘fine-to-fee’ ratio was variable. Under a constant ‘fine-to-fee’ ratio, the relationship between the take rate and the punishment rate was also negative but markedly weaker. This brings us to the following result which supports Hypothesis 1.

[Table 3 about here]

Result 1. Consistent with Hypothesis 1, under a variable ‘fine-to-fee’ ratio, 70% (-2.727/3.885) of the punishment triggered by the take rate was attributable to the multiplier effect.

Under a constant ‘fine-to-fee’ ratio, the punishment rates were generally higher compared to a variable ‘fine-to-fee’ ratio. This was captured by the positive and significant coefficient of the dummy $d_{\sigma=\bar{\sigma}}$. Hence, when the take rates were low, punishment was higher under the constant ‘fine-to-fee’ ratio and, when the take rates were high, it was higher under

²⁰ There were 87 left-censored observations and 19 right-censored. We also tried a logit regression where the dependent variable was a dichotomous variable taking value 1 if the responder destroyed and 0 otherwise. The results were similar to those presented in this paper. However, this approach omits much of the information about the punishment rate and, therefore, is less preferred than the approach based on the Tobit model. The results of the logit are reported in the on-line appendix.

²¹ The data for ‘experience’ was collected from the final questionnaire provided to the subjects. In particular, subjects were asked to indicate whether they had previously participated in “0”, “1”, “2” or “3 or more than 3” experiments.

the variable ‘fine-to-fee’ ratio. This is shown in Figure 3, where we plotted the predicted punishment rate for the constant and variable ‘fine-to-fee’ ratios respectively against the take rates. In particular, for take rates lower than 56%, the predicted punishment was higher under the constant ‘fine-to-fee’ ratio, whereas, for take rates higher than 56%, the predicted punishment was higher under the variable ‘fine-to-fee’ ratio.²²

[Figure 3 about here]

Role of emotions. We now turn to the analysis of the emotions experienced by the responders. We will initially consider all the emotions (positive, negative and neutral), to check whether there existed any similarity between them and to investigate which emotions were driven by the take rate. We will then focus on anger, irritation and contempt, to study the emotional basis of the punishing behavior.

First, it is worth pointing out that, as seen in the literature, different emotions capture similar underlying emotional states. In particular, anger was strongly positively correlated to irritation (Spearman $\rho = 0.75$, $p = 0.000$), envy to jealousy ($\rho = 0.83$, $p = 0.000$) and happiness to joy ($\rho = 0.79$, $p = 0.000$). The full list of correlations is displayed in Table 4. To obtain an idea of the underlying structure of the discrete emotions assessed, we also conducted an exploratory factor analysis. This yielded two factors which appear to capture the *valence* (positive or negative) of the emotions considered. The factor loadings are displayed in Table 5.²³ Fear, envy, anger, sadness, shame, irritation, contempt and jealousy form a ‘negative valence factor’, while happiness and joy comprise a ‘positive valence factor’. The eigenvalue for the negative valence factor is 3.539, while for the positive valence factor it is 1.889.

[Table 4 about here]

[Table 5 about here]

In order to study whether and which emotions were driven by the take rate, we first looked at the patterns of the locally weighted smoothed regression lines between the intensity of each emotion and the take rate, for each of the treatments (Figure 4). There seemed to be no differences across treatments on how the take rate impacted on each emotion. This was

²² Among the other explanatory variables, the only statistically significant coefficient was the one for age. In particular, older subjects punish less than younger ones.

²³ We consider factor loadings with a value above 0.45 based on Hair *et al.*'s (1998) rules of thumb. The only emotion that did not pass this threshold was *surprise*. The Kaiser-Meyer-Olkin measure of sampling adequacy (0.751) indicates that factor analysis is appropriate. The on-line appendix contains a plot of the eigenvalues against the number of factors in their order of extraction (scree test), a scatter plot of the factor loadings, and a full list of the factor loadings.

confirmed in non-parametric tests where, for each class of the take rates, we compared the intensity of each emotion between the variable and constant ‘fine-to-fee’ ratios. None of these comparisons were statistically significant (Mann-Whitney test, $p > 0.1$). Figure 4 also provides some preliminary evidence of a positive relationship between take rate and negative emotions, and of a negative relationship between take rate and positive emotions.

[Figure 4 about here]

We tested for this using ordered logit regressions, one for each emotion. The dependent variable was the emotion of interest (s_i), whereas the independent variables were the take rate received from the take authority, a dummy variable which took the value of 1 when a constant ‘fine-to-fee’ ratio was employed, experience, gender (Male = 1 for male subjects), age, nationality of the subjects (non-UK = 1 for non-UK subjects), and two interaction terms, one between the dummy for the constant ‘fine-to-fee’ ratio and nationality, and another between the dummy for the constant ‘fine-to-fee’ ratio and the take rate.²⁴ The results of these regressions are shown in Table 6.

[Table 6 about here]

Negative emotions (in particular, anger and irritation) were significantly positively related to the take rate. Similarly, happiness was significantly negatively related to the take rate. This evidence is consistent with previous PTTG studies and the theory presented earlier.²⁵ It is also robust regardless of the background of the subjects (UK versus non-UK students) and the type of punishment technology employed.

