Inducing good behavior

van der Veen, A.

3. Inducing Good Behavior: Bonuses versus Fines in Inspection Games[1]

3.1. Introduction

There are many situations where authorities have preferences over individuals' choices. A tax authority wants taxpayers to truthfully report income, an employer wants an employee to work hard, a regulator wants a factory to comply with pollution regulations, police want motorists to observe speed limits, etc. A fundamental problem for authorities is how to induce compliance with desired behavior when individuals have incentives to deviate from such behavior. A standard approach is to monitor a proportion of individuals and penalize those caught misbehaving. To further encourage compliance, the authority may consider rewarding an individual who was inspected and found complying. For example, in 2003 the National Tax Service (NTS) of Korea introduced a system of bonuses for taxpayers found to have high compliance levels: bonuses included benefits such as providing a three-year exemption from tax audit and preferential treatment from financial institutions, e.g. reduced interest rates on loans (NTS, 2004, p. 31). Alternatively, the authority may consider increasing the sanctions on individuals who, upon inspection, are found not complying. For example, the Dutch government decided to increase the fine for undeclared savings from 100% to 300% in May 2009 (Tweede Kamer, 2009). In this chapter we study which of these two mechanisms is most successful in promoting good behavior. The essence of such situations is captured by the 'inspection game', which we describe in Section 3.2. In this game an authority chooses to inspect or not, and an individual chooses to comply or not, and the unique Nash equilibrium is in mixed strategies, with positive probabilities of inspection and non-compliance. Perhaps unsurprisingly, fines for non-compliant behavior increase the equilibrium probability of compliance. On the other hand, and perhaps paradoxically, bonuses for compliant behavior reduce the equilibrium probability of compliant behavior. Thus, according to standard game theoretical reasoning, fines, and not bonuses, should be used to encourage compliance in such settings. Previous experiments have revealed limited success of the Nash equilibrium for predicting behavior in games where the equilibrium is in mixed strategies (Ochs, 1995; Potters and Winden, 1996; Goeree and Holt, 2001; Goeree, Holt, and Palfrey, 2003). One

[1]This chapter is based on the identically titled paper joint with Daniele Nosenzo, Theo Offerman, and Martin Sefton and benefited from helpful comments of Daniel Seidmann, participants at the 2010 ESA Conference in Copenhagen, the 2010 CREED-CeDEx-CBESS Meeting in Amsterdam, and seminar audiences in Amsterdam. We are grateful to CREED programmer Jos Theelen for programming the experiment.

of the reasons why the Nash equilibrium does not provide an accurate description of behavior in these types of games is that it fails to capture 'own-payoff effects': players do change their behavior in response to changes in their own payoff, whereas the mixed strategy Nash equilibrium predicts that they will not. In the case of the inspection game, the own-payoff effect of introducing fines reinforces the theoretically expected effect: fines make non-compliance less attractive to the individual, and so the own-payoff effect points toward more compliance. However, the own-payoff effect of introducing bonuses for compliant behavior reduces the probability of non-compliance. Thus, Nash equilibrium and own-payoff effects point in different directions in this case, and so it is unclear whether the theoretical prediction that fines outperform bonuses in encouraging compliance will be supported in practice. We describe our experiment for comparing the effectiveness of bonuses and fines in Section 3.3. Our inspection game is framed as an employer-worker scenario where an employer can either inspect or not and a worker can either supply high or low effort. We designed three experimental treatments, each consisting of two parts. The first part was identical across treatments: subjects played a control version of the inspection game where the employer pays the worker a flat wage, unless she is inspected and found supplying low effort in which case the wage is not paid. In the second part of the BONUS treatment, subjects played a version of the game where the employer paid an additional bonus to the worker when the employer inspected and the worker supplied high effort. In the second part of the FINE treatment, subjects played a version of the game where the worker paid a fine to the employer if the employer inspected and the worker supplied low effort. Finally, in the second part of the CONTROL treatment, subjects continued playing the same game as in the first part. This design allows us to examine whether bonuses or fines are more effective in encouraging working/discouraging shirking. In addition, we are able to compare the efficiency properties of rewarding versus punishing mechanisms. We report our results in Section 3.4. We find that fines are more effective than bonuses in encouraging working and in raising combined earnings. This is in line with standard game theoretic predictions. However, the prediction that bonuses discourage working receives little support: although subjects shirk slightly more in the BONUS treatment than CONTROL the difference is small and not statistically significant. Moreover, the prediction that introducing bonuses will reduce combined earnings is not supported: the losses to employers are almost exactly offset by gains to workers. In general, standard comparative static predictions work well when own-payoff effects point in the same direction, but not otherwise. We show that observed deviations from Nash equilibrium predictions can be explained quite well by behavioral theories that incorporate loss aversion and can accommodate own payoff effects: Impulse Balance Equilibrium (Selten and Chmura, 2008) and an augmented version of Quantal Response Equilibrium (McKelvey and Palfrey, 1995). In Section 3.5 we discuss these results in relation to the existing literature and conclude.

