# Can Higher Bonuses Lead to Less Effort? Incentive Reversal in Teams

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

March 4, 2011

**Abstract**

Conventional wisdom suggests that an increase in monetary incentives should induce agents to exert higher effort. In this paper, however, 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 incentives (either because rewards 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.

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

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

*Klor: The Hebrew University of Jerusalem and CEPR, Economics Department, Mt. Scopus, Jerusalem 91905, eklor@mscc.huji.ac.il. Kube: University of Bonn, Department of Economics, Adenauerallee 24-42, 53113 Bonn, kube@coll.mpg.de. Winter: The Hebrew University of Jerusalem, Center for the Study of Rationality, Givat Ram, Jerusalem 91904, mseyal@mscc.huji.ac.il. Zultan: University College London, Cognitive, Perceptual and Brain Sciences, 26 Bedford Way, London WC1H 0AP, r.zultan@ucl.ac.uk. We are deeply grateful to The Israeli Foundation Trustees (IFT) and the German Research Foundation (DFG) for funding this project. The authors thank seminar participants at The Center for the Study of Rationality, The Max Planck Institute for Research on Collective Goods, 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.

# 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 as it occurs, for example, in team production settings. Particularly, in the context of sequential team production, incentive reversal might occur — even 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 for the occurrence of the counterintuitive relationship between monetary incentives and motivation.

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 surprisingly 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 $j$,

[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).

(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. By contrast, if there is substitution among agents' efforts, the argument above does not hold. That is, if the effort of agent $i$ pays off when agent $j$ is exerting effort as well, it pays off even more when agent $i$ expects $j$ to shirk.

Simple as it may seem, it is not clear whether the argument for incentive reversal is empirically sound on three grounds. First, incentive reversal is a puzzling and a rather counter-intuitive phenomenon precisely because we tend to think about monetary incentives and motivation as moving in the same direction in a fully rational environment. Second, incentive reversal requires non-trivial backward induction reasoning.[2] Finally, and perhaps most importantly, social preferences (and in particular the presence of reciprocity) may eliminate the prospects of incentive reversal. Indeed, if an individual who detects the shirking of his peer is inclined to retaliate by shirking as well, even if from a strictly monetary standpoint it is rational for her to exert effort, the observed individual (anticipating reciprocal behavior) would be reluctant to shirk. In this event, incentive reversal might thus not be observable.[3]

Whether incentive reversal in teams actually occurs or not is, of course, ultimately an empirical question. Moreover, theoretical predictions strongly rely on having sufficiently precise knowledge about the shape of the produc-

[2] As Johnson et al. (2002) show, naïve subjects are not likely to behave in line with backward induction, even when playing with computerized partners who are known to follow the backward induction path; although with instruction and practice, subjects learn to follow backward induction reasoning. For other experiments studying backward induction in multi-stage bargaining games see Harrison and McCabe (1996), Binmore et al. (2002), and Carpenter (2003). 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 (e.g. Clark and Sefton, 2001; Fischbacher et al., 2001; Falk and Fischbacher, 2002) and between periods when the game is repeated (e.g. Guttman, 1986; Fischbacher and Gächter, 2010).

tion technology, the move structure and information set of each player, as well as the potential rewards and individuals' costs of exerting effort. To this end, 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 team's joint effort. In both experiments we create experimental treatments with either high or low incentives that are susceptible to incentive reversal. In the first experiment, the incentives are manipulated by changing the costs of exerting effort and in the second experiment by manipulating the promised rewards.

In order to be able to attribute an incentive reversal effect to the process described above, we take two different 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 is eliminated. In the second experiment, we use the strategy method instead of the play method to obtain counterfactual data. Thus, by observing subjects' decisions in each node of the game tree we can test for incentive reversal by looking at behavior along the theoretical equilibrium path. Additionally, we can carry out a direct and clean within-subject analysis of reciprocal behavior by exploring behavior off the equilibrium path.

