Centre de Referència en Economia Analítica

Barcelona Economics Working Paper Series

Working Paper nº 115

A Change Would Do You Good...
An Experimental Study on How to Overcome Coordination Failure in Organizations

Jordi Brandts and David J. Cooper

Barcelona Economics WP nº 115

March, 2004

A Change Would Do You Good . . .
An Experimental Study on How to Overcome Coordination Failure in Organizations

JORDI BRANDTS* and DAVID J. COOPER**

March 10, 2004

* Institut d'Anàlisi Econòmica (CSIC)

** Case Western Reserve University

Abstract: Many organizations suffer poor performance because individuals within the organization fail to coordinate on efficient patterns of behavior. Using controlled laboratory experiments, we study how financial incentives can be used to find a way out of such performance traps. Our experiments are set in a corporate environment where subjects' payoffs depend on coordinating at high effort levels; the underlying game being played repeatedly by employees is a weak-link game. In an initial phase, the benefits of coordination are low relative to the cost of increased effort. Play in this initial phase typically converges to an inefficient outcome with employees failing to coordinate at high effort levels. The experimental design then explores the effects of varying the financial incentives to coordinate at a higher effort level. We find that an increase in the benefits of coordination leads to improved coordination, but, surprisingly, large increases have no more impact than small increases. Once subjects have coordinated on a higher effort level, reductions in the financial incentives to coordinate have little effect on behavior. Hence, a "shock therapy" of temporary increases in incentives to coordinate can lead to permanent improvements in an organization's performance.

Keywords: Incentives, Coordination, Experiments, Organizations

JEL Classification Codes: C92, D23, J31, L23, M52

Acknowledgements: The authors thank the NSF (SES-0214310), the Spanish *Ministerio de Educación y Cultura* (SEC2002-01352) and the Barcelona Economics Program of CREA for financial help, Bethia Cullis, Adam Malinowski, and David Rodríguez for skillful research assistance, and Eric Bettinger, Colin Camerer, Vince Crawford, Charlie Holt, Jim Rebitzer, Mari Rege and seminar participants at the European ESA meetings and the Public Choice Society meetings for useful comments. We are grateful to Colin Camerer, Teck Ho, and Juin Kuan Ko who shared their software for fitting EWA with us.

Authors

Jordi Brandts	David J. Cooper
Institut d'Anàlisi Econòmica (CSIC) Campus UAB 08193 Bellaterra (Barcelona) Spain	Department of Economics Weatherhead School of Management Case Western Reserve University 10900 Euclid Avenue Cleveland, OH 44106-7206 USA
Phone: 34-935806612 Fax: 34-935801452 jordi.brandts@uab.es	Phone: 1-216-3684294 Fax: 1-216-3685039 djc13@weatherhead.cwru.edu

1. Introduction

Coordination failure can cause corporations and other organizations to get trapped in situations that are unsatisfactory for all involved even though preferable outcomes are possible and would be stable if ever reached. Even if the benefits of improved coordination are obvious, any process designed to bring about a change for the better faces substantial obstacles. As an archetypical example, imagine a firm producing via an assembly line where the slowest worker determines the speed of the entire line. All the workers are exerting minimal effort, but could be better off if all tried harder and the line became more productive. However, any one worker who unilaterally begins to work harder wastes his effort if slow work persists elsewhere. Only if our hypothetical worker is reasonably certain that others will also be working harder should he be willing to increase his effort. Thus, overcoming coordination failure is a question of coordinated change. Achieving the necessary coordinated change may be particularly difficult if communication among individuals is minimal and the inefficient situation has persisted for some time. Thinking again of our hypothetical worker, imagine how much more difficult it may be to convince him that others will be working harder if he has no means of talking with the other workers and has observed a long history of laggardly behavior. In this paper we study controlled laboratory experiments that simulate difficult environments of this sort. Our goal is to explore how a history of coordination failure can best be overcome using financial incentives.

The answer to this question is of real importance since coordination failure as described above plays a central role in a number of important economic settings. The research presented here is primarily motivated by the problems presented by turning around a failing corporation. Organizational change has been a topic of interest for scholars in economics and management for a long time.[1] One of the insights emerging from this literature is that the presence of complementarities may be at the root of many organizational problems. For example, consider Knez and Simester's (2002) case study about the successful turnaround of Continental Airlines in the mid 1990s. The critical element in Continental's success was the introduction of an incentive program designed to improve on-time arrival, a key determinant of airline profitability. Knez and Semester stress the importance of interdependencies among autonomous groups of employees in determining on-time arrival: "When a flight departs late, gates, employees and

[1] Foster and Ketchen (1998), Weick and Quinn (1999), and Pettigrew, Woodman and Cameron (2001) present surveys of recent work in the organizational behavior and strategy literatures on change.

equipment are unavailable to service other flights arriving and departing from the same airport. The problem is further compounded when flights carry connecting passengers since departing flights may have to be delayed to allow passengers to make their connections." They posit that the global nature of Continental's incentive plan played a central role in its success, assuring employees that their increased effort would be matched by colleagues in other units. In other words, coordinated change was necessary to improve Continental's situation. As another example, Ichniowski, Shaw, and Prennushi (1997) find similar results in a study of productivity in steel plants. The type of steel production they study takes place in an assembly line setting with productivity largely determined by unscheduled downtime. This implies that one employee who is doing a poor job (leading to breakdowns on his part of the assembly line) can largely destroy the efficiency of the entire line. Improving performance at one point in this production process will do little good if performance lags elsewhere.

Similar issues play an important role in other areas of economics, especially so in development economics. An idea going back to Rosenstein-Rodan (1943) and Hirschman (1958) is that underdevelopment can be seen as a large-scale coordination problem. For instance, countries may fail to develop when the simultaneous industrialization of many sectors of an economy can be profitable for each of them but no sector can break-even industrializing alone.[2] The question in this context is what the government can do to produce a "big push" which takes an economy from an underdeveloped state to one of greater prosperity. In our terms, the hope is to produce coordinated change.

The Continental example suggests that the use of financial incentives can play an important role in overcoming coordination failure. While the ultimate goal is to overcome coordination failure in field settings, laboratory experiments can play an important, complementary role in understanding how changes in incentives can lead to improved coordination. Taking advantage of the controlled nature of laboratory experiments, we can introduce exogenous variation in both the timing and intensity of an incentive scheme without altering any other features of the decision making environment. Laboratory experiments also make it possible to generate multiple observations with the same time path of incentives, allowing us to separate systematic effects of the incentives from peculiarities of time or place.

[2] Murphy, Shleifer and Vishny (1989) and Ciccone and Matsuyama (1996) present specific models of economies with these features.

We therefore feel that the controlled laboratory experiments presented below give us a valuable new view on the role on incentives in reversing coordination failures.

We study an experimental environment which we label the "corporate turnaround game." These experiments are designed to simulate a corporate environment in which coordination failure has occurred. We then systematically study the effects of changing incentives. More specifically, the corporate turnaround game involves repeated play of a game between four "employees" of a "firm". The productivity and profitability of the firm is determined by the minimum effort level chosen by its four employees, a very strong form of complementarity. The key variable in the experiments is a bonus rate set by an exogenous manager. This bonus rate determines the fraction of the firm's profits transferred to the employees and hence governs the benefit to the four employees of coordinating at a high effort level. Once the bonus rate is known, the game played by the employees for any one round of the corporate turnaround game is a weak-link game. Initially, employees face a low bonus rate and, by extension, low incentives to coordinate. Under these circumstances the experimental "firms" typically become coordinated at the least efficient possible outcome. We then study how changes in the bonus rate can lead to a change for the better. Our design focuses on four questions: (1) Can an increase in the bonus rate enable the firm to overcome its coordination failure? (2) Does the magnitude of the bonus rate increase matter or is the simple fact of an increase effective as such? (3) If an increase in the bonus rate brings about improved coordination, can the bonus rate increase be revoked without affecting the improved outcome? (4) Does the length of time a firm has been underperforming affect the impact of the increase?

The experimental data yields some unexpected answers to these questions. As most economists would expect, increasing the bonus rate leads to an improvement in coordination among employees. Surprisingly, the magnitude of the bonus rate increase doesn't matter as large changes in the bonus rate lead to no greater improvement in coordination than smaller increases. In cases where a bonus increase leads to improved coordination, the bonus can subsequently be reduced without a significant impact on behavior. In other words, the temporary application of positive incentives – "shock therapy" – can have a permanent impact. The experimental data exhibit oddly asymmetric history dependence – it is relatively easy to overcome a history of coordination failure by strengthening incentives to coordinate, but reducing incentives to coordinate will generally not overturn a history of successful coordination. Finally, we find that the length of time a firm has experienced coordination failure weakens but does not eliminate the impact of a bonus rate increase.

To understand these results, we need to correctly formulate the coordination problem facing subjects. It may be tempting to focus on the coordination problem within the framework of a single round and a single play of the weak-link game. However, the "corporate turnaround game" is a repeated game with many rounds and changing bonus rates. In an inherently dynamic environment like the turnaround game, no single round can be considered in isolation. We propose that the central problem facing employees isn't *whether* to coordinate but rather *how and when* to coordinate. The weak-link game isn't a particularly complicated game to understand. After a few periods of play, undoubtedly most subjects realize that all would be better off if they could coordinate at the highest effort level. However, one employee acting alone stands little chance of success. Breaking the trap of coordination failure requires a coordinated move to higher effort levels. An increase in the bonus rate provides the necessary coordinating device. If we look at the individual level data, virtually all employees respond to an increase in the bonus rate by bumping up their effort levels regardless of the size of the increase. If enough employees make a large increase to their effort level, other employees will generally follow their lead and the firm will overcome its history of coordination failure. The bonus increase is more important as a trigger starting the process of change rather than as a determinant of the eventual quantitative outcome of this process.

In an attempt to characterize behavior in the turnaround game, we turn to several recent game theoretic models that incorporate elements of bounded rationality: quantile response equilibrium (McKelvey and Palfrey, 1995), experience weighted attraction learning (Camerer and Ho, 1999), and experience weighted attraction learning with forward-looking types (Camerer, Ho, and Chong, 2002). Although these models have successfully tracked data from other experiments, none prove to be capable of tracking the main qualitative features of our experimental data, particularly the strong positive response to increases in the bonus rate. This failure illustrates the dynamic nature of play in the turnaround game. Even in the learning model with forward looking types, firms are doomed to coordination failure by players who focus on single plays of the weak-link game rather than considering the game as a whole. Simulated employees who initially don't respond to an increase in the bonus rate also tend to choose low effort levels in later rounds, failing to recognize that an increase in effort, while costly in the short run, can lead to better outcomes in the long run. These laggards pull their firms back to coordination failure. In contrast, employees in the experimental data who do not initially respond to an increase in bonus rate often increase their effort levels strongly in subsequent rounds. These subjects may not start the process of change, but they provide the momentum to keep it going. This contagious element of overcoming coordination failure suggests that many subjects who fail to initially think in terms of re-coordinating at a better outcome quickly realize the benefits of this course of action. Unlike the simululated employees, experimental subjects who do not initially act in a forward looking manner quickly learn to do so.

Translating the experimental results to field settings, we can make several managerial/policy recommendations on how to overcome coordination failure. All of these recommendations are driven by a single insight; the primary role of increased incentives is to serve as a coordinating device for change. Given that increasing incentives for coordination is costly, our results suggest that large increases in incentives are not needed. The size of the increase is probably less important than the knowledge that all are receiving increased incentives. Explicitly stating that the increased incentives are intended to foster additional coordination is also likely to be helpful, emphasizing the use of increased incentives as a coordinating device. Likewise, our results indicate that it is not necessary to maintain an expensive incentive scheme indefinitely. Once the process of change has taken place, coordination will tend survive a reduction in incentives. Encouragingly, our results suggest that while sooner is better in trying to overcome coordination failure, all is not lost if the imposition on an incentive scheme must be delayed. Finally, our results indicate that while even small changes in incentives overcome coordination failure, a mere restart of the game does not. This indicates there are limits on our ability to overcome coordination failure – it is not such a fragile phenomenon that any intervention will be successful.

Section 2 presents in detail the turnaround game that we use in the experiments and relates the results presented here to previous work with weak-link games. Section 3 outlines the experimental design and our preliminary hypotheses about the data. Section 4 presents the procedures. Section 5 summarizes the experimental results and relates them to our preliminary hypotheses. Section 6 explores the ability of several game theoretic models to capture the main qualitative features in the data. In section 7 we present some final remarks.

2. The Corporate Turnaround Game

The "corporate turnaround game" played by subjects in our experiments can be thought of as presenting a stylized model of a firm. This firm consists of a number of employees who choose among different effort levels. The firm's overall productivity (as well as profitability) is determined by the effort of its employees. To incentivize these employees, the firm uses a bonus system that transfers some of the firm's profits back to employees. In this stripped down environment, altering the bonus structure is the sole tool managers have available to "turn around" an underperforming firm.

The corporate turnaround game embodies three basic design choices or assumptions. First, the firm's technology has a weak-link structure, with productivity depending on the minimum effort chosen by an employee. Second, the firm manager only observes the minimum effort selected, since it determines the output, but employees can observe all effort levels selected. Third, the firm manager rewards employees with bonuses based on the minimum effort observed and is able to change the bonus rate but cannot otherwise influence the employees' choices. In what follows we first discuss the rationale behind these three design choices in more depth and then introduce the details of the turnaround game.

In an organization with a "weak-link" structure the individual (or unit) doing the worst job, either due to a lack of native skill or low effort, determines the overall productivity of an organization. Kremer (1993, p. 551) describes the problem nicely:

> "Many production processes consist of a series of tasks, mistakes in any of which can dramatically reduce the product's value. The space shuttle *Challenger* has thousands of components: it exploded because it was launched at a temperature that caused one of these components, the O-rings, to malfunction. "Irregular" garments with slight imperfections sell at half price. Companies can fail due to bad marketing, even if the product design, manufacturing, and accounting are excellent."

Starting from this observation, Knez and Camerer (1994) argue that the game played within many firms takes on the form of the "minimum game" introduced by Van Huyck, Battalio, and Beil (1990). Knez and Camerer refer to these games by the more evocative term "weak-link games," a terminology that we follow. These are a type of coordination game. Each player simultaneously chooses a strategy that can be thought of as an effort level. Each player's payoff is a decreasing function of his own effort and an increasing function of the minimum effort chosen by the players in the group. Payoffs are set up so that it is worthwhile for a player to raise his effort level if *and only if* it will increase the minimum effort for the group. Coordinating on any of the available effort levels is a Nash equilibrium for a weak-link game, but results from earlier experiments suggest that play typically evolves towards the payoff dominated equilibrium where all players choose the lowest possible effort level.[3] Having reached this worst of all possible outcomes, improvement can only occur if all players increase their effort levels.

