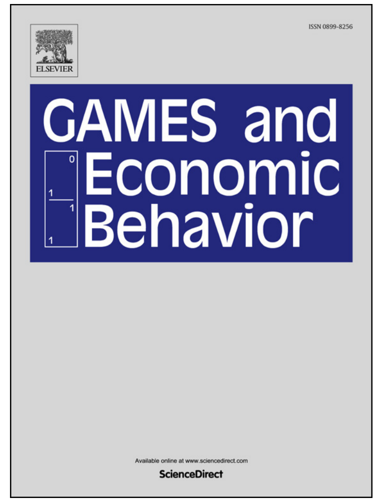
Accepted Manuscript



Endogenous incentive contracts and efficient coordination

David J. Cooper, Christos A. Ioannou, Shi Qi

PII: S0899-8256(18)30122-2
DOI: https://doi.org/10.1016/j.geb.2018.07.008
Reference: YGAME 2887

To appear in: *Games and Economic Behavior*

Received date: 18 July 2017

Please cite this article in press as: Cooper, D.J., et al. Endogenous incentive contracts and efficient coordination. *Games Econ. Behav.* (2018), https://doi.org/10.1016/j.geb.2018.07.008

Endogenous Incentive Contracts and Efficient Coordination

David J. Cooper Christos A. Ioannou Shi Qi

July, 2018

Abstract

We examine the effects of endogenous assignment to incentive contracts on worker productivity. Assignment to high performance pay via a market mechanism is roughly twice as effective as imposing the same contract exogenously. This positive effect is largely offset by a negative effect for workers that endogenously choose low performance pay. We decompose the positive effect of endogenous assignment to high performance pay into effects due to selection and strategic anticipation, and find that selection has a greater effect than strategic anticipation. We use a Reverse Sort treatment to show that the effect of selection is sufficiently strong to overcome the direct effect of lower performance pay, yielding coordination at high effort levels in spite of low incentives.

Keywords: Experiment, coordination, incentive contracts, selection
JEL Classification Codes: C92, J31, M52, C73, C51

Acknowledgements: We would like to thank the NSF (SES-1127704) for financial support. We thank Joe Ballard, Brent Davis, Phil Brookins and Ellis Magee for their work as research assistants on this project. We thank Jordi Brandts, Vince Crawford, Florian Englmaier, Tony Kwasnica, Arno Riedl, Ernan Haruvy and Katya Sherstyuk for helpful discussion of this paper as well as seminar participants at the SEA meetings, Florida State, Heidelberg, Southampton, Monash, Sydney, Oxford, Nottingham, Mannheim, Amsterdam, GATE, Pittsburgh, Penn State, and the NBER. Finally, we would like to thank the editors and referees for their insightful comments, which significantly improved the paper. Any errors are solely our own.

David J. Cooper	Christos A. Ioannou	Shi Qi
Department of Economics Florida State University, Tallahassee, FL 32306-2180, USA School of Economics University of East Anglia, Norwich Research Park Norwich NR4 7TJ UK Phone : 850-644-7097 djcooper@fsu.edu	Universite Paris 1 Pantheon - Sorbonne Centre d'Économie de la Sorbonne Maison des Sciences Economiques, 106-112 Boulevard de l'Ĥopital, 75647 Paris Cedex 13, France christos.a.ioannou@gmail.com	Economics Department College of William and Mary, Williamsburg, VA 23187, USA Phone: 757-221-4645 sqi01@wm.edu

1 Introduction

The performance of many work groups is constrained by a single weak link. A paper cannot be completed until all co-authors finish their assigned sections, a meeting cannot start until all critical personnel are present, and an assembly line moves no faster than its slowest worker. If pay is based on group performance, as must be the case when only output rather than individual effort is observed, the strong complementarities generated by a weak-link technology can cause *productivity traps* where pessimistic beliefs create a self-fulfilling prophecy: nobody works hard in the (correct) expectation that any effort will be wasted given that no individual can unilaterally improve productivity. Escaping such a trap is difficult since coordinated change by *all* members of the group is needed to increase productivity.