Figure 3.1.: Inspection Games

Canonical Game

	H	L
I	$v-w-h$ $w-c$	$-h$ 0
N	$v-w$ $w-c$	$-w$ w

Game with Fines

	H	L
I	$v-w-h$ $w-c$	$\boldsymbol{f}-h$ $-\boldsymbol{f}$
N	$v-w$ $w-c$	$-w$ w

Game with Bonuses

	H	L
I	$v-w-\boldsymbol{b}-h$ $w+\boldsymbol{b}-c$	$-h$ 0
N	$v-w$ $w-c$	$-w$ w

Notes: Employer is the ROW player, Worker is the COLUMN player. Within each cell, the Employer's payoff is shown at the top and the Worker's payoff at the bottom.

3.2. Inspection Games

We study a simple simultaneous move inspection game. An employer can either inspect (I) or not inspect (N), and a worker can supply either high (H) or low (L) effort. The employer incurs a cost of h from inspecting, and high effort results in the worker incurring a cost of c and the employer receiving revenue of v. The employer pays the worker a wage of w, unless the worker supplies low effort and the employer inspects. The resulting payoffs are shown in the leftmost panel of Figure 3.1. We assume that all variables are positive and $v > c$, $w > h$, $w > c$. Note that joint payoffs are maximized when the worker supplies high effort and the employer does not inspect. Following Fudenberg and Tirole (1992, p. 17), we refer to this as the canonical version of the game. For a review of the theory of inspection games see Avenhaus, Von Stengel, and Zamir (2002).

The canonical game has a unique Nash equilibrium where the employer inspects with probability $p_c = c/w$ and the worker chooses low effort ("shirks") with probability $q_c = h/w$. In this equilibrium the employer's expected payoff is $\pi_c^{employer} = v–w–hv/w$, the worker's expected payoff is $\pi_c^{worker} = w–c$, and joint expected payoffs are $\pi_c = v–c–hv/w$. We now compare two possibilities for encouraging high effort relative to the canonical version of the game: imposing an additional fine on workers caught supplying low effort, versus paying a bonus to workers who are inspected and found supplying high effort. Suppose an additional fine f is imposed on a worker caught shirking, resulting in the payoff matrix shown in the middle panel of Figure 3.1. Note that the fine is a transfer between the worker and the employer. Now the unique Nash equilibrium has the employer inspect with probability $p_f = c/(w+f)$ and the worker shirk with probability $q_f = h/(w+f)$. Thus, according to Nash equilibrium, fines discourage both inspections and shirking. In Nash equilibrium expected payoffs are $\pi_f^{employer} = v–w–hv/(w+f)$, and $\pi_f^{worker} = w–c$, and so the employer benefits from the introduction of fines, while the worker's expected payoff is independent of fines. According to Nash equilibrium, fines enhance efficiency because joint expected payoffs are reduced by low effort and/or inspection, and both of these are discouraged by a fine on workers caught shirking. Next, we examine the case where the employer pays a bonus b to a worker who is inspected and found to have chosen high effort. The payoff matrix for this game is shown in the rightmost panel of Figure 3.1. Now in equilibrium the employer inspects

