CAN HIGHER REWARDS LEAD TO LESS EFFORT? INCENTIVE REVERSAL IN TEAMS

Esteban Klor, Sebastian Kube,
Eyal Winter and and Ro'i Zultan

Discussion Paper No. 13-09

November 2013

Monaster Center for
Economic Research
Ben-Gurion University of the Negev
P.O. Box 653
Beer Sheva, Israel

Fax: 972-8-6472941
Tel: 972-8-6472286

Can Higher Rewards Lead to Less Effort? Incentive Reversal in Teams[*]

Esteban F. Klor, Sebastian Kube, Eyal Winter and Ro'i Zultan[†]

Abstract

Conventional wisdom suggests that a global increase in monetary rewards should induce agents to exert higher effort. In this paper we demonstrate that this may not hold in team settings. In the context of sequential team production with positive externalities between agents, incentive reversal might occur: an increase in monetary rewards (either because bonuses increase or effort costs decrease) may lead agents to exert lower effort in the completion of a joint task — even if agents are fully rational, self-centered money maximizers. Herein we discuss this seemingly paradoxical phenomenon and report on two experiments that provide supportive evidence.

Keywords: Incentives, Incentive Reversal, Team Production, Externalities, Laboratory Experiments, Personnel Economics.

JEL: C92, D23, J31, J33, J41, M12, M52

[*] Paper forthcoming in *Journal of Economic Behavior & Organization*; doi: 10.1016/j.jebo.2013.10.010.

[†] Corresponding author: e-mail: zultan@bgu.ac.il; postal address: Ben-Gurion University of the Negev, Department of Economics, P.O.B. 653, Beer-Sheva 84105; telephone: +972(0)86472306; fax: +972(0)86472941.

1 Introduction

Most economists would presumably agree to the statement that, basically, economics is all about incentives.[1] The statement is regularly understood to be about monetary payments, in the sense that high monetary rewards equal strong incentives, and vice versa. This simplification applies to many economic situations. However, it does not necessarily apply to environments in which individuals interact in groups and their individual rewards are affected by others' actions, e.g., in team production settings. In the context of sequential team production, incentive reversal might occur – in particular for rational individuals whose main objective is the maximization of their own monetary income. In this paper, we illustrate under which circumstances this might happen and report corresponding experimental results.

Following Winter (2009), who introduced the theoretical foundations for incentive reversal, we consider simple strategic environments involving team production with moral hazard. In this context, incentive reversal refers to situations in which an increase of promised rewards to all team members results in fewer agents exerting effort. Incentive reversal is caused by the existence of externalities among peers that arise from the team's production technology, and builds on two properties that are descriptive of many team environments: i) Some agents have internal information about the effort level of others (which requires a certain extent of sequencing in the production process), and ii) agents' efforts are complements in the team's production technology. Given these assumptions, the line of reasoning behind incentive reversal is straightforward. Since the underlying production technology involves complementarity in terms of team members' efforts, moderate rewards can generate an implicit threat against shirking, in the sense that agent i chooses to exert effort only if his peer, agent j, (whose effort is observable by i) has done so as well. A substantial increase to agent i's rewards may induce this agent to exert effort as a dominant strategy (regardless of what agent j is doing). This in turn eliminates the implicit threat that was present in the outset and induces agent j to shirk even though his promised reward increased as well.

[1] A statement which, for example, has been made by Aumann (2006) in his Nobel prize lecture in 2005. Aumann recounted the following story about Jim Tobin: "The discussion was freewheeling, and one question that came up was: Can one sum up economics in one word? Tobin's answer was 'yes'; the word is incentives" (p. 351).

Simple as it may seem, it is not clear whether the argument for incentive reversal is empirically sound for at least two reasons: cognitive limitations and other-regarding preferences. We tend to think about monetary rewards and motivation as moving in the same direction. Thus, when the rewards of all agents in a team are increased, they may respond "heuristically" with high effort to the increase in their own reward, without considering the strategic implications of the increase in their peers' rewards. Such heuristic responses might be facilitated if individuals are not able, or expect others not to be able, to follow the backward induction reasoning underlying incentive reversal.[2]

Even if cognitive limitations do not apply, other-regarding preferences (and in particular the presence of reciprocity) may eliminate incentive reversal. If an individual who detects the shirking of his peer is inclined to retaliate by shirking as well, regardless of the monetary incentives, the observed individual (anticipating reciprocal behavior) would be reluctant to shirk. In this event, incentive reversal will be washed out.[3]

Given the above considerations, whether incentive reversal in teams actually occurs or not ultimately remains an empirical question. Moreover, theoretical predictions strongly rely on having sufficiently precise knowledge about the shape of the production technology, the move structure and information set of each agent, as well as the potential rewards and individuals' costs of exerting effort. We conducted two separate experiments that allowed a sufficient degree of control over these factors to clearly test for incentive reversals. Both experiments involve teams of agents who work on a joint team project. Agents decide on their individual effort level (with effort being costly) and are paid as a function of the

[2] The experimental literature casts doubts on the ability and propensity of people to follow backward induction (e.g., Binmore et al., 2002; Carpenter, 2003; Harrison and McCabe, 1996). Johnson et al. (2002) found that information gathering strategies of subjects in three-stage bargaining games reflect forward looking reasoning rather than backward induction. Game strategies did not converge to the backward induction prediction even when subjects knew that they were playing with computerized partners who follow the backward induction path. Bone et al. (2009) provide evidence that people do not use backward induction even in non-strategic risky situations.

[3] The literature on social dilemmas provides ample evidence that people choose reciprocal strategies even when those entail playing strictly dominated strategies, both within a round with sequential moves and between periods when the game is repeated (e.g. Clark and Sefton, 2001; Falk and Fischbacher, 2002; Fischbacher and Gächter, 2010; Fischbacher et al., 2001; Gächter et al., 2010; Guttman, 1986; Meidinger and Villeval, 2002; Potters et al., 2007; Varian, 1994). See, however, Glöckner et al. (2011) for an experimental study where an increase in monetary rewards weakens reciprocal reactions and reduces voluntary cooperation in a social dilemma.

team's joint effort. In both experiments we look at situations that are susceptible to incentive reversal. Keeping the environment (in particular the production technology) fixed, we explore how subjects behave under high, respectively under low rewards.

The two experiments differ in several respects, allowing us to establish the behavioral validity of the incentive reversal phenomenon across specific features of the decision environment. The first experiment implemented a two-agent game in a laboratory setting using a labor framing. Subjects in this experiment were provided with the explicit payoff structure of the game and second movers made their decisions after observing the decisions of first movers. The level of incentives was manipulated within subjects between rounds by setting different reward levels, as in Winter (2009). The second experiment implemented a three-agent game in a classroom environment. The game was presented as a money game, the payoff structure of which was not explicitly provided, but could be extrapolated from the basic rules. All decisions were collected simultaneously, with second- and third-mover strategies conditional on previous movers' decisions. The incentive level manipulation was implemented by varying the costs of effort, rather than the rewards, in a between-subjects design. Put together, the two experiments provide a robust test for the existence of incentive reversal.

In order to be able to ascertain that any observed incentive reversal effect is indeed driven by the hypothesized mechanism, we take two complementary approaches. In the first experiment, we add two control treatments that correspond to the experimental treatments in all but one aspect: the subjects choose their actions simultaneously rather than sequentially. Thus, while we retain the payoff structure, the strategic structure which gives rise to incentive reversal in the sequential games is eliminated in the simultaneous games. In the second experiment, we use a strategy method instead of a direct-response method to obtain counterfactual data. By observing subjects' decisions in each node of the game tree we can test for incentive reversal by looking at behavior along and off the theoretical equilibrium path.

Our experimental data provide clear evidence that incentive reversal in teams can occur. When the comparative-static predictions of the theoretical analysis predict incentive reversal, we do observe it. In the first experiment, increasing the second-mover's rewards has the negative effect of reducing the first-mover's incentive to exert effort as this agent chooses to free-ride on the second-mover's effort – but only in sequential games and not in simultaneous games. These behavioral patterns are indeed observed: the average effort provided by the first-movers drops by almost 50 percent when rewards are increased under the sequential protocol, whereas the average effort stays constant in the simultaneous protocol. Incentive reversal is observed in our second experiment as well. The average team output is significantly higher under high costs than under low costs. For example, first-movers' average effort is increased by almost 130 percent when costs are increased. Moreover, subjects' subsequent choices along the equilibrium path are well in line with the predictions from incentive reversal. Interestingly, this holds true even though we observe reciprocal behavior in both treatments, which underlines the relative importance of incentive reversal in such an environment.

2 Theoretical Framework

The theoretical framework we consider is based on Winter (2009). Winter (2009) analysed the possibility of incentive reversal in a general theoretical framework. He showed that when the production technology has positive externalities among peers and agents choose sequentially the amount of effort that they exert on their individual tasks, the set of agents who exert effort in (subgame-perfect) equilibrium may decrease if the principal increases the agents' rewards. This effect is purely driven by monetary incentives, and is not caused by other behavioral considerations or income effects. Winter's framework uses a stochastic technology function whereby the probability of success of a given project increases in the total amount of agents' effort. Here we provide an illustration of the main intuition behind incentive reversal with a deterministic technology that is also employed in our experimental design.[4]

As an example, let us analyze a team of two agents working on a joint project. The agents choose whether to exert effort or shirk, with effort being costly. We denote this decision by e_i, with $e_i = 1$ when agent i exerts effort and $e_i = 0$ when he shirks. Agents move sequentially and information is perfect. Agent i's

[4] Our experimental design replaces the probabilistic setup with a deterministic one to abstract from the possibility that agents' risk attitudes may affect their choices. A similar approach is used, for example, in Goerg et al. (2010) and Steiger and Zultan (2011).

payoff function is given by

$$U_i(e_i, e_j) = r_i P(e_i + e_j) - e_i C_i, \tag{1}$$

where r_i is the reward that agent i receives per unit produced, P denotes the amount of units produced as a function of total effort exerted, and C_i is agent i's positive cost of exerting effort. We assume that the production function P is strictly convex on the sum of efforts, i.e., the effort of one agent increases the marginal productivity of the other agent. For the two-agent case being examined this implies

$$P(2) - P(1) > P(1) - P(0); \tag{2}$$

that is, the technology has complementarities across agents' efforts. Thus, an agent's effort creates positive externalities on the other agent's productivity.