We now consider the punishing behavior of the subjects and to what extent it can be explained by negative emotions. In the previous analysis of the relationship between punishing behavior and take rate, we found that, for take rates below the median, the punishment rate was higher under the constant ‘fine-to-fee’ ratio, whereas for take rates above the median, the punishment rate was higher under the variable ‘fine-to-fee’ ratio. This means that the bias in the predictive power of negative emotions that may characterize a PTTG with a variable ‘fine-to-fee’ ratio may be negative when the take rates are below the

²⁴ In the regressions, we employed robust standard errors to control for heteroscedasticity. Due to some subjects failing to report all the emotions, we have 1 missing observation for sadness, shame and envy (140 observations instead of 141), and 3 missing observations for contempt (138 observations instead of 141). The qualitative results do not change if we do not include the demographic variables among the explanatory variables. The results of the regressions without the demographics are reported in the on-line appendix.

²⁵ The coefficients of the other explanatory variables were mostly not significant. We briefly mention here those which were significant. In particular, subjects with increasing experience in laboratory experiments felt less fear ($p = 0.012$) and joy ($p = 0.027$). Older subjects experienced less fear ($p = 0.044$), anger ($p = 0.043$), sadness ($p = 0.055$) and jealousy ($p = 0.049$). Non-UK subjects under a constant ‘fine-to-fee’ ratio were, on average, sadder ($p = 0.041$). Finally, male subjects experienced more sadness ($p = 0.052$) and jealousy ($p = 0.056$).

median, and positive when take rates are above the median, compared to a PTTG with a constant ‘fine-to-fee’ ratio. It is thus important to distinguish between the role played by emotions when the take rates are low, and their role when the take rates are high. To do so, we estimated, for each emotion, the following model:

$$p_i = \beta_0 + \beta_1 \times s_i + \beta_2 \times d_{t>\tilde{t}} + \beta_3 \times s_i \times d_{t>\tilde{t}} + \beta_4 \times s_i \times d_{\sigma=\bar{\sigma}} + \beta_5 \times s_i \times d_{t>\tilde{t}} \times d_{\sigma=\bar{\sigma}} \\ + \beta_6 \times nonUK + \beta_7 \times experience + \beta_8 \times nonUK \times d_{\sigma=\bar{\sigma}} + \beta_9 \\ \times experience \times s_i + \beta_{10} \times age + \beta_{11} \times Male$$

where p_i is the punishment rate, s_i the intensity of the emotion of interest, $d_{t>\tilde{t}}$ a dummy which is equal to 1 when the responder experiences a take rate below the median take rate (\tilde{t}), $d_{\sigma=\bar{\sigma}}$ the dummy which identifies the treatment with a constant ‘fine-to-fee’ ratio, $nonUK$ a dummy for the nationality of the subject (= 1 for non-UK subjects), $experience$ the experience of the subject in economic experiments, and age and $Male$ the age and gender (Male = 1 for male subjects) respectively of the subject. To estimate the model, we ran a battery of Tobit regressions, one for each emotion.²⁶ Table 7 displays the results of the regressions for anger, irritation and contempt. We focused on these emotions as they were the emotions which more likely predicted the punishing behavior in previous PTTG studies. The on-line appendix contains the analysis for each of the other emotions.

[Table 7 about here]

For low take rates, an increase in the intensity of contempt induced responders to punish more ($\beta_1 = 45.42$) under both a variable and a constant ‘fine-to-fee’ ratio.²⁷ None of the other two negative emotions seem to explain the punishing behavior for low take rates. This brings us to the following result which, with respect to contempt, supports Hypothesis 2 and rejects Hypothesis 3.

Result 2. In line with Hypothesis 2, but in contrast to Hypothesis 3, contempt was the only negative emotion that explained the punishing behavior of the responders for low take rates, and its effect was similar under both a variable and a constant ‘fine-to-fee’ ratio.

Looking now at high take rates we note that, in line with previous literature, subjects who experienced higher anger, irritation and contempt punished more. In particular, under a variable ‘fine-to-fee’ ratio, a one-unit increase in the intensity of anger, irritation and contempt produced a 28.37%, 32.04% and 45.42% increase respectively in the punishment

²⁶ We used robust standard errors to control for heteroscedasticity. If we exclude the demographic variables from the explanatory variables, the results remain qualitatively the same. The results of these regressions are reported in the on-line appendix.

²⁷ The interaction between contempt and the dummy for the constant ‘fine-to-fee’ ratio is positive (as expected) but not significant.

rate.²⁸ However, much of this increase was due to the multiplier effect caused by the ‘fine-to-fee’ ratio. Indeed the impact of irritation, anger and contempt on the decision to punish in response to high take rates was hugely attenuated when we employed a constant ‘fine-to-fee’ ratio. In particular, a one-unit increase in irritation, anger and contempt raised the punishment by only 2.96%, 4.45%, and 13.41% respectively when the ‘fine-to-fee’ ratio was constant. This evidence supports both Hypotheses 2 and 3. We now present Results 3 and 4.

Result 3. In line with Hypothesis 2, anger, irritation and contempt explain the punishing behavior, even when we neutralize the multiplier effect caused by the variable ‘fine-to-fee’ ratio.

Result 4. Consistent with Hypothesis 3, the multiplier effect causes a bias when the take rates are high, and this bias accounts for 90.76%, 87.31% and 70.48% respectively of how much irritation, anger and contempt explain the punishing behavior in the PTTG with a variable ‘fine-to-fee’ ratio.²⁹

Looking at the other covariates, the emotional response of UK and non-UK subjects had a similar impact on the decision to punish, and was not affected by which technology punishment was employed. In contrast, subjects with increasing experience in economic experiments were more able to cope with contempt, as they punished significantly less when they experienced such an emotion compared to the inexperienced or less experienced subjects.³⁰ Note that experience in previous economic experiments did not eliminate the effect of contempt on punishing behavior, but only reduced it.