By studying a production technology with a weak-link structure, we focus on a worst-case scenario. Presumably many organizations face coordination problems in more forgiving settings where a change for the better is easily achieved. However, if we can understand how to overcome coordination failure in organizations with a weak-link structure, a tough environment, it should be even easier to accomplish in less difficult circumstances.

The second critical assumption of our experimental environment is that while employees can observe the effort levels of all other employees, the firm manager can only observe the minimum effort chosen. By extension, the manager lacks the necessary information to tailor bonuses to the effort put forth by individuals and can only offer bonuses based on the minimum effort over all employees. In other words, limiting the information available to the firm manager has the effect of limiting the tools available for overcoming coordination failure.

There is no particular reason to believe that our game is either more or less realistic for restricting the information about employee effort available to managers. We decided to limit the information available to managers for two reasons. Most importantly, the goal is to study a setting where overcoming coordination failure is difficult. By limiting the instruments available to change employees' behavior, we make it tougher to turn around a failing firm. Presumably the lessons learned from such a harsh environment will also be valuable in more forgiving

---

[3] For example, Van Huyck *et al* (1990) ran 7 sessions of a minimum game with 14 - 16 inexperienced subjects. By the end of the 10th round, the minimum strategy chosen for all 7 sessions was the lowest possible choice. In the 10th round, 72% of the subjects chose the lowest possible choice. Knez and Camerer (1994 and 2000) report similar results. These results are sensitive to the group size. For example, with two player groups the Pareto dominant outcome, coordination on the highest effort level, is almost always observed after 10 rounds. See also the high coordination rates in three-player groups reported in Weber, Camerer and Knez (2004).

settings. As an additional benefit, limiting the manager's information greatly simplifies the environment (and even the environment studied here presents substantial complexities). Although environments where the manager can observe all employees' choices raise interesting issues – a major theme in the literature discussed above is the relative efficacy of global incentive schemes versus more targeted plans – we feel that understanding the simpler environment studied here is a necessary first step in understanding more complicated settings.

Our third crucial assumption is that the *only* instrument of change controlled by managers is the ability to change a bonus rate based on the minimum effort of employees. In reality, financial incentives are just one of many tools available to overcome coordination failure: better communication, building trust, etc. However, our goal is to study a simple environment where only one variable changes. Financial incentives are an apt choice as a tool for change, since they can easily and rapidly be altered by managers. From our point of view, financial incentives also have the advantage of being well suited to controlled experiments and being easily modeled using game theoretic tools. Only after understanding how financial incentives in isolation can lead to turnarounds can we begin to study the interactions between changes in financial incentives and other possible interventions.[4]

Turning to the specifics of the turnaround game, the players in our turnaround game are the manager and four employees of a firm. For all the experimental sessions reported below, the experimenter plays the role of the firm manager while subjects fill the roles of the four employees.[5] Even though the manager's choices are exogenous, for expositional purposes it is useful to treat the manager as a player in the game. The game starts with the firm manager setting a flat wage (W) that each employee receives regardless of the outcome and a bonus rate (B) that determines how much additional pay each employee receives for each additional unit increase in the minimum effort. All four employees observe W and B and then simultaneously choose effort levels, where E_i is the effort level chosen by the i^{th} employee. Effort levels must be non-negative. The payoffs given by Equations 1 and 2, where K_F, v_F, and c_E are positive constants.

[4] In Brandts and Cooper (2004b) we study the impact of allowing managers to communicate with employees.

[5] Our primary goal is to study the role of changing financial incentives in overcoming a history of coordination failure. Making the manager exogenous allowed us to control how the bonus rate changed over time rather than being dependent on random variation in the bonus rates set by subjects acting as managers. What bonus rates would be set by subject-managers and how employees' responses are affected by the use of a subject as the manager are important but separate questions which we examine in a companion paper (Brandts and Cooper, 2004b).

$$\text{Firm: } \pi_F = K_F + \left((v_F - 4B) \times \min_{i \in \{1,2,3,4\}} \left(E_i \right) \right) - 4W \quad \textbf{(eq. 1)}$$

$$\text{Employee i: } \pi_e^i = W - c_E E_i + \left(B \times \min_{j \in \{1,2,3,4\}} \left(E_j \right) \right) \quad \textbf{(eq. 2)}$$

The firm's profits depend on the minimum effort contributed by its employees, consistent with our assumption that the firm's production technology has the weak-link property. We only allow the firm manager to award bonuses based on the minimum effort, as implied by the asumption that the manager cannot observe individual efforts. As can be seen in Equation 1, the bonus transfers a portion of the firm's profits to its employees. This illustrates why we are interested in the ability of temporary increases in the bonus rate to permanently increase employees' efforts – high bonuses may be effective at increasing the firm's revenues, but this move will be self-defeating if these increased revenues accrue largely to the employees as increased bonuses.

All payoffs are denominated in "experimental pesetas" which were converted to monetary payoffs at a rate of 1 dollar or 1 euro equals 500 experimental pesetas. For all of the experiments reported below, we set $K_F = 900$, $v_F = 60$, $c_E = 5$, and $W = 200$. Fixing these variables lets us focus on changes in the bonus rate, B. We rewrite the payoff functions as follows:

$$\text{Firm: } \pi_F = 100 + \left((60 - 4B) \times \min_{i \in \{1,2,3,4\}} \left(E_i \right) \right) \quad \textbf{(eq. 3)}$$

$$\text{Employee i: } \pi_e^i = 200 - 5E_i + \left(B \times \min_{j \in \{1,2,3,4\}} \left(E_j \right) \right) \quad \textbf{(eq. 4)}$$

We restrict an employee's effort to be in ten hour increments: $E_i \in \{0,10, 20, 30, 40\}$. Intuitively, employees spend forty hours per week on the job, and effort measures the number of these hours that they actually work hard rather than loafing.[6]

Given that the manager is exogenous in the experiments reported below, we now focus on the proper subgame where employees choose effort levels. For all values of the bonus rate, B, used in our experiments the resulting game is a weak-link game. Coordinating on any of the five available effort levels is a Nash equilibrium.

(Insert Table 1 here)

[6] In the experiments, terms with strong connotations such as "effort" and "loafing" were not used. Employees were told that they spend 40 hours per week on the job, and that these hours could either be allocated to Task A or Task B. Task A plays the role of effort and Task B plays the role of loafing.

To understand why overcoming coordination failure is so difficult in this environment, consider the game induced by a bonus value of B = 6 (see Table 1). Suppose that the employees have previously all chosen effort level 0. The likelihood of spontaneously overcoming this coordination failure is low since the incentives to increase effort beyond the minimum are quite weak and the risks are high. Consider an employee who thinks about raising his effort from 0 to 10. He knows that his payoff will certainly be reduced 50 pesetas due to increased effort. If all the other subjects follow his lead, his total gain is only 10 pesetas beyond the 200 pesetas he gets without risk by choosing 0. For the proposed increase to have a positive expected profit, the employee must believe the probability of all three other employees raising their efforts from 0 to 10 equals 5/6. Treating the other three employees as statistically independent, this translates into requiring a 94% chance of increased effort for each of the other three employees.[7] In other words, our fictitious employee must be almost certain that the other employees will increase their efforts for such an increase to be worthwhile for him. Given these grim incentives, overcoming coordination failure is highly unlikely.

Now imagine that a new manager takes over the firm. Determined to shake the firm out of its underperforming ways, he decides to raise the bonus rate to B = 14. (Looking at Equation 3, this is the highest bonus rate at which the firm earns a profit.) This yields the payoff table shown in Table 2.

(Insert Table 2 here)

The incentives to increase effort are now much stronger for the employees. Once again, suppose we start with all four employees choosing effort level 0. Consider again an employee who is thinking of increasing his effort from 0 to 10. While the certain losses remain 50 pesetas, the potential gain is now 90 pesetas. The probability that all three other employees will increase their efforts required to make this change break even is now only 5/11. Assuming the other three employees are independent, this translates into requiring a 76% chance that each employee increases his effort. While still daunting, these are better odds than we saw with B = 6. One can imagine employees at least attempting to overcome coordination.

The preceding examples illustrate the central features of our experimental design. Our focus is *not* on comparative static results. It is quite possible that subjects with no previous

[7] To derive this probability, solve for p such that $200 = 150*(1-p^3) + 210*p^3$. Given the linear payoff structure the same trade-off arises for one-step increases of effort starting at a level higher than 0, as well as for two or more step increases when feasible.

experience will generally converge to a more efficient equilibrum in the turnaround game with B = 14 than in the game with B = 6.[8] This, however, isn't our point. Instead, we want to know what happens for players who have already experienced a history of coordination failure. With low benefits for coordination, as with B = 6, overcoming coordination failure seems hopeless. Greater benefits for coordination, such as B = 14, make the situation less hopeless. However, even with this improvement in incentives it seems unlikely that any one employee acting unilaterally can overcome a history of coordination failure.

3. Experimental Design and Hypotheses

Subjects played thirty rounds in fixed groups ("firms") of four subjects ("employees"). During a session the bonus rate changed in a predetermined way. Other than the bonus rate, no detail of the experimental environment was varied between rounds or between sessions. The bonus rate and resulting payoff-matrices were announced at the beginning of each of the three ten-round blocks and were fixed during that time frame. While playing in a block with a particular bonus rate, subjects did not know what the bonus rate would be in subsequent ten-round blocks.

The bonus rate was always fixed at B = 6 for the first ten-round block. This is the lowest (integer) bonus rate that does not make choosing positive effort levels a dominated strategy. The goal was to get a high percentage of firms coordinated on the inefficient outcome with minimum effort equal to zero.

(Insert Table 3 here)

The treatments vary the bonus rates for the second and third blocks of ten rounds. The intent is to explore how financial incentives can best be used as a tool to overcome coordination failure. The experimental design, as summarized in Table 3, focuses on the four questions presented in the introduction: (1) can firms be extricated from an initial bad outcome by increasing the bonus rate, (2) does the size of the increase matter, (3) does the bonus increase need to be permanent, and (4) does delaying the bonus increase reduce its efficacy?

Both as a tool for explaining the rationale underlying the experimental design and as a means of organizing the results, we now present four *ex ante* hypotheses about the experimental

[8] Results from Brandts and Cooper (2004c) support this hypothesis.

data and relate each of these hypotheses to the purposes of the experimental design. We justify these hypotheses on an intuitive basis – see Section 6 for a discussion of the predictions made by an assortment of game theoretic models.

The first hypothesis addresses a necessary condition for the experimental design to be of interest: the first block of ten rounds must generally lead to an inefficient outcome. If no coordination failure occurs, it would be difficult to study overcoming coordination failure. Based on the results of earlier experiments with weak-link games (Van Huyck et al, 1990; Knez and Camerer, 1994 and 2000), it would be surprising if Hypothesis 1 did not hold.

HYPOTHESIS 1: In rounds 1-10 average minimum effort will be close to zero.

Hypothesis 2 focuses on choices in the second block of ten rounds. Assuming Hypothesis 1 is correct, many of the firms will have converged to an effort level of zero. We now increase the bonus rate in Cells 1 – 5 to see if the firms can be extricated from this inefficient outcome. We vary the size of the bonus increase between cells to determine if the magnitude of the increase matters. For cells 1, 4, and 5, the bonus rate increases to B = 14 for rounds 11 – 20, the maximum amount at which the firm still earns some profit. Cells 2 and 3 employ smaller increases, B = 10 and B = 8 respectively. Cell 6, with no change in the bonus rate for rounds 11 - 20, acts as a control. This control is needed because one possible cause of changing effort levels following an increase in the bonus rate for Cells 1 – 5 is a pure restart effect. In other words, the act of stopping play and announcing a new bonus rate might cause the effect rather than the change in the bonus rate. Even though the bonus rate isn't changing, play also stops for Cell 6 prior to round 11 and the (unaltered) bonus rate is announced. If no increase in effort levels is observed for rounds 11 – 20 of Cell 6, this eliminates pure restart effects as a cause for changing effort levels in the other cells. Hypothesis 2 presents an optimistic view of how improving the bonus rate affects effort. Intuitively, increasing the bonus rate improves the incentives for coordination. Hypothesis 2 posits that these improved incentives will overcome any hysteresis in the data and that this will not be purely the result of a restart effect.

HYPOTHESIS 2: (a) Increasing the bonus rate in round 11 will cause the average minimum effort to increase. (b) The increase in average minimum effort will be increasing in the new bonus rate. (c) There will be no increase in average minimum effort for rounds 11 – 20 in Cell 6, the controls.

Hypothesis 3 focuses on behavior in rounds 21 – 30. Suppose Hypothesis 2 is borne out by the data. On the one hand, the firm has done well in increasing its productivity. However, much of the resulting increase in revenue may be departing the firm via increased bonuses. Ideally, the firm would like to keep the improved effort levels without having to keep bonuses at a high level. We therefore explore whether the bonus rate can be decreased without sacrificing productivity. We focus on the three cells, Cells 1, 4, and 5, in which the bonus rate was increased to B = 14 for rounds 11 – 20. In Cell 4, the bonus rate is lowered to B = 10 for rounds 21 – 30. Cell 5 is more extreme, returning the bonus rate to its original level, B = 6, for rounds 21 – 30. Cell 1 serves as a control, keeping the bonus rate fixed at B = 14. Hypothesis 3 represents an optimistic view of how subjects will respond to a decrease in the bonus rate. Intuitively, we posit that history dependence will be sufficiently strong to overcome the weakened incentives to coordinate. In some sense Hypotheses 2 and 3 are mutually incompatible, as the former credits incentives with greater strength while the latter bestows this honor on history dependence.

HYPOTHESIS 3: Decreasing the bonus rate in rounds 21 – 30 will not lead to a decrease of average minimum effort to the level of rounds 1 – 10.

Hypothesis 4 studies the impact of delaying the increase in the bonus rate. In Cell 6, the bonus rate remains at B = 6 for rounds 11 – 20 and then increases to B = 14 for rounds 21 – 30. This gives firms an additional ten rounds, as compared with the other five cells, to get coordinated at an inefficient outcome. Hypothesis 2 holds that an increase in incentives to coordinate will overcome any history dependence in the data. It seems reasonable however that history dependence should become stronger the longer subjects have been stuck in a bad outcome. We therefore hypothesize that delaying an increase of the bonus rate will lead to less impact on effort levels. To test this hypothesis we compare rounds 21 – 30 of Cell 6 with rounds 11 – 20 of Cells 1, 4, and 5.

HYPOTHESIS 4: Increasing the bonus will have a smaller effect after twenty rounds with B = 6 than after just ten rounds.