For the purposes of this example, let us consider one of the sets of parameters that we use in our first experiment. Suppose that the production function is given by $P(2) = 100$, $P(1) = 70$ and $P(0) = 50$, and that effort costs are $C_1 = C_2 = 1000$. The rewards per unit produced are $r_1 = 28$ for the first mover and $r_2 = 43$ for the second mover. For these parameters, there exists a unique Subgame-Perfect Equilibrium where on the equilibrium path both agents choose to exert effort. Thus, total effort exerted equals 2.

Suppose now that the principal increases both agents' rewards such that $r_1 = 31$ and $r_2 = 60$, with the remaining parameters (costs- and production-function) unchanged. Under these new (higher) rewards, exerting effort becomes a dominant strategy for agent 2. Agent 1 realizes this and chooses to shirk in equilibrium. Therefore, the increase in rewards for the two agents causes a decrease in total effort (see also the equilibrium prediction in Table 1).

Intuitively, under the scheme with low rewards, agent 1 has to exert effort to motivate agent 2 to exert effort as well (because agent 1's effort increases agent 2's marginal productivity). With high rewards, agent 2 is willing to exert effort regardless of agent 1's strategy. This allows agent 1 to free-ride on agent 2's effort while saving his own effort cost. Consequently, shirking becomes agent 1's equilibrium strategy under the new incentive scheme. In addition to the particular properties of the production technology,[5] information about the effort

[5] See Winter (2010) for an analysis of efficient rewards schemes for different production technologies and information structures.

exerted by peers plays a crucial role for incentive reversal to occur (and we are going to use this as a treatment variation in our first experiment to clearly identify the effect of incentive reversal in our data). When agent 2 is uninformed of the strategic choice of agent 1, the sequential game described above turns into a simultaneous game from an informational point of view. When rewards are low, both agents shirk in the unique Nash equilibrium of the game. By contrast, when rewards are high, agent 1 shirks whereas agent 2 exerts effort, the same equilibrium strategies of the sequential game. Therefore, while an increase in rewards causes a decrease of total effort in the sequential game, it causes an increase of total effort in the simultaneous game.

3 Experiment 1

This section presents the results of the first set of tests in a controlled lab environment that potentially allows for incentive reversal to occur.

3.1 Experimental Design and Procedure

The underlying game of the first experiment was essentially the one described in the previous section. Each of two agents in a team makes an effort decision e_i. The payoff is determined according to Equation 1, based on the individual reward factor r_i, the individual cost of effort C_i, and the production function P. Parameters used in the experiment included three different combinations of cost and production function and two reward levels for each such *parameter set*, such that the individual reward in the *high rewards* is higher than that in the *low rewards* for each of the agents.

We chose the parameters such that the two reward levels lead to incentive reversal under a sequential protocol, in which the second agent is informed of the decision of the first agent before deciding whether to exert effort. That is, in the Subgame-Perfect Equilibrium of the game both agents exert effort under low rewards, but only the second mover exerts effort under high rewards. In the control simultaneous-move treatment, all parameters are kept as in the sequential treatment, with the difference that the two agents decide simultaneously. In total, our 2 (protocols) x 3 (parameter sets) x 2 (reward levels) design induces 12 different games. Table 1 summarizes the experimental design and the equilibrium predictions.

Table 1: Parameters for Experiment 1

	Set of parameters		
	I	II	III
	Units produced $P(e_1 + e_2)$		
Total Effort = 0	30	70	50
Total Effort = 1	60	80	70
Total Effort = 2	100	100	100
	Cost of effort C_i		
Agent 1	2500	1000	1000
Agent 2	1100	400	1000
	Rewards per unit produced r_i		
- Low rewards treatment			
Agent 1	48	35	28
Agent 2	31	35	43
- High rewards treatment			
Agent 1	49	40	31
Agent 2	51	45	60
	Equilibrium predictions (e_1, e_2)		
Sequential & Low rewards	(1,1)	(1,1)	(1,1)
Sequential & High rewards	(0,1)	(0,1)	(0,1)
Simultaneous & Low rewards	(0,0)	(0,0)	(0,0)
Simultaneous & High rewards	(0,1)	(0,1)	(0,1)

Note: Over the six rounds, subjects played the three different parameter sets (I, II and III) and two different reward schemes (Low, High) in the following order: I-Low, II-High, III-Low, I-High, II-Low, III-High.

Different groups of subjects played under the sequential and under the simultaneous protocol. The parameter set and reward level were varied within subjects, so that, for each protocol, subjects played under each of the six games defined by the three parameter sets and two reward levels in six consecutive rounds. The procedure was similar in all treatments. At the beginning of each session, subjects were randomly assigned to a role as either agent 1 (first mover) or agent 2 (second mover). Roles remained fixed throughout the entire session.

Each session consisted of six independent rounds. Subjects were re-matched in a stranger design, i.e., in each round with a randomly selected subject in the opposing role, to address potential drawbacks of reputation building. Additionally, no feedback was provided between rounds. This also kept the decisions of the first movers in the session strictly independent.[6] At the beginning of each round, all subjects observed the relevant parameters for that particular round. The sequential protocol presented the parameters in the form of a game tree whereas the simultaneous protocol presented the parameters using a matrix.[7] Although there might be order effects in the data, this should not affect our results as the order is the same in all treatments and the data is averaged over different rounds for each subject.

The use of different parameter sets allows us to generalize our results beyond a particular specification. Moreover, the design allows us to examine the behavior of the same subject as the reward scheme changes between low and high rewards, keeping the individuals' characteristics fixed and abstracting from the specific parameters used in different rounds.

The computerized sessions were conducted at the RatioLab - The Center for Rationality and Interactive Decision Theory at The Hebrew University of Jerusalem. We recruited 60 students from various academic backgrounds out of the RatioLab subject pool, which consisted of approximately 3,000 subjects at the time. Thirty-six subjects participated in 3 sessions in the sequential treatments, and 24 subjects participated in 2 sessions in the simultaneous treatments. In each session, twelve subjects were admitted into the lab and received written instructions, which were then read out aloud by the experimenter.[8] Throughout the experiment we ensured anonymity and effectively isolated each subject in a cubicle to minimize any interpersonal influence. Communication among subjects was not allowed throughout the session. Each subject received a base payment of 300 experimental points at the beginning of each round (80 experimental points equal NIS 1). Subjects' subsequent earnings were determined by their payoffs of a randomly selected round. Average earnings were equal to NIS

[6] There is, of course, dependence between the decisions of the second movers in the sequential protocol sessions, as they receive information within the round by observing the corresponding first-mover's choice.

[7] The corresponding game trees and matrices are available in the online appendix.

[8] The instructions included an example with a parameter set different from the ones used in the actual experiment. An English translation of the instructions is available in the online appendix. The original instructions in Hebrew are available from the authors upon request.

63.[9] Each session lasted about 45 minutes.

3.2 *Results*

To test for incentive reversal, we first compute for each subject the number of times he chooses to exert effort under high, respectively under low rewards. Figure 1 depicts the average propensity of the subjects to exert effort (with 95 percent confidence intervals based on individuals as independent observations), separately for each treatment.

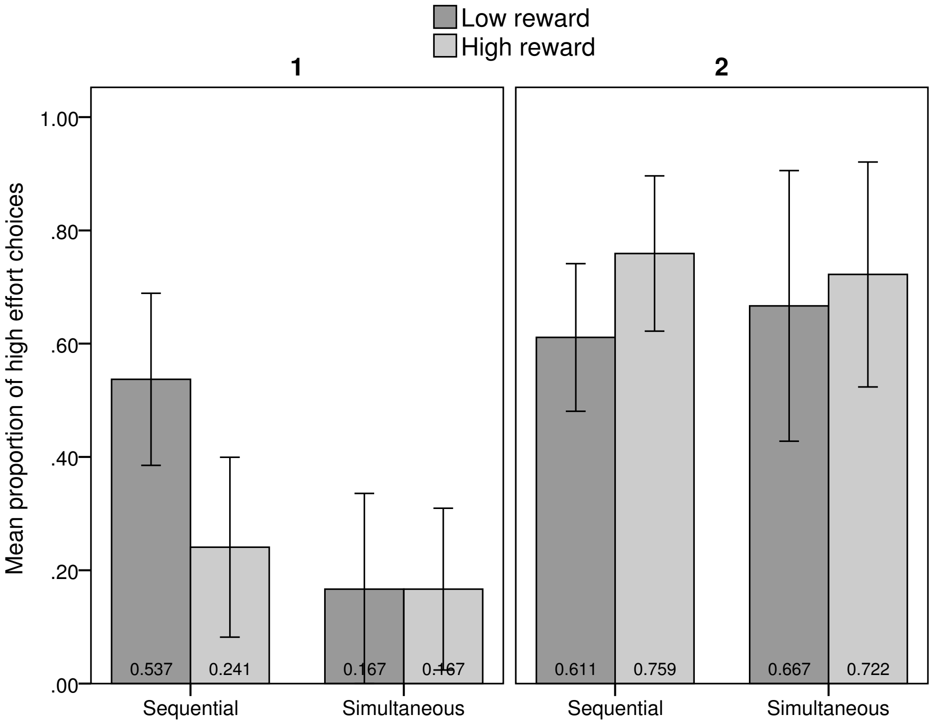


Figure 1: Experiment 1 — Effort decisions

Note: bars represent 95 percent confidence intervals based on 18 and 12 subjects each in the role of agent 1 (left panel) and agent 2 (right panel) in the sequential and simultaneous treatments, respectively.

Since the incentive reversal predicted by theory hinges on the decisions of agent 1, let us first focus on the behavior of subjects in the role of agent 1 (de-

[9] This is more than three times the hourly minimum wage in Israel, which was slightly below NIS 20 at the time we ran the experiment. Therefore, the amounts involved in the experiment are significant amounts considering the time the subjects devoted to the experiment. The current exchange rate is slightly above NIS 3.5 per U.S. dollar.