Our experimental data provide clear support for the empirical relevance of incentive reversal in teams. The increase in rewards in the first experiment and the decrease in effort costs in the second experiment cause a significant decline in effort provision. 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. This behavior is prominent in sequential games but not in simultaneous games — as theory predicts. The average effort provided by the first-movers drops by almost 50 percent when incentives 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 (i.e., under low incentives) than under low costs. For example, first-movers' average effort is increased by almost 130 percent when costs are increased (i.e., immediate incentives are decreased). Moreover, subjects' subsequent choices along the equilibrium path are well in line with the predictions from incentives reversal. Interestingly, this holds true even though we observe some tendency for reciprocal behavior in both treatments, which underlines the relative importance of incentive reversal in such an environment.

Our findings complement the existing literature studying the impact of monetary incentives on individuals' behavior. In fact, 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).[4] These results, however, build on the behavioral dissonance between intrinsic and extrinsic motivations (see also Bowles (2009) for a brief overview or Frey and Jegen (2001) for a comprehensive survey of empirical evidence for motivation crowding-out). So, while in the instances studied by the above articles it is the absence of money-maximizing individuals that cause incentives to 'backfire', the incentives reversal described in our paper is due to the presence of fully rational, self-centered, money-maximizing individuals.

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 work-

[4]See Benabou and Tirole (2006) for an interesting theoretical model that accounts for the lack of monotonicity between monetary incentives and motivation.

ing 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 (in press)). 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, however, 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.

To sum up, incentive reversal in teams is an important manifestation of second (or higher) degree incentives. It highlights 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, the 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 and joint decisions of committees are all relevant environments in which incentive reversal may emerge.

The remainder of the paper proceeds as follow. Section 2 presents the theoretical framework behind the experimental design. In section 3 we describe the experimental design of Experiment 1 and the results from this experiment. Section 4 describes the experimental design and results for Experiment 2. We conclude in Section 5.

## 2 Theoretical Framework

The theoretical framework we consider is based on Winter (2009). Winter (2009) analyses the possibility of incentive reversal in a general theoretical framework. He shows 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 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. Hereby we provide an illustration of the main intuition behind incentive reversal with a deterministic technology that is also employed in our experimental design.[5]

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 $e$, with $e = 1$ when an agent exerts effort and $e = 0$ when he shirks. Agents move sequentially and information is perfect. Agent $i$'s payoff function is given by

$$U_i(e_1, e_2) = r_i P(e_1 + e_2) - e_i C_i, \tag{1}$$

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 function $P$ is strictly convex on the sum of effort. For the two-agent case being examined this implies that

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

that is, the technology has complementarities across agents' efforts since the effort of one agent increases the marginal productivity of the other agent. In other words, the technology is such that an agent's effort creates positive externalities on the other agent's productivity.

For the purposes of this example, let us consider the set III of parameters that we use in our first experiment (see Table 1 below). In particular, suppose that the rewards are $r_1 = 28$ and $r_2 = 43$, and the costs are $C_1 = C_2 = 1,000$. Finally, let us set $P(2) = 100$, $P(1) = 70$ and $P(0) = 50$. For these parameters, there exists a unique Subgame-Perfect Equilibrium whereby on the equilibrium path both agents choose to exert effort. Thus,

[5] 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).

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 rest of the parameters 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 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. 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 cost associated with exerting effort. Consequently, shirking becomes agent 1's equilibrium strategy under the new incentive scheme. In addition to the particular properties of the production technology, information about the effort exerted by peers plays a crucial role for incentive reversal to occur.[6] When agent 2 is uninformed of the strategic choice of agent 1, the sequential game described above basically turns into a simultaneous game. 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 aimed at establishing how well the theoretical predictions of incentive reversal reflect actual behavior in the laboratory.

## 3.1 Experimental Design and Procedure

In the experiment, teams of two agents work on a joint project, under either a simultaneous or sequential protocol. We ran three sessions with a sequential

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

protocol and two sessions with a simultaneous protocol. Both protocols use a similar procedure. In each session, twelve subjects were admitted into the lab and received written instructions, which were then read out aloud by the experimenter.[7]

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. Throughout the experiment we ensured anonymity and effectively isolated each subject in a cubicle to minimize any interpersonal influence that could stimulate uniformity of behavior. Communication among subjects was not allowed throughout the session.