Role of expectations. As in previous PTTG studies, we also considered the role played by expectations in driving behavior and emotions.³¹ First, we note that there was no statistically significant difference in the way in which responders reported their expectations under constant and variable ‘fine-to-fee’ ratios (Mann-Whitney $p = 0.422$). Figure 5 shows that responders who punished were generally subjects who expected lower take rates than the actual ones (dots above the 45° line). This is consistent with previous findings on the PTTG.

²⁸ Note that the interaction between contempt and the dummy for high take rates is not significant, meaning that the effect of contempt for high take rates is the same as for low take rates.

²⁹ The bias is calculated as 23.92/28.37 for anger, 29.08/32.04 for irritation, and 32.01/45.42 for contempt.

³⁰ The differences in contempt due to experience may be attributed to a difference in expectations concerning the take rate rather than to a difference in coping. To test for this possibility, we also conducted a Tobit regression, where we included the expected take rate among the explanatory variables. The result did not change. Hence, we concluded that more experienced subjects punish less because they are able to cope better with contempt, and not because they have different expectations compared to less experienced subjects.

³¹ As in previous PTTG studies, expectations were not incentivized. This was because we did not want to introduce any distortion that would have limited the comparability of our study with previous literature. For a discussion of the reliability of measuring expectations without financial incentives, see Bosman *et al.* (2005).

In particular, 37 of the 77 optimistic responders (48.05%) punished the take authority, whereas only 17 of the 64 pessimistic responders (26.56%) punished. The difference achieved statistical significance ($\chi^2 = 6.83$, $p = 0.009$).³²

[Figure 5 about here]

We also studied whether emotions and behavior were affected by expectations. In particular, we included expectations for each regression of Table 6.³³ It turns out that expectations have a significant positive impact on envy and jealousy ($p = 0.031$ and 0.015 respectively). In other words, responders who expected higher take rates from the take authority, especially in comparison to the actual take rate, were more envious and jealous when they learnt about the decision of their counterpart. Finally, we included expectations in the regression of Tables 3 and 7. However, they do not seem to play a significant role in explaining the punishment behavior of the responders.

5. Conclusion

This study contributes to the experimental literature in economics that looks at the role of emotions on the decision to punish. In particular, we investigated whether previous findings about emotions and behavior in the PTTG were explained by the punishment technology adopted, and to what extent punishment can be truly attributed to negative emotions.

Our results provide clear-cut evidence that, in the standard PTTG, as much as 70% of the punishment triggered by the take rate is attributable to the multiplier effect caused by the variable ‘fine-to-fee’ ratio. This confirms Hypothesis 1. When we turn to the role played by negative emotions we find that, consistent with Hypothesis 2, they are still important predictors of the punishing behavior in the PTTG. In particular, anger, irritation and contempt appeared to be important driving forces for the punishing behavior of responders, especially in response to high take rates, and even once we controlled for the punishment technology. However, as postulated by Hypothesis 3, in the PTTG with variable ‘fine-to-fee’ ratio, their effect is magnified by as much as 90% for high take rates.

³² The result holds if we conduct a separate test for the treatment with a variable ‘fine-to-fee’ ratio ($\chi^2 = 3.544$, $p = 0.060$) and one for the treatment with a constant ‘fine-to-fee’ ratio ($\chi^2 = 3.077$, $p = 0.079$). Since expectations were elicited at the end of the experiment, it might be possible that subjects wrongly reported them. In particular, over-optimistic responders might have found it difficult to admit that they were wrong. If such bias exists, we should observe a correlation between expected take rates and actual take rates. However, this correlation was low and not significant (Spearman $\rho = 0.025$, $p = 0.772$).

³³ We tried different specifications to account for expectations. In one specification, we simply added among the explanatory variables the expected take rate. In another specification, we included the difference between the take rate and the expected take rate. The results were the same.

Acknowledgements

This research was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program “Investissements d’Avenir” (ANR-11-IDEX-007) operated by the French National Research Agency (ANR). I wish to thank seminar participants at the University of East Anglia, and in Lyon (5th Annual Meeting of ASFEE), Enrique Fatas, Abhijit Ramalingam, Elke Renner, Daniel J. Zizzo, and two anonymous reviewers for useful feedback and advice, Melanie Parravano for experimental assistance in the laboratory, and the Centre for Behavioural and Experimental Social Science (CBESS) and the University of East Anglia for financial support. The usual disclaimer applies.

References

- Albert, M., & Mertins, V. (2008). Participation and decision making: a three-person power-to-take experiment. *Justus-Liebig-University Giessen Discussion Paper Series in Economics* 05-2008.
- Alos-Ferrer, C., & Strack, F. (2014). From dual processes to multiple selves: Implications for economic behavior. *Journal of Economic Psychology*, 41, 1-11.
- Andreoni, J. (1995). “Cooperation in public-goods experiments: kindness or confusion?” *American Economic Review*, 85(4), 891-904.
- Averill, J. R. (1983). Studies on anger and aggression: implications for theories of emotion. *American Psychologist*, 38, 1145-1160.
- Ben-Shakhar, G., Bornstein, G., Hopfensitz, A., & van Winden, F. (2007). Reciprocity and emotions in bargaining using physiological and self-report measures. *Journal of Economic Psychology*, 28(3), 314-323.
- Bosman, R., & Riedl, A. (2004). Emotions and economic shocks in a first-price auction: an experimental study. *University of Amsterdam Working Paper*.
- Bosman, R., & van Winden, F. (2002). Emotional hazard in a power-to-take experiment. *Economic Journal*, 112(476), 147-169.
- Bosman, R., Hennig-Schmidt, H., & van Winden, F. (2006). Exploring group decision making in a power-to-take experiment. *Experimental Economics*, 9(1), 35-51.
- Bosman, R., Sutter, M., & van Winden, F. (2005). The impact of real effort and emotions in the power-to-take game. *Journal of Economic Psychology*, 26(3), 407-429.
- Brocas, I., & Carrillo, J. D. (2014). Dual-process theories of decision-making: a selective survey. *Journal of Economic Psychology*, 41, 45-54.
- Carpenter, J. P. (2007). The demand for punishment. *Journal of Economic Behavior & Organization*, 62(4), 522-542.