4. Procedures

Sessions were run both at Pompeu Fabra and at Case. In both cases, a computerized lab was used to run the sessions. Each treatment contains data for five firms at each of the two locations, so that the sample is balanced. Our focus here is not on the cross-country comparison. The econometric analysis controls for any location effect, which turns out to be rather secondary.

Subjects were recruited from the undergraduate populations at Pompeu Fabra and Case using newspaper ads, posters, and classroom announcements. Subjects were only allowed to participate in a single session. Due to record keeping errors, two subjects participated in a second session. All data for these two subjects and their firms has been dropped from the dataset.[9]

For the most part, the experimental procedures were quite standard. At the beginning of each session subjects read the instructions directly from their computer screens. Before beginning play, all subjects were asked to complete a short quiz about the payoffs and the rules of the experiment. The full text for the instructions and quiz is given in the appendix.

One slightly unusual feature of the instructions is that rather than using abstract terminology we employ a corporate context. For example, the four players are explicitly referred to as "employees" and are told that they are working for a "firm." We avoided the use of terms with strong connotations. For example, instead of asking subjects to choose a level of "effort" they are asked to allocate time between "Activity A" and "Activity B," with Activity A playing the role of effort. We used a corporate context to make the instructions easier to understand for participants, an important issue for some subject pools used in the broader design.[10] While we doubt that the use of a corporate context causes any demand induced effects, it is possible that the use of meaningful context per se has some impact on subjects' choices.[11]

At the beginning of each ten-round block employees were informed of the bonus rate for that block. Employees were not told what bonus rates would be in subsequent blocks. In each round the four employees of a firm simultaneously chose their effort levels for the round. The screen where employees made this decision displayed the current bonus rate and the formula for the firm manager's payoff. The latter information is irrelevant here, but was displayed to maintain parallelism with sessions where another subject played as the firm manager. The employees were also shown a payoff table, similar to the ones displayed in Tables 1 and 2,

[9] One affected firm was in Cell 2 and the other was in Cell 4.

[10] We have run related experiments using corporate executives (Brandts and Cooper, 2004b). For this population, having a concrete setting for the experiment is important in helping subjects to understand the instructions.

[11] Cooper and Kagel (2003) find that the use of meaningful context can speed the development of strategic play.

showing their payoff as a function of their own effort level and the minimum effort level chosen by the other employees. This payoff table was automatically adjusted to reflect the current bonus rate. At the end of each round, each employee was told their effort level, the minimum effort for their firm, their payoff for the round, and their running total payoff for the experiment. Separate windows on the feedback screen showed them a summary of results from earlier rounds and the effort levels selected for all four employees in their firm. These effort levels were sorted from highest to lowest and did not include any identifying information about which employee was responsible for which effort level. Note that we gave subjects more feedback about the choices of others than is typical; preceding minimum game experiments have only shown subjects the minimum effort level chosen for their group. In a related paper, we show that only giving subjects' information about the minimum effort has little impact on the likelihood of coordination failure emerging initially but substantially reduces the likelihood that a successful turnaround occurs when the bonus rate is increased (Brandts and Cooper, 2004c). While we want to study a challenging environment for overcoming coordination failure, a reasonable chance of success is necessary to generate any interesting results.

Subjects played the game in a fixed cohort of four employees. These cohorts remained constant during the course of the experiment, a fact that was stressed in the instructions. For experimenters who are used to worrying about repeated game effects, this may seem like a strange design choice. However, the field settings that motivated these experiments involve repeated interactions among the same agents. Repeated game effects and strategic teaching are presumably quite natural in these settings. Moreover, the use of fixed groups has the effect of intensifying history dependence, a central issue in our design.[12] As such, we think that our experiments must incorporate repeated interactions to be a useful tool.[13]

At the end of the session, each subject was paid in cash for all rounds played plus a show-up fee. Payoff was done on an individual and private basis. All payoffs are in "experimental pesetas." As mentioned previously, these were converted to dollars at a rate of one dollar or one euro equals five hundred experimental pesetas. This yielded slightly higher marginal payoffs to coordination for Spanish sessions. Since the conversion rate between dollars and euros was very

[12] See Weber (2002) for an elegant illustration of the importance of history dependence in weak-link games.

[13] Most existing studies of the weak-link game use fixed matching as we do. For one notable exception see Van Huyck et al (1990) – their Treatment C compares play for fixed and random pairings in two player weak-link games. For the final round of play, effort levels are significantly higher for the fixed pairings than for the random pairings.

close to 1 to 1 during the time period when sessions were run (Fall, 2002), we doubt this had any impact on the results. Subjects in Cleveland received a show-up fee of ten dollars while the show-up fee was only five euros in Barcelona. The larger show-up fee in Cleveland was deemed necessary to insure an adequate supply of subjects. There is no reason to believe that the differing show-up fees affected our results. The average total payoff was $25.83 in Cleveland and 20.94 € in Barcelona. Once we account for the differing show-up fees, the average earnings are almost identical across the two locations. These earnings were sufficiently large to generate a good supply of subjects in both locations.

5. Results

We begin this section by giving an overview of the main results. We then turn to regression analysis to buttress our primary claims.

Before outlining the results, it is useful to reiterate our terminology. "Employee" refers to an individual subject in the experiment while "firm" refers to a fixed grouping of four employees. Thus, "employee-level" data consist of four separate effort levels per round per firm, one choice per employee. "Firm-level" data consists of a single observation per round per firm, the minimum effort chosen by the four employees of the firm. Whenever we refer to "effort" or "average effort," we are referring to employee-level data. If we refer to "minimum effort" or "average minimum effort," we mean firm-level data.

A. Overview of Results: Our overview of the data begins by examining results from rounds 1 – 10. Recall that the goal for these ten rounds, played with B = 6, was to get firms stuck in a bad outcome – only then can we meaningfully examine overcoming coordination failure. We therefore start by confirming that play moves towards the least efficient outcome over the first ten rounds. The minimum effort is low throughout. It is zero for 71% of the observations in the first ten rounds, with this being the modal outcome in all ten rounds. Average minimum effort changes little over the first ten rounds. It is 6.72 in round 1, compared with an average of 5.86 in round 10. However, these averages hide a great deal of underlying movement. Figure 1 compares minimum effort distributions in rounds 1 and 10. There is a clear bifurcation in the data. Most firms move downward to the minimum of zero, but a small minority moves up to a minimum effort of 40. The frequencies of all the intermediate effort levels diminish. Since the increase is larger per firm going to 40 than the decrease per firm going to zero, the overall effect on the average minimum effort is the small decrease noted above.[14]

(Insert Figure 1 here)

Most of the upward movement observed in later rounds comes from the large majority of groups that have a minimum effort of zero in round 10. Increasing the difficulty of turning around a "failing" firm, most firms with a minimum effort of zero have multiple employees choosing zero. Of the 45 (out of 58) firms with a minimum effort of zero in round 10, 43 have more than one employee choosing zero and 26 have all four employees choosing zero.

Regularity 1: In rounds 1 - 10, the modal outcome throughout is a minimum effort of zero. This outcome becomes more common over time. These results confirm Hypothesis 1.

Having trapped many of the experimental firms in the worst possible outcome, we now turn to the task of overcoming this coordination failure. Figure 2 shows average minimum effort levels in rounds 11 – 20 as a function of the bonus rate in these rounds.

(Insert Figure 2 here)

Focusing on the cells where the bonus rate has increased, two central features of the data can be observed. First, an increase in the bonus rate leads to an increase in the minimum effort. This effect is visible for all three bonus rates used in rounds 11 – 20. It is not a pure restart effect since for Cell 6, the control treatment in which the bonus rate remains at B = 6 for rounds 11 – 20, there is no analogous effect. Instead, the average minimum effort in Cell 6 is flat throughout rounds 11 – 20 and generally equals the level in round 10.[15]

Second, there does *not* appear to be a positive relationship between the magnitude of the bonus increase and its long-run impact on minimum efforts. The highest bonus, B = 14, actually generates the lowest minimum efforts in rounds 16 - 20! Effort levels are roughly the same for B = 8 and B = 10 in rounds 16 – 20. If anything, performance appears to be the best with B = 10 given that this cell had the lowest average minimum effort prior to the bonus increase.[16] The

[14] The convergence to an inefficient outcome over rounds 1 – 10 is more dramatic if we consider employee-level data. The average effort declines from 20.86 in round 1 to 8.28 in round 10.

[15] There is a pure restart effect in the employee-level data. Average effort in Cell 6 increases from 9.00 in round 10 to 16.25 in round 11. The problem is that this increase at the employee-level is too weak to generate any movement at the firm-level. By round 15, the average efforts have fallen back to their original level. By way of contrast, the average effort rises from 8.13 in round 10 to 24.32 in round 11 for firms with bonus rate increases for rounds 11 – 20 (Cells 1 – 5).

[16] The downward spike for the final round of B=10 is driven by a small number of individuals who for inexplicable reasons drop from chosing 40 to choosing 0 in the final round.

occurence of an increase to the bonus rate seems to matter far more in overcoming coordination failure than the magnitude of the increase.

One possible explanation for the preceding result is a ceiling effect – if virtually all firms coordinated at high effort levels with an increase to a bonus of B = 8, there would be little room for greater bonuses to perform better. This is not the case. Consider the firms most in need of a turnaround, the 38 firms that had a minimum effort of 0 in round 10 and saw a bonus increase in round 11. In round 20, only 15 of these firms had all four employees choose effort level 40 and 11 firms still had a minimum effort of 0.[17] For the highest bonus rate used in round 11 – 20, B = 14, only 7 firms had all four employees choose effort level 40 in round 20 out of 24 firms with a minimum effort of 0 in round 10. This fraction is lower than the 8 out of 14 for the lower bonus rates, B = 8 and B = 10, pooled. There is plenty of room for B = 14 to improve performance over the lower bonus rates, but it clearly fails to do so.

Regularity 2: *(a) Increasing the bonus in rounds 11 – 20 significantly increases minimum effort levels. No similar increase is observed in rounds 11 – 20 if the bonus remains at B = 6. (b) The increase in the minimum effort level is not monotonically related to how large the increase in the bonus rate is. These results confirm Hypotheses 2a and 2c. Hypothesis 2b is not supported by the data.*

To better understand this process of change, Figure 3 displays employee-level data from rounds 11 – 20. To focus on the cases with a history of coordination failure, this figure only includes data from the 38 firms that had a minimum effort of zero in round 10 and then had a bonus increase in round 11. We pool data from sessions with B = 8, B = 10, and B = 14 for rounds 11 – 20 since behavior doesn't seem to be sensitive to the size of the bonus increase.

(Insert Figure 3 here)

Examining Figure 3, the immediate response to an increase in the bonus rate is relatively modest. While virtually all employees move away from effort level 0, they don't necessarily move far. For round 11, effort level 40 is the modal outcome, but almost as many employees choose effort levels 10 and 20. A bifurcation then emerges over time. In some groups, the employees who have moved to higher effort levels draw their more cautious partners after them. This is reflected in the steady movement away from intermediate effort levels toward effort level 40. In other groups, the employees who don't raise their effort level following the bonus increase

[17] Looking at all 48 firms that had bonus rate increases for round 11 – 20, 22 had all four employees choose effort level 40 in round 20 while 11 had a minimum effort of zero.

drive the process, pulling other employees back to themselves as can be seen from the increasing weight on effort level 0.

(Insert Table 4 here)

Which side of this bifurcation a firm finds itself on depends on how many of its employees initially respond strongly to the bonus hike. We label an employee as a "strong responder" if they raise their effort by *at least two levels* between rounds 10 and 11 following the bonus increase. All 38 groups that had a minimum effort of zero in round 10 followed by a bonus increase for round 11 included at least one employee who was a strong responder. Table 4 shows the relationship between the number of strong responders in these firms and their long run response to the bonus increase. There is a clear relationship between the number of strong responders and average effort levels – the more employees who respond strongly to the bonus rate increase in round 11, the higher the firm's minimum effort (on average) in round 20.[18] This result seems unsurprising until one realizes that no similar relationship exists between the minimum effort in round 11 and the minimum effort in round 20[19] or between the number of employees who increase their effort, by just one or more levels, between rounds 10 and 11 and the minimum effort in round 20.[20] Overcoming coordination failure requires a *strong* positive response to the bonus increase from multiple employees.

We have established that increases of the bonus rate lead to higher effort levels – of the 48 firms that see a bonus increase in rounds 11 – 20, 33 have higher minimum effort levels in round 20 than round 10.[21] However, these gains come at the cost of higher bonus payments. We therefore would like to know whether the bonus can be reduced without causing a collapse back to the original minimum effort level. To answer this question, Figure 4 shows average minimum effort in rounds 21 – 30 for Cells 1, 4, and 5, the cells that had a bonus rate of B = 14 for rounds

---

[18] This relationship has moderate statistical significance – running a regression of minimum effort in round 20 on the number of strong responders, the relevant parameter just misses significance at the 5% level.

[19] Of these 38 firms, the 15 firms with a minimum effort level of 0 in round 11 have an average minimum effort in round 20 of 22.7. The 17 firms with a minimum effort level of 10 have an average minimum effort of 19.4. The remaining six firms have an average minimum effort of 31.7 in round 20.

[20] For all but one of the 38 firms, more than one employee increases his/her effort level in round 11. The average minimum effort in round 20 equals 20.0 if two employees increase their effort (4 firms), 23.8 if three employees increase their effort (16 firms), and 23.5 if all four employees increase their effort (17 firms).

[21] This number becomes more impressive after noting that 4 of the 48 firms had all four employees choose effort level 40 in round 10 and therefore could not show an increase.

11 – 20. The figure also shows, as a point of comparison, the average minimum effort for rounds 1 – 10 of these three cells.

(Insert Figure 4 here)

One can see that a cut in the bonus rate does not lead to a collapse back to the initial effort level. A cut to B = 10 actually yields an increase in average minimum efforts! Cutting the bonus rate to B = 6 causes the average effort to fall sharply, but not back to its original levels. Responses to the bonus rate reduction are typically extreme – firms tend to either not change at all or change a lot. Suppose we compare minimum efforts in round 20 with those in round 29.[22] Of the 19 firms that see a decrease in the bonus rate for rounds 21 – 30, 10 have the same effort level in round 29 as in round 20.[23] Among the 9 firms that see changes, 6 see changes of at least two effort levels. The relatively good performance of firms that have their bonus reduced to B = 6 is almost entirely due to firms that didn't respond to the change – there were four firms in Cell 5 that increased their minimum effort between rounds 10 and 20 but did not change their minimum effort in response to the bonus cut for rounds 21 – 30. Generally, effort levels show history dependence in only one direction – it is easy to move firms to higher effort levels, harder to move them back to lower effort levels.