Table 2: Logit regression on first mover choice

	Coefficient	Robust standard error	p-value
Constant	-1.15	0.41	0.005
Low reward	1.30	0.41	0.002
Simultaneous protocol	-0.46	0.61	0.450
Low reward x Simultaneous protocol	-1.30	0.50	0.010

Note: Robust standard errors clustered on subjects. N=180, prob $> \chi^2$ = 0.0006

picted in the left panel of Figure 1). The results show that rewards do not affect the effort exerted by these subjects in the simultaneous protocol. The subjects' mean effort level (0.167) is identical under both reward schemes. By contrast, the reward structure does affect subjects' behavior in the sequential protocol. Here, we observe that first-movers are significantly more likely to exert effort in rounds with low rewards compared to rounds with high rewards (mean of 53.7 percent versus 24.1 percent across the different parameter sets; Wilcoxon Matched-Pairs Signed Ranks test, $Z = 2.713$, $p < 0.01$, two-sided, based on individual averages). This also holds true if we look at each parameter set separately. The effect is not only qualitatively significant. It is also quantitatively important as agent 1's effort more than doubles when rewards are low. Table 2 uses a regression analysis to support these conclusions. In particular, Table 2 shows that increasing rewards causes a significant decrease in first mover effort, only in the sequential treatment. These conclusions are summarized in Result 1.

Result 1. *In accordance with incentive reversal, i) the increase in rewards leads to a reduction in effort of agent 1 under the sequential protocol; and ii) the increase in rewards does not lead to a reduction in effort of agent 1 under the simultaneous protocol.*

Let us now turn to the behavior of subjects in the role of agent 2. A large majority of these subjects exerts effort. The mean effort level ranges from 0.611 (in the sequential protocol with low rewards) to 0.759 (in the sequential protocol with high rewards). Effort levels are higher in the high reward rounds, though the difference is only statistically significant in the sequential-protocol treatments

Table 3: Marginal effects of increasing rewards on second mover and team outcomes

Treatment	(1) Second mover effort[a]	(2) Team production[b]	(3) Team salary[b]	(4) Team payoff[b]
Sequential	14.8%** (7.4%)	-6.67*** (2.52)	824*** (233)	1225*** (88)
Simultaneous	5.6% (4.4%)	2.22* (1.33)	1436*** (64)	1480*** (7)
N	180	180	180	180

Notes: robust standard errors clustered on matching groups (sessions) in parentheses.
[a] Logit regression. [b] OLS regression. *,**,*** Significance at the 0.1, 0.05, and 0.01, respectively.

(see Column (1) in Table 3).[10]

We conjecture that this difference is caused by the fact that exerting effort is a dominant strategy for agent 2 when rewards are high, but it is only a best response to agent 1's exerting effort when rewards are low. This leads agent 2 to occasionally shirk under low rewards, namely if he observes agent 1 shirking. Indeed, we find that agent 2's behavior is contingent on agent 1's behavior in the low reward rounds. The proportion of decisions to exert effort is .86 if agent 1 exerted effort compared to .32 if agent 1 shirked. A similar tendency is apparent in the high reward rounds, however this is much weaker with proportions of decisions to exert effort of .92 and .71 following effort exertion and shirking by agent 1, respectively.

Although most subjects in the role of agent 2 choose to exert effort, we also observe some instances where they shirk. In particular in the sequential protocol, this might be interpreted as an indication of reciprocity. Under low rewards, second-movers choose to shirk when they observe the first-mover shirking in 68 percent of the cases (17/25). Yet, this behavior cannot be clearly attributed to reciprocal preferences, because strategies of payoff-maximization and reciprocity coincide here. The second-movers' best response is to shirk whenever

[10] Note that high effort levels in the simultaneous/low rewards treatment are not part of the equilibrium prediction that both agents shirk. Possibly, participants were willing to risk a small loss in trying to cooperate on reaching the efficient all-work outcome, which Pareto dominates the all-shirk equilibrium outcome.

Table 4: Marginal effect of first mover's effort on second mover's effort

Treatment	Marginal effect	Robust standard error	95% CI	
			Low	High
Low rewards	54.2%	8.1%	38.4%	70.0%
High rewards	21.6%	6.5%	8.9%	34.3%

Note: Logit regression with standard errors clustered on matching groups (sessions). N=108.

the first mover shirks under low rewards. However, shirking in response to shirking is also observed under high rewards, even though it is not agent 2's payoff-maximizing response in this treatment. In these instances, which account for approximately 29 percent (12/41) of the cases in which the second mover shirks, second-movers behavior does reveal clear signs of reciprocity.

To test the effect of first mover decisions on second mover decisions, we ran a logit regression with second mover choices as the dependent variable and first mover choice and reward level as dependent variable. To control for the dependencies within matching groups we applied robust standard errors clustered on sessions (that is, matching groups). The results are presented in Table 4 and summarized in the following result.

Result 2. *When rewards are low, second movers reciprocate the actions of first movers. This tendency is less pronounced when rewards are high and exerting effort is dominant for second movers. In these cases, some second movers reciprocate shirking with shirking, but most second movers exert effort.*

Given that second movers do not reciprocate perfectly with low rewards and reciprocate to some degree with high rewards, it is interesting to verify whether incentive reversal, i.e., reducing effort in response to an increase in rewards, is indeed optimal for first movers. It turns out that this is not the case. If we calculate first-movers' expected payoffs based on second-movers' actual effort choices (instead of the theoretically predicted), we see that in all three parameter sets, the payoff-maximizing action for first movers would have been the same in both reward levels – which implies that incentive reversal is not optimal given second-movers' actual behavior.

We proceed now to analyze the effects of increasing the rewards on the team level. The observed incentive reversal has interesting implications on total production, especially if we keep in mind that the production function is convex. Table 5 depicts the distribution of total team effort, the average amount of units produced by the teams and the teams' average payoffs.

Table 5: Experiment 1 – Distribution of Effort and Total Units Produced.

	Treatment			
	Sequential		Simultaneous	
	Low rewards	High rewards	Low rewards	High rewards
Amount of team's total effort				
0	17	12	10	10
1	12	30	22	20
2	25	12	4	6
Total	54	54	36	36
Average number of team's units produced	79.3	72.6	67.5	69.7
Average team's salary paid by principal	6,400	7,224	5,517	6,953
Average team's payoff	5,037	6,263	4,689	6,170

Notes: Average team's payoffs include the costs the subjects incurred while choosing to exert effort. The average team's salary paid by the principal only takes into account the number of units produced and the rewards promised for each unit produced, in addition to the base payment of 300 points.

Let us first focus on the sequential treatments. The table shows that, when rewards are low, subjects are more likely to coordinate on an extreme level of effort, whereby total team effort equals 2 or 0. In the low rewards treatment, teams exert the maximum level of effort over 45 percent of the time (25/54 instances). On the contrary, in the treatment with high rewards incentive reversal occurs. We observe that a total team effort of one is the most frequent outcome. A multinomial logistic regression with robust standard errors clustered on sessions with protocol and reward level as independent variables confirms that, in the sequential treatments, increasing the rewards leads to a higher probability of exactly one agent in a team exerting effort ($z = 5.69, p < 0.001$) and a lower probability of both agents exerting effort ($z = -3.27, p < 0.001$). The effect of rewards on the probability of both agents shirking is not significant ($z = -1.26, p = 0.209$), as are the differences between the simultaneous treatments ($p > 0.500$ for all three possible outcomes).

The difference in the level of team effort induced by the reward scheme is amplified by the convex production technology necessary for incentive reversal to occur. As a consequence of these two effects, the mean number of units produced by a team when rewards are low is 79.3 compared to a mean production of 72.6 units when rewards are high. This difference is significant ($p = 0.008$, see Column (2) in Table 3). This important difference in units produced is not reflected in the costs of production faced by the principal. A team's average pay equals 5037 (NIS 75.6) when rewards are low and 6263 (NIS 88.9) when rewards are high ($p < 0.001$, see Column (3) in Table 3). That is, when rewards are high, even though the principal pays more money overall, she receives a lower amount of units produced. Agents, on the contrary, are better off in the high rewards treatment – in addition to receiving higher rewards they also save the costs of exerting effort ($p < 0.001$, see Column (4) in Table 3).

Result 3. *As a consequence of the observed incentive reversal under the sequential protocol, average production output and principals' payoff are reduced.*

The right panel of Table 5 presents summary statistics for the simultaneous treatment. The results in the table show that the effect of increasing rewards on effort and production is marginal. If anything, it seems that higher rewards induce higher effort, but, as noted above, the difference in the distribution of total effort is far from being significant and the difference in production is only significant at the 10-percent level (Column (2) in Table 3).

Summarizing, the results of Experiment 1 provide clear evidence in support of incentive reversal. Increasing agent 2's rewards has the negative effect of reducing agent 1's incentive to exert effort as this agent chooses to free-ride on agent 2's effort. This behavior is prominent in sequential games but not in simultaneous games – which suggests that the incentive reversal effect can be attributed to the process described in Winter (2010). In particular it rules out considerations of inequality aversion as a potential explanation, because for a given parameter set the payoff consequences are the same between the simultaneous and the sequential protocol.

4 Experiment 2

Experiment 1 provided clear evidence in support for incentive reversal. To complement and check for the robustness of these findings, we ran an additional experiment which again featured a sequential team production problem. In addition to exploring additional design specification as a robustness check, Experiment 2 made the backward induction process more strenuous compared to Experiment 1, as it involved teams of three agents and subjects were not given a graphical representation of the game, but had to extrapolate it from the instructions (if they desired to do so).

The experiment was conducted in a classroom environment. It was known that all subjects were from the same class and were likely to know each other, however the identities of the specific team members were kept unknown. The decision was one-shot, using the strategy method over roles and histories to provide counterfactual data, allowing us to carry out a direct and clean within-subject analysis of reciprocal attitudes.

Treatment manipulations were between subject groups, i.e., between classrooms. Between treatments, the reward schemes and the production function were kept constant but effort costs were changed (in contrast to Experiment 1, where effort costs were fixed but rewards were manipulated). Incentive reversal is thus manifested in higher efforts when the costs change from low to high. The game was framed as a simple monetary game, for which the rules were provided in the instructions.