Thirty six subjects participated in 3 sessions in the sequential treatment, and 24 subjects participated in 2 sessions in the simultaneous treatment. 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. At the beginning of each round all the 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. In the sequential protocol we informed subjects in the role of second movers of the corresponding first mover's choice before they were able to choose an option. Otherwise no feedback was given between rounds, so that first movers were informed of the corresponding second mover's choices only at the end of the session. In the simultaneous protocol both agents choose an option without knowing the option chosen by the other agent, with all agents being informed of their partners' decisions only at the end of the experimental session. Each session lasted about 45 minutes. Each subject received a base payment of 300 experimental points at the beginning of each

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

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 63.[8]

Each experimental session entailed six independent rounds. In each round, the subjects were (commonly known to be) re-matched in a stranger design, i.e., with a randomly selected subject. Subjects knew that their decisions and earnings in one round were independent from their decisions in another round. We used three different sets of parameters to generalize our results beyond a particular specification. Each subject played all three sets of parameters twice over the six rounds, once with low rewards and once with high rewards, with a different partner in each round. The order of the parameter sets was predetermined and stayed constant in all sessions and for all subjects.[9] This design allows us to examine the behavior of the same subject as the rewards scheme changes from low to high bonuses, abstracting from the specific parameters used in different rounds. Table 1 presents all the parameter sets used in experiment 1 as well as the equilibrium payoffs and strategies for the sequential and simultaneous treatments.

In each session, every subject played all three sets of parameters twice, once with low rewards and once with high rewards, always with a different partner. The order of the parameter sets was predetermined and stayed constant in all sessions and for all subjects.

### 3.2 Results

To test for incentive reversal, we first compute for each subject the number of times he chooses to exert effort differentiating between rounds with high and low rewards. Figure 1 depicts the average propensity of the subjects to exert effort separately for every protocol with standard errors.

[8] This is more than three times the 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 below NIS 3.7 per U.S. dollar.

[9] Over the six rounds, subjects played with 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 (cf. Table 1). Notice that no feedback was given between

Table 1: Parameters for Experiment 1.

| | Set of parameters | | | | | |
|---|---|---|---|---|---|---|
| | I | | II | | III | |
| Units produced when total effort equals: | | | | | | |
| 0 | 30 | | 70 | | 50 | |
| 1 | 60 | | 80 | | 70 | |
| 2 | 100 | | 100 | | 100 | |
| | Cost of effort | | | | | |
| Agent 1 | 2,500 | | 1,000 | | 1,000 | |
| Agent 2 | 1,100 | | 400 | | 1,000 | |
| | Rewards per unit produced | | | | | |
| - 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 strategies | | | | | |
| | Sequential | | | Simultaneous | | |
| - Low rewards treatment | | | | | | |
| Agent 1 | $e_1 = 1$ | | | $e_1 = 0$ | | |
| Agent 2 | $e_1 = 1$ | | | $e_1 = 0$ | | |
| - High rewards treatment | | | | | | |
| Agent 1 | $e_1 = 0$ | | | $e_1 = 0$ | | |
| Agent 2 | $e_1 = 1$ | | | $e_1 = 1$ | | |
| | Equilibrium payoffs | | | | | |
| Protocol | Sequential | Simultaneous | Sequential | Simultaneous | Sequential | Simultaneous |
| - Low rewards treatment | | | | | | |
| Agent 1 | 2,600 | 1,740 | 2,800 | 2,750 | 2,100 | 1,700 |
| Agent 2 | 2,300 | 1,230 | 3,400 | 2,750 | 3,600 | 2,450 |
| - High rewards treatment | | | | | | |
| Agent 1 | 3,240 | 3,240 | 3,500 | 3,500 | 2,470 | 2,470 |
| Agent 2 | 2,260 | 2,260 | 3,500 | 3,500 | 3,500 | 3,500 |

Note: Equilibrium Payoffs include base payment of 300 points given at the beginning of each round.