with probability $p_b = c/(w+b)$ and the worker shirks with probability $q_b = (h+b)/(w+b)$. According to Nash equilibrium bonuses reduce the probability of inspection and increase the probability of shirking. The workers equilibrium expected payoff is $\pi_b^{worker} = w–c+cb/(w+b)$, increasing in b, while the employer's is $\pi_b^{employer} = v–w–v(h+b)/(w+b)$, decreasing in b. Overall, bonuses reduce joint expected payoffs because the beneficial effect of less frequent inspection is outweighed by the detrimental effect of increased shirking. As is well known, comparative static predictions based on mixed strategy Nash equilibrium can often be counter-intuitive. This is because a player's equilibrium probability must keep her opponent indifferent among actions, and so a player's own decision probabilities are determined by the opponent payoffs and not by own payoffs. Consider, for example, how the introduction of a bonus affects own-payoffs from the perspective of the worker. Introducing the bonus has no effect on the expected payoff from shirking, but increases the expected payoff from choosing high effort (for a given inspection probability). Based on this own-payoff effect, one might expect the worker to shirk less frequently following the introduction of bonuses. However, the Nash equilibrium prediction goes in the opposite direction: bonuses lead to an increase in the equilibrium shirking probability. Previous experimental work (e.g., Ochs, 1995; Goeree and Holt, 2001; Goeree, Holt, and Palfrey, 2003) shows that counterintuitive Nash equilibrium predictions are often rejected by the data: changing a player's own payoff does have an impact on that player's decision probabilities. Goeree and Holt (2001) observe own-payoff effects in one-shot games; Ochs (1995) and Goeree, Holt, and Palfrey (2003) observe own-payoff effects even after players have had ample opportunities to learn. Note that own-payoff effects may either reinforce or counteract equilibrium forces. Introducing fines into the inspection game generates an own-payoff effect that pulls workers' behavior in the same direction as Nash equilibrium predictions: introducing fines does not change the expected payoff from choosing high effort but does reduce the expected payoff from shirking. Thus the own-payoff effect discourages shirking, and this is consistent with the Nash equilibrium comparative static prediction. Similarly, own-payoff effects reinforce Nash equilibrium predictions about inspection probabilities in the inspection game with bonuses, but counteract Nash equilibrium predictions in inspection games with fines. In summary, given the evidence on the importance of own-payoff effects in previous experiments, it is not clear that experimental evidence will support the standard game theoretical analysis outlined above. In particular, the own-payoff effects arising when bonuses are paid to workers who are inspected and found supplying high effort may make them a more effective tool for encouraging effort than suggested by standard theory.

3.3. Experimental Design and Procedures

The experiment consisted of fifteen sessions at the University of Nottingham. Ten subjects participated in each session. Subjects were recruited from a campus-wide distribution list and

Figure 3.2.: Parameterization of the Inspection Games Used in the Experiment

Canonical Game

	H	L
I	52 25	12 20
N	60 25	0 40

Game with Fines

	H	L
I	52 25	**32** **0**
N	60 25	0 40

Game with Bonuses

	H	L
I	**32** **45**	12 20
N	60 25	0 40

Notes: Employer is the ROW player, Worker is the COLUMN player. Within each cell, the Employer's payoff is shown at the top and the Worker's payoff at the bottom.

no subject participated in more than one session.[2] No communication between subjects was permitted throughout a session. At the beginning of a session subjects were randomly assigned to computer terminals and were informed that the experimental session would consist of two parts, during each of which they could earn 'points'. Subjects were also told that their cash earnings for the session would be based on all points accumulated in both parts of the experiment. Instructions for Part One were then distributed and read aloud. At the end of these subjects had to answer a series of questions to test their comprehension of the instructions. A monitor checked the answers and dealt with any questions in private. We did not continue with the experiment until all subjects had correctly answered all the questions. Part One then consisted of 40 rounds. At the beginning of the first round subjects learned their role: five subjects were assigned the role of 'Employer' and five the role of 'Worker'. Subjects kept these roles for the entire session (i.e. for both Part One and Part Two). Across rounds subjects were randomly matched in pairs consisting of one Employer and one Worker, and in each round each pair played the canonical inspection game shown in the leftmost panel of Figure 3.2.[3] At the end of each round subjects were informed of their own and their opponents' choices and point earnings. Subjects were also shown their accumulated point earnings and a table with the distribution of choices across all subjects in the session for the previous twenty rounds.

At the end of Part One subjects were given instructions for Part Two, which were then read aloud. These explained that the second part consisted of another 80 rounds, again with pairings randomly determined at the beginning of each round. In our five CONTROL sessions these rounds used the same earnings table as in Part One. In our five FINE sessions the earnings table was as in Part One except that the worker would pay a fine of 20 points to the employer if the worker chose low effort and the employer chose to inspect. Thus in Part Two of the

[2] Subjects were recruited through the online recruitment system ORSEE (Greiner, 2004). Instructions are available in Appendix C.

[3] Point earnings were derived from the game described in the previous section (see Figure 1) with $v = 60$, $c = 15$, $h = 8$, $w = 20$, and with 20 points added to all outcomes to ensure that subjects could not make losses in any of the games used in the experiment. These parameters were chosen so that Nash equilibrium probabilities are not too close to 0, 0.5 or 1 (all probabilities lie in the intervals $[0.2, 0.4]$ or $[0.6, 0.8]$). We also sought separation between games with and without bonuses or fines so that, where a change in behavior is predicted by standard theory, the predicted change in probabilities across games is at least 20 percentage points.