4.1 Experimental Design and Procedure

The experimental game involves teams of $n = 3$ interacting agents. Each team receives an initial team endowment E of NIS 30 (approximately \$8). Agents move sequentially. Conditional on the decision(s) of the predecessors, each agent i individually decides whether to exert effort (e_i=1) or shirk (e_i=0). Shirking is costless, while exerting effort entails an individual fixed cost c_i, which differs across agents and treatments. The team's endowment is doubled for each agent who chooses to exert effort. Note that this is a convex technology, which implies

Table 6: Experiment 2 – Treatments and Equilibrium Predictions

Costs of exerting effort for...	Low costs	High costs
...agent 1 (c_1)	55	60
...agent 2 (c_2)	50	55
...agent 3 (c_3)	5	25
Equilibrium strategies (e_1, e_2, e_3)	(0,0,1)	(1,1,1)
Equilibrium Payoffs (π_1, π_2, π_3)	(20,20,15)	(20,25,55)

that it has complementarity on agents' efforts. The resulting final endowment is equally divided between all the team members at the end of the experiment. Hence, an agent's final payoff is given by:[11]

$$\pi_i = \frac{E}{n} \cdot 2^K - c_i e_i \tag{3}$$

with $K = \sum_{k=1}^{n} e_k$ denoting the sum of effort provided by all team members.

Depending on the cost structure (low or high), the production technology may lead to incentive reversal. This factor is varied between treatments. The costs schemes we used were c^L = (55, 50, 5) and c^H = (60, 55, 25). Since agents move sequentially, when effort costs are high (c^H) each agent should exert effort (i.e., double the team's endowment) if, and only if, she observes all previous movers exerting effort. In the unique SPE of the game all agents choose to exert effort in the high-costs treatment. Conversely, when effort costs are low (c^L), it is a dominant strategy for the last mover to exert effort. Solving the game using backward induction, the first two movers then choose e_i=0 along the equilibrium path. Thus, incentive reversal occurs: a reduction in costs (which implies that agents' potential payoffs are increased) leads to a reduction in overall efforts. Table 6 summarizes the treatment parameters and the treatments' equilibrium predictions.

The subjects that participated in this experiment were undergraduate students at the Hebrew University of Jerusalem. All subjects participated on the same day, with each group playing only a single treatment. None of the subjects had

[11] Negative payoffs were ignored, so that if for an agent who chose to exert effort the costs were higher than his final share of the endowment, we set his final payoffs equal to zero. Subjects knew this feature of the game in advance. Importantly, the restriction that final payoffs are non-negative does not alter the equilibrium of the game.

participated in our first experiment. The experimenter entered the classroom at the end of the exercise lesson, and offered the students to participate in a short experiment, to which most of the students responded positively (78 out of approximately 90). Once only those students who volunteered to participate in the experiment remained in the classroom, the instructions were handed and read out aloud. Instructions were framed neutrally, avoiding loaded terms (e.g., we spoke of "doubling the team's endowment" rather than of exerting effort or shirking). Subjects then had to answer control questions in order to ensure understanding of the instructions.[12] Afterwards, subjects marked their choices on the designated form. We used the strategy method (Selten, 1967), so that each subject decided for each information set (history) of each role (first, second, and third mover), making seven decisions in total. Once all forms were collected, the payoffs were calculated in the following way: The subjects in each treatment were randomly assigned to teams of three subjects, and randomly assigned roles within each team. The decisions corresponding to the assigned role and previous movers' decisions determined the team members' payoffs. The subjects did not receive any feedback regarding the identity or decisions of their team members. Payments were made in private and subjects were identified by the last four digits of their ID number, which they wrote on the decision sheet. The average payoff was NIS 24 (approximately $6.5).

4.2 Results

Table 7 presents all the subjects' decisions contingent on the previous choices of the other subjects, as obtained from the strategy method.

Let us first focus on the behavior along the equilibrium paths. According to the theoretical prediction of incentive reversal, fewer first movers should exert effort under low costs than under high costs. In support of this prediction, we observe that the proportion of subjects who exert effort as first movers when costs are high is significantly higher than the proportion of subjects who do so when costs are low (54.1 percent versus 23.7 percent; $\chi^2 = 7.291$, $p = 0.07$, two-

[12] An English translation of the instructions is available in the online appendix. The original instructions in Hebrew are available from the authors upon request. Out of the 78 subjects, 3 students failed to answer correctly the control questionnaire. We removed from the analysis below these students' answers, although their inclusion would not qualitatively change any of the results.

Table 7: Experiment 2 – Description of Subjects' Chosen Strategies

	Low costs							
Number of subjects:	38							
Percent of Agents 1	$e_1 = 1$ 23.7				**$e_1 = 0$** **76.3**			
Percent of Agents 2	$e_2 = 1$ 73.7		**$e_2 = 0$** **26.3**		$e_2 = 1$ 10.5		**$e_2 = 0$** **89.5**	
Percent of Agents 3	**$e_3 = 1$** **100.0**	$e_3 = 0$ 0.0	**$e_3 = 1$** **97.4**	$e_3 = 0$ 2.6	**$e_3 = 1$** **97.4**	$e_3 = 0$ 2.6	**$e_3 = 1$** **89.5**	$e_3 = 0$ 10.5
	High costs							
Number of subjects:	37							
Percent of Agents 1	**$e_1 = 1$** **54.1**				$e_1 = 0$ 45.9			
Percent of Agents 2	**$e_2 = 1$** **81.1**		$e_2 = 0$ 18.9		$e_2 = 1$ 13.5		**$e_2 = 0$** **86.5**	
Percent of Agents 3	**$e_3 = 1$** **94.6**	$e_3 = 0$ 5.4	$e_3 = 1$ 24.3	**$e_3 = 0$** **75.7**	$e_3 = 1$ 27.0	**$e_3 = 0$** **73.0**	$e_3 = 1$ 2.7	**$e_3 = 0$** **97.3**

sided). Given that the first mover shirks under low costs, also the second mover should do so, which is true for 89.5 percent of all corresponding decisions that we observe. Analogously, under high costs the second mover should provide effort along the equilibrium path if he observes the first mover exerting effort as well. We observe this behavior in 81.1 percent of all corresponding cases. Finally, also the choices of the third-movers along the equilibrium path are well in line with the predictions from incentive reversal: 89.5 percent (94.6 percent) exert effort under low (high) costs.

The increased efficiency when costs are higher is also evident when we consider the resulting productivity. Since data were collected using the strategy method, we do not look at the actual realization (which would be arbitrary) but rather at the expected realizations, i.e., the decisions weighted by the corresponding observed distribution of previous movers' decisions.[13] Table 8 reports the expected number of subjects choosing to exert effort, as well as the expected costs and productivity for each treatment.

Similarly to the results in the sequential protocol of Experiment 1, we observe that with high costs subjects are more likely to coordinate on an extreme strategy whereby the number of subjects exerting effort is either 0 or 3. In particular, in this treatment the most frequent outcome is for all of the team's subjects to exert effort (chosen over 41 percent of the time). On the contrary, low costs lead to incentive reversal, because most of the times only one agent exerts effort while the other two shirk (61.5 percent of the time).

The convex technology of production amplifies the difference in teams' total effort levels between high and low costs treatments when we look at the expected teams' costs and productivity. Team productivity is considerably higher for the high costs treatment (NIS 127) compared to the low costs treatment (NIS 97.6). That is, a decrease in the associated costs of production causes a substantial decrease in units produced. As a result, the principal receives less output.

Result 4. *In accordance with incentive reversal, reducing the effort costs leads to lower efforts and consequently to a lower production output.*

[13] For example, in Table 7 we see that under low costs, 23.7 percent of agent 1 decide to exert effort, and 73.7 percent of agent 2 state that they want to exert effort if agent 1 does, and 100 percent of agent 3 would want to exert effort if both the previous agents exerted effort. Therefore, the expected frequency for the case that all three agents in a team exert effort is given by $0.237 \cdot 0.737 \cdot 1 \approx 0.175$; as it is displayed in the corresponding cell in Table 8 (first column, fourth row). All the other values in Table 8 are derived analogously.

Table 8: Experiment 2 – Expected Distribution of the Sum of Team Effort, Costs, and Payoffs

Percent of Teams with a Team Effort of...	Low costs	High costs
... $K = 0$	.072	.386
... $K = 1$	.615	.133
... $K = 2$	.139	.065
... $K = 3$	.175	.415
Expected Team Effort	1.42	1.51
Expected Team Cost (NIS)	30.4	83.1
Expected Team Production (NIS)	97.6	127.0
Expected Team Payoff (NIS)	67.2	43.9

Notes: The ex-post probabilities reported in the table reflect the proportions of effort choices weighted by the corresponding observed distribution of previous movers' decisions. For example, $K = 0$ is only possible if all agents in the team shirk (i.e., along the branches on the far right of the game trees in Table 7). Under low costs, the percentage of agents 1 (2, 3) shirking at these nodes is 0.76 (0.895, 0.105), which leads to the ex-post probability of $.76 \cdot .895 \cdot .105 = .072$ reported here. Expected team effort is then given by summing up the products of potential team efforts and corresponding ex-post probabilities. Expected team payoff is given by the difference between expected team production and team costs, which are both calculated analogously to the expected team effort.

A game with three agents provides more situations in which reciprocity does not coincide with monetary rewards.[14] Furthermore, using the strategy method enables us to easily study those situations and identify reciprocal strategies more clearly. In fact, at those decision nodes where reciprocal and money-maximizing actions diverge, we observe that some subjects show a tendency to reciprocate the decisions of the previous mover(s). For example, under high costs the last movers frequently exert effort when they see that at least one of the previous movers exerted effort as well (24.3 percent when the first mover exerted effort but the second one did not, and 27 percent when the second mover exerted effort but the first one did not). Another example would be that under low costs, 10.5 percent of the third movers shirk if both previous movers shirked as well. The case where the reciprocal effect is most pronounced is under low costs when

[14] For example, the two-agent case does not allow to disentangle positive reciprocity from money-maximizing behavior. In Experiment 1, agent 2 always maximizes his monetary payoff when he exerts effort after observing agent 1 exerting effort. This no longer holds when there is a third agent.

the first mover exerted effort. In that case, 73.7 percent of the second movers choose to exert effort rather than to maximize their monetary payoff by shirking. Interestingly, however, the same subjects who would reciprocate as second or third movers do not anticipate this behavior from their partners when deciding as first movers; hence the overall low cooperation and productivity when costs are low, and the perseverance of the incentive reversal effect.