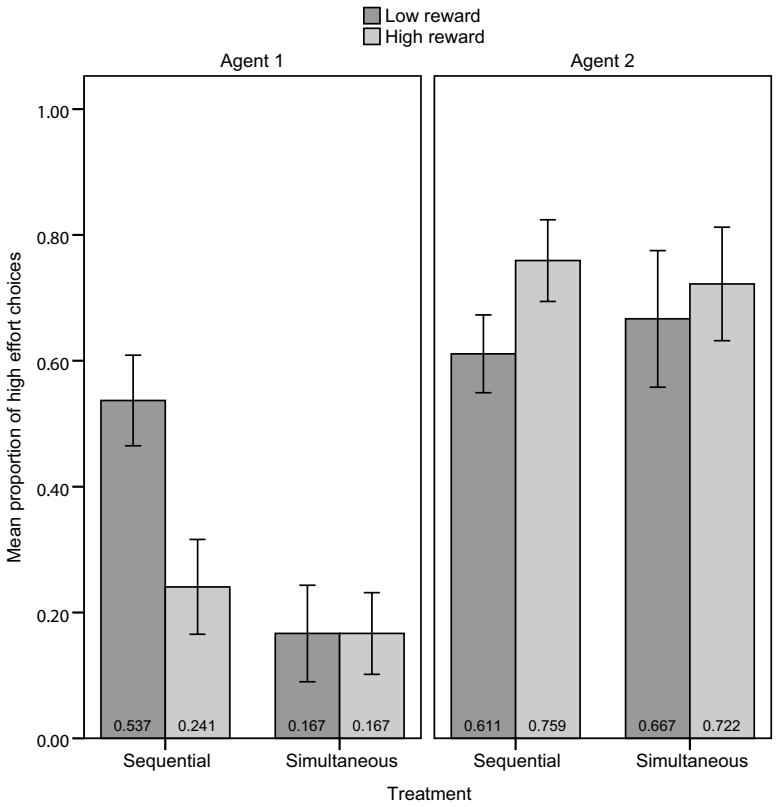



Figure 1: Experiment 1 – Effort decisions

Let us focus first on the behavior of subjects in the role of agent 1 depicted in the left panel of the figure. 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 protocols. The mean is thus not substantially different from the Nash equilibrium of the game, which prescribes that subjects should not exert effort in the simultaneous protocol.[10]

By contrast, the reward structure does affect subjects' behavior in the sequential protocol. Here, we observe that first-movers are significantly more

rounds.

[10]The estimated standard error of the proportion is 0.062.

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 Signed Ranks test, $Z = 2.769$, $p < 0.01$, two-sided). A repeated-measures model for testing the interaction between protocol and reward level also reveals a significant interaction ($F = 7.314$, $p < 0.05$, two-sided). This effect is not only qualitatively significant. It is also quantitatively important as subjects' effort more than doubles when rewards are low. As effort decisions only differ between the sequential protocol-low reward treatment and the other three treatments, the higher effort in this treatment implies significant main effects as well ($F = 7.314$, $p < 0.05$ for reward level; $F = 5.781$, $p < 0.05$ for protocol; both two-sided).[11]

Let us now turn to the behavior of subjects in the role of agent 2. In accordance with the theoretical predictions, a large majority of subjects exerts effort while in the role of agent 2. 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 seem to be higher in the high reward rounds, though the difference in efforts is not statistically significant between high- and low-rewards rounds.[12] 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 exhibit reciprocal behavior to agent 1's strategy only when rewards are low.[13]

The observed incentive reversal has interesting implications on total pro-

[11] These tests are carried for the participants in the role of agent 1, for whom the model has different predictions as a function of the rewards scheme. Analyses for agent 2's decisions are limited, of course, due to inter-subject dependencies.

[12] Strict testing for agent 2 decisions is weak because the observations are not independent. If we take subjects as independent observations, we obtain a weakly-significant effect in the sequential treatment (Wilcoxon Signed Ranks test, $Z = 1.721$, $p = 0.085$, two-sided). No significant difference is apparent in the simultaneous treatment, even under these relaxed assumptions (Wilcoxon Signed Ranks test, $Z = 0.574$, $p = 0.566$, two-sided).

[13] A likelihood ratio test provides evidence that agent 2's behavior is highly contingent on agent 1's behavior in the low reward rounds (likelihood ratio equals 17.56 with $p < 0.001$). This is not the case in the high rewards round (likelihood ratio equals 2.99 with $p > 0.05$). Note that these tests treat each observation as independent, so our test statistics reported here potentially overestimate significance levels.