<u>Regularity 3</u>*: After having B = 14 for rounds 11 – 20, reducing the bonus down to either B = 6 or B = 10 does not reduce the minimum effort back to its original level. The data supports Hypothesis 3.*

We can gain some insight into why some firms stay at high effort levels following a bonus rate cut while others do not by looking at the employee-level data. Consider the 10 firms in Cell 5, the most extreme treatment where the bonus rate drops back to B = 6. Eight of these ten firms have minimum effort levels greater than zero in round 20.[24] For two of these firms, no employee changes their effort level in round 21. Both remain coordinated at the payoff dominant equilibrium (all employees choose effort level 40) throughout rounds 21 – 30. In the remaining six firms at least one employee reduces their effort level in round 21 below the firm's minimum

[22] Due to a computer error, we lost the round 30 data for 6 of the 29 groups considered here. We use round 29 rather than round 30 to avoid having to drop these 6 groups.

[23] As a point of comparison, among the 10 firms that have B = 14 for rounds 11 – 30, 9 have no change in the minimum effort from round 20 to round 29.

[24] Unsurprisingly, the two firms with minimum effort levels of 0 in round 20 remain at this level for rounds 21 – 30.

effort in round 20. Four of the six firms converge to lower minimum effort levels while the other two eventually return to mimimum effort level they achieved in round 20.

The primary difference between the firms that recover from an initial drop in the minimum effort and those that don't is how the other employees respond to having someone cut the minimum effort. In the four firms that don't recover, at least one employee (and usually more) who didn't reduce their effort level in round 21 responds to the reduction in minimum effort in round 21 by cutting their own effort in round 22. In the two firms that recover, the employees who do not cut their effort in round 21 maintain their high effort in subsequent rounds. Thus, a negative response to the bonus cut involves a chain reaction – one or more employees initially cutting their efforts triggers effort reductions among the other employees. If there is a cohort of employees who hold steady, the employees who originally react negatively to the bonus cut eventually recover to their original effort levels.

One puzzling feature of the data that deserves a note is the weak performance with B = 14. Both for rounds 11 – 20 following an increase from B = 6 and for rounds 21 – 30 following play with B = 14 in rounds 11 – 20, B = 10 produces substantially higher effort levels than B = 14. If we examine the data from rounds 21 – 30 for Cells 1, 2, and 3 (the three cells where the bonus rate is constant for rounds 11 – 30) the lowest minimum efforts are achieved with B = 14.[25] Even though no standard economic model would predict this effect, it seems too robust to be dismissed as a statistical anomaly. Satisficing might be playing a role here – subjects with B = 14 generally are making high payoffs and might be disinclined to spend much effort figuring out how to get even higher payoffs – but this is pure speculation.

Regularity 2 indicates that increasing the bonus rate yields higher minimum effort levels, but Regularity 3 implies the existence of substantial history dependence in the data.[26] This leads to our final question: given the presence of hysteresis, will it be harder to turn around a firm that has a longer history of coordination failure? Specifically, does delaying the bonus rate increase from B = 6 to B = 14 until rounds 21 – 30 make it harder to raise a firm's minimum effort level? Figure 5 compares average minimum effort when the bonus is raised to B = 14 after ten rounds (Cell 1, 4, and 5) versus twenty rounds (Cell 6) of play with B = 6. The x-axis in this graph

[25] For rounds 21 – 30, the average minimum effort is 22.8 for Cell 1 (B = 14), 36.5 for Cell 2 (B = 10), and 28.6 for Cell 3 (B = 8).

[26] Subjects in rounds 21 – 30 of Cell 5 behave differently than subjects in rounds 1 – 10. Both sets of subjects face B = 6, but the former have previous experience with B = 14.

displays the first ten rounds following the bonus increase, rounds 11 – 20 for Cells 1, 4, and 5 and rounds 21 – 30 for Cell 6. Initially, average minimum efforts appear to be unaffected by how much time passes before the bonus increase. However, looking over rounds 6 – 10 following the bonus increase, a modest difference emerges, with lower effort levels when the increase has been delayed. To help understand the forces underlying this effect, Figure 5 also graphs the average effort across all employees, broken down by whether the bonus rate increase occurs after 10 or 20 rounds. As with the firm-level data, the employee-level data is not affected by the timing of the bonus hike for the first few periods. Instead, the negative effect of delay is driven by differences in the willingness of employees who initially respond positively to the bonus hike to remain at these high effort levels. When the switch to B = 14 occurs in round 11, employees who initially move to higher effort levels tend to stay there and eventually pull lower effort employees up to their level. In contrast, when the increase is delayed until round 21, employees who at first move to higher effort levels tend to retreat back to lower effort levels, dampening a continued increase in minimum effort levels.[27] It appears that an extended history of bad outcomes makes employees pessimistic. With a late increase of the bonus rate, there still exists a cohort of employees who try to move to higher effort levels. However, when others don't rapidly follow their lead, they give up and cut their own effort levels.

(Insert Figure 5 here)

<u>Regularity 4</u>: *Delaying the increase from B = 6 to B = 14 by ten rounds has a small negative effect on efficiency. This provides support, albeit weak, for Hypothesis 4.*

B. Regression Analysis: The preceding overview of the data relied entirely on descriptive statistics. In this subsection we attempt to put our conclusion on a firmer statistical footing through the use of regression analysis.

We run regressions both on firm-level data and on employee-level data. While we are ultimately interested in the firm-level outcomes, the employee-level regressions serve as a robustness check. In particular, because a firm's outcome depends on the minimum effort chosen by an employee of the firm, the firm-level data tends to accentuate the impact of outliers. Using the employee-level data allows us to ascertain whether the treatment effects are broad based or

[27] Note that the average efforts for the two treatments diverge before the average minimum efforts. This indicates that the problem is *not* a failure by employees who choose lower effort levels to respond to their compatriots' choices of higher efforts.

only impact a relatively small fraction of the population. In firm-level regressions, the dependent variable is the firm's minimum effort. Each play by a firm counts as a single observation. In employee-level data, the dependent variable is the effort level selected by the employee. Each play by a firm generates four observations, one for each of the four employees. All of the regressions reported below are ordered probits. Given the categorical nature of the data, this is a natural choice.

A critical point in this analysis is how we control for repeated observations of the same employees or the same firms. In controlling for repeated observations, we don't believe there is a "right" approach. Instead, there is a tradeoff between minimizing the likelihood of Type 1 errors (false rejection of the null) versus minimizing the probability of Type 2 errors (failing to correctly reject the null). While Type 1 errors are usually the primary concern, both types of error are worrisome in our analysis. Some of our conclusions could be incorrect due to Type 1 errors (Regularities 3 and 4) while others could be fallacious due to Type 2 errors (Regularity 2). We therefore report regressions using a relatively conservative clustering approach to correcting for repeated observations, due to Liang and Zeger (1986),[28] as well as regressions using a more powerful random effects specification. These should be viewed as setting bounds on what is actually in the data, with the former approach more likely to yield Type 2 errors and the latter more likely to yield Type 1 errors.[29]

For both the firm-level and employee-level regressions, the random effects regressions are estimated using maximum-likelihood techniques. While the random effects in the firm-level data regressions are standard, those in the employee-level regressions are not. In this case we need to account for correlation between two observations from the same employee as well as correlation between two observations from different employees who are in the same firm. We therefore have a random effect at the employee-level nested within a random effect at the firm-level. Let y_{ijt} be the latent dependent variable for employee j in firm i in round t, let X_{ijt} be the vector of independent variables for employee j in firm i in round t, let μ_i be a firm specific error term, let ν_j be an employee specific error term, and let ε_{ijt} be an i.i.d. error term for employee j in

[28] Readers may be familiar with this as the "Cluster" option in Stata. For firm-level *and* employee-level regressions, each firm is treated as a separate cluster. This is appropriate for the employee-level regressions since observations from employees in the same firm cannot be treated as independent.

[29] Underlying the tradeoff between Type 1 and Type 2 errors is a tradeoff between efficiency and the strength of assumptions about the distribution of errors.

firm i in round t. Assume that $\mu_i \sim N(0,v_G^2)$, $\nu_j \sim N(0,v_I^2)$, and that $\varepsilon_{ijt} \sim N(0,1)$. Then y_{ijt} is given by the following equation where α is a scalar and β is a vector of parameters:

$$y_{ijt} = \alpha + \beta X_{ijt} + \mu_i + \nu_j + \varepsilon_{ijt}$$

The dependent variable is derived in the standard way for an ordered probit given the latent variable and cutoffs between categories. The maximum likelihood estimation fits values for the cutoffs, α, β, v_G, and v_I. Thus, instead of having one correlation term as in a standard random effects specification, there are two.

Throughout this paper, all tests of significance for individual parameters are two-tailed z-tests. All tests of joint significance use log likelihood ratio tests. The cutoffs and random effects parameters are not reported in any of our tables since these estimates are of little economic interest. In all cases these omitted parameter estimates are statistically significant.

(Insert Table 5 here)

Table 5 shows ordered probit regression results that back our claims in Regularity 2. In particular we wish to establish that increasing the bonus rate for rounds 11 – 20 significantly increases effort levels, but that the size of the effect is not an increasing function of the size of the bonus rate increase. The data for the regressions shown in Table 5 includes data from all six cells for rounds 11 – 20. Cell 6, where the bonus rate remains at B = 6 throughout this time period serves as the base.

The primary independent variables in these regressions are dummies for the bonus rates (other than B = 6) used in rounds 11 – 20: B = 8, B = 10, and B = 14. To capture how the effect of a bonus increases over time (without imposing a linear trend), the dummies for the bonuses are interacted with dummies for rounds 11 – 15 and round 16 – 20. We include two additional control variables. First, even though the sample is balanced between Cleveland and Barcelona, a dummy for sessions conducted in Barcelona is included to absorb some of the noise. Second, we include the firm's minimum effort in round 10.[30] This lagged dependent variable allows us to control for firm's different starting points prior to the increase in the bonus rate. In some sense this is giving us additional control over the firm effects in the data.

[30] Instead of using the firm's minimum effort in round 10 as an independent variable, the employee-level regressions could have used the individual employee's effort in round 10. We have tried this alternative, and found that it yields somewhat worse fits. The primary qualitative results are not affected.

The primary conclusions to be drawn from the regressions in Table 5 are robust to exactly how we control for repeated measures and what data is used. Even in rounds 11 – 15, all three bonus rates yield significantly greater effort, both on the employee and firm-levels, than in the controls. This improvement becomes even larger in rounds 16 – 20. Which bonus increase is used has some impact on behavior, but the response is non-monotonic. In making this statement, we focus on rounds 16 – 20 when a bonus increase has had enough time to yield its full impact. Regardless of whether we examine firm-level data or employee-level data and regardless of how we control for individual effects, the difference between effort levels with B = 14 and those with B = 10 is negative and statistically significant, albeit weakly.[31] The effort levels with B = 14 are also lower than those with B = 8, but this is only statistically significant when the more powerful random effects specification is used to control for individual effects.[32] Significant differences are never found between B = 8 and B = 10. To summarize, the econometric results reinforce Regularity 2 – subjects respond positively to an increase in the bonus rate, but larger increases do not yield larger responses. Our puzzling observation that the highest bonus rate increase, B = 14, yields the lowest effort levels is supported statistically.

Turning to secondary features of the data, effort levels in Barcelona are lower than in Cleveland, but the difference is modest and generally not statistically significant. Not surprisingly, there is positive and statistically significant correlation between a firm's minimum effort in round 10 and efforts (both by the firm and the individual employees) in rounds 11 – 20.

(Insert Table 6 Here)

The regressions shown in Table 6 provide statistical support for Regularity 3. The aim is to establish that when bonus rates are first raised for rounds 11 – 20 and then lowered for rounds 21 – 30, the effort levels remain significantly above those prevalent prior to the initial increase in the bonus rate. Data is taken from the three cells, Cells 1, 4 and 5, where the bonus rate increases to B = 14 for rounds 11 – 20. Data from rounds 6 – 10, the last five rounds with B = 6

---

[31] To test this proposition, we reran the regressions on Table 4 with B = 14 differenced from B = 10 for rounds 16 – 20. For the firm-level regressions, the relevant parameter is significant at the 10% level using clustering (z = 1.77, p = .077) and at the 1% level using random effects (z = 3.818, p < .01). For the employee-level regressions, the relevant parameter is significant at the 10% level using either clustering (z = 1.86, p = .063) or nested random effects (z =1.72, p = .089).

[32] Rerunning the regressions on Table 4 with B = 14 differenced from B = 8 for rounds 16 – 20, the relevant parameter is not significant at the 10% using clustering for either the firm-level data (z = 0.57, p = .57) or the employee-level data (z = 0.73, p = .46). With random effects, the relevant parameter becomes significant at the 10% level either using firm-level data (z = 1.93, p = .054) or employee-level data (z =1.80, p = .071).

preceding the increase in the bonus rate, and rounds 21 – 30 are included. The data from rounds 6 – 10 serves as the base. The primary independent variables in these regressions are dummies for the bonus rates used in rounds 21 – 30: B = 6, B = 10, and B = 14. To capture how the impact of decreasing the bonus rate develops over time, the dummies for the bonuses are interacted with dummies for rounds 21 – 25 and rounds 26 – 30. A key feature of these regressions is that the variables for B = 10 and B = 14 are differenced from B = 6 for the equivalent time period. For example, the parameter estimate for (Rounds 21 – 25)*(B = 6) captures the difference between behavior with B = 6 in rounds 21 – 25 (Cell 5) and the base while the parameter estimate for (Rounds 21 – 25)*(B = 10) captures the difference between behavior with B = 10 in rounds 21 – 25 (Cell 4) and behavior with B = 6 in rounds 21 – 25 (Cell 5). Once again we include a dummy for the location of the session. To control for differences in firm's starting points, we include the firm's minimum effort in rounds 5 and 20 as lagged dependent variables.[33]

The regressions shown in Table 6 strongly support the conclusions reported as Regularity 3. The key parameters are (Rounds 21 – 25)*(B = 6) and (Rounds 26 – 30)*(B = 6). Looking at either firm-level data or employee-level data and regardless of how we control for repeated observations, these parameter estimates are always statistically significant at the 10% level and, with only one exception, at the 1% level. Cutting the bonus rate from B = 14 back to B = 6 leads to effort significantly above levels prior to the bonus increase. This isn't to say that cutting the bonus rate so sharply doesn't have some negative impact as a cut to B = 10 yields significantly higher efforts than a cut to B = 6. The latter conclusion holds true with both firm and employee-level data and is not sensitive to how we control for repeated observations. Once again, performance with B = 14 is surprisingly poor. While effort levels are higher than when the bonus rate is cut to B = 6, the difference is generally not statistically significant.