Result 5. *Incentive reversal occurs despite the unequivocal evidence for reciprocal reactions of late movers.*

Taken together, the results of Experiment 2 corroborate and expand those of Experiment 1. We find some evidence for reciprocal behavior in the second experiment. Nevertheless, incentive reversal occurs even in this strong social context of small natural groups. Since one might argue that small natural groups are more representative of actual work environments, it further underlines the strength and relevance of incentive reversal. That is, a global increase of salaries may lead agents at the beginning of the production process to free ride on the effort of agents choosing their strategies at the end of the process.

5 Conclusion

In this paper we report on two experiments designed to directly test for incentive reversal – the seemingly paradoxical inverse relationship between monetary rewards and incentives. Our results provide strong support for the emergence of incentive reversal. In particular, we observe in both experiments that when rewards increase (or costs decrease), late movers become less responsive to the observed history. Under these circumstances, early movers seem to anticipate that late movers will always exert effort. As a consequence, first movers shirk and free ride on the effort of late movers.

Although our lab experiments establish the empirical validity of incentive reversal, they abstract from features that are potentially important for team environments in the field, but are outside the scope of the theoretical model in Winter (2009). Most importantly, we did not allow for learning and reputation building, which come with repeated interactions. Our experimental results on reciprocal behavior suggest that incentive reversal depends on one-shot interactions and might not persist in a stable environment that allows for repeated interactions and concerns about reputations. In Experiment 1, we find that the optimal actions of first movers given the observed choices made by second movers do not entail incentive reversal. With repetition, first movers would be able to adapt to the behavior of second movers, thus eliminating incentive reversal. In particular, second movers in the high rewards treatment were willing to forgo part of their payoff in order to punish shirking by first movers. This behavior is likely to be strengthened if it can serve to build a 'tough' reputation, leading to high efforts by first movers regardless of the reward level. In Experiment 2, a similar pattern emerges. The fact that participants who choose reciprocal strategies as late movers do not anticipate reciprocity from others when deciding as first movers suggests that, with learning, behavior could converge to high effort choices even with low costs.

Our findings complement the existing literature studying the impact of monetary rewards on individuals' behavior. There is substantial evidence based on laboratory and field experiments showing that individuals' willingness to exert effort may not monotonically increase with monetary rewards. For example, parents' late pickup at daycare centers turns more severe after imposing a fine on late arrival, and scouts' performance in door-to-door collection of donations deteriorates when these children are offered to keep a share of the raised donations for themselves (Gneezy and Rustichini, 2000a,b). Similarly, opting to fine untrustworthy behavior actually increases such behavior (Fehr and List, 2004; Houser et al., 2008). These results, however, build on the behavioral dissonance between intrinsic and extrinsic motivations (see also Bowles, 2008, 2009 for brief overviews or Frey and Jegen, 2001 for a comprehensive survey of empirical evidence for motivation crowding-out). Whereas in the articles above it is the absence of money-maximizing individuals that causes incentives to 'backfire', the incentive reversal phenomenon described in our paper is due to the presence of rational, self-centered, money-maximizing individuals. Another difference is that the crowding-out literature highlights the potential adverse effects of *small* monetary rewards, which crowd out intrinsic motivation but are not large enough to substitute for it. The phenomenon studied here, in comparison, emerges when potential rewards are *large* enough to eliminate the role of actions taken by previous movers.

Along these lines, there exist also some closely related studies that analyze dysfunctional behavioral responses without relying on the discrepancy between intrinsic and extrinsic rewards. For example, Camerer et al. (1997) find a negative elasticity of New York City cabdrivers' number of working hours with respect to realized earnings per hour. They argue that this is due to income effects, i.e., drivers having daily income targets (but see also Farber, 2008, and Crawford and Meng, forthcoming). Another example would be Fehr and Schmidt (2004), who demonstrate that in an environment with multidimensional effort where only one effort dimension is contractible, piece-rate contracts are outperformed by fixed-wage contracts. In contrast to our work, these studies usually focus on individual decision problems rather than on team relationships. Moreover, they put forward different reasons for the occurrence of incentive reversal.

In essence, incentive reversal in teams is a manifestation of second (or higher) degree incentives. It follows a long tradition in game theory by highlighting the fact that individuals respond not only to direct incentives but also take into account the incentives of others with whom they interact. As such, potential implications of incentive reversal go beyond the workplace and the labor market. It applies to a variety of team environments and suggests that increasing all team members' stakes in the success of the joint activity may (though not necessarily shall) be counter effective. Political campaigns, commercial ventures, fundraising, joint decisions of committees and allocation in public-private partnerships (Athias and Soubeyran, 2013) are all relevant environments in which incentive reversal may emerge.

While incentive reversal is a "rational" phenomenon, our findings also have "behavioral" implications. Substantial experimental and empirical evidence reveals the role of reciprocity in teams (e.g. Falk and Ichino, 2006; Fehr and Fischbacher, 2003; Gould and Winter, 2009; Ichino and Maggi, 2000; Mas and Moretti, 2009). Team members are psychologically reluctant to exert effort or contribute when they detect shirking by their peers. This reluctance is, in fact, very important for the functioning of teams, as it generates an implicit threat against shirking. Our findings about incentive reversal suggest that high-powered monetary incentives may be counter-effective as they may jeopardize the credibility of this implicit threat.

We believe that the findings reported here are of interest for theorists and practitioners alike. They show that even the well intentioned introduction of (additional) rewards may occasionally backfire. For example, granting a pay rise to the workforce or offering job-training opportunities that reduce workers' effort costs might not always lead to an increase in performance (with the caveat regarding repeated interactions discussed above). On the contrary, actions that are meant to motivate workers may actually lead to incentive reversal – resulting in an effort reduction and higher costs to the principal. While this possibility depends on the exact characteristic of the environment at hand, principals should be aware of it and take it into account when designing contracts.

Acknowledgments

We are deeply grateful to The Israeli Foundation Trustees (IFT), the German Research Foundation (DFG), the German-Israeli Research Foundation and Google for funding this project. The authors thank Dirk Engelmann, the associate editors, two reviewers, and seminar participants at The Center for the Study of Rationality, The Max Planck Institute for Research on Collective Goods, University of Bonn, University of Mannheim, University College London, University of Warwick and at ESA meetings in Lyon and Haifa for comments and suggestions. Esteban Klor thanks the NBER and Boston University for their hospitality while he was working on this project.

References

Athias, L., Soubeyran, R., 2013. Demand risk allocation in incomplete contracts: The case of public private partnerships Unpublished manuscript, University of Lausanne.

Aumann, R.J., 2006. War and peace, in: Grandin, K. (Ed.), Nobel Prizes 2005: Les Prix Nobel. Almqvist & Wiksell Intl, Stockholm, pp. 350–358.

Binmore, K., McCarthy, J., Ponti, G., Samuelson, L., Shaked, A., 2002. A backward induction experiment. Journal of Economic Theory 104, 48–88.

Bone, J., Hey, J.D., Suckling, J., 2009. Do people plan? Experimental Economics 12, 12–25.

Bowles, S., 2008. Policies designed for self-interested citizens may undermine" the moral sentiments": Evidence from economic experiments. Science 320, 1605–1609.

Bowles, S., 2009. When economic incentives backfire. Harvard Business Review March, 22–23.

Camerer, C., Babcock, L., Loewenstein, G., Thaler, R., 1997. Labor supply of New York City cabdrivers: One day at a time. The Quarterly Journal of Economics 112, 407–441.

Carpenter, J.P., 2003. Bargaining outcomes as the result of coordinated expectations. Journal of Conflict Resolution 47, 119.

Clark, K., Sefton, M., 2001. The sequential prisoner's dilemma: evidence on reciprocation. The Economic Journal 111, 51–68.

Crawford, V.P., Meng, J., forthcoming. New York City cabdrivers' labor supply revisited: Reference-dependence preferences with rational-expectations targets for hours and income. American Economic Review .

Falk, A., Fischbacher, U., 2002. "Crime" in the lab — detecting social interaction. European Economic Review 46, 859–869.

Falk, A., Ichino, A., 2006. Clean evidence on peer effects. Journal of Labor Economics 24, 39–57.

Farber, H.S., 2008. Reference-dependent preferences and labor supply: The case of New York City taxi drivers. American Economic Review 98, 1069–1082.

Fehr, E., Fischbacher, U., 2003. The nature of human altruism. Nature 425, 785–791.

Fehr, E., List, J.A., 2004. The hidden costs and returns of incentives-trust and trustworthiness among CEOs. Journal of the European Economic Association 2, 743–771.

Fehr, E., Schmidt, K.M., 2004. Fairness and incentives in a multi-task principal–agent model. Scandinavian Journal of Economics 106, 453–474.

Fischbacher, U., Gächter, S., 2010. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. American Economic Review 100, 541–556.

Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? evidence from a public goods experiment. Economics Letters 71, 397–404.

Frey, B.S., Jegen, R., 2001. Motivation crowding theory. Journal of Economic Surveys 15, 589–611.

Gächter, S., Nosenzo, D., Renner, E., Sefton, M., 2010. Sequential vs. simultaneous contributions to public goods: Experimental evidence. Journal of Public Economics 94, 515–522.

Glöckner, A., Irlenbusch, B., Kube, S., Nicklisch, A., Normann, H.T., 2011. Leading with(out) sacrifice? a public-goods experiment with a super privileged player. Economic Inquiry 49, 591–597.

Gneezy, U., Rustichini, A., 2000a. A fine is a price. The Journal of Legal Studies 29, 1–17.

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

Goerg, S., Kube, S., Zultan, R., 2010. Treating equals unequally: Incentives in teams, workers' motivation, and production technology. Journal of Labor Economics 28, 747–772.

Gould, E.D., Winter, E., 2009. Interactions between workers and the technology of production: evidence from professional baseball. The Review of Economics and Statistics 91, 188–200.

Guttman, J.M., 1986. Matching behavior and collective action: Some experimental evidence. Journal of Economic Behavior & Organization 7, 171–198.

Harrison, G.W., McCabe, K.A., 1996. Expectations and fairness in a simple bargaining experiment. International Journal of Game Theory 25, 303–327.

Houser, D., Xiao, E., McCabe, K.A., Smith, V., 2008. When punishment fails: Research on sanctions, intentions and non-cooperation. Games and Economic Behavior 62, 509–532.