Table 2: Experiment 1 – Distribution of Effort and Total Units Produced.

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

Note: 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.

duction, especially if we keep in mind that the production function is convex. Table 2 depicts the distribution of total team effort, the average amount of units produced by the teams and the teams' average payoffs.

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. On the contrary, in the treatment with high rewards incentive reversal occurs, so that we observe not only a lower average level of effort but also that a total team effort of one is the most frequent outcome. The difference between the two distributions is significant ($\chi^2 = 13.144, p < 0.005$).[14]

The difference in the level of team effort induced by the rewards 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 (Wilcoxon Signed Ranks test, $Z = 2.032$, $p < 0.05$, two-sided). This important difference in units produced is not reflected in the costs of production faced by the principal. A team's average pay equals NIS 75.6 when rewards are low and NIS 88.9

[14]The tests on the team statistics reported in the bottom half of Table 2 take the averages for subjects in the role of agent 1 in the Low- and High-rewards rounds as the unit of observation. Note that each subject in the role of agent 2 is equally represented in the two rewards levels, thus alleviating the problem of interdependencies.

when rewards are high (Wilcoxon Signed-Ranks test, $Z = 2.678$, $p < 0.01$, two-sided). 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 (Wilcoxon Signed-Ranks test, $Z = 3.724$, $p < 0.001$, two-sided).

The right panel of Table 2 presents summary statistics for the simultaneous treatment. The results in the table show that, as expected, the difference between the high and low rewards regimes is marginal. If anything, it seems that higher rewards induce higher effort (Wilcoxon Signed-Ranks test, $Z = 0.789$, $p = 0.430$, two-sided).

Summarizing, the results of Experiment 1 provide clear evidence in support for incentive reversal. Accordingly, 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 indeed 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

To complement and check for the robustness of the findings of Experiment 1, we ran an additional experiment which again featured a sequential team production problem. The new experiment introduced several innovations compared to Experiment 1. We conducted the experiment in a classroom environment, in which it was known that all subjects are from the same class and are likely to know each other, although the identity of the specific team members of each subject was kept unknown. We employed the strategy method in order to obtain counterfactual data, enabling us to carry out a direct and clean within-subject analysis of reciprocal attitudes. The decision

was one-shot. Incentive-level treatments were manipulated between subject groups (i.e., between classrooms), allowing for rigorous analyses at the team level. The game was framed as a simple monetary game, for which the rules were provided in the instructions. Contrary to Experiment 1, we did not explicitly use a specific game form. Instead, subjects in Experiment 2 had to extrapolate the game form from the instructions (if they desired to do so). Thus, applying the model to more than two agents without providing the subjects the exact game form enables us to study whether incentive reversals arises in more complex social interactions where higher levels of reasoning are required. Furthermore, the treatment manipulation is on effort costs, while the reward schemes are constant across treatments. Thus, incentive reversal is manifested in higher efforts when the costs change from low to high (in contrast to Experiment 1, where rewards were manipulated, and thus incentive reversal resulted in higher effort when rewards were lower).

### 4.1 Experimental Design and Procedure

Each game consists of a team of $n$=3 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 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:[15]

$$\pi_i = \frac{E}{n} \cdot 2^{\sum_{k=1}^{n} e_k} - c_i e_i \tag{3}$$

[15]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-prediction of the game.

Table 3: Experiment 2 – Treatments and Equilibrium Predictions

| Costs of doubling endowment for agent i: | Low costs | High costs |
|---|---|---|
| $c_1$ | 55 | 60 |
| $c_2$ | 50 | 55 |
| $c_3$ | 5 | 25 |
| | | |
| Equilibrium strategies | (ND,ND,D) | (D,D,D) |
| Equilibrium Payoffs | (20,20,15) | (20,25,55) |

Note: The equilibrium strategies and payoffs relate to Nash equilibrium in the simultaneous game and subgame perfect equilibrium in the sequential game. ND refers to the strategy of choosing not to double the endowment whereas D refers to the strategy of choosing to double the endowment.

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 players 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 this 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 rewards are increased) leads to a reduction in overall efforts. Table 3 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 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 money-making experiment, to which most of the students responded positively (78 out of approximately 90). Once only those students who volunteered to participate in the exper-

iment 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.[16] 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 of each role, making seven decisions in total. Once all forms were collected, the payoffs were calculated in the following way: The participants 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. Payoffs 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).