As for the secondary results, we again observe a consistently negative sign for the Barcelona dummy. Unlike previously, these are typically significant. The lagged dependent variables, particularly the minimum effort for round 20, have significant explanatory power.

(Insert Table 7 Here)

[33] Various combinations of these variables, including interactions with the time periods were tried. Which particular specification was used does not impact our qualitative conclusions.

Table 7 addresses the most tenuous of our conclusions, Regularity 4. The goal is to establish that the modest difference in effort levels observed with an early versus a late increase in the bonus rate is statistically significant. Data is taken from rounds 11 – 20 of Cells 1, 4, and 5, the three cells with early increases to B = 14, as well as rounds 21 – 30 of Cell 6, the cell with a late increase to B = 14. Data from rounds 11 – 15 of cells with an early increase to the bonus rate serves as the base. The primary independent variables are dummies for when the bonus rate increase took place interacted with dummies for five round blocks. The critical parameter, the dummy for a late increase interacted with a dummy for rounds 26 – 30 is differenced from the parameter estimate for an early increase in rounds 16 – 20. This parameter captures the effect of a late increase in the bonus rate after some time for adjustment. The regressions again include a dummy for whether the session took place in Barcelona as well as a control for the firms' differing starting points, the firm's minimum effort in the last round prior to the bonus rate increase to B = 14.

The results reported in Table 7 provide limited support for Regularity 4. The key parameter estimate, the coefficient for rounds 26 – 30 (late increase), is consistently negative but only significant when the more powerful random effects specifications are used to control for repeated observations. This is true for both the firm-level data and the employee-level data.[34] When significant, the estimated Barcelona effect is again negative. Finally, the firm's starting point has a large impact on its eventual effort level.

In summary, the regressions provide solid statistical support for the conclusions reported as Regularities 2 and 3. These conclusions are robust to differing methods of controlling for repeated observations as well as the use of firm versus employee-level data. As such, we can conclude with some confidence that increases in the bonus rate will yield increases in effort levels, that the size of the bonus rate increase is not monotonically related to the response in the effort level, and that temporary increases to the bonus rate can yield permanent increases in effort levels. Regularity 4 gets some statistical support as well, but only when relatively powerful means are used to control for repeated observations. This underscores our observation that lengthening the time before a bonus increase only has a moderate impact on likelihood that a failing firm can be turned around.

[34] The coefficient estimates for rounds 16 – 20 are statistically significant for the firm-level data but not the employee-level data. The reflects the adjustment process with an early increase in the bonus rate, as low effort individuals are pulled up to higher effort levels.

6. Game Theoretic Models and the Turnaround Game

Our data analysis reveals a number of strong regularities. This raises an obvious question – does there exist a unified model that can explain these regularities? Our goal in the current section is to show that a number of likely candidates drawn from the behavioral game theory literature fail to provide a good explanation for the observed data.[35] We focus particularly one of the most puzzling regularities in the data, the non-monotonic effect of increasing the bonus rate for rounds 11 – 20. We conclude this section by discussing what these failures teach us about the forces that allow groups to overcome coordination failure.

A) Nash Equilibrium and Subgame Perfection: As noted previously, the weak-link game played in any single round of the turnaround game has five Nash equilibria with employees coordinating at any of the five available effort levels. Considering the full thirty periods of the turnaround game, any sequence of the five equilibria is a subgame perfect equilibrium. Thus, as theories for explaining the evolution of play in the turnaround game both Nash equilibrium and subgame perfection are vacuous. Any sequence of outcomes is consistent with subgame perfection – increases in the minimum effort level following a hike in the bonus rate are no more or less consistent with the theory than decreases.[36]

B) Quantal Response Equilibrium: Moving to a stochastic choice framework, the quantal-response equilibrium (QRE) model introduced by McKelvey and Palfrey (1995) yields more predictive power than subgame perfection.[37] In this model, players are noisy best responders. The probability of a strategy being played is an increasing function of its expected payoff, but all strategies are played with positive probability. Anderson, Goeree, and Holt (2001) show that weak link games like those considered here have a unique (logit) QRE. While this result does not directly extend to the games being considered here,[38] as we have a discrete rather than a continuous choice space, we have numerically confirmed that the result holds for our games as well. The QRE model predicts that an increase in the bonus rate will yield higher average effort levels. Intuitively, holding the probability distribution over others' action fixed, an increase in

[35] As all of these models have been described in detail elsewhere, we eschew presenting the technical details in favour of brief intuitive descriptions.

[36] In a narrow sense, neither Nash equilibrium nor subgame perfection is supported by the data. In only 52.8% of rounds are employees playing a pure strategy Nash equilibrium.

[37] Specifically, we are considering the QRE with a logit choice rule.

[38] Personal communication, Charlie Holt, 11/24/2003.

the bonus rate has a larger positive effect on the expected payoffs for high effort levels than for low effort levels. Since players are noisy best responders, this shifts play in the direction of higher effort.

To see if the QRE model explains subjects' responses to an increase in the bonus rate (as summarized in Regularity 2), we first calibrate the QRE model based on data for rounds 1 – 10 and then predict effort levels for rounds 11 – 20. Specifically, we have used maximum likelihood estimation to fit the noise parameter for the pooled data from rounds 1 - 10. This estimation exercise should be taken with a grain of salt, since we treat observations from the same employees as independent and also ignore the changing nature of the effort distribution. Nonetheless, the estimated value for the noise parameter should be in the right ballpark for the "true" parameter value.[39] Using the estimated value of the noise parameter, we calculate the QRE for the four different bonus rates used in rounds 11 – 20. Based on these results we then predict average effort levels as a function of the bonus rate. Table 8 compares the predicted average effort levels from the QRE model with those observed in the data.

(Insert Table 8 here)

The predicted and actual average effort levels are almost identical for rounds 1 – 10. This should not come as a great surprise, since the model was fit to the data from rounds 1 – 10. As noted above, the QRE model predicts that higher bonus rates will yield higher effort levels. However, the predicted difference between the average effort with B = 6 and higher bonus rates is quite small compared to the differences observed in the data.[40] Moreover, the model predicts a monotonic response to increases in the effort level. As noted in Regularity 2, the data does not yield any such monotonicity. The poor performance of the QRE model becomes even clearer when Regularity 3 is considered – because of its static nature the QRE model cannot replicate the history dependence observed in the experimental data.

C) Variations on EWA: An obvious explanation for the poor performance of the QRE model is that we are using a static model to predict behavior in a dynamic setting. The data display a high degree of history dependence, as reflected in Regularity 3. Indeed, the experiments were designed to emphasize the role of history dependence in fostering or hindering change. It

[39] The qualitative results are not sensitive to using a subset of the data from rounds 1 – 10 (e.g. only round 1 data, only round 10 data, etc.).

[40] Some readers may be surprised by the small predicted shift here given the large effect observed by Goeree and Holt (2001). The key difference is that they consider a two-person game while we study a four-person game.

therefore seems reasonable that a learning model explicitly incorporating history dependence will better track the data than a static model like QRE.

To explore this conjecture, we use the experience-weighted attraction (EWA) learning model introduced by Camerer and Ho (1999). There exist many learning models in the literature, and it remains unclear which (if any) of these models has the best explanatory power for experimental data. EWA is a good place to start, however, because it has the useful property of nesting both belief-based learning models and reinforcement learning models.

Over the years, a number of variants on the basic EWA model have been advanced. The model that we present here, which includes sophisticated learners and strategic teaching, is close (but not identical) to the model presented in Camerer, Ho, and Chong (2002). In describing the model, it eases the exposition to first describe a simpler model that only contains "unsophisticated" learners. Each player starts out with weights ("attractions" in Camerer and Ho's nomenclature) for each of his strategies. The probability of a strategy being played is given by a logit rule and is therefore an increasing function of its weight. At the end of each round, the weight for each of the strategies is updated. As in a reinforcement learning model, this updating depends in part on the payoff received by the realized strategy – higher payoffs lead to a greater frequency of play in future rounds. Like belief-based models, unused strategies also get their weights updated based on what they would have earned if they had been employed. The relative weights of realized payoffs and hypothetical payoffs are determined by a parameter fitted from the data. It is this inclusion of both realized and hypothetical payoffs that allows the EWA model to nest both reinforcement and belief-based models of learning.[41]

To give the model its best chance to track the data, we modify it in three fashions. First, there are fairly obvious restart effects in the data between ten period blocks, even in cells where the bonus rate doesn't change. For example, the average effort by an employee jumps from 9.00 in round 10 of Cell 6 to 16.25 in round 11 even though the B = 6 in both rounds. As one avenue

[41] The nesting of belief-based models comes with a caveat. EWA does not directly include payoff maximization as a belief-based model (e.g. fictitious play) does. This becomes important if the payoff function changes between rounds as in our experiments. Even without adjusting beliefs, a belief-based model will instantaneously respond to a payoff change as the payoff maximization problem has been altered. Unless we force attractions to change when the payoffs change, the basic EWA model will only respond to changing payoffs with a delay. Intuitively, EWA only responds to payoffs that happened or could have happened *in the past*. This makes it hard for the EWA model to capture employees' immediate reactions to a change in the bonus rate. The inclusion of sophisticated and forward-looking types allows us to work around this issue; these new types are true payoff maximizers, and therefore respond immediately to a change in payoffs. We therefore don't think that this peculiarity of EWA plays an important role in driving our conclusions.

through which the model can capture restart effects, at the end of each ten round block the attractions that govern choices in the subsequent block are set equal to a weighted average of the initial attractions and the attractions at the end of the ten period block. This weight is a parameter fit from the data.

Second, we modify the model to include "sophisticated" learners. These are players who anticipate the learning of others. The structure of the experiment gives each subject all the information they need to calculate the attractions of others (if they knew the parameters for the model). We therefore model sophisticated learners as keeping a running record of other players' attractions and therefore having the ability to predict their actions in the current round. Sophisticated learners' choices are a noisy best response to the choice probabilities of unsophisticated EWA learners.[42]

Finally, and perhaps most importantly, we add forward-looking learners to the population. A forward-looking learner is similar to a sophisticated learner, but also anticipates the impact that the current period's outcome will have on the learning of other players and hence on their actions in the next round. For simplicity, forward-looking learners are assumed to only look one period into the future.[43] A forward-looking learner then assigns each available action an attraction equal to the weighted sum of expected payoffs for the current period and the next period, where the weight is a parameter fitted from the data. Thus, the choices of a forward-looking learner are a noisy best response to the anticipated choices of unsophisticated learners in the current round and the immediately following round. By including such players we allow for the possibility of strategic teaching – some employees will raise their effort levels in an attempt to move the entire group to coordination at a higher effort level.[44]

[42] Our formulation of sophisticated learning draws Camerer, Ho, and Chong's (2002) model of sophisticated EWA as well as related learning models by Stahl (1999) and Cooper and Kagel (2004). Unlike Camerer *et al*, we assume sophisticated learners believe that *all* other players are unsophisticated learners. This assumption, while inelegant, greatly simplifies use of the model.

[43] Even looking only a single period into the future, the computational complexity of this model is formidable. Because all outcomes can happen with positive probability, a forward-looking learner must calculate expected payoffs for the next period conditional on every possible history for the current period – all 625 of them. The number of calculations increases exponentially with the number of period the player looks forward; a two-period window forces the consideration of 390,625 possible histories. Given that it takes about a week for the program fitting the MLE with a one-period window to converge, we estimate it would take about 12 years to fit the model with a two-period window.

[44] See Camerer *et al* (2002) for a lengthy discussion of strategic teaching and the EWA model as well as a thorough literature review. Other than slight differences due to the differing structures of the experiments being studied, our specification of forward-learning types is the same as that proposed in Camerer *et al*.

In evaluating the performance of the EWA model with sophisticated and forward-looking types, we focus on its ability to track behavior following an increase in the bonus rate. Recall that increasing the bonus rate for round 11 – 20 causes a large increase in effort levels, but the size of the bonus rate increase has little impact on effort levels. To generate predictions for the model, we first calibrate the model by fitting it to the first ten rounds of data from Cells 1 – 5 and the first twenty rounds of data from Cell 6. This extra data from Cell 6 is included so the reset parameter can be estimated. We do not use additional rounds from the other cells – the goal is to predict behavior following an increase in the bonus rate based on behavior prior to the bonus rate increase. All the parameters for the full model are fit using standard maximum likelihood techniques. This includes the proportions of unsophisticated, sophisticated, and forward-looking learners, the weight forward-looking learners put on future payoffs, and the weight on existing attractions at the end of each ten period block. Using these parameter estimates, we then simulate the model for the first twenty rounds of the experiment. For all simulations the bonus rate equals 6 for rounds 1 – 10. For rounds 11 – 20, we run 1000 simulations each with bonus rates of 6, 8, 10, and 14. Figure 6 reports the results of this simulation exercise, comparing the average effort in the experimental data with the average effort in the simulations for round 10 – 20. Note that this is employee-level data rather than firm-level data – the simulations generally do a better job of tracking the employee-level data. Also note that the vertical scales are different on the two panels of the figure.

(Insert Figure 6 here)

To point out the obvious, the simulations do a poor job of replicating the experimental data. The average effort for round 10 differs little between the simulations and the experimental data and both show an increase in effort levels for round 11, albeit much smaller in the simulations than the experimental data. It is in round 12 that the simulations begin to dramatically differ from the experimental data. In the experimental data, average effort levels continue to increase gradually over time. In the simulations, round 12 is a disaster as most of the gains of round 11 are immediately lost. Effort levels fall steadily over time, eventually reaching even lower levels than in round 10. In the simulations, bonus rate increases fail to overcome a history of coordination failure.

The most obvious reasons that the EWA model with sophisticated and forward-looking types might fail to track the experimental data are technical. For example, the model does not allow for heterogeneity among individuals of the same type – e.g. the parameter values used by all unsophisticated learners are identical.[45] Allowing for heterogeneity increases the average effort level for rounds 1 – 10 by making it more likely that a firm includes multiple employees who persistently choose high effort levels at the beginning of the experiment and thus eventually coordinate at a high effort level.[46] Unfortunately, adding heterogeneity to the model does not improve the simulated employees' weak response to a bonus rate increase. Intuitively, the positive impact of a bonus rate increase is driven by firms that are stuck at low minimum effort levels. Adding heterogeneity decreases the likelihood of suffering coordination failure prior to the bonus rate, but actually makes the behaviour of such firms harder to change. As another possibility, we suspect that the fitted values of the reset parameter and proportion of forward-looking types are too small.[47] To determine if the simulation results are robust to allowing for higher values of these parameters, we ran simulations where the reset parameter was doubled and the proportions of sophisticated and forward looking types were tripled. Increasing these parameters produces a larger increase in effort for round 11 but doesn't prevent the downward spiral that follows. A lack of sophisticated and forward-looking types does not explain the poor performance of the learning model. Beyond these two possibilities, we have explored a variety of other ways in which the model might be altered. Although it is possible to somewhat improve the model's ability to track the data, the steady rise in effort level following a bonus rate hike is never replicated. We therefore feel that the failure of the simulations cannot be attributed to a narrow technical issue.