- Casari, M. (2005). On the design of peer punishment experiments. *Experimental Economics*, 8(2), 107-115.
- Cubitt, R. P., Drouvelis, M., & Gächter, S. (2011). Framing and free riding: emotional responses and punishment in social dilemma games. *Experimental Economics*, 14(2), 254-272.
- Damasio, A. R. (1994). *Descartes error: emotion, reason and the human brain*. New York: G.P. Putnam & Sons.
- de Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, 305(5688), 1254-1258.
- Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, 47(2), 268-298.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137-140.
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3), 817-868.
- Frank, R. H. (1988). *Passion within reason*. New York: W.W. Norton and Company.
- Frijda, N. H. (1986). *The emotions*. Cambridge: Cambridge University Press.
- Frijda, N. H. (2007). What might emotions be? Comments on the comments. *Social Science Information*, 46(3), 433-443.
- Friedman, J. W. (1969). On experimental research in oligopoly. *The Review of Economic Studies*, 36(4), 399-415.
- Galeotti, F. (2013). An experiment of waiting time and punishing behavior. *Economics Bulletin*, 33(2), 1383-1389.
- Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., Fischbach, A., & Preckel, F. (2014). "My questionnaire is too long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology*, 39(3), 188-205.
- Hair J. F., Tatham R. L., Anderson R. E., & Black, W. (1998). *Multivariate data analysis* (5th ed.). London: Prentice-Hall.
- Hopfensitz, A., & Reuben, E. (2009). The importance of emotions for the effectiveness of social punishment. *Economic Journal*, 119(540), 1534-1559.
- Joffily, M., Masclet, D., Noussair, C. N., & Villeval, M. C. (2013). Emotions, sanctions and cooperation. *Southern Economic Journal*, 80(4), 1002-1027.

- Kräkel, M. (2008). Emotions in tournaments. *Journal of Economic Behavior & Organization*, 67(1), 204-214.
- Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.
- LeDoux, J. E. (1996). *The emotional brain*. New York: Simon and Schuster.
- Loewenstein, G. F. (2000). Emotions in economic theory and economic behavior. *American Economic Review*, 90(2), 426-432.
- Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, 127(2), 267-286.
- Nikiforakis, N., & Normann, H. T. (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics*, 11(4), 358-369.
- Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. New York: Cambridge University Press.
- Phelps, E. A. (2009). The study of emotion in neuroeconomics. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics* (pp. 233-250). London: Elsevier.
- Picard, R. (1997). *Affective computing*. Cambridge: The MIT Press.
- Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, anger, and spite: emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes*, 68(3), 208-224.
- Rabin, M. (1993). Incorporating fairness into game-theory and economics. *American Economic Review*, 83(5), 1281-1302.
- Reuben, E., & van Winden, F. (2008). Social ties and coordination on negative reciprocity: the role of affect. *Journal of Public Economics*, 92(1-2), 34-53.
- Reuben, E., & van Winden, F. (2010). Fairness perceptions and prosocial emotions in the power to take. *Journal of Economic Psychology*, 31(6), 908-922.
- Rick, S., & Loewenstein, G. F. (2008). The role of emotion in economic behavior. In M. Lewis, J. M. Haviland-Jones, & L. Feldman-Barrett (Eds.), *The handbook of emotion* (3rd ed.). New York: Guilford.
- Schulz, J. F., Fischbacher, U., Thöni, C., & Utikal, V. (2014). Affect and fairness: dictator games under cognitive load. *Journal of Economic Psychology*, 41, 77-87.
- Sutter, M., Bosman, R., Kocher, M. G., & van Winden, F. (2009). Gender pairing and bargaining - beware the same sex! *Experimental Economics*, 12(3), 318-331.
- Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1), 75-98.

Tables and Figures

Table 1: Take rates

Treatment	n.	Mean	St. dev.
Variable σ	70	64.2	25.20
Constant σ	71	58.7	27.85

Table 2: Punishment rates

Treatment	Very Low t	Low t	High t	Very High t
Constant σ	36.17 (18.6)	58.43 (2.7)	76.6 (5.15)	93 (7.05)
Variable σ	31.39 (19.27)	59.12 (1.96)	73.08 (5.96)	97.5 (4.27)

Note: Standard deviations are in parenthesis.