Table 3.1.: Choice Proportions, Average by Treatment

	Part One			Part Two		
	CONTROL	FINE	BONUS	CONTROL	FINE	BONUS
Proportion of Shirking	0.39	0.52	0.45	0.44	0.23	0.50
Nash	*0.40*	*0.40*	*0.40*	*0.40*	*0.20*	*0.70*
Proportion of Inspecting	0.80	0.77	0.78	0.81	0.62	0.45
Nash	*0.75*	*0.75*	*0.75*	*0.75*	*0.375*	*0.375*

Notes: table shows the proportion of shirking/inspecting decisions in the last 20 rounds of each Part of the experiment.

experiment subjects in the FINE sessions played the inspection game shown in the middle panel of Figure 3.2 on the preceding page. In our five BONUS sessions the earnings table was as in Part One except that the employer would pay a bonus of 20 points to the worker if the worker chose high effort and the employer chose to inspect (rightmost panel of Figure 3.2). At the end of Part Two subjects were paid in cash according to their accumulated point earnings from all rounds using an exchange rate of £0.004 per point. Sessions took about 40 minutes on average and earnings ranged between £10.2 and £23.1, averaging £14.9 (approximately US$24 at the time of the experiment).

3.4. Results

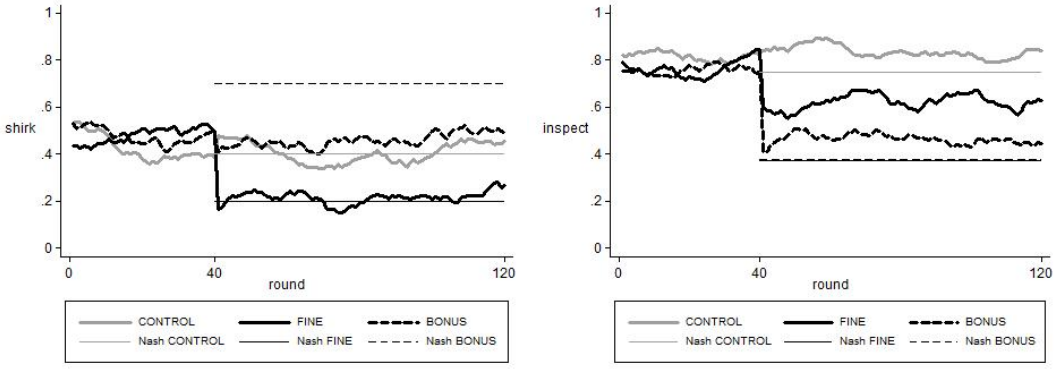
3.4.1. Inspecting and Shirking Probabilities

Figure 3.3 on the next page displays the smoothed proportions of inspecting and shirking decisions across all the rounds of the experiment. For some cases there is a clear change in behavior in round 41, following the transition from Part One to Part Two and the introduction of fines or bonuses, but otherwise the observed proportions appear quite stable across rounds. Table 3.1 reports the proportions of shirking and inspecting over the last 20 rounds of each Part of the experiment. The Nash equilibrium predictions for choice probabilities are also reported for comparison. The first 40 rounds of the experiment (Part One) are common to the three treatments, and we do not find any significant differences in the proportions of shirking or inspecting across treatments (Kruskal-Wallis test p-values are 0.37 for shirk and 0.78 for inspect).[4] Averaged across all sessions the observed proportion of shirking decisions is 45% and the observed proportion of inspecting decisions is 78%: both statistics compare favorably with predictions made by Nash equilibrium (40% and 75%, respectively).[5]

[4] Our non-parametric analysis is based on two-tailed tests applied to 5 independent observations per treatment. We consider data from each session as one independent observation. Tests are applied to averages based on the last 20 rounds of each Part of the experiment. The data analysis does not lead to different results if we focus on all rounds.

[5] Treating data from each session as an independent observation and using a one-sample sign test, we cannot reject the hypothesis that in Part One the proportions of shirking and inspecting across the 15 sessions are equal to Nash equilibrium predictions ($p = 1.00$ for shirking and $p = 0.18$ for inspecting).

Figure 3.3.: Proportions of Shirking (left panel) and Inspecting (right panel) across Treatments



Notes: for each round, the average of the proportions in the interval [round – 5, round + 5] is displayed.