To better understand why the EWA model is failing, consider the critical point where the model definitively fails to replicates the main features of the data – round 12 after a bonus rate increase. Recall that effort continues to rise in round 12 of the experimental data but falls sharply in the simulations. Examining the experimental data at this point illuminates what is missing from the model. We focus on the toughest cases, firms in Cells 1 – 5 that had a

---

[45] See Wilcox (2003) for a discussion of the impact of heterogeneity on estimates of EWA parameters.

[46] These statements are based on results from fitting a two segment model like that presented in Camerer and Ho (1999) to the data.

[47] The fitted reset parameter is .385 – this is the weight put on the original attractions versus the attractions immediately preceding the bonus rate increase. The fitted proportions of sophisticated and forward-looking types are 4.1% and 8.2% respectively. These numbers are statistically significant, but of dubious economic significance. As a point of comparison, Cooper and Kagel (2004) report 19% sophisticated learners among inexperienced subjects and 32% among experienced subjects.

minimum effort of 0 in round 10. For these firms, average effort only increases slightly between rounds 11 and 12, rising from 22.4 to 23.2. More striking is the behavior in round 12 of employees who choose low effort levels in round 11: employees who choose an effort of 0 in round 11 choose an average effort of 14.2 in round 12 and those who choose an effort of 10 in round 11 choose an average effort of 15.8 in round 12. This positive movement cannot be attributed solely to employees best responding to the previous round's outcome – if we only look at employees who did not uniquely determine the minimum effort for their group (and hence can't be playing a best response to round 11 outcomes by increasing their effort), employees who choose an effort of 0 in round 11 choose an average effort of 11.3 in round 12 and those who choose an effort of 10 in round 11 choose an average effort of 13.7 in round 12. It is this delayed positive response which is missing in the simulations. The employees in the simulations who choose low values in round 11 tend to stay at these low values in round 12: simulated employees who choose an effort of 0 in round 11 choose an average effort of 2.6 in round 12 and those who choose an effort of 10 in round 11 choose an average effort of 10.4 in round 12. The worst offenders are unsophisticated learners. Unlike sophisticated and forward-looking types, who can (and do) respond to a change in the bonus rate by moving to higher effort levels, unsophisticated learners are trapped by their history. Once they have learned to choose a low effort level, it is extremely difficult to teach them to choose anything else. The strength of learning models, their ability to incorporate history dependence, is a weakness here.

The EWA model's failure to track the experimental data teaches us an important lesson – to overcome coordination failure, it is critical not just to have strong leaders (as noted in Section 5) but also to have responsive followers. There is a large cohort of experimental subjects who lack the wit or nerve to lead a movement towards coordination when the bonus rate increases. However, many of these subjects strongly increase their effort levels after observing others' positive responses to the bonus hike. These responsive followers reinforce the momentum started by strong leaders. Thinking more deeply about what is missing from the learning model, we conjecture that sophistication in real subjects isn't as rigid as in the learning model. Seeing a dramatic shift in others' effort levels, the lightbulb may come on for a hitherto unsophisticated subject. One can imagine them thinking, "I wonder why they just raised their effort levels? Maybe they are trying to move us to a higher effort level – that would let me make more money.

Perhaps I should try to help out." In other words, sophistication may grow with experience, not exogenous but endogenously in response to sophisticated play by other subjects.

7. Final remarks

Coordination failure is a serious problem that besets many sorts of organization. While one must always exercise caution in translating experimental results to field settings, our experiments suggest that organizations can use financial incentives to overcome a history of coordination failure. Given that small increases in incentives are just as effective as large increases and given that incentives only need to be increased on a temporary basis, it seems that successful coordination can be accomplished on the cheap.

In understanding why an incentive increase is effective, we believe that understanding the nature of the coordination problem is essential. We don't think that employees in firms experiencing coordination failure are unable to read the payoff table or fail to realize that everyone could be better off if all chose effort level 40. The trick, giving the riskiness of unilateral increases in effort, is figuring out how and when to get everyone to change their behavior together. Thinking of the turnaround game as a single thirty-round game rather than a sequence of thirty games, this is an issue of equilibrium selection. The bonus rate increases then serve as a focal point in the sense of Schelling (1960), triggering a coordinated change to higher effort levels.[48] This also provides an additional explanation for why the learning models perform poorly here, as these models treat the experiment as a sequence of separate games and ignore the larger equilibrium selection issue.

Focusing on the use of financial incentives as a coordination device for change allows us to better understand how an effective incentive scheme ought to be devised. First, any effective scheme needs to be global in nature. If the goal is to get all agents (or at least many of them) to change behavior simultaneously, a piecemeal approach is unlikely to generate the needed fraction of strong responses. This matches well with the conclusions of the empirical literature

[48] In Crawford's (1991) discussion of Van Huyck, Battalio, and Beil's (1991) work on median games he notes that any change in treatment tends to lead to coordination at higher effort levels. Crawford informally refers to a "bell-ringing effect," where the changes in treatments serve as a coordinating device similar to the changes in bonus rates for our experiments. While this result is from a substantially different environment and uses different manipulations than those studied here, it suggests that the value of bonus hikes as a trigger for coordination is part of a more general empirical regularity. Interestingly, these results don't necessarily extend to the harsher environment of the weak-link game. Even with a history of successful coordination in a closely related game, Van Huyck, Battalio, and Beil (1990) find that switching to the minimum game causes a quick collapse in effort levels (p. 243). This is consistent with our failure to find pure restart effects.

on organizational change. Second, the launch of an effective scheme needs to be highly public. Schelling's classic example on focal points, that individuals wanting to get together in New York City would naturally coordinate on Grand Central Station as a meeting place, works because everyone (at the time) would have known where Grand Central Station was. Without common knowledge of its existence, an incentive scheme is unlikely to generate the coordinated change needed to overcome coordination failure.

The results of our work suggest a number of directions in which future research will be extended. First, we plan on exploring how relaxing the simplifying assumptions of our paper affects the results. While the use of an exogenous manager greatly simplified the experimental environment and allowed us to focus on issues of coordination, it remains a point of obvious interest to examine behavior with an endogenous manager. Likewise, concentrating on financial incentives greatly simplified our experimental design, but we plan on exploring other channels through which change can take place, particularly communication between managers and employees. Second, while the learning models give us some insight into how subjects overcome coordination failure, this insight comes through their failure to explain subjects' behavior. We speculate that a model that incorporated endogenous growth of forward-looking behavior might succeed in tracking behavior. The development of such a model is beyond the scope of this paper (nor is it clear that this is the best environment for examining this issue), but this remains a topic of future research.

References

Anderson, Simon, Jacob Goeree and Charles A. Holt (2001), "Minimum-Effort Coordination Games: Stochastic Potential and Logit Equilibrium", Games and Economic Behavior, 34, 2, 177-199.

Brandts, Jordi and David Cooper (2004b), "Managers, Employees and Communication in Experimental Firms" in preparation.

Brandts, Jordi and David Cooper (2004c), "The Effect of Feedback Information on Decisions in Experimental Coordination Games," in preparation.

Camerer, Colin F., and Teck-Hua Ho (1999), "Experience-Weighted Attraction in Games," Econometrica, 67, 827-874.

Camerer, Colin F., Juin-Kuan Chong and Teck-Hua Ho (2002), "Sophisticated Experience-Wieghted Attraction Learning and Strategic Teaching in Repeated Games," Journal of Economic Theory, 104, 1, 16-47.

Ciccone, Antonio and K. Matsuyama (1996), "Start-up Costs and Pecuniary Externalities as Barriers to Economic Development", Journal of Development Economics, 49, 33-59.

Cooper, David J. and John H. Kagel (2003), "The Impact of Meaningful Context on Strategic Play in Signaling Games", Journal Of Economic Behavior and Organization, 50, 311-337.

Cooper, David J. and John H. Kagel (2004), "Learning and Transfer in Signaling Games," mimeo, Case Western Reserve University.

Crawford, Vince (1991), "An 'Evolutionary' Interpretation of Van Huyck, Battalio, and Beil's Experimental Results on Coordination", Games and Economic Behavior, 3, 25-59.

Feltovich, Nicholas J. (2000), "Reinforcemnet-Based vs. Beliefs-Based Learning In Experimental Asymmetric Information Games," Econometrica, 68, 605-641.

Foster, Lawrence W. and David Ketchen (1998), Advances in Applied Business Strategy. Turnaround Research: Past Accomplishments and Future Challenges, Stamford, Connecticut: JAI Press.

Goeree, Jacob K. and Charles A. Holt (2001), "Ten Little Treasures of Game Theory and Ten Intuitive Contradictions", American Economic Review, 91, 5, 1402-1422.

Hirschman, A. O. (1958), The Strategy of Economic Development, New Haven, CT: Yale University Press.

Ichniowski, C., K. Shaw, and G. Prennushi (1997), "The Effects of Human ResourceManagement Practices on Productivity: A Study of Steel Finishing Lines," American Economic Review, **87**, 3, pp. 291-313.

Knez, Marc and Colin Camerer (1994), "Creating Expectational Assets in the Laboratory: Coordination in 'Weakest-Link' Games," Strategic Management Journal, 15, 101-119.

Knez, Marc and Colin Camerer (2000), "Increasing Cooperation in Prisoner's Dilemmas by Establishing a Precedent of Efficiency in Coordination Games," Organizational Behavior and Human Decision Processes, 82, 2, 194 – 216.

Knez, Marc and Duncan Simester (2002), "Form-Wide Incentives and Mutual Monitoring At Continental Airlines", Journal of Labor Economics, 19, 4, 743-772.

Kremer, Michael (1993), "The O-Ring Theory of Economic Development," Quarterly Journal of Economics, 107, 551- 575.

Liang, Kung-yee, and Scott L. Zeger (1986), “Longitudinal Data Analysis Using Gerealized Linear Models,” Biometrika, 73 , 13-22.

McKelvey, Richard M. amd Thomas R. Palfrey (1995), “Quantal Response Equilibria in Normal Form Games,” Games and Economic Behavior, 7, 6-38.

Murphy, Kevin M., Andrei Shleifer and Robert W. Vishny (1989), “Industrialization and the Big Push”, The Journal of Political Economy, 97, 5, 1003-1026.

Pettigrew, Andrew M., Richard W. Woodman and Kim S. Cameron (2001), Studying Organizational Change and Development: Challenges for Future Research”, Academy of Mangement Journal, 44, 4, 697-713.

Rosenstein-Rodan, Paul (1943), “Problems of Industrialization of Eastern and South-eastern Europe”, Economic Journal, 53, 202-211.

Salmon, Timothy C. (2001), “An Evaluation of Economtric Models of Adaptive Learning,” Econometrica, 69, 1597-1628.

Schelling, T. (1960), The Strategy of Conflict, Cambridge, MA: Harvard University Press.

Stahl, Dale O. (1999), “Sophisticated Learning and Learning Sophistication”, mimeo.

Van Huyck, John B., Raymond Battalio and Richard Beil (1990), “Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure”, Amercian Economic Review, 80, 1, 234-248.

Van Huyck, John B., Raymond Battalio and Richard Beil (1991), “Strategic Uncertainty, Equilibrium Selection, and Coordination Failure in Average Opinion Games”, 106, 3, 885-911.

Weber, Roberto (2002), “Managing Growth to achieve efficient coordination in large groups,” mimeo.

Weber, Roberto A., Colin F. Camerer and Marc Knez (2004), “Timing and Virtual Observability in Ultimatum Bargaining and ‘Weak Link’ Coordination Games,” Experimental Economics, 7, 25-48.

Weick, Karl E. and Robert E. Quinn (1999), “Organizational Change and Development”, Annual Review of Psychology, 50, 361-386.

Wilcox, Nathaniel T. (2003), “Heterogeneity and Learning Principles”, mimeo.

APPENDIX
INSTRUCTIONS FOR CELL 2 OF THE EXPERIMENTAL DESIGN

General information: The purpose of this experiment is to study how people make decisions in a particular situation. From now on and till the end of the experiment any communication with other participants is not permitted. If you have a question, please raise your hand and one of us will come to your desk to answer it.

You will receive 500 pesetas for showing up on time for the experiment. In addition, you will make money during the experiment. Upon completion of the experiment the amount that you make will be paid to you in cash. Payments are confidential; no other participant will be told the amount you make.

Parts, Rounds and Groups: This experiment will have several *parts*. In Part 1 there will be 10 *rounds*. After these ten rounds have finished, we will give you instructions for the next part of the experiment. In each round you will be in a *group* with 3 other participants. The participants you are grouped with will be the same in all rounds.

Description of the Decision Task(*s*) in Part 1 of the Experiment: You and the other members of your group are employees of a firm. You can think of a round of the experiment as being a workweek. In each week, each of the employees in each firm spends 40 hours at the firm. You have to choose how to allocate your time between two activities, Activity A and Activity B. Specifically, you will be asked to choose how much time to devote to Activity A. The available choices are 0 hours, 10 hours, 20 hours, 30 hours and 40 hours. Your remaining hours will be put towards Activity B. For example, if you devote 30 hours to Activity A, this means that 10 hours will be put towards Activity B.

For each round of the experiment you will receive a flat wage and a bonus that depends on the *minimum* number of hours spent on Activity A by a member of your group. For all rounds of this experiment, the flat wage equals 200 pesetas. The bonus rate, B, may vary between rounds. The bonus rate is selected by the firm manager. In this experiment, the firm manager is being played by the computer. We will always let you know the bonus rate before you choose how many hours to devote to Activity A.

Payoffs: The payoff that an employee receives in a round depends on the number of hours he chooses to spend on Activity A, the number of hours chosen by the others in his firm to spend on Activity A, and the bonus rate B selected by the firm manager. The payoff for the i^{th} employee of the firm, π_i, is given given by the formula below where H_i is the number of hours spent by the i^{th} employee of the firm on Activity A and $\min(H_A)$ is the *smallest* number of hours an employee of the firm spends on Activity A. You do not need to memorize this formula – the computer program will give you payoff tables at any point where you need to make a decision.

$$\pi_i = 200 - (5 * H_i) + (B * \min(H_A))$$

The firm manager's payoff depends on the number of hours spent on Activity A by the employees of the firm and the bonus rate B. The firm manager's payoffs, π_F, is given by the following formula. Again, $\min(H_A)$ is the *smallest* number of hours an employee of the firm spends on Activity A. (Recall that the firm manager is actually being played by the computer. Nobody is actually receiving these payoffs.) Do not worry about memorizing this formula, as the program displays the firm's payoff function any time you make a decision and the computer automatically calculates the firm manager's payoffs for you as a part of the feedback you receive after each round.

$$\pi_F = 100 + (60 - 4B) * \min(H_A)$$

Playing a Round: For each round of the experiment, the computer will display a screen like the one shown below. The payoffs shown in the payoff table will be adjusted for the changing values of B. For the example below, we set B = 8. Notice that this is displayed above the payoff table.