Table 3: Tobit regression on punishment rates

	Regression 1			Regression 2		
	B	se	p	B	se	P
t_i	3.885***	1.077	0	4.009***	1.005	0
$d_{\sigma=\bar{\sigma}} \times t_i$	-2.727**	1.126	0.017	-2.841***	1.063	0.008
$d_{\sigma=\bar{\sigma}}$	162.285**	79.792	0.044	158.868**	74.332	0.034
$d_{\sigma=\bar{\sigma}} \times \text{non-UK}$	-6.019	39.41	0.879	3.95	39.673	0.921
non-UK	26.465	29.38	0.369	30.978	29.276	0.292
Experience				-5.528	8.564	0.52
Male				-30.789	19.387	0.115
Age				-3.951**	1.972	0.047
Constant	-285.573***	81.39	0.001	-172*	89.096	0.056
Obs	141			141		
Pseudo R-Square	0.075			0.085		
Df	136			133		
Prob > F	0.001			0.002		

Note: Tobit regression with robust standard errors. t_i and p_i are expressed in percentage. Hence, the beta coefficients identifies percentages. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Correlations of the emotions

	Fear	Envy	Anger	Sadness	Happiness	Shame	Irritation	Contempt	Joy	Jealousy	Surprise
Fear	1										
Envy	0.40	1									
Anger	0.27	0.54	1								
Sadness	0.40	0.48	0.59	1							
Happiness	0.04	-0.18	-0.25	-0.05	1						
Shame	0.26	0.13	0.39	0.39	0.27	1					
Irritation	0.24	0.46	0.75	0.45	-0.28	0.33	1				
Contempt	0.20	0.23	0.47	0.37	0.11	0.42	0.48	1			
Joy	0.12	-0.13	-0.16	0.05	0.79	0.30	-0.16	0.19	1		
Jealousy	0.38	0.83	0.53	0.48	-0.11	0.17	0.47	0.28	-0.13	1	
Surprise	0.06	-0.16	0.09	0.05	0.25	0.08	0.02	0.07	0.29	-0.16	1

Note: The table displays Spearman's rank correlation coefficients. Coefficients in bold are values above 0.7.

Table 5: Factor loadings for factor analysis

Emotion	Negative valence factor	Positive valence factor	Uniqueness
fear	0.518		0.732
envy	0.699		0.438
anger	0.811		0.332
sadness	0.699		0.513
happiness		0.837	0.245
shame	0.487		0.645
irritation	0.747		0.424
contempt	0.538		0.684
joy		0.843	0.278
jealousy	0.722		0.431

Note: Loadings for factor analysis using oblique rotation. We consider factor loadings greater than ± 0.45 (see Hair *et al.*, 1998). *Surprise* is not included as it does not pass this cut-off. A full list of the loadings is reported in the on-line appendix.

Table 6: Ordered logit regressions on emotions

	Fear			Envy			Anger			Sadness			Happiness			Shame		
	b	se	p	b	se	P	b	se	p	b	se	p	b	se	p	b	se	p
t_i	0	0.01	0.63	0.01	0.01	0.49	0.04***	0.01	0	0	0.01	0.66	-0.03**	0.01	0.01	0.02	0.02	0.31
$d_{\sigma=\bar{\sigma}}$	0.11	0.86	0.9	-0.61	0.92	0.51	0.48	0.81	0.56	-1.07	0.79	0.18	0.66	1.09	0.54	1.19	1.39	0.39
$d_{\sigma=\bar{\sigma}} \times t_i$	0	0.01	0.89	0.01	0.01	0.62	-0.01	0.01	0.52	0.01	0.01	0.58	-0.01	0.02	0.51	-0.02	0.02	0.41
non-UK	1.09*	0.57	0.06	0.17	0.51	0.74	0.34	0.48	0.48	0.6	0.46	0.19	0.29	0.52	0.57	0.63	0.63	0.32
Experience	-0.37**	0.15	0.01	-0.12	0.14	0.41	-0.02	0.16	0.88	-0.19	0.15	0.2	-0.07	0.19	0.72	-0.11	0.19	0.56
$d_{\sigma=\bar{\sigma}} \times$ non-UK	0.05	0.77	0.95	-0.08	0.65	0.9	0.34	0.67	0.61	1.37**	0.67	0.04	-0.11	0.7	0.88	0.5	0.84	0.55
Male	-0.28	0.4	0.48	0.2	0.35	0.57	0.42	0.38	0.27	0.66*	0.34	0.05	0.29	0.4	0.47	0.44	0.44	0.32
Age	-0.08**	0.04	0.04	-0.07	0.05	0.13	-0.07**	0.03	0.04	-0.06*	0.03	0.05	0.01	0.04	0.89	0.04	0.04	0.38
Obs.	141			140			141			140			141			140		
Pseudo R-Square	0.05			0.02			0.07			0.06			0.08			0.05		
Prob > F	0.01			0.31			0			0			0			0.31		

	Irritation			Contempt			Joy			Jealousy			Surprise		
	b	se	p	b	se	p	b	se	p	b	se	p	b	se	p
t_i	0.03***	0.01	0	0.01	0.01	0.19	-0.01	0.01	0.24	0.01	0.01	0.2	-0.01	0.01	0.64
$d_{\sigma=\bar{\sigma}}$	0.14	0.7	0.84	0.19	0.92	0.84	1.06	1.08	0.33	0.16	0.78	0.84	0.79	1.01	0.43
$d_{\sigma=\bar{\sigma}} \times t_i$	0	0.01	0.83	0	0.01	0.97	-0.02	0.02	0.3	-0.01	0.01	0.6	-0.01	0.01	0.44
non-UK	-0.47	0.45	0.3	0.48	0.46	0.3	0.27	0.51	0.6	0.33	0.51	0.52	0.19	0.47	0.69
Experience	-0.08	0.14	0.59	0.05	0.15	0.74	-0.35**	0.16	0.03	-0.08	0.13	0.55	-0.16	0.13	0.22
$d_{\sigma=\bar{\sigma}} \times$ non-UK	-0.08	0.67	0.91	0	0.66	0.99	0.14	0.68	0.84	0.36	0.64	0.57	0.38	0.61	0.53
Male	0.52	0.36	0.14	0.19	0.35	0.58	0.15	0.37	0.67	0.63*	0.33	0.06	-0.23	0.33	0.49
Age	-0.07	0.04	0.12	0.01	0.04	0.85	0.01	0.04	0.74	-0.09**	0.05	0.05	0.01	0.04	0.74
Obs	141			138			141			141			141		
Pseudo R-Square	0.07			0.02			0.04			0.03			0.02		
Prob > F	0			0.4			0.09			0.2			0.05		

Note: Ordered logit regressions with robust standard errors. * p<0.1, ** p<0.05, *** p<0.01.