In Part Two of the experiment the proportions of shirking and inspecting diverge significantly across treatments (Kruskal-Wallis test: $p = 0.02$ for shirk, and $p = 0.01$ for inspect).[6] Clearly, the changes in payoff matrices introduced in Part Two of the different treatments caused subjects to adjust their behavior. For pair-wise statistical comparisons between treatments we use Mann-Whitney rank-sum tests. As predicted, we find less shirking in FINE (23%) than in CONTROL (44%), and the difference is statistically significant ($p = 0.02$). Although Nash equilibrium predicts workers will shirk considerably more in BONUS than in CONTROL (70% vs. 40%), shirking in BONUS is only slightly higher than in CONTROL (50% vs. 44%), and the difference is not statistically significant ($p = 0.55$). As for inspection probabilities, these are significantly lower in FINE than CONTROL ($p = 0.01$) and BONUS than CONTROL ($p = 0.01$). We also note, however, that the inspection probability in FINE is considerably higher than predicted (62% vs. 37.5%), while the proportion of inspections in BONUS is closer to the theoretical level (45% vs. 37.5%). In fact, whereas Nash equilibrium predicts that introducing bonuses and fines have the same effect on inspection probabilities, we find a statistically significant difference in the proportions of inspections between FINE and BONUS ($p = 0.01$).

3.4.2. Earnings

Table 3.2 reports average earnings per game across treatments in the last 20 rounds of Part Two of the experiment. Nash equilibrium predictions are also reported for comparison.

In principle, joint earnings can range from 32 points (when the employer inspects and the worker shirks) to 85 (when the employer does not inspect and the worker works). Theory predicts

[6] According to one-sample sign tests, the proportion of shirking is significantly different from the equilibrium prediction in Part Two of BONUS ($p = 0.06$), but not in FINE ($p = 0.37$) or CONTROL ($p = 1.00$). The proportion of inspecting in Part Two of the experiment differs significantly from the Nash prediction in FINE and BONUS ($p = 0.06$ in both cases), but not in CONTROL ($p = 0.37$). These p-values are each based on five independent sessions so insignificant results should be treated with caution.

Table 3.2.: Earnings in Part Two, Average by Treatment

	Part Two		
	CONTROL	FINE	BONUS
Joint Earnings	58.7 (5.75)	69.6 (2.64)	58.9 (2.40)
Nash	*61.0*	*73.0*	*50.5*
Worker Earnings	24.2 (1.08)	22.5 (1.38)	32.7 (1.01)
Nash	*25.0*	*25.0*	*32.5*
Employer Earnings	34.5 (5.11)	47.1 (1.35)	26.1 (2.30)
Nash	*36.0*	*48.0*	*18.0*

Notes: table shows average point earnings per game (last 20 rounds only). Standard deviations based on session averages in parentheses.

that joint earnings are equal to 61 points in the game used in CONTROL. In the experiment, earnings in our CONTROL sessions are close to this, averaging 58.7 points across the last 20 rounds of Part Two. Theory also predicts that fines are beneficial and bonuses are detrimental for efficiency. Using Mann-Whitney rank-sum tests, we find that, consistent with these predictions, joint earnings in FINE are higher than in CONTROL, and the difference in the distributions is statistically significant ($p = 0.01$). On the contrary, we find no evidence that bonuses hamper efficiency: in fact, introducing bonuses slightly increases on average joint earnings relative to CONTROL, although the effect is not statistically significant ($p = 0.85$). A second aspect of our data is worth discussing: while according to Nash equilibrium the introduction of fines is Pareto improving, as it is predicted to leave the workers' earnings unchanged relative to CONTROL and to increase the employer's payoff, we find that fines are in fact detrimental for workers. In FINE, workers earn about 1.5 points per game less than in CONTROL ($p = 0.06$). Fines are instead beneficial for the employer as predicted ($p = 0.01$). Thus, the introduction of fines has distributive consequences that are not fully accounted for by standard theory: employers are better off when fines are introduced, but this occurs at the expenses of workers who are worse off relative to CONTROL, although the latter effect is small in magnitude and only weakly statistically significant. The introduction of bonuses has instead the predicted distributive consequences: it significantly increases the worker's payoff and decreases the employer's payoff ($p = 0.01$ and $p = 0.02$ respectively).

3.4.3. Explaining Observed Behavior

Whereas Nash equilibrium predictions seem to capture well the comparative static effects of fines on shirking behavior and bonuses on inspecting behavior, they do not capture observed effects of fines on inspections or bonuses on effort. It is notable that the instances where Nash predictions fail are those where own-payoff effects, as discussed in Section 3.2 on page 29, work in the opposite direction to equilibrium effects. Table 3.3 on the facing page contains predicted choice probabilities made by two alternative concepts: Quantal Response Equilibrium (QRE)