## 4.2 Results

Table 4 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, first movers should shirk under low costs, but provide effort under high costs. In support of the theoretical predictions, 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-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

[16] An English translation of the instructions appears in the appendix. The original instructions in Hebrew are available from the authors upon request. Out of the 78 participants, 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 4: Experiment 2 – Description of Subjects' Chosen Strategies

| | Low costs | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Number of subjects: | 38 | | | | | | | |
| Percent of Agents 1 | D 23.7 | | | | ND 76.3 | | | |
| Percent of Agents 2 | D 73.7 | | ND 26.3 | | D 10.5 | | ND 89.5 | |
| Percent of Agents 3 | D 100.0 | ND 0.0 | D 97.4 | ND 2.6 | D 97.4 | ND 2.6 | D 89.5 | ND 10.5 |
| | **High costs** | | | | | | | |
| Number of subjects: | 37 | | | | | | | |
| Percent of Agents 1 | D 54.1 | | | | ND 45.9 | | | |
| Percent of Agents 2 | D 81.1 | | ND 18.9 | | D 13.5 | | ND 86.5 | |
| Percent of Agents 3 | D 94.6 | ND 5.4 | D 24.3 | ND 75.7 | D 27.0 | ND 73.0 | D 2.7 | ND 97.3 |

Note: D represents the decision to double the endowment and ND represents the decision not to double the endowment.

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 incentives 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 but rather at the expected realizations, i.e., the decisions weighted by the corresponding observed distribution of previous movers' decisions.[17] Table 5 reports the expected number of subjects choosing to exert effort, as well as the expected costs and productivity for each treatment.

[17]For example, in Table 4 we see that under low costs, 23.7 percent of player 1 decide to exert effort, and 73.7 percent of player 2 state that they want to exert effort if player 1 does, and 100 percent of player 3 would want to exert effort if both the previous players 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 5 (first column, fourth row). All the other values in Table 5 are derived analogously.

Table 5: Experiment 2 – Expected Distribution of Decisions to Double the Endowment, Costs, and Payoffs

| Percent of Teams that choose to double | Low costs | High costs |
|---|---|---|
| 0 | 7.2 | 38.6 |
| 1 | 61.5 | 13.3 |
| 2 | 13.9 | 6.5 |
| 3 | 17.5 | 41.5 |
| Expected number | 1.42 | 1.51 |
| | | |
| Expected Team Cost (NIS) | 30.4 | 83.1 |
| Expected Team Productivity (NIS) | 97.6 | 127.0 |
| Expected Team Payoff (NIS) | 67.2 | 43.9 |

Note: The ex-post probabilities reported in the table reflect the doubling proportions weighted by the corresponding observed distribution of previous movers decisions. The productivity is the expected final endowment in NIS, before deducting doubling costs.

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 strategy 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 times).

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 substantial decrease in the associated costs of production causes a substantial decrease in units produced, a counterintuitive result caused by incentive reversal. As a result, the principal receives less output but agents' payoffs increase.

### 4.3 Discussion

The second experiment provides a more comprehensive view of the incentive reversal phenomenon, as testing the model in small natural groups provides an appropriate environment to potentially observe social behavior. In addition, a game with three agents provides more situations in which reciprocity is not dictated by monetary incentives.[18] Furthermore, using the strategy method enables us to 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 mover 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 high costs, 10.5 percent of the third movers shirk if both the previous movers shirked as well. The case where the reciprocal effect is most pronounced is under low costs when 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.

Taken together, we find some evidence for reciprocal behavior in the second experiment. Nevertheless, incentive reversal occurs even in this strong social context. The subjects are not experienced participants recruited from an existing pool of volunteers, are not used to making money in experiments, and did not expect to participate in this experiment in advance. Hence, the created environment implies that monetary oriented motivations are rela-

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

tively low. On the other hand, the subjects know that their partners in the experiment are recruited from among their classmates, implying stronger social preferences than we would expect in a laboratory setting. The circumstance that the observed social effects are not very strong in our data hints at the relative importance of reciprocity in this setup. Since some people might argue that this setup resembles more an actual work environment, it further underlines the strength and relevance of incentive reversal. An increase in 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. Importantly, we added to the related literature by designing an experiment where incentive reversal arises when agents are fully rational (monetary maximizers) and in the absence of income effects. 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) exerting effort becomes a dominant strategy for late movers. Therefore, they cannot condition their effort on first movers' actions. As a consequence, first movers shirk and free ride on the effort of late movers.