Ichino, A., Maggi, G., 2000. Work environment and individual background: Explaining regional shirking differentials in a large italian firm. Quarterly Journal of Economics 115, 1057–1090.

Johnson, E.J., Camerer, C., Sen, S., Rymon, T., 2002. Detecting failures of backward induction: Monitoring information search in sequential bargaining. Journal of Economic Theory 104, 16–47.

Mas, A., Moretti, E., 2009. Peers at work. American Economic Review 99, 112–145.

Meidinger, C., Villeval, M.C., 2002. Leadership in teams: Signaling or reciprocating? GATE Working Paper 10–3.

Potters, J., Sefton, M., Vesterlund, L., 2007. Leading-by-example and signaling in voluntary contribution games: an experimental study. Economic Theory 33, 169–182.

Selten, R., 1967. Die strategiemethode zur erforschung des eingeschränkt rationalen verhaltens im rahmen eines oligopolexperimentes, in: Sauermann, H. (Ed.), Beiträge zur experimentellen Wirtschaftsforschung. Mohr Siebeck, Tübingen, pp. 136–168.

Steiger, E.M., Zultan, R., 2011. See no evil: Information chains and reciprocity in teams. Jena Economic Research Papers 2011-40.

Varian, H.R., 1994. Sequential contributions to public goods. Journal of Public Economics 53, 165–186.

Winter, E., 2009. Incentive reversal. American Economic Journal: Microeconomics 1, 133–147.

Winter, E., 2010. Transparency and incentives among peers. Rand Journal of Economics 41, 504–523.

Appendix A: Materials for Experiment 1

General instructions

Welcome to the experiment.

From now on you are not allowed to talk to the other participants.

During the experiment you can accumulate points according to the decisions you will make.
The points will be converted to money at a rate of **80 points = 1 NIS**.
At the end of the experiment please wait for the experimenter to call you for payment.

We will distribute the instructions for the experiment and read them out loud. If you have any questions, please wait until we have finished reading the experiment, and then raise your hand. The experimenter will come to you to answer your question.

Instructions for the experiment

(*Text in square brackets appeared only in the sequential treatment*)

In this experiment you will participate in six rounds. In each round, the participants in the experiment will be allocated to groups of two. Each participant has to decide whether to work **Normal** or **Hard**. The more participants in a group choose to work hard, the more units the group produces. The following table provides an example:

No. of participants working hard	0	1	2
No. of units produced	10	50	100

In this example, if all participants in the group work normal, 10 units are produced. If one participant works normal and the other hard, 50 units are produced. If both participants work hard, 100 units are produced.

In each group, one of the participants is randomly chosen to be **Participant 1**, and the other to be **Participant 2**. [First, Participant 1 decides whether to work normal or hard. Next, **Participant 2 observes the decision of participant 1**, and decides whether to work normal or hard.]

For each unit produced by your group, you will receive a certain number of points, which differs between participants. The following table provides an example:

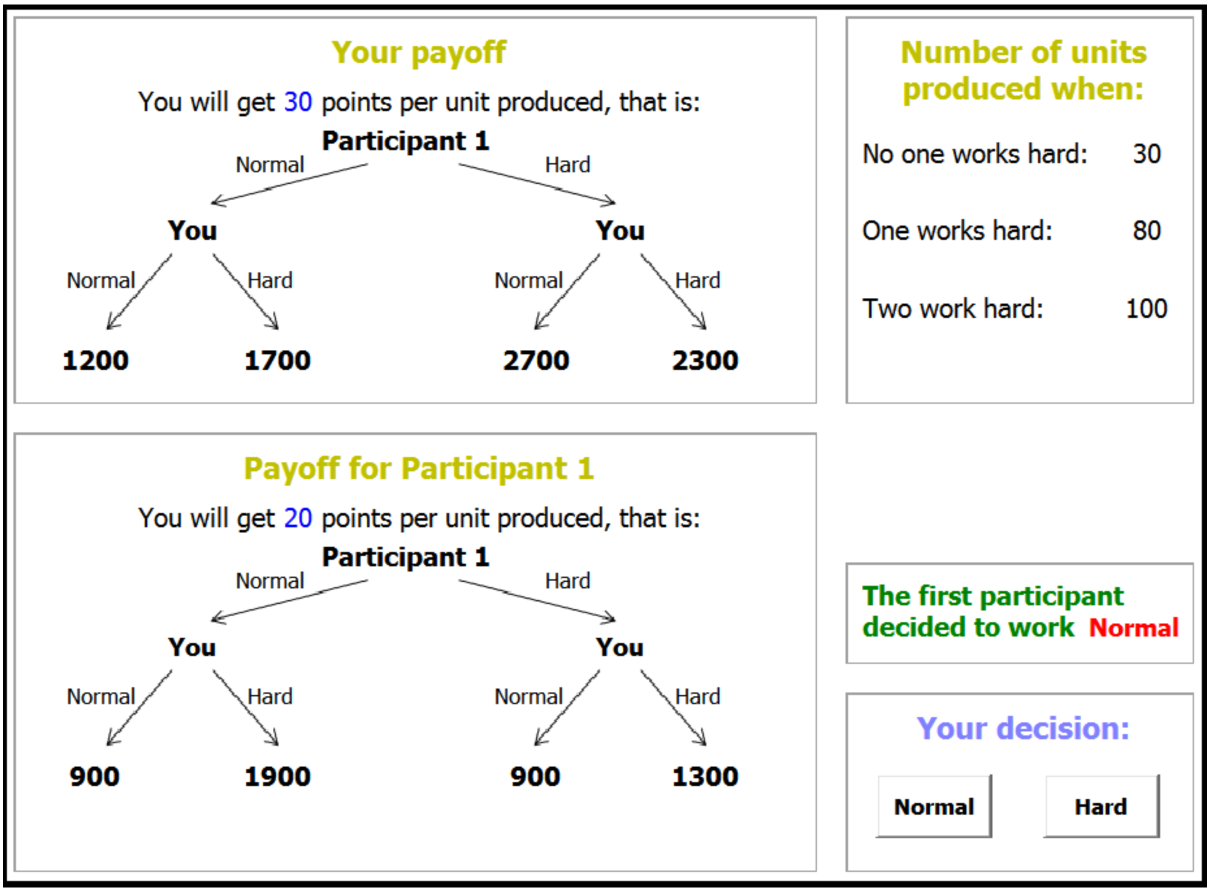
Participant 1	Participant 2
16 points per produced unit	30 points per produced unit

In addition, you will receive a base wage of 300 points.

In the beginning of the experiment you will learn whether you are chosen to be Participant 1 or Participant 2. You will remain in this role for all six rounds of the experiment

Sequential treatment

In each round of the experiment you will see a screen similar to the following screen. Note that the numbers provided in this example are different from the numbers appearing in the tables in the previous page.



Your payoff

You will get 30 points per unit produced, that is:

Number of units produced when:

No one works hard:	30
One works hard:	80
Two work hard:	100

Payoff for Participant 1

You will get 20 points per unit produced, that is:

The first participant decided to work Normal

Your decision:

The top left panel details the number of units produced based on the number of participants who work hard, similar to the table in the previous page.

The top right panel details the number of points you will earn depending on your and the other participant's decision. the panel below it details the number of points that the other participant will earn. In the screen that Participant 1 sees for this example, the position of the two panels is switched, so that each participant sees their own possible earnings in the top panel and the possible earnings of the other participant in the bottom panel.

As this example provides the screen for Participant 2, the decision that Participant 1 made appears on the left. Naturally, this information does not appear in the screen for Participant 1.

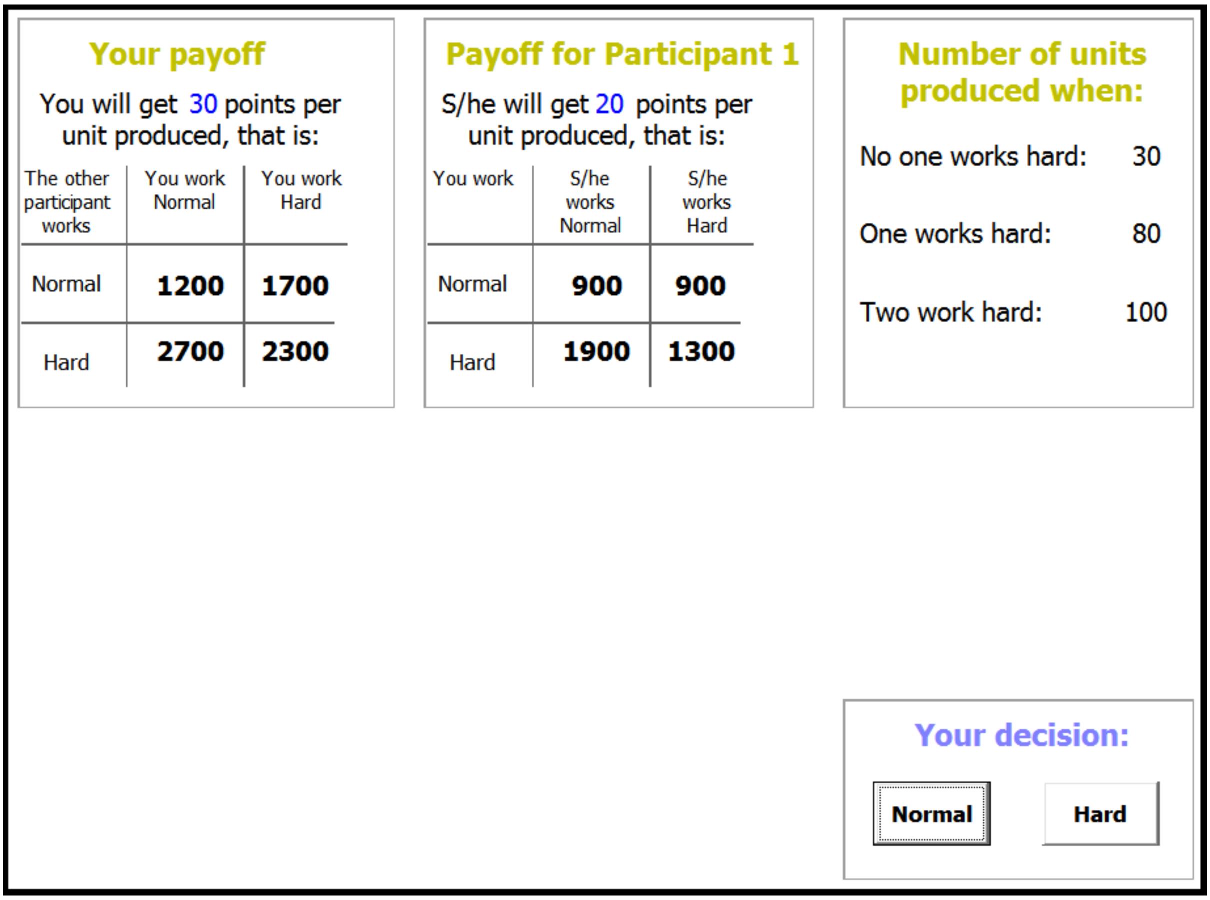
The buttons for making your decision appear at the bottom left of the screen.