A number of mechanisms have been proposed to achieve the coordinated increase in effort needed to escape a productivity trap (defined as coordination at low effort levels). A simple option that has consistently proven effective is increasing incentives to coordinate at high rather than low effort. This has been shown to help groups escape productivity traps in lab studies of the weak-link game (Brandts and Cooper, 2006; Hamman, Rick, and Weber, 2007; Brandts, Cooper, and Weber, 2014) as well as in field settings (Knez and Simester, 2001). The preceding lab studies feature a common element: assignment of individuals to incentive contracts is random and exogenous. This was a natural first step for the literature, but in labor markets, where workers frequently choose between jobs with different incentive contracts, assignment to incentive contracts is often endogenous. The primary purpose of the experiments presented below is to explore the effect of having the assignment of individuals to incentive contracts take place endogenously through a market mechanism.

All of our experiments feature an initial phase in which subjects repeatedly play a weak-link game under an incentive contract featuring high fixed pay and low incentives to coordinate at high effort levels ("low performance pay"). This reliably induces a productivity trap. In the second phase of the experiment, half of the subjects continue with the initial contract and half are assigned to a new incentive contract with lower fixed pay and larger incentives to coordinate at high effort levels ("high performance pay"). Treatments vary by whether the assignment of subjects to incentive contracts for the second phase is random or endogenous through a market mechanism.

The positive effect of introducing high performance pay almost doubles with endogenous assignment. However, a matching negative effect from endogenous assignment to low performance pay largely offsets the positive effect of endogenous assignment to high performance pay. The *total* effect of endogenous assignment to incentive contracts, averaging over both contracts, is virtually zero.

Beyond identifying the effects of endogenous assignment to incentive contracts, a major goal of our study is to understand why high performance pay is more effective when assigned endogenously. In work environments with the weak-link property, strategic uncertainty makes incentive contracts "fragile." Individuals may fail to respond to high performance pay if they fear others won't increase effort in response. Endogenous assignment to high performance pay facilitates coordination at the highest possible effort level ("efficient coordination") by reducing strategic uncertainty through two channels, selection and strategic anticipation. "Selection" refers to the tendency of market mechanisms to assign high performance pay to individuals ("optimists") who have inherently optimistic prior beliefs about the chance of coordination at the efficient outcome *independent of how individuals are assigned to groups*. Because long-run outcomes in coordination games are largely driven by initial beliefs (see Van Huyck, Battalio, and Beil, 1990 and 1991), systematic assignment of optimists to high performance pay increases the chance of efficient coordination. Beyond selection, participants can infer that assignment to high performance pay is *not* random based on the information available in a market mechanism. "Strategic anticipation" is the ability of an individual to (correctly) anticipate that optimists are more likely to be assigned high performance pay by a market mechanism than with random assignment.[1] Individuals with strategic anticipation understand the effects of selection and become more optimistic under high performance pay *when contract assignment is endogenous*. This increased optimism improves the odds of efficient coordination with high performance pay.

We use an innovative "Sort" treatment to measure how much of the positive effect of endogenizing assignment to high performance pay is due to selection rather than strategic anticipation. In this treatment, subjects' characteristics and initial choices are used to predict which incentive contract they would have been assigned by the market mechanism. We then exogenously imple-

[1] This is related to forward induction, as both rely on individuals understanding that past choices reveal information about others' beliefs, but is not based on iterated removal of dominated strategies (Ben-Porath and Dekel, 1992).

ment the predicted contract assignments to imperfectly reproduce the market's outcome. The Sort treatment preserves the effects of selection as it inherits the market's tendency to assign optimists to high performance pay, but eliminates the effects of strategic anticipation by depriving subjects of any information that allows them to anticipate the selection process. We find that selection plays a larger role than strategic anticipation in increasing efficient coordination with endogenous assignment to high performance pay.

The effect of selection is so strong that it can overcome the direct effect of lower incentives to coordinate at high effort levels. We demonstrate this through a "Reverse Sort" treatment that flips the contract assignments from the Sort treatment, switching the sign of the selection effect. This results in higher effort levels with low performance pay than high performance pay!

Our work contributes to two strands of the existing literature. Both lab and field studies have examined how endogenous choice of performance pay schemes (e.g. piece-rate systems) affects performance in individual tasks. These studies generally find that such schemes improve productivity and that the majority of this positive effect is due to selection of more able individuals (i.e. Lazear, 2000; Cadsby, Song, and Tapon, 2007; Eriksson and Villeval, 2008; Dohmen and Falk, 2011; Bandiera, Guiso, Prat, and Sadun, 2015). Our work has similar findings – endogenous choice of high performance pay improves productivity with selection accounting for most of the effect – but differs from the existing literature because of the critical role played by strategic uncertainty. Selection is based on individuals' beliefs rather than their ability to perform a task. The market mechanism assigns those who are inherently optimistic about the likelihood of efficient coordination to groups with high performance pay. Group productivity in field settings generally does not just depend on skill, but also depends on being able to work together. It makes sense that selection takes place along the latter dimension as well as the former.