Table 3.3.: Predicted Choice Probabilities

	Probability of Shirking			Probability of Inspecting		
	CONTROL	FINE	BONUS	CONTROL	FINE	BONUS
Results	0.44	0.23	0.50	0.81	0.62	0.45
Nash	0.40	0.20	0.70	0.75	0.375	0.375
QRE ($\lambda = 0.989$)	0.46	0.19	0.68	0.76	0.41	0.35
IBE	0.41	0.16	0.43	0.68	0.61	0.40
Nash$^{\text{with loss-aversion}}$	0.25	0.11	0.54	0.60	0.23	0.33
QRE$^{\text{with loss-aversion}}$ ($\lambda = 0.289$)	0.42	0.10	0.46	0.69	0.47	0.36

Notes: Results shows the proportion of shirking/inspecting decsisions in the last 20 rounds of the second part; The other rows give the predicitions according to the different equilibrium concepts.

and Impulse Balance Equilibrium (IBE).[7] The predictions are for our Part Two data. In QRE players' choices are stochastic. Better responses (i.e. yielding a higher expected payoff) are predicted to be played more frequently than worse responses, but not with 100% certainty. The degree of precision λ with which players choose their responses determines the extent to which QRE predictions deviate from Nash equilibrium predictions. When $\lambda = 0$ players choose actions equi-probably and in the limit as λ approaches ∞ players always choose their best-response. Part One data is used to estimate the QRE precision parameter λ in our experimental setting.[8] For the estimated value of λ QRE predictions are generally close to Nash equilibrium predictions.

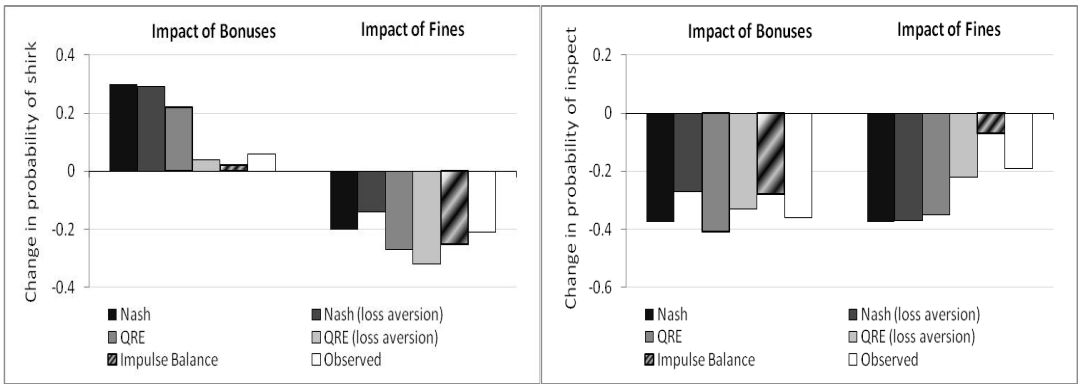
IBE is based on the idea that players look at forgone payoffs when they adjust their decision probabilities: choosing an option that yields a lower payoff than the alternative option generates an 'impulse' in the direction of the non-chosen option. Impulses generated by foregone payoffs that represent a 'loss' relative to a player's security payoff level (her pure strategy maximin value) weigh twice as much as foregone 'gains'. In equilibrium, players choose the decision probabilities such that the impulses of foregone payoffs are equal across options. IBE predictions differ markedly from Nash equilibrium when own payoff and Nash equilibrium effects are in conflict: the IBE predicted probability of shirking in BONUS is 43% (versus the 70% Nash prediction) and the predicted probability of inspecting in FINE is 61% (versus 37.5%). The fact that Nash equilibrium and QRE are not augmented by loss-aversion while IBE is has generated a recent debate about whether the incorporation of loss-aversion is what drives the observed differences in performance across these equilibrium concepts (see Selten and Chmura, 2008; Brunner, Camerer, and Goeree, 2011; Selten, Chmura, and Goerg, 2011). To examine this possibility, Table 3.3 also reports predictions made by Nash equilibrium and QRE when these concepts are augmented with loss-aversion.[9] Incorporating loss-aversion into the concepts generally improves the performance

[7] Appendix D contains details on the procedures used to derive the equilibrium predictions for IBE and QRE.

[8] As in Selten and Chmura (2008) and Brunner, Camerer, and Goeree (2011), we calculate the best fitting overall estimate for λ in our data by minimizing the sum of mean squared distances of the predicted QRE probabilities from the observed session-averaged choice probabilities in the experiment. This yields an estimated λ of 0.989. This estimated value of λ was obtained using data from Part One as this allows us to make out-of-sample predictions for behavior in the games used in Part Two of the experiment.