We believe that the findings reported here are not only of interest for theorists, but also for practitioners. They underline that the introduction of (additional) incentives, which maybe was well-intentioned in the beginning, can 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. As our results suggest, such actions which are meant to motivate workers can 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 consider whether it should be taken into account in specific situations.

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. Ichino and Maggi, 2000; Fehr and Fischbacher, 2003; Falk and Ichino, 2006; Gould and Winter, 2009; 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 and in particular the presence of second degree incentives, suggest that high power monetary incentives may be counter effective as they may destroy this implicit threat. This form of behavioral incentive reversal, which shares the very same logic of our fully rational incentive reversal is applicable to almost any team environment without relying on complementarity among agents. Therefore, it is important to take it into account when designing incentive schemes for teams.

## References


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

Benabou, R. and Tirole, J. (2006). Belief in a just world and redistributive politics, *Quarterly Journal of Economics* **121**(2): 699–746.

Binmore, K., McCarthy, J., Ponti, G., Samuelson, L. and Shaked, A. (2002). A backward induction experiment, *Journal of Economic Theory* **104**(1): 48–88.

Bone, J., Hey, J. D. and Suckling, J. (2009). Do people plan?, *Experimental Economics* **12**(1): 12–25.

Bowles, S. (2009). When economic incentives backfire, *Harvard Business Review* .

Camerer, C., Babcock, L., Loewenstein, G. and Thaler, R. (1997). Labor supply of New York City cabdrivers: One day at a time, *The Quarterly Journal of Economics* **112**(2): 407–441.

Carpenter, J. P. (2003). Bargaining outcomes as the result of coordinated expectations, *Journal of Conflict Resolution* **47**(2): 119.

Clark, K. and Sefton, M. (2001). The sequential prisoner's dilemma: evidence on reciprocation, *The Economic Journal* **111**(468): 51–68.

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

Falk, A. and Fischbacher, U. (2002). "Crime" in the lab-detecting social interaction, *European Economic Review* **46**(4-5): 859–869.

Falk, A. and Ichino, A. (2006). Clean evidence on peer effects, *Journal of Labor Economics* **24**(1): 39–57.

Farber, H. S. (2008). Reference-dependent preferences and labor supply: The case of New York City taxi drivers, *American Economic Review* **98**(3): 1069–1082.

Fehr, E. and Fischbacher, U. (2003). The nature of human altruism, *Nature* **425**(6960): 785–791.

Fehr, E. and List, J. A. (2004). The hidden costs and returns of incentives-trust and trustworthiness among CEOs, *Journal of the European Economic Association* **2**(5): 743–771.

Fehr, E. and Schmidt, K. M. (2004). Fairness and incentives in a multi-task principal–agent model, *Scandinavian Journal of Economics* **106**(3): 453–474.

Fischbacher, U. and Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments, *American Economic Review* **100**(1): 541–556.

Fischbacher, U., Gächter, S. and Fehr, E. (2001). Are people conditionally cooperative? evidence from a public goods experiment, *Economics Letters* **71**(3): 397–404.

Frey, B. S. and Jegen, R. (2001). Motivation crowding theory, *Journal of Economic Surveys* **15**(5): 589–611.

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

Gneezy, U. and Rustichini, A. (2000b). Pay enough or don't pay at all, *Quarterly Journal of Economics* **115**(3): 791–810.

Goerg, S., Kube, S. and Zultan, R. (2010). Treating equals unequally: Incentives in teams, workers' motivation, and production technology, *Journal of Labor Economics* **28**(4): 747–772.

Gould, E. D. and Winter, E. (2009). Interactions between workers and the technology of production: evidence from professional baseball, *The Review of Economics and Statistics* **91**(1): 188–200.