Our experiments also relate to the literature on buying the right to play a coordination game, notably Van Huyck, Battalio, and Beil (1993). Van Huyck *et al.* study repeated play of a two-stage game where players bid for the right to play a median game. They find that winning bids converge to the payoff from the efficient equilibrium and winning subjects' play in the median game converges to the efficient equilibrium. Van Huyck *et al.* attribute the latter result to forward induction, while Crawford and Broseta (1998) argue that it reflects an interaction between learning, forward induction, and an optimistic subject effect analogous to selection. Our result that endogenous assignment to high performance pay improves the likelihood of efficient coordination is obviously related, but we reach different conclusions about the source of this effect. Using the Sort treatment, we show that the effect of endogenizing contract assignment is primarily due to selection rather than strategic anticipation (taking the place of forward induction). This has an important implication for the interpretation of our results. Selection effects do not rely on workers understanding the selection process, and hence are more likely to carry over to field settings.

A feature that differentiates our experiments from the existing literature is that the outside option for individuals who are not assigned high performance pay by the auction is continued play with *low* performance pay contracts. For experiments with endogenous assignment to performance pay, the outside option is typically either a flat payment or individual work under pre-existing incentives (usually flat-rate pay). The same forces of selection and strategic anticipation that push groups assigned high performance pay towards efficient coordination also lead groups assigned low performance pay to coordinate at inefficient outcomes. The negative effect of endogenous assignment to low performance pay offsets the positive effect of endogenous assignment to high performance pay, making the total effect of endogenous assignment neutral. This counterbalancing effect cannot occur with a fixed outside option.

The preceding implies a link between the total effect of endogenous assignment to incentive contracts and labor mobility. Our experiments model a "closed" labor market, meaning that workers can easily move between firms within the market, but cannot easily enter and exit the market. High performance pay works in large part by attracting optimists to a firm. In a closed market, those optimists must come from other firms within the industry, causing a negative externality. Contrast this with an "open" market where workers can easily move into and out of the market. Optimists can move into the market and pessimists can move out in response to changing incentive contracts, reducing the negative externality caused by a firm adopting high performance pay. We conjecture that many labor markets are closed in the short run and open in the long run.[2] This suggests that the total effect of endogenous assignment to high performance pay becomes more positive over time. More broadly, our results illustrate the importance of considering what happens to individuals who do *not* select into high performance pay as well as what happens to those who do.

[2]For example, consider a market containing jobs that require a large amount of specialized training. In the short run it is hard to train new workers, and those who leave the market sacrifice the value of their specialized skills. In the long run more workers can be trained and brought into the market.

We study a simple environment with a single mechanism for improving productivity – high performance pay. This is not meant to dismiss the importance of other mechanisms that have been studied as tools for overcoming productivity traps. Given that escaping a productivity trap is non-trivial, it presumably makes sense to use more than one instrument. We show that increased performance pay is a more powerful instrument than previous experiments indicate.

This paper is organized as follows. Section 2 describes related research. Section 3 lays out the experimental design, procedures, and hypotheses. Section 4 presents the main results of the paper. Section 5 concludes.

2 Related Literature

Since the seminal work of Van Huyck, Battalio, and Beil (1990), it has been well known that achieving efficient coordination in weak-link games is unlikely when the number of players is large and no mechanism exists for promoting efficient coordination. Using performance pay to increase incentives for efficient coordination has proven a reliable method of escaping productivity traps. Many other mechanisms intended to promote efficient coordination have also been shown to be effective. These include costless pre-play communication (Cooper, De Jong, Forsythe, and Ross, 1992; Duffy and Feltovich, 2002; Duffy and Feltovich, 2006; Blume and Ortmann, 2007; Brandts and Cooper, 2007; Cason, Sheremeta, and Zhang, 2012; Brandts, Cooper, and Weber, 2014), inter-generational advice (Chaudhuri, Schotter, and Sopher, 2009), competition between groups (Bornstein, Gneezy, and Nagel, 2002; Myung, 2012), gradual growth of groups (Weber, 2006), help with commitment (Brandts, Cooper, Fatas, and Qi, 2016), and endogenous group formation (Riedl, Rohde, and Strobel, 2016; Salmon and Weber, 2017).