In this example you are participant 2, and receive 30 points per produced unit, whereas Participant 1 receives 20 points per produced unit. In addition, you can see that Participant 1 chose to work normal. If, for example, you will choose to work hard, 80 units will be produced, for which you will receive 80x30=2,400 points, of which 1,000 points will be deducted for the hard work, and to which 300 points will be added as base wage. In total you will earn 1,700 points.

Participant 1 will receive in this case 80x20=1,600 points, in addition to the base wage of 300 points, and in total 1,900 points. Participant 1 does not pay 1,000 points as he chose to work normal.

Simultaneous treatment

In each round of the experiment you will see a screen similar to the following screen. Note that the numbers provided in this example are different from the numbers appearing in the tables in the previous page.



Your payoff

You will get 30 points per unit produced, that is:

The other participant works	You work Normal	You work Hard
Normal	**1200**	**1700**
Hard	**2700**	**2300**

Payoff for Participant 1

S/he will get 20 points per unit produced, that is:

You work	S/he works Normal	S/he works Hard
Normal	**900**	**900**
Hard	**1900**	**1300**

Number of units produced when:

No one works hard:	30
One works hard:	80
Two work hard:	100

Your decision:

The top left panel details the number of units produced based on the number of participants who work hard, similar to the table in the previous page.

The right panel details the number of points you will earn depending on your and the other participant's decision. the panel to the left of it details the number of points that the other participant will earn. In the screen that Participant 1 sees for this example, the position of the two panels is switched, so that each participant sees their own possible earnings in the right-hand panel and the possible earnings of the other participant in the left-hand panel.

The buttons for making your decision appear at the bottom left of the screen.

In this example you are participant 2, and receive 30 points per produced unit, whereas Participant 1 receives 20 points per produced unit. If, for example, you will choose to work hard and Participant 1 decides to work normal, 80 units will be produced, for which you will receive 80x30=2,400 points, of which 1,000 points will be deducted for the hard work, and to which 300 points will be added as base wage. In total you will earn 1,700 points.

Participant 1 will receive in this case 80x20=1,600 points, in addition to the base wage of 300 points, and in total 1,900 points. Participant 1 does not pay 1,000 points as he chose to work normal.

Appendix B: Materials for Experiment 2

(*The costs and payments correspond to the low-costs treatment.*)

Instructions

In this experiment, we will let you play a game for three participants: Participant 1, Participant 2, and Participant 3. In the game, you may win money, as explained below.

Rules of the Game

The three participants in the game constitute a group. A budget of NIS 30 is made available to the group. Each participant, in turn, may choose to double the group's budget for a certain price that he or she will pay at the end of the game. Participant 1 decides first, followed by Participant 2 and finally Participant 3. Each participant knows what the preceding participants have chosen.

At the end of the game, the final budget is divided equally among the three members of the group, and the member who chose to double it will pay the price of his or her decision from his or her share.

The following table shows the participants' payments in accordance with their decisions. Note that if a participant chooses to double the budget, his or her final profit will be his or her share in the budget (in accordance with the table) less the price of having doubled the budget (not shown in the table).

Number of participants who choose to double the budget	Budget obtained	Each participant's share in the budget
0	NIS 30	NIS 10
1	NIS 60	NIS 20
2	NIS 120	NIS 40
3	NIS 240	NIS 80

The prices that each participant must pay for doubling the budget are the following:

Participant 1: NIS 55 Participant 2: NIS 50 Participant 3: NIS 5

For example, if all members of the group decide not to double the budget, each member will be left with NIS 10. If all of members decide to double the budget, each member will accumulate NIS 80, from which the price of having doubled the budget will be subtracted at the end, ultimately leaving Participant 1 with NIS 25, Participant 2 with NIS 30, and Participant 3 with NIS 75.

If a participant is left with a negative sum at the end of the game, he or she will not have to pay anything; he or she will simply remain with 0.

How the experiment will take place

We will be handing out a sheet of paper. On one side of the sheet, you are asked to record your decisions. On the other side, several questions appear, the purpose of which is to make sure that you understood the instructions. **If you fail to answer these questions correctly, we will not be able to take your data into account and, accordingly, you will not be paid.**

You must decide what you would do in the "shoes" of each participant and record your decision on the page. After we collect all the pages, we will aggregate them randomly into three-person groups and conduct a draw within each group to determine who will be Participant 1, who will be Participant 2, and who will be Participant 3. Then we will play the game, in such a way each participant will play on the basis of what he or she recorded on the page. In this manner, each player's earnings will be determined.

For us to pay you what you are owed, you must record the last four digits of your ID number in the appropriate place on the page. We will use this information to identify you in order to pay you.

After you record your decision on the page, please return both pages to the experimenter. Thank you for participating in the experiment!

Decision sheet

ID no:______

Please record your decision in each of the following cases:

If I am Participant 1, I will choose:

○ To double the sums to NIS 20 per person, and then it is Participant 2’s turn.

○ To leave the sums at NIS 10 per person, and then it is Participant 2’s turn.

If I am Participant 2, then…

If Participant 1 chooses not to double the budget, I will choose:

○ To double the sums to NIS 20 per person, and then it is Participant 3’s turn.

○ To leave the sums at NIS 10 per person, and then it is Participant 3’s turn.

If Participant 1 chooses to double the budget, I will choose:

○ To double the sums to NIS 40 per person, and then it is Participant 3’s turn.

○ To leave the sums at NIS 20 per person, and then it is Participant 3’s turn.

If I am Participant 3, then…

If the two previous participants choose not to double the budget, I will choose:

○ To double the sums to NIS 20 per person, and then the game ends.

○ To leave the sums at NIS 10 per person, and then the game ends.

If only Participant 1 chooses to double the budget, I will choose:

○ To double the sums to NIS 40 per person, and then the game ends.

○ To leave the sums at NIS 20 per person, and then the game ends.

If only Participant 2 chooses to double the budget, I will choose:

○ To double the sums to NIS 40 per person, and then the game ends.

○ To leave the sums at NIS 20 per person, and then the game ends.

If both of the previous participants choose to double the budget, I will choose:

○ To double the sums to NIS 80 per person, and then the game ends.

○ To leave the sums at NIS 20 per person, and then the game ends.

Control questions

Please answer the following questions: Reminder: the price of doubling the budget is NIS 55 for Participant 1, NIS 50 for Participant 2, and NIS 5 for Participant 3.

1. How much will each participant ultimately receive if Participant 1 chooses to double the budget, Participant 2 chooses not to double it, and Participant 3 chooses to double it?

   Participant 1 will receive NIS ______.
   Participant 2 will receive NIS ______.
   Participant 3 will receive NIS ______.

2. How much will each participant ultimately receive if Participant 1 chooses not to double the budget, Participant 2 chooses to double it, and Participant 3 chooses not to double it?

   Participant 1 will receive NIS ______.
   Participant 2 will receive NIS ______.
   Participant 3 will receive NIS ______.

3. How much will each participant ultimately receive if Participant 1 chooses not to double the budget, Participant 2 chooses to double it, and Participant 3 chooses to double it?

   Participant 1 will receive NIS ______.
   Participant 2 will receive NIS ______.
   Participant 3 will receive NIS ______.

4. How much will each participant ultimately receive if Participant 1 chooses not to double the budget, Participant 2 chooses to not double it, and Participant 3 chooses to double it?

   Participant 1 will receive NIS ______.
   Participant 2 will receive NIS ______.
   Participant 3 will receive NIS ______.

Appendix C: Game Matrices and Trees

The payoffs for Experiment 1 are excluding the base wage of 300 points.

Parameter Set I

		II	
		$e_2 = 0$	$e_2 = 1$
I	$e_1 = 0$	$(\underline{1470},\underline{1530})$	$(\underline{2940},1960)$
	$e_1 = 1$	$(440,3060)$	$(2400,\underline{4000})$

Parameter Set II

		II	
		$e_2 = 0$	$e_2 = 1$
I	$e_1 = 0$	$(\underline{2800},3150)$	$(\underline{3200},\underline{3200})$
	$e_1 = 1$	$(2200,3600)$	$(3000,\underline{4100})$

Parameter Set III

		II	
		$e_2 = 0$	$e_2 = 1$
I	$e_1 = 0$	$(\underline{1550},3000)$	$(\underline{2170},\underline{3200})$
	$e_1 = 1$	$(1170,4200)$	$(2100,\underline{5000})$

Figure C1: Experiment 1 – Simultaneous Protocol, High Rewards

Parameter Set I

		II	
		$e_2 = 0$	$e_2 = 1$
I	$e_1 = 0$	($\underline{1440}$,$\underline{930}$)	($\underline{2880}$,760)
	$e_1 = 1$	(380,1860)	(2300,$\underline{2000}$)