[9] As in Selten and Chmura (2008) we incorporate loss aversion by transforming payoffs above the security level as follows. If x is the payoff and m is the security level, any payoff $x > m$ is transformed into $x' = m + (x - m)/2$.

Figure 3.4.: Changes in Shirk (left) and Inspect (right) after Introduction of Bonuses and Fines.



Notes: in each round, the average is displayed of the proportions of (max) 5 previous rounds, the current round and (max) 5 future rounds.

of QRE, but not the performance of Nash equilibrium. Overall, the comparative static effects observed in our experiment are generally better captured by IBE and QRE with loss-aversion than by Nash equilibrium analysis or by QRE without loss-aversion. This is summarized in Figure 3.4. The Figure shows how the introduction of bonuses and fines affect the probability of shirking and inspecting relative to CONTROL according to the three solution concepts, as well as in the data for the last 20 rounds of Part Two.

When Nash equilibrium effects and own-payoff effects work in the same direction (i.e. for the impact of fines on shirking and the impact of bonuses on inspections) there is little to choose among the various solution concepts. When Nash equilibrium effects and own payoff effects work in opposite directions (i.e. for the impact of fines on inspecting and the impact of bonuses on shirking), Nash equilibrium (with or without loss-aversion) is outperformed by the alternative concepts. Among these, IBE and QRE augmented by loss-aversion perform better than QRE without loss-aversion. Nash equilibrium predicts that bonuses increase shirking by 30% relative to CONTROL, whereas shirking only increases by about 6% in our data. This observed effect compares quite favorably with the comparative static predictions made by IBE (a predicted 2% increase in shirking) and QRE augmented by loss-aversion (a predicted 4% increase), but not with the comparative static predictions made by QRE without loss-aversion (a predicted 22% increase). Similarly, Nash equilibrium predicts that fines reduce inspection rate by about 37% relative to CONTROL, whereas inspection rates actually fall by about 19%. QRE without loss-aversion predicts a decrease in inspecting by 35%, whereas the predicted magnitude of the decrease is smaller in IBE and QRE with loss-aversion (about 20% or less).

The exact procedure is discussed in Appendix D.

3.5. Conclusion

We compare the effectiveness of bonuses and fines as instruments for encouraging compliance in inspection games. In our setting the incentive for a worker to work is given by the monitoring activity of an employer and the costs/benefits incurred by the worker when she is inspected and found to have worked or shirked. The unique Nash equilibrium of the game is in mixed strategies with positive probabilities of inspection and shirking. We find that bonuses targeted at those inspected and found working are not effective in encouraging working: in fact, subjects in our experiment shirk slightly more often when bonuses are present, although the effect is not statistically significant. On the other hand, we find that introducing harsher fines for shirkers is an effective tool for encouraging working. The question of whether rewards or punishments are a better tool for inducing socially desirable behavior has been addressed in previous experimental work. Most of the literature has used two-stage games where in the second stage, after having observed choices made in the first stage, players can incur costs to punish or reward other players. Players are not predicted to use costly rewards or punishments if they are solely concerned about own earnings, but they might if they have preferences for reciprocity. In fact, a large experimental literature documents the willingness of some people to eschew private interests and react positively toward those that treat them well (positive reciprocity) or negatively toward those that treat them poorly (negative reciprocity). In particular, early studies of games that allow for both positive and negative reciprocity found that the latter has a particularly strong impact (Abbink, Irlenbusch, and Renner, 2000; Offerman, 2002; Charness and Rabin, 2002). These findings are echoed in Andreoni, Harbaugh, and Vesterlund (2003) who investigate the effects of rewards and punishments in a proposer-responder game where the proposer chooses an amount to transfer to the responder and the responder can then either punish or reward the proposer. They find that proposers' transfers are particularly sensitive to the threat of punishment, although rewards have also positive effects. Similarly, Sefton, Shupp, and Walker (2007) examine the effect of rewards and punishments on contributions in a repeated public good game and find that punishments help subjects to sustain higher cooperation levels compared to a control game with no reward/punishment opportunities, whereas the possibility of rewards has only a transient effect.[10] Our research differs from these studies in that we do not study discretionary, or informal, rewards and punishments, but we rather focus on formal bonuses and fines that are automatically triggered after specific combinations of actions chosen by the