The experiments described below feature a market mechanism that assigns subjects endogenously to different incentive contracts. A similar mechanism is used by Van Huyck, Battalio, and Beil (1993), the first paper to show that paying for the right to play a coordination game leads to efficient coordination. In Van Huyck *et al.*, subjects play in fixed groups of 18. The treatment of interest has a two-stage game in each round. In the first stage, subjects bid in an English clock auction for the right to play a nine-player median game (Van Huyck, Battalio, and Beil (1991)). The nine winners play the median game in the second stage. Within a few rounds, play in the median game reliably converges to the efficient equilibrium. Median games are not as challenging an environment as weak-link games, but it is nevertheless highly unlikely that the efficient equilibrium would emerge spontaneously. Van Huyck *et al.* attribute the positive effect of the auctions to forward induction. Later papers (Broseta, Fatas, and Neugebauer, 2003 and Sherstyuk, Karmanskaya, and Teslia, 2014) report similar results.[3]

Our work differs from Van Huyck *et al.* along a number of dimensions. The most important is that our experiments are designed to disentangle why endogenous assignment to groups via a market mechanism increases efficient coordination. We show that strategic anticipation, a closely related concept to forward induction, explains little of the effect. Another important difference is that subjects who don't win the auction still participate in a coordination game. This lets us measure the total effect of endogenous assignment to incentive contracts.

Crawford and Broseta (1998) develop a structural model to study the learning process in Van Huyck *et al.* This is different from the learning process in our experiments, because the Van Huyck *et al.* design features many auctions rather than one, leading to a feedback loop between the auctions and the coordination game. Forward induction and selection of optimistic subjects increase the effort chosen by auction winners. This causes higher future bids as subjects' expectations about the outcome of the coordination game become more optimistic. Crawford and Broseta attribute the efficient coordination outcome mainly to two major components, forward induction and a combination of an "optimistic subjects" effect and a "robustness" effect. The optimistic subjects effect captures the tendency of subjects who are optimistic *ex ante* about the outcome of the coordination game to win the auction, and the robustness effect captures the interaction between learning and strategic uncertainty. Crawford and Broseta estimate the impacts of forward induction, the optimistic subjects effect, and the robustness effect by fitting their model to the data and then running simulations. They attribute roughly half of the improvement with auctions to forward induction and half to the combination of the optimistic subjects and robustness effects. For technical reasons it is not possible to separate the latter two effects. As previously mentioned, our experimental design lets us directly measure the effects of selection (similar to the optimistic subjects effect) and strategic anticipation (in lieu of forward induction). Our results place relatively low weight on strategic anticipation.

Cachon and Camerer (1996) argue that loss aversion plays a major role in the Van Huyck *et*

[3]Kogan, Kwasnica, and Weber (2011) show that incorporating futures markets harms efficiency in the coordination game, with the effect coming through beliefs rather than direct incentive effects.

al. results. Our design does not have "pay to play," limiting the role of loss aversion. As such, our results do not speak to the issues they raise.

The market mechanism in our paper sorts individuals by their beliefs about the likelihood of efficient coordination rather than their abilities to perform certain tasks. Kosfeld and von Siemens (2009, 2011) present a related theoretical model that predicts sorting into or out of workplaces based on being a conditional cooperator. Their theory involves strategic uncertainty, as conditional cooperators would like to be matched with other conditional cooperators, as well as sorting via monetary incentives.

Endogenous group assignment and strategic uncertainty are important features of our experimental environment. For the most part, strategic uncertainty and group interactions do not play much of a role in the literature on endogenous choice of incentive contracts. One exception is Dohmen and Falk (2011) who consider interactive settings (revenue sharing and tournaments) as well as piece-rate compensation. Strategic uncertainty and beliefs play a role in their results, albeit differently from our experiments, as self-assessment of relative ability has a significant effect on selection into tournaments.