Table 7: Tobit regressions on punishment for anger, irritation and contempt

	Anger			Irritation			Contempt		
	b	se	p	B	se	p	b	se	p
$d_{t>\bar{t}}$	40.29	38.31	0.29	44.76	38.78	0.25	130.84***	40.41	0
non UK	46.54	28.48	0.1	54.57*	28.17	0.05	40.99	27.56	0.14
$d_{\sigma=\bar{\sigma}} \times \text{nonUK}$	-26.22	34.97	0.45	-16.79	33.68	0.62	-3.15	35.58	0.93
Experience	1.39	16.24	0.93	19.46	19.32	0.32	33.51*	17.75	0.06
Male	-35.21*	20.07	0.08	-29.1	20.12	0.15	-22.16	19.39	0.26
Age	-2.21	1.86	0.24	-3.02	1.89	0.11	-4.27**	1.83	0.02
Anger	0.03	12.8	1						
$d_{t>\bar{t}} \times \text{Anger}$	28.37**	11.59	0.02						
$d_{\sigma=\bar{\sigma}} \times d_{t>\bar{t}} \times \text{Anger}$	-23.92**	10.15	0.02						
$d_{\sigma=\bar{\sigma}} \times \text{Anger}$	11.96	9.16	0.19						
Experience \times Anger	-1.33	3.68	0.72						
Irritation				5.14	14.52	0.72			
$d_{t>\bar{t}} \times \text{Irritation}$				32.04***	12.01	0.01			
$d_{\sigma=\bar{\sigma}} \times d_{t>\bar{t}} \times \text{Irritation}$				-29.08***	10.77	0.01			
$d_{\sigma=\bar{\sigma}} \times \text{Irritation}$				14.93	9.36	0.11			
Experience \times Irritation				-5.95	4.33	0.17			
Contempt							45.42**	17.55	0.01
$d_{t>\bar{t}} \times \text{Contempt}$							1.76	13.01	0.89
$d_{\sigma=\bar{\sigma}} \times d_{t>\bar{t}} \times \text{Contempt}$							-32.01***	11.66	0.01
$d_{\sigma=\bar{\sigma}} \times \text{Contempt}$							10.01	9.06	0.27
Experience \times Contempt							-13.44**	5.27	0.01
Constant	85.83***	12.58	0	88.19***	12.92	0	84.38***	12.23	0
Obs	141			141			138		
Pseudo R-Square	0.09			0.09			0.09		
Df	130			130			127		
Prob > F	0			0			0		

Note: Tobit regressions with robust standard errors. p_i is expressed in percentage. Hence, the beta coefficients identify percentages. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: Patterns of variable and constant ‘fine-to-fee’ ratios