[10] More recent research has shown that the effectiveness of rewards and punishments in settings such as this depends on the rewarding/punishing technology. Sutter, Haigner, and Kocher (2010) find that when the benefit/cost of receiving reward/punishment is three times larger than the cost of delivering it (i.e. with a 3:1 technology), both mechanisms are effective in encouraging contributions. Similarly, Rand, Dreber, Ellingsen, Fudenberg, and Nowak (2009) find that rewards are as effective as punishments in sustaining cooperation in a repeated public good game experiment with unknown time horizon and with a 3:1 reward/punishment technology. Gürerk, Irlenbusch, and Rockenbach (2006) study a public good game where the rewarding mechanism displays a 1:1 technology and a punishment mechanism displays a 3:1 technology. They find that only the latter have an impact on contributions. Gürerk, Irlenbusch, and Rockenbach (2009) use a public goods game where one group member (the 'leader') can reward or punish the other contributors. Although both rewarding and punishment mechanisms display a 3:1 technology, they find that contributions are higher when punishments are used.

players.[11] Moreover, we study bonuses and fines that are pure transfers from one party to another, and so have no direct efficiency implications. Thus, bonuses or fines can only enhance performance to the extent that they succeed in inducing behavior that is more aligned with the group interest. Finally, unlike previous research on the effect of rewards/ punishments in social dilemmas, in our game standard theory predicts that bonuses and fines will affect performance. As far as we are aware there have only been two experimental studies of inspection games. Dorris and Glimcher (2004) observe the behavior of human and monkey subjects in inspection games with different parameterizations of the inspection cost.[12] In some experiments they had humans playing against humans, whereas in others they had humans or monkeys in the role of Worker playing against a computer in the role of Inspector. They find that (human and monkey) Workers' behavior is close to Nash equilibrium predictions only for high inspection costs. Dorris and Glimcher (2004) do not study the impact of bonus or fines in their setup. Rauhut (2009) studies the impact of the severity of the punishment in an inspection game. His set up differs from ours in that the punishment hurts the inspectee but does not affect the payoff of the inspector in any way. A consequence is that an increase in the punishment decreases the probability of inspection but leaves the probability of shirking unaffected in the Nash equilibrium. Nevertheless, he finds that inspectees shirk less often when the punishment is increased, in agreement with the own-payoff effect.[13] Our study differs from his also in that we study reward as well as punishment. As far as we are aware ours is the first study to compare positive and negative incentives in inspection games. Our study also contributes to a recent literature evaluating different solution concepts for predicting behavior in games with mixed strategy equilibria (e.g., Selten and Chmura, 2008; Brunner, Camerer, and Goeree, 2011; Selten, Chmura, and Goerg, 2011). Standard game theoretical analysis applied to the game used in our experiment yields the perhaps paradoxical result that introducing bonuses increases considerably the probability that the employee will shirk. While in our experiment we do observe a slight increase in shirking in the presence of bonuses, this effect is much smaller than predicted by Nash equilibrium and is not statistically significant. This is more in line with the predictions made by alternative concepts such as Impulse Balance Equilibrium and Quantal Response Equilibrium (although, for our data, the latter concept performs better than Nash equilibrium only if it incorporates loss aversion). More generally, our results show that when Nash equilibrium and alternative predictions diverge we find more support for the latter than for the former. In this study we have focused on the case where rewards and punishments are simple transfers between the interacting parties (e.g. monetary fines for misconduct or bonuses for good conduct). This seems to be a useful starting point as the connections between incentives, behavior, and earnings

[11] There have been public good game experiments where rewards/punishments are automatically assigned to players depending on how their contributions compare with others. Dickinson (2001) assigns rewards/punishment points to the highest/lowest contributor in the group, and Falkinger, Fehr, Gächter, and Winter-Ebmer (2000) assigns rewards/punishments to those who contribute more/less than average.

[12] See also Glimcher, Dorris, and Bayer (2005).

[13] In fact, Rauhut studies a game where two inspectors interact with two inspectees who are involved in a prisoners' dilemma. Under some assumptions, this expanded game has the same characteristics as an inspection game.

are straightforward to interpret: bonuses and fines have no direct efficiency consequences unless they induce a change in behavior. We find that fines, but not bonuses, enhance efficiency. An interesting extension would be one where the costs and benefits of rewarding/being rewarded are asymmetric (e.g., when bonuses consist of medals and prizes, that may have more value for the person receiving them than for the person awarding them). If the bonus remains equally costly to the inspector while it becomes more beneficial to the inspectee, our results suggest that the inspectee will shirk less often because of the enhanced own-payoff effect of working. Thus, in such a setup bonuses may have a positive effect on inspectees' good behavior. Also, in this study we examine the performance of exogenously imposed mechanisms. In our experiment, workers chose whether to work or shirk and employers chose whether to inspect or not inspect. Fines and bonuses were then triggered automatically in response to the actions chosen by the players. Another interesting avenue for further research would be to explore the endogenous choice of punishing and rewarding mechanisms.