Endogenous group assignment and strategic uncertainty also play an important role in the literature on endogenous assignment to groups and VCMs. Endogenous assignment to groups has been shown to generally increase contributions to public goods (e.g. Kosfeld, Okada, and Riedl, 2009; Brekke, Hauge, Lind, and Nyborg, 2011; Charness, Cobo-Reyes, and Jiménez, 2011; Aimone, Iannaccone, Makowsky, and Rubin, 2013). Our work is related to this literature, but differs along several dimensions: (1) weak-link games raise different strategic issues than VCM games, (2) group assignment occurs via pricing within a market mechanism (rather than direct choice of groups), and (3) we focus on the relative importance of selection and strategic anticipation in determining the effect of endogenous assignment to incentive contracts.[4]

The results of our Sort treatment are related to earlier papers studying VCM public goods games. Specifically, Gunnthorsdottir, Houser, and McCabe (2007) find that a subject's initial public contribution is a useful measure of cooperative disposition (independent from any history effects), and Ones and Putterman (2007) show that the initial cooperative or non-cooperative

[4]Related papers showing effects of endogenous assignment to groups, roles, or institutions include Bohnet and Kübler (2005); Cabrales, Miniaci, Piovesan, and Ponti (2010); Dal Bó, Foster, and Putterman (2010); and Lazear, Malmendier, and Weber (2012).

tendencies that group members bring to a collective action situation have a *persistent* effect on cooperation. Similar to our Sort treatment, these papers use initial behavior to assign individuals to groups and show that there is a persistent long-run effect consistent with subjects having an innate type.

3 Experimental Design

3.1 The Corporate Turnaround Game: The turnaround game (Brandts and Cooper, 2006) is designed to study how a group caught in a productivity trap can escape (i.e. coordinate at a higher effort level). We use a variation of the turnaround game from Brandts and Cooper consisting of twenty rounds of a weak-link coordination game, split into two ten-round blocks. Subjects play in groups of four experimental subjects. They are randomly assigned to a group for the first ten-round block, and then are assigned to a new group for the final ten rounds. The method of assigning subjects to new groups for the final block varies by treatment as explained below.

The structure of the turnaround game played follows from three basic design choices. First, the firm's technology has a weak-link structure. As described by Kremer (1993), for many organizations the individual (or unit) doing the worst job – the "weak link" – determines the overall productivity of an organization. Imposing a weak-link structure creates a worst-case scenario for escaping a productivity trap since a unanimous increase in effort is necessary to improve group output. Presumably, many organizations face more forgiving environments where positive change is easier, but if we can understand how to escape productivity traps in tough environments, our insights should carry over to less difficult circumstances.

Second, we assume the firm can observe group output, equivalent to observing minimum effort in the absence of noise, but cannot observe any individual employee's effort level. It therefore cannot implement an incentive system based on individual effort rather than group output. Making individual effort unobservable simplifies the environment while increasing the difficulty of escaping a productivity trap.

Finally, the only instrument used to help escape a productivity trap is a simple linear incentive contract. We have eliminated more complex contracts as well as other instruments, such as communication, that have previously been shown to yield an increase in efficient coordination (see

Footnote 1 for relevant citations). Our goal is to study a simple environment that isolates the effect of incentive contracts, while also increasing the difficulty of escaping a productivity trap. In all likelihood, organizations facing a productivity traps will employ more than one method for escaping, and a topic for future research is how the various instruments interact.

At the beginning of each ten-round block, an incentive contract is announced for the group. This contract specifies a subject's compensation in experimental currency units (ECUs) as a function of the minimum effort across the four subjects in the group. A contract consists of a flat base wage (W) that each subject receives regardless of the outcome of the game and a bonus rate (B) that determines the amount of additional pay each subject receives per unit increase in the minimum effort. Higher values of B provide greater incentives to coordinate at high effort levels.

The incentive system in our experiment is a linear revenue sharing scheme. Straightforward revenue sharing systems are common in practice and well-suited to our experimental design for a couple of reasons: (1) This is a very simple incentive system. Subjects can easily understand how the incentives work as well as the implications of their assignment to an incentive contract. (2) There is an obvious link between the resolution of strategic uncertainty and workers' choices between incentive contracts. Workers should be willing to accept lower fixed wages in exchange for a higher value of B if and only if they believe that higher values of B are associated with a higher probability of coordination at high effort levels. This link is central to our experimental design.