Each employee will choose a number of hours to spend on Activity A using the buttons on the right hand side of the screen. You may change your choices as often as you like, but once you click on "Enter" your choice is final. Note that when you make your decision you will *not* know what the other employees in your firm are doing in the round.

At no point in time will we identify the other employees in your firm. In other words, the actions you take in this experiment will remain confidential.

[Insert Screen Dump Here]

<u>Information that you will receive</u>: After each round you will be informed about the number of hours you have spent on Activity A, the lowest number chosen by all of the employees in your firm, the firm's payoff, your payoff for the latest round, and your accumulated payoffs through the current round. You will also be shown the decisions by you and the decisions of all the other employees of your group from the current and previous rounds.

<u>Payment</u>: At the end of the experiment you will be paid, in cash, the sum of the payoffs that you will have earned in the rounds of the experiment. As noted previously, you will be paid privately and we will not disclose any information about your actions or your payoff to the other participants in the experiment.

Payoff Quiz

Before we begin the experiment, please answer the following questions. For all of these questions, assume that B = 8. This gives employees the payoff table shown below. We will go through the answers to a sample problem before you do the rest of the quiz. Please raise your hand if you are having trouble answering one of the questions.

Bonus Rate = 8

Firm Payoff = 100 + 28*min(H_A)

		Minimum Hours Spent on Activity A by Other Employees				
		0	10	20	30	40
My Hours on Activity A	0	200	200	200	200	200
	10	150	230	230	230	230
	20	100	180	260	260	260
	30	50	130	210	290	290
	40	0	80	160	240	320

Sample Question: Suppose you choose to spend 10 hours on Activity A. The other employees choose to spend 30, 20, and 40 hours on Activity A.

The minimum number of hours an employee of the firm spends on Activity A is ______

Your payoff is ________ pesetas.

1) Suppose you choose to spend 20 hours on Activity A. The other employees choose to spend 30, 0, and 10 hours on Activity A.

The smallest number of hours an employee of the firm spends on Activity A is ______

Your payoff is ________ pesetas.

2) Suppose you choose to spend 0 hours on Activity A. The other employees choose to spend 20, 30, and 10 hours on Activity A.

The smallest number of hours an employee of the firm spends on Activity A is ______

Your payoff is ________ pesetas.

3) I am grouped with the same three individuals for all thirty rounds of the experiment (True/False)? ________

4) My actions and payoffs will be confidential (True/False)? ________

Instructions for Part 1

For the next ten rounds, the bonus rate B equals **6**.

This bonus rate was set by the computer, not by another human participant.

Instructions for Part 2

For the next ten rounds, the bonus rate B equals **10**.

This bonus rate was set by the computer, not by another human participant.

Instructions for Part 3

For the next ten rounds, the bonus rate B equals **10**.

This bonus rate was set by the computer, not by another human participant.

Table 1
Employee i's Payoff Table, B = 6

		Minimum Effort by Other Employees				
		0	**10**	**20**	**30**	**40**
Effort By Employee i	**0**	200	200	200	200	200
	10	150	210	210	210	210
	20	100	160	220	220	220
	30	50	110	170	230	230
	40	0	60	120	180	240

Table 2
Employee i's Payoff Table, B = 14

		Minimum Effort by Other Employees				
		0	**10**	**20**	**30**	**40**
Effort By Employee i	**0**	200	200	200	200	200
	10	150	290	290	290	290
	20	100	240	380	380	380
	30	50	190	330	470	470
	40	0	140	280	420	560

Table 3
List of Treatments

	Cell 1	Cell 2	Cell 3	Cell 4	Cell 5	Cell 6
Bonus Rate Rounds 1 – 10	6	6	6	6	6	6
Bonus Rate Rounds 11 – 20	14	10	8	14	14	6
Bonus Rate Rounds 21 – 30	14	10	8	10	6	14

Table 4
Effect of Immediate Reaction to Bonus Increase

Number of Strong Responders in Round 11	Number of Observations	Average Minimum Effort Round 20
1	9	15.56
2	15	20.67
3	9	28.89
4	5	30.00

Table 5
Ordered Probit Regressions on Data from Cells 1 - 6, Rounds 11 - 20

Data Type	Firm-Level Data Minimum Effort		Employee-Level Data Effort	
Controls for Individual Effects	Clustering	Random Effects	Clustering	Nested Random Effects
Rounds 11 – 15 * B = 8	1.468*** (.416)	1.395*** (.376)	1.257*** (.385)	1.702*** (.227)
Rounds 11 – 15 * B = 10	1.427*** (.336)	1.505*** (.373)	1.249*** (.227)	1.244*** (.212)
Rounds 11 – 15 * B = 14	1.470*** (.311)	1.575*** (.345)	1.266*** (.246)	2.157*** (.213)
Rounds 16 – 20	-.052 (.041)	-.103 (.391)	-.404*** (.111)	-.751*** (.151)
Rounds 16 – 20 * B = 8	2.137*** (.587)	2.985*** (.419)	1.685*** (.541)	3.282*** (.252)
Rounds 16 – 20 * B = 10	2.567*** (.466)	3.422*** (.408)	2.120*** (.444)	3.240*** (.234)
Rounds 16 – 20 * B = 14	1.796*** (.346)	2.360*** (.352)	1.270*** (.291)	2.938*** (.219)
Barcelona	-.009 (.261)	-.233 (.173)	-.064 (.242)	-.440*** (.109)
Minimum Effort Round 10	.067*** (.016)	.157*** (.013)	.058*** (.014)	.121*** (.008)
Log-likelihood	-708.86	-528.37	-2821.90	-2073.39
# Observations	580	580	2320	2320

*** Significant at 1% level
** Significant at 5% level
* Significant at 10% level

Table 6
Ordered Probit Regressions on Data from Cells 1, 4, and 5, Rounds 6 – 10 and 21 – 30

Data Type	Firm-Level Data Minimum Effort		Employee-Level Data Effort	
Controls for Individual Effects	Clustering	Random Effect	Clustering	Nested Random Effects
Rounds 21 – 25 * B = 6	1.646*** (.527)	3.600*** (.361)	1.279*** (.316)	1.751*** (.125)
Rounds 21 – 25 * B = 10	1.276*** (.447)	1.882*** (.344)	.819** (.352)	.540*** (.151)
Rounds 21 – 25 * B = 14	1.123** (.507)	.937*** (.329)	.706** (.358)	.874*** (.170)
Rounds 26 – 30 * B = 6	1.707*** (.548)	3.870*** (.378)	.799* (.410)	1.095*** (.123)
Rounds 26 – 30 * B = 10	2.155*** (.605)	3.678*** (.437)	1.855*** (.596)	1.912*** (.167)
Rounds 26 – 30 * B = 14	.675 (.525)	.092 (.330)	.566 (.428)	.779*** (.169)
Barcelona	-.731** (.340)	-.429** (.176)	-.529*** (.235)	-.137 (.108)
Minimum Effort Round 5	.046* (.025)	.029*** (.010)	.031 (.022)	.002 (.005)
Minimum Effort Round 20	.075*** (.014)	.047*** (.009)	.055*** (.006)	.009* (.005)
Log-likelihood	-335.90	-255.26	-1681.55	-1509.22
# Observations	429	429	1716	1716

*** Significant at 1% level
** Significant at 5% level
* Significant at 10% level

Table 7
Ordered Probit Regressions on Cells 1, 4, and 5, round 11 – 20 and Cell 6, rounds 21 - 30

Data Type	Firm-Level Data Minimum Effort		Employee-Level Data Effort	
Controls for Individual Effects	Clustering	Random Effect	Clustering	Nested Random Effects
Rounds 21 – 25 (Late Increase)	-.197 (.365)	-.026 (.232)	.002 (.134)	-.073 (.135)
Rounds 16 – 20 (Early Increase)	.326** (.129)	.741*** (.154)	-.169 (.293)	.028 (.079)
Rounds 26 – 30 (Late Increase)	-.465 (.442)	-.606** (.245)	-.478 (.427)	-.746*** (.143)
Barcelona	.029 (.320)	-.386** (.169)	.051 (.299)	-.278*** (.104)
Minimum Effort Last B = 6 Round	.049*** (.015)	.164*** (.016)	.041*** (.013)	.071*** (.008)
Log-likelihood	-563.04	-390.89	-2189.30	-1554.54
# Observations	389	389	1556	1556

*** Significant at 1% level
** Significant at 5% level
* Significant at 10% level

Table 8
Comparison of QRE Predictions w/ Data, Average Effort Level

Bonus Rate	QRE Prediction	Rounds 1 – 10	Rounds 11-15	Rounds 16 – 20	Rounds 11 - 20
6	12.28	12.31	14.60	8.40	11.50
8	12.62	---	29.40	32.60	31.00
10	13.01	---	24.22	34.67	29.44
14	14.06	---	26.69	25.72	26.21

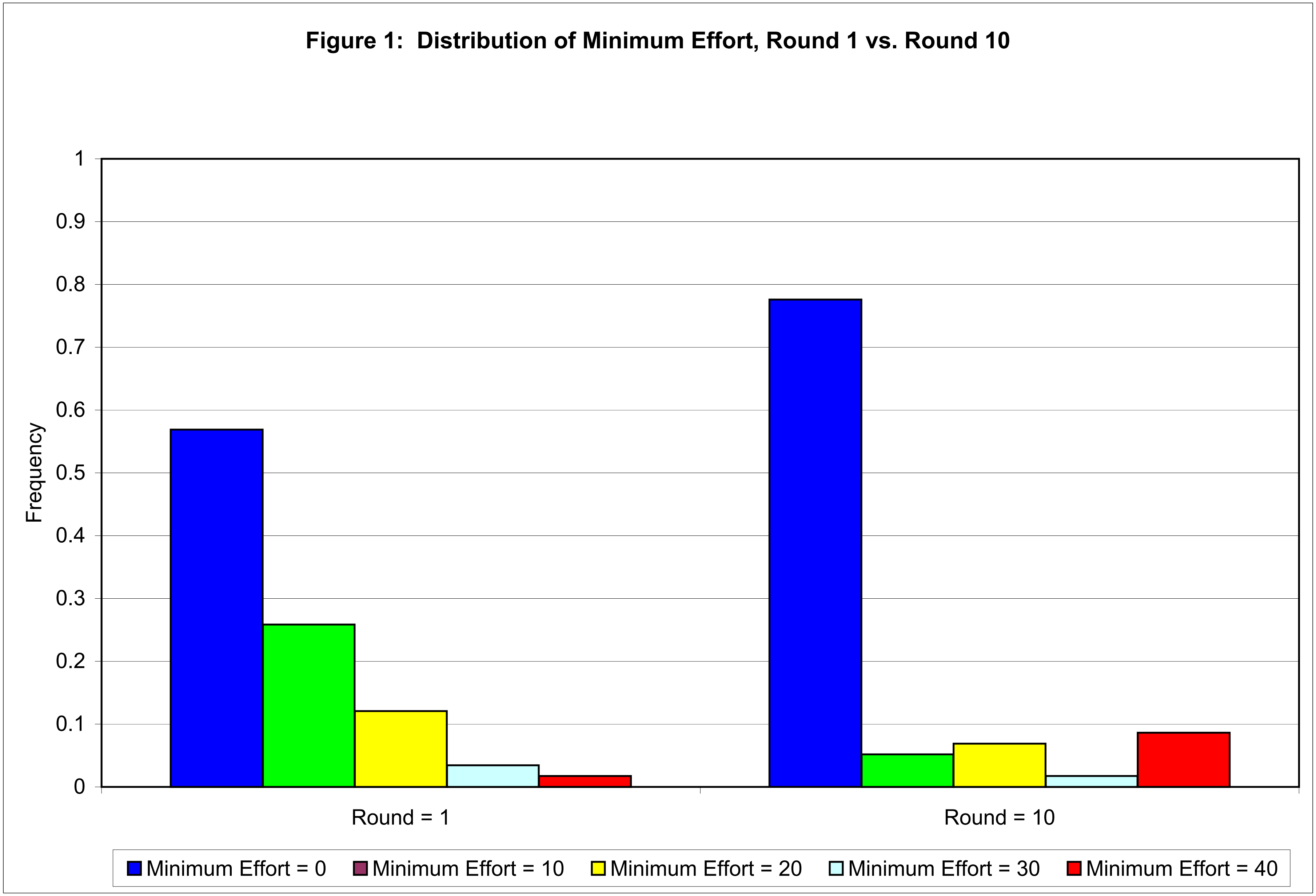


Figure 1: Distribution of Minimum Effort, Round 1 vs. Round 10
Frequency
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Round = 1
Round = 10
Minimum Effort = 0
Minimum Effort = 10
Minimum Effort = 20
Minimum Effort = 30
Minimum Effort = 40