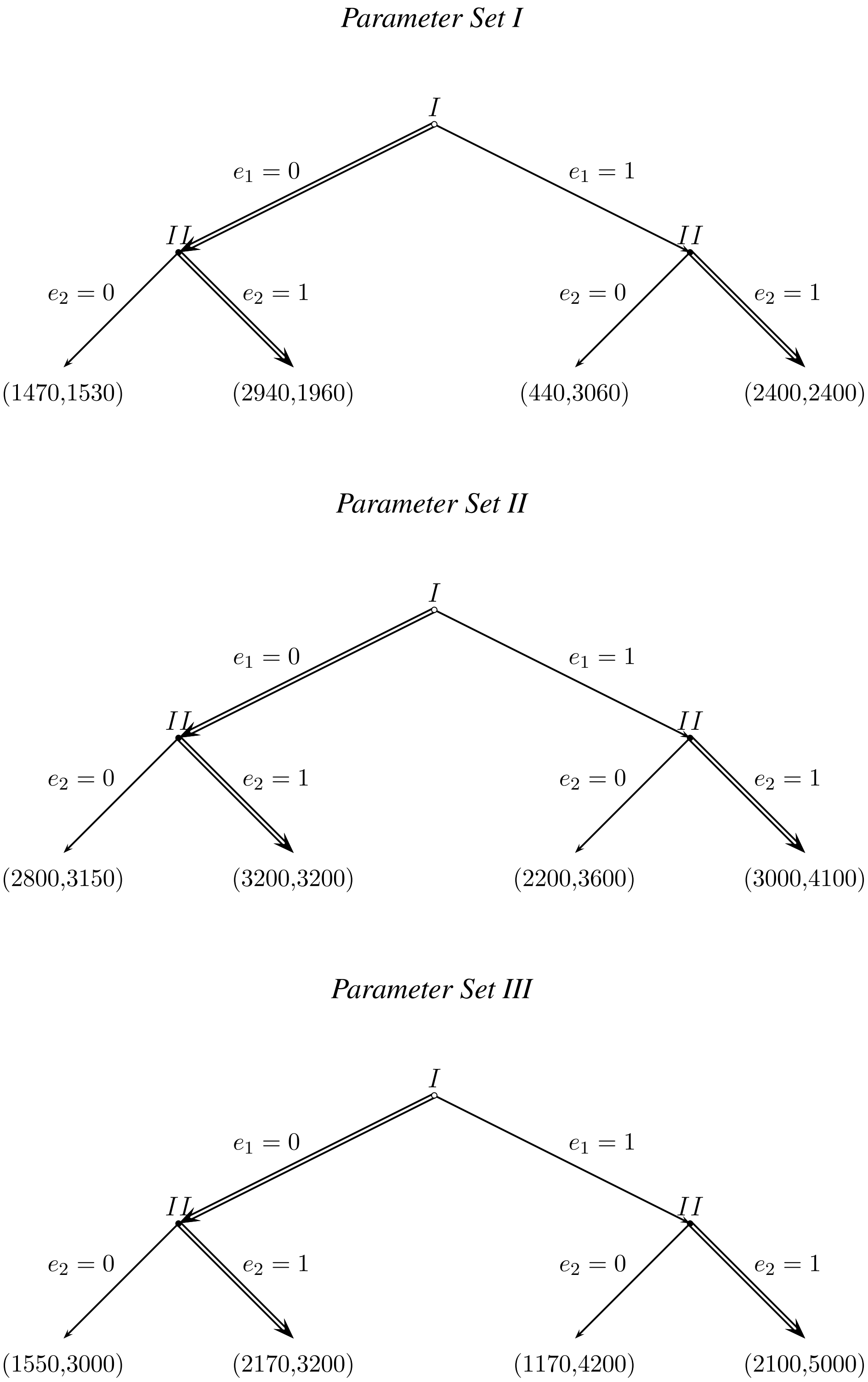
Parameter Set II

		II	
		$e_2 = 0$	$e_2 = 1$
I	$e_1 = 0$	($\underline{2450}$,$\underline{2450}$)	($\underline{2800}$,2400)
	$e_1 = 1$	(1800,2800)	(2500,$\underline{3100}$)

Parameter Set III

		II	
		$e_2 = 0$	$e_2 = 1$
I	$e_1 = 0$	($\underline{1400}$,$\underline{2150}$)	($\underline{1960}$,2010)
	$e_1 = 1$	(960,3010)	(1800,$\underline{3300}$)

Figure C2: Experiment 1 – Simultaneous Protocol, Low Rewards

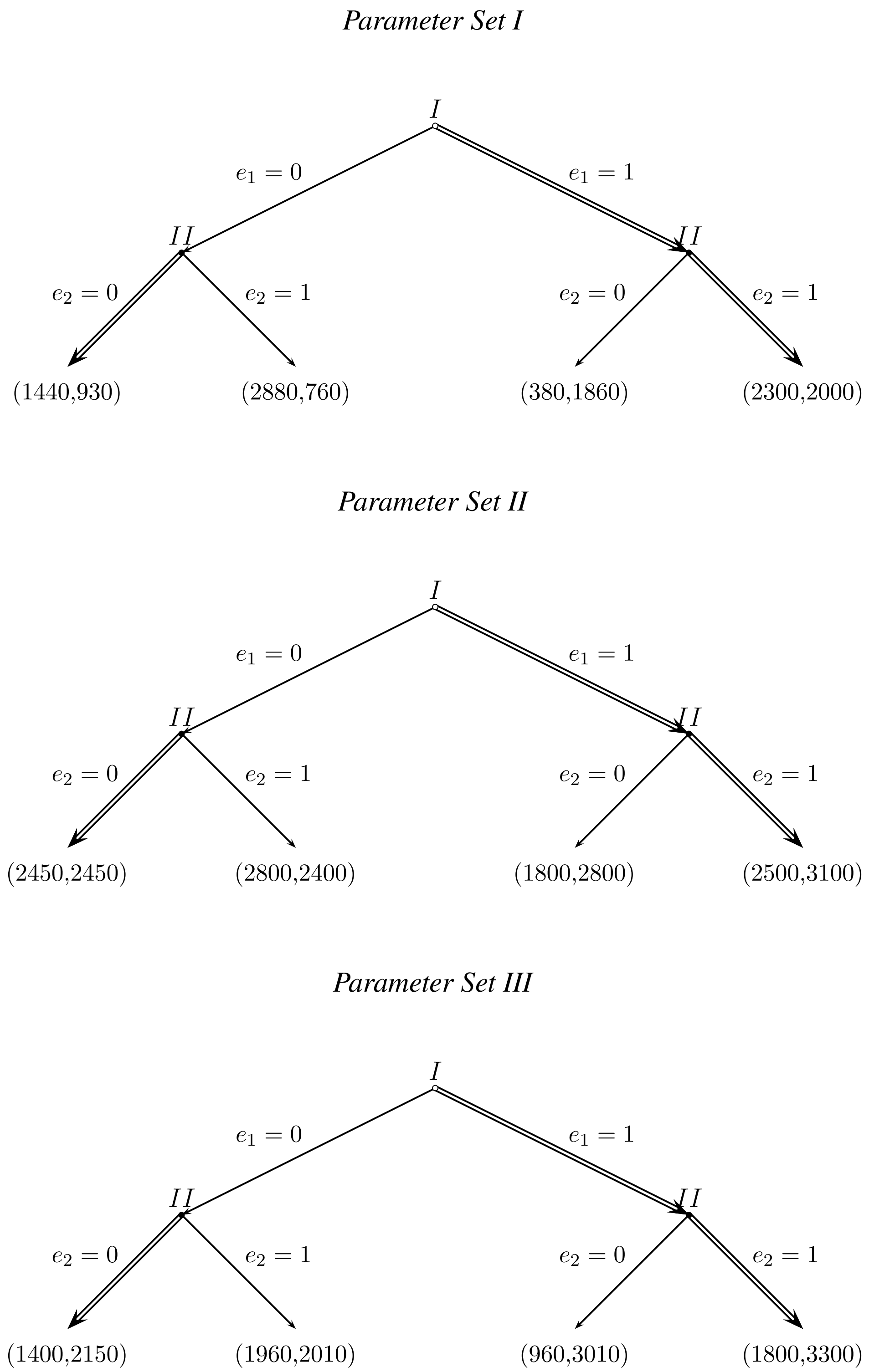


Parameter Set I

Parameter Set II

Parameter Set III

Figure C3: Experiment 1 – Sequential Protocol, High Rewards



Parameter Set I

Parameter Set II

Parameter Set III

Figure C4: Experiment 1 – Sequential Protocol, Low Rewards

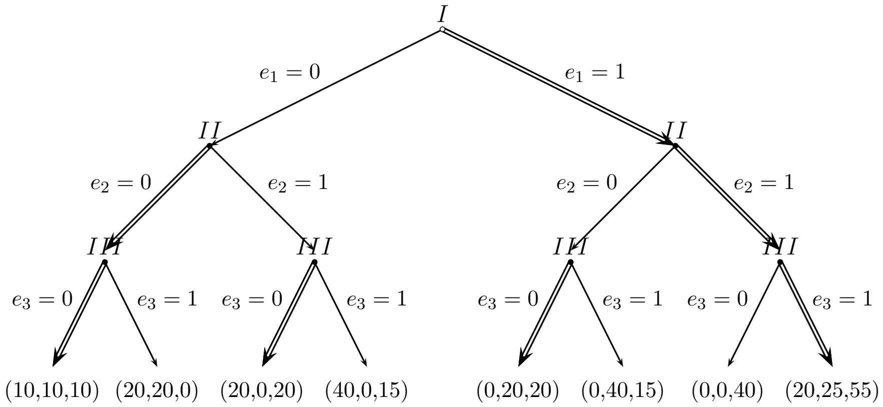


Figure C5: Experiment 2 – High Cost

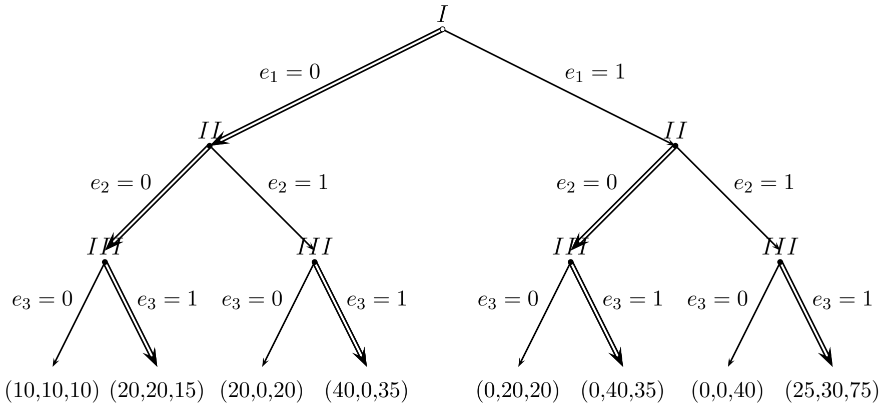


Figure C6: Experiment 2 – Low Cost