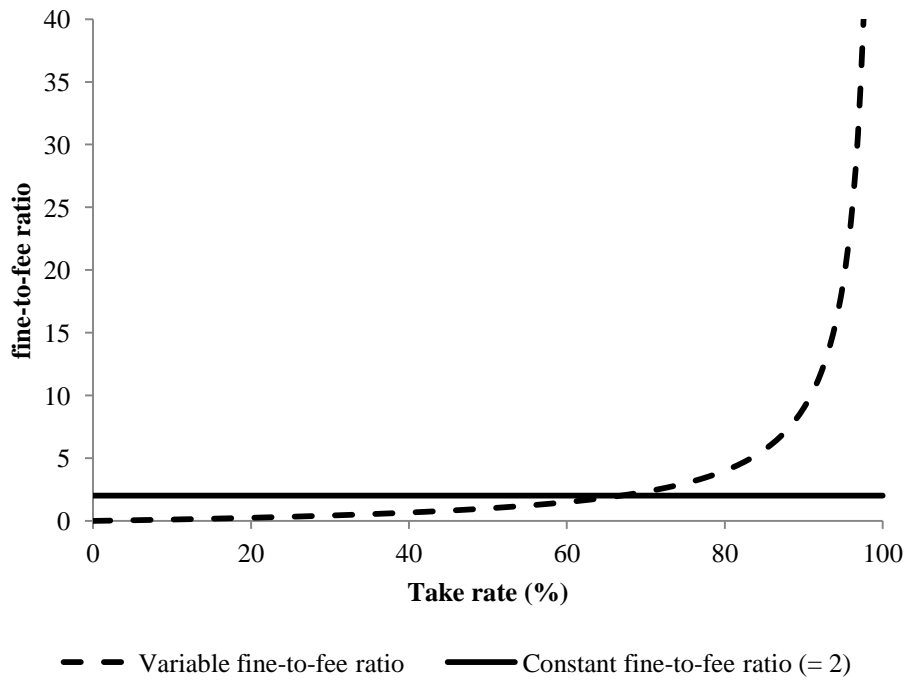
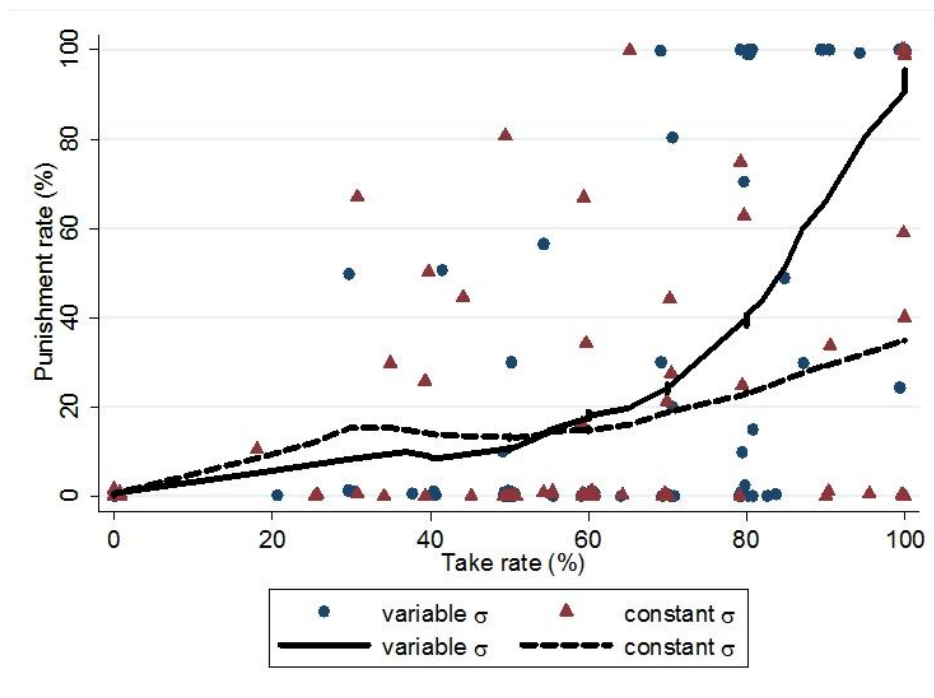
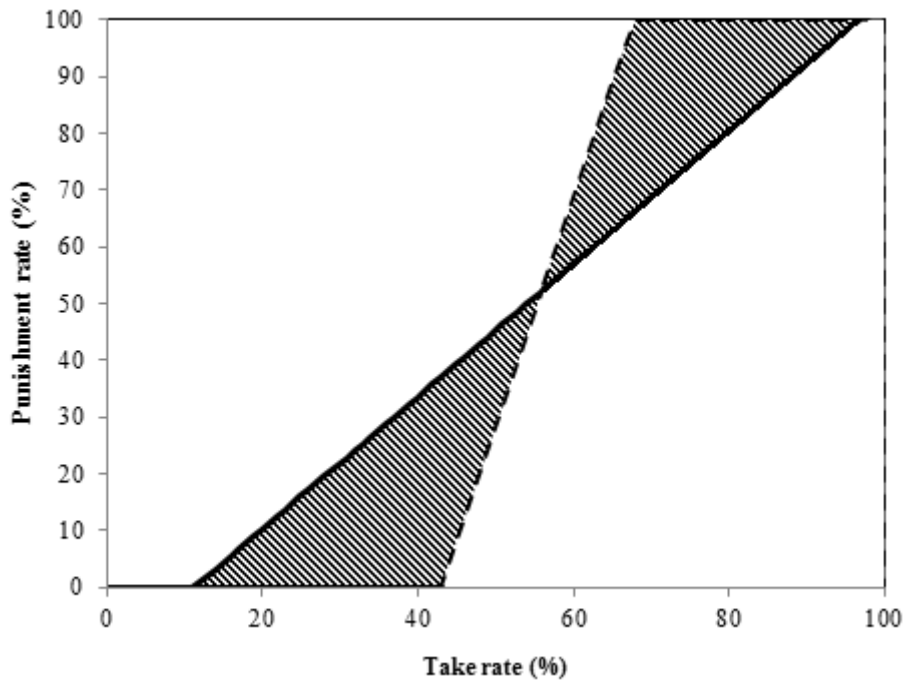