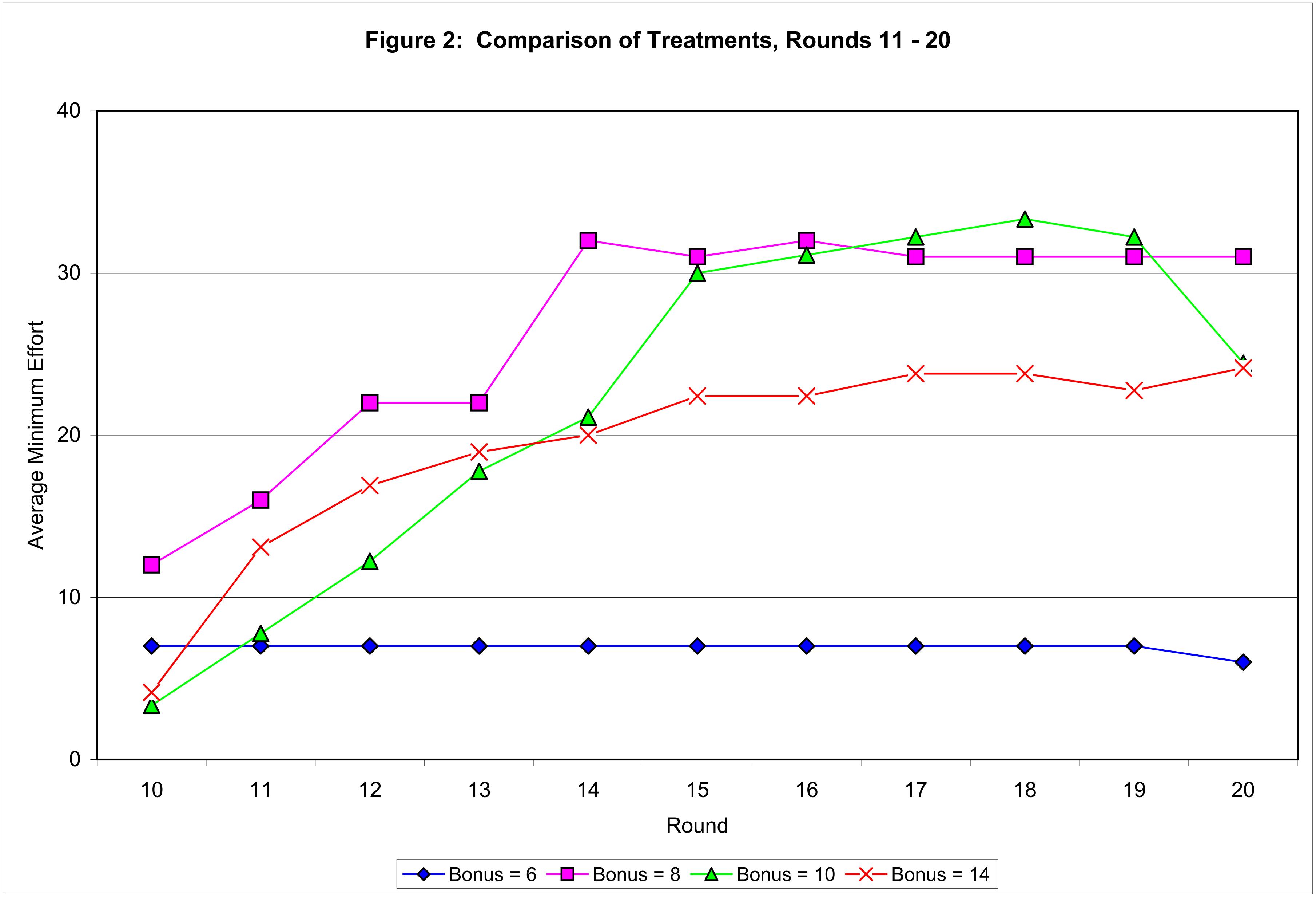


Figure 2: Comparison of Treatments, Rounds 11 - 20
Average Minimum Effort
40
30
20
10
0
10
11
12
13
14
15
16
17
18
19
20
Round
Bonus = 6
Bonus = 8
Bonus = 10
Bonus = 14

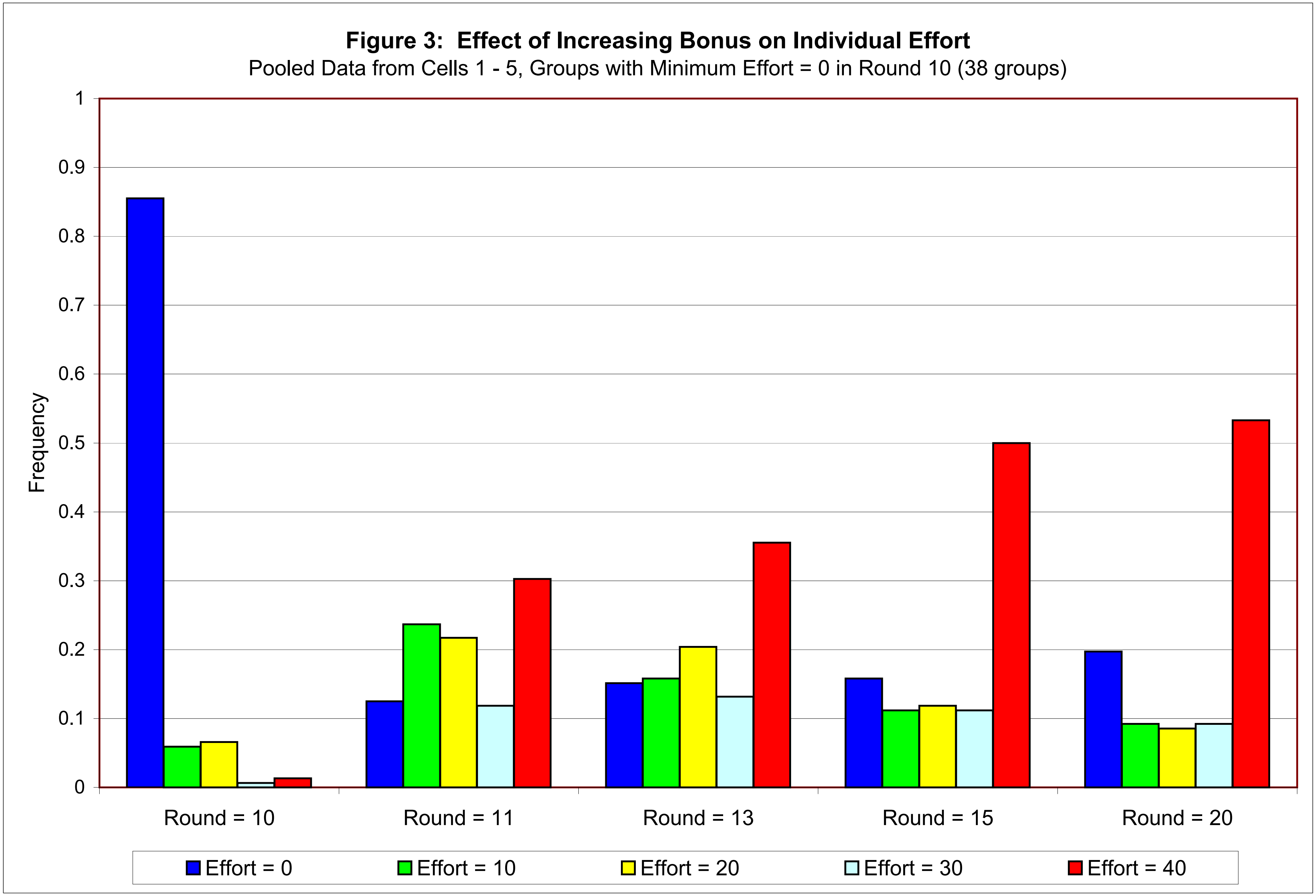
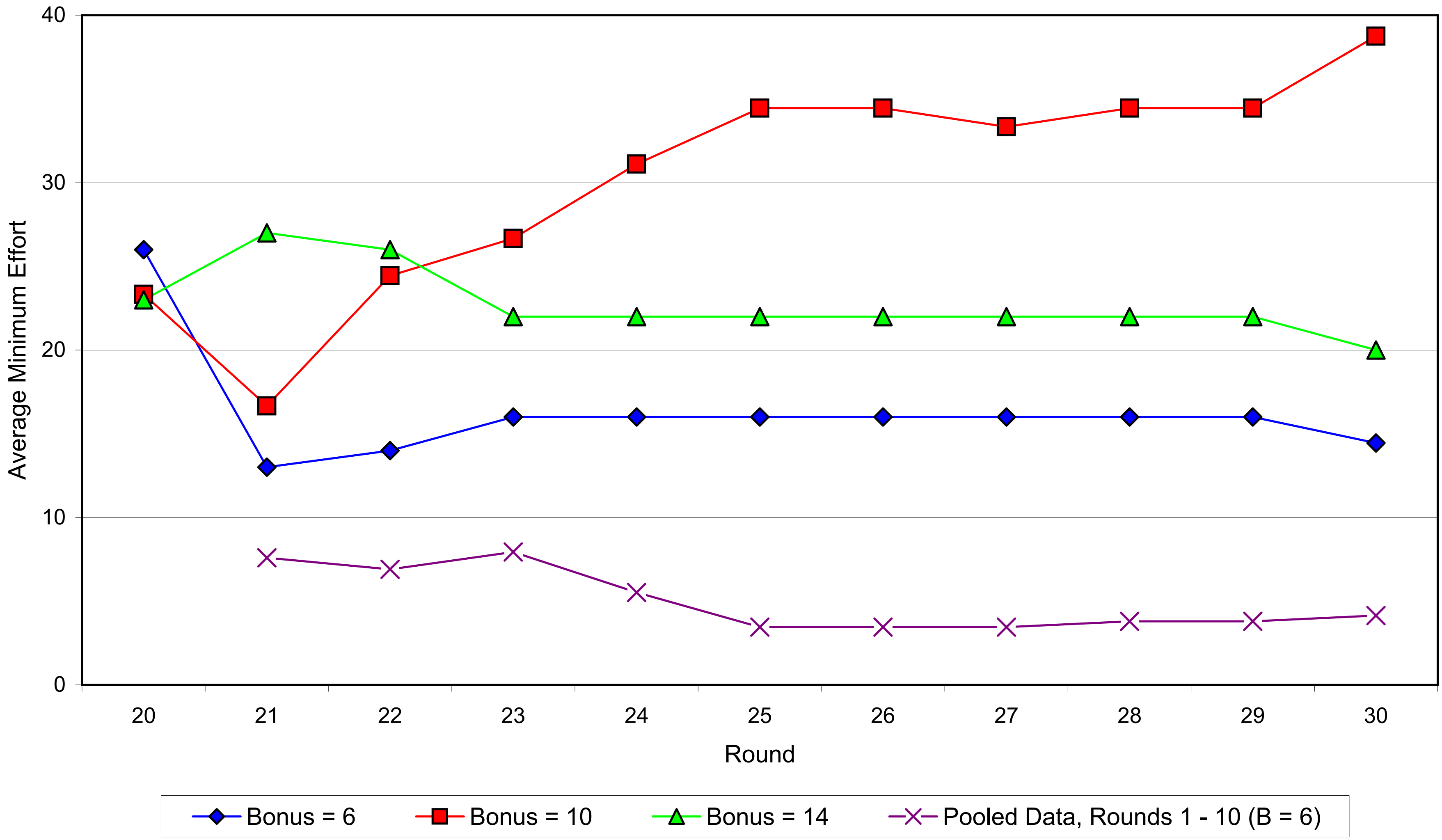


Figure 3: Effect of Increasing Bonus on Individual Effort
Pooled Data from Cells 1 - 5, Groups with Minimum Effort = 0 in Round 10 (38 groups)
Frequency
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Round = 10
Round = 11
Round = 13
Round = 15
Round = 20
Effort = 0
Effort = 10
Effort = 20
Effort = 30
Effort = 40

Figure 4: Comparison of Treatments, Rounds 21 - 30 and Control



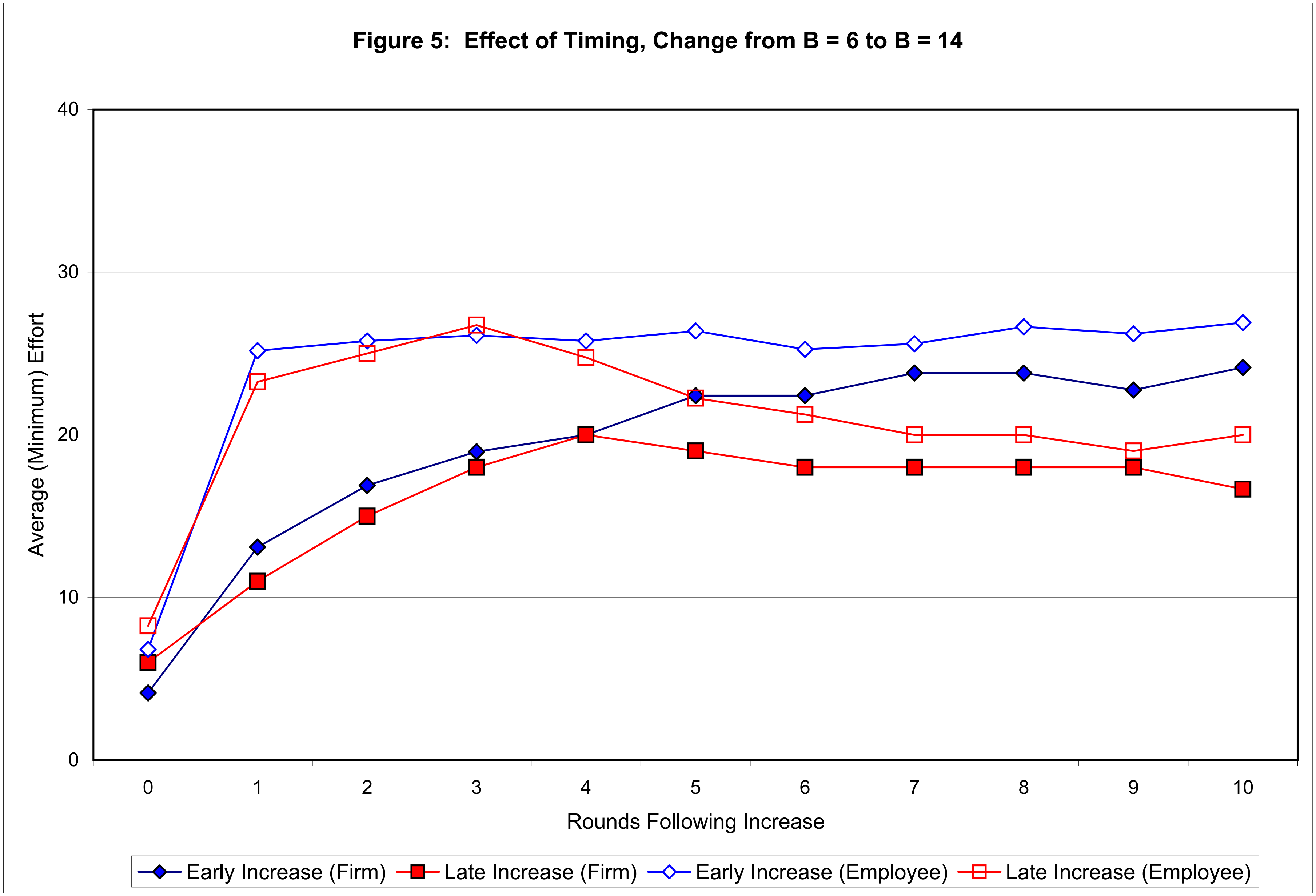


Figure 5: Effect of Timing, Change from B = 6 to B = 14
Average (Minimum) Effort
0
10
20
30
40
0 1 2 3 4 5 6 7 8 9 10
Rounds Following Increase
Early Increase (Firm)
Late Increase (Firm)
Early Increase (Employee)
Late Increase (Employee)

Figure 6
Comparison of EWA with Sophisticated Types Simulations with Experimental Data