In each round, the subjects simultaneously choose effort levels, where E_i is the effort level chosen by the i^{th} subject ($i \in \{1, 2, 3, 4\}$). Effort choices are restricted to be in ten-unit increments: $E_i \in \{0, 10, 20, 30, 40\}$. To make effort costly, payoffs are reduced by 5 ECUs per unit of effort expended. The payoff π_i for subject i is given by the following equation:

$$\pi_i = W - 5E_i + (B \times \min_{j \in \{1,2,3,4\}} \{E_j\}).$$

If $B > 5$, as is always the case in our experiments, the resulting stage game is a weak-link coordination game. Coordination by all four subjects on any of the five available effort levels is a Nash equilibrium. The five equilibria are Pareto ranked. The most desirable equilibrium involves coordination *at the highest possible effort (i.e. efficient coordination).*

Two different types of contracts are used in our experiment. Contract 1 always offers $W = 300$ and $B = 6$. The payoff matrix in the top panel of Table 1 results from Contract 1. Contract 2 increases the bonus rate from $B = 6$ to $B = 10$. Thus, Contract 2 features higher incentives to coordinate at high effort levels than Contract 1. The precise manner in which W is determined for Contract 2 is treatment dependent as explained in Section 3.2. For a given base wage W, the payoff matrix resulting from Contract 2 is shown in the bottom panel of Table 1.

Table 1: Payoff Matrices by Contract

Contract 1

		Minimum Efforts by Subjects of the Group				
		0	10	20	30	40
Effort by Subject i	0	300	-	-	-	-
	10	250	310	-	-	-
	20	200	260	320	-	-
	30	150	210	270	330	-
	40	100	160	220	280	340

Contract 2

		Minimum Efforts by Subjects of the Group				
		0	10	20	30	40
Effort by Subject i	0	W	-	-	-	-
	10	$W-50$	$W+50$	-	-	-
	20	$W-100$	W	$W+100$	-	-
	30	$W-150$	$W-50$	$W+50$	$W+150$	-
	40	$W-200$	$W-100$	W	$W+100$	$W+200$

Notes: Contract 1 offers a base wage $W = 300$ and a bonus factor $B = 6$. For Contract 2, W is the base wage determined by the Auction, and the bonus factor is $B = 10$. When the base wage $W \in (140, 300)$, the payoff from coordinating at 0 is always higher in Contract 1 and the payoff from coordinating at 40 is always higher in Contract 2.

Contract 1 features low incentives for coordination at a high effort level. Subjects receive a riskless payoff of 300 ECUs by choosing 0. Attempting to coordinate at higher effort levels is risky with little potential upside. For simplicity, assume that all subjects choose either 0 or 40. Relative

ACCEPTED MANUSCRIPT

to a choice of 0, increasing effort to 40 incurs a sunk effort cost of 200 ECUs in exchange for a potential net gain of only 40 ECUs. For this to have positive expected value, the probability of *all* of the three other group members choosing 40 must be greater than 5/6.

Contract 2 offers higher incentives to take the risk of trying to coordinate at a higher effort level. Consider again the situation facing a subject choosing between 0 and 40. Relative to a choice of 0, choosing 40 still incurs a sunk effort cost of 200 ECUs, but the potential net gain is increased to 200 ECUs. The probability of *all* of the other group members choosing effort level 40 only needs to be 1/2 for choosing 40 to yield higher expected payoff than choosing 0.

We did not restrict *a priori* the base wage of Contract 2 to be in the range of $140 < W < 300$, but the realized values were always within this range. These two conditions imply that the payoff from coordinating at 0 was always higher in Contract 1 than in Contract 2 and the payoff from coordinating at 40 was always higher in Contract 2 than in Contract 1.

3.2 Design and Treatments: We began all sessions by gathering measures of individual characteristics. We used the method of Eckel and Grossman (2008) to measure risk attitudes (see Appendix B for a full description).

Subjects were also asked to complete a questionnaire collecting information on their cognitive and demographic characteristics. As a measure of intelligence, we used self-reported SAT scores (mathematics and comprehensive).[5] The SAT is the primary college entry exam in the United States. There is high correlation between SAT scores and IQ scores (Frey and Detterman, 2004). We measured a subject's personality traits using a brief version of the Big-Five Personality Test provided in Gosling, Rentfrow, and Swann (2003). Survey questions from Glaeser, Laibson, Scheinkman, and Soutter (2000) were used to measure how trusting/trustworthy a subject was. We also gathered basic demographics (age, gender, race, and major).

Because beliefs play a critical role in coordination games, we considered eliciting beliefs prior to Blocks 1 and 2. We chose not to do so due to the risk of demand-induced effects. Forcing subjects to think about others