Guttman, J. M. (1986). Matching behavior and collective action: Some experimental evidence, *Journal of Economic Behavior & Organization* **7**(2): 171–198.

Harrison, G. W. and McCabe, K. A. (1996). Expectations and fairness in a simple bargaining experiment, *International Journal of Game Theory* **25**(3): 303–327.

Houser, D., Xiao, E., McCabe, K. A. and Smith, V. (2008). When punishment fails: Research on sanctions, intentions and non-cooperation, *Games and Economic Behavior* **62**(2): 509–532.

Ichino, A. and Maggi, G. (2000). Work environment and individual background: Explaining regional shirking differentials in a large italian firm, *Quarterly Journal of Economics* **115**(3): 1057–1090.

Johnson, E. J., Camerer, C., Sen, S. and Rymon, T. (2002). Detecting failures of backward induction: Monitoring information search in sequential bargaining, *Journal of Economic Theory* **104**(1): 16–47.

Mas, A. and Moretti, E. (2009). Peers at work, *American Economic Review* **99**(1): 112–145.

Selten, R. (1967). Die strategiemethode zur erforschung des eingeschrnkt rationalen verhaltens im rahmen eines oligopolexperimentes, *in* H. Sauermann (ed.), *Beitrge zur experimentellen Wirtschaftsforschung*, Mohr Siebeck, Tbingen, pp. 136–168.

Winter, E. (2009). Incentive reversal, *American Economic Journal: Microeconomics* **1**(2): 133–147.

Winter, E. (2010). Transparency and incentives among peers, *Rand Journal of Economics* **41**(3): 504–523.

# Appendix A: Instructions 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.

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.

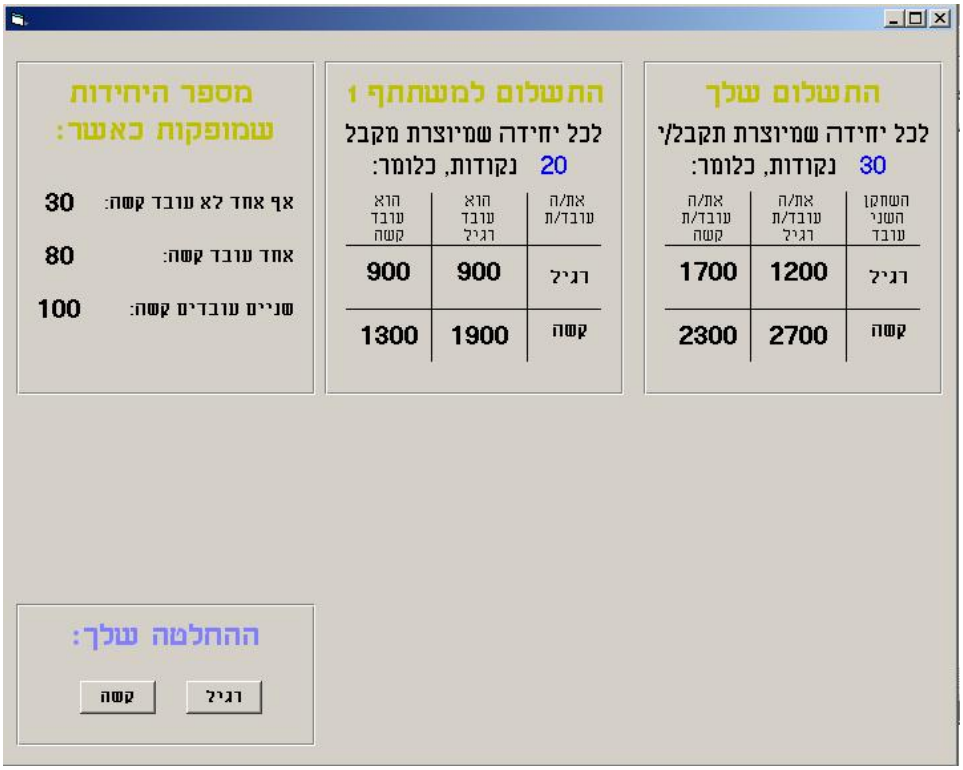
The buttons for making your decision appear at the bottom left of the screen.

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.




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 _____.