Figure 2: Relationship between punishment rate and take rate



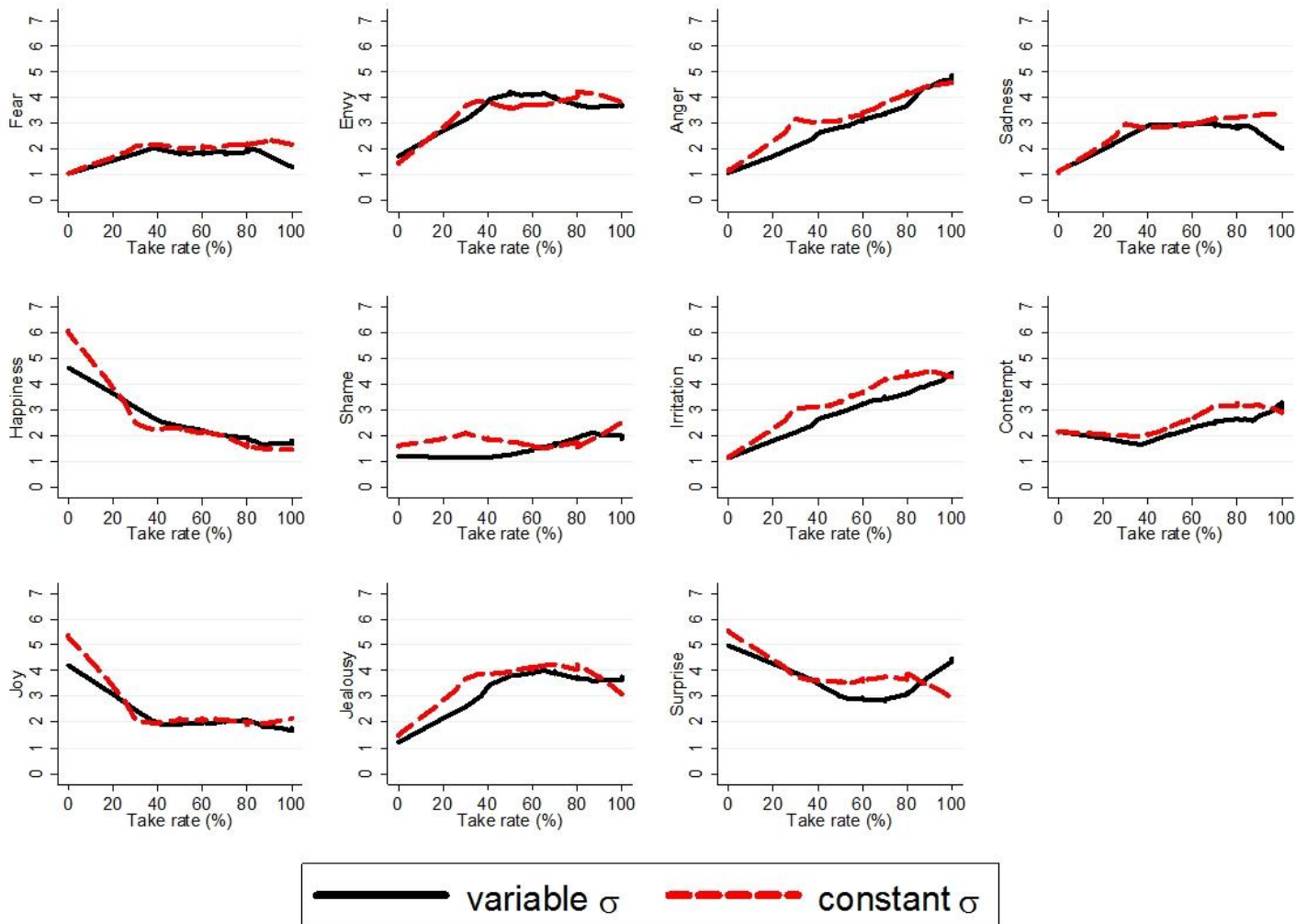
Note: The locally weighted regressions were computed using a bandwidth of 0.8 (80% of the data).

Figure 3: Predicted punishment rate against take rate



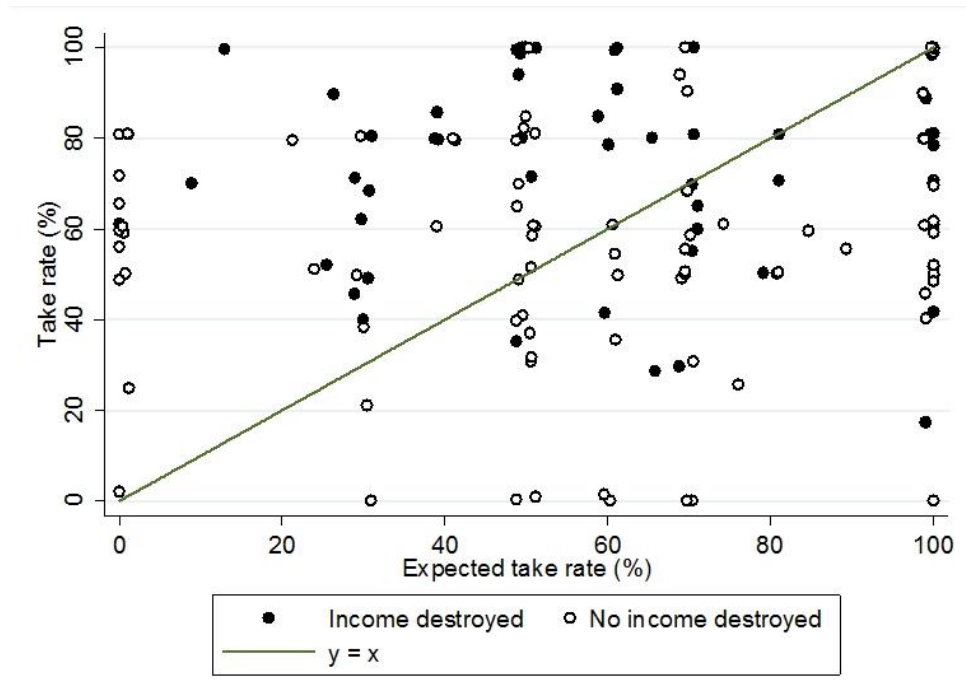
Note: The dashed line is the predicted punishment rate under a variable 'fine-to-fee' ratio, the solid line is the predicted punishment rate under a constant 'fine-to-fee' ratio. The crossed area measures the extent to which the punishment was inflated ($t > 56\%$) or deflated ($t < 56\%$) under the variable 'fine-to-fee' ratio compared to the constant 'fine-to-fee' ratio.

Figure 4: Relationship between emotions and take rate



Note: The locally weighted regressions were computed using a bandwidth of 0.8 (80% of the data).

Figure 5: Scatter plot of expected and actual take rates



Note: optimistic responders (who expected a higher take rate than the actual take rate) are identified by dots above the 45° line; pessimistic responders (who expected a lower take rate than the actual take rate) are identified by dots below the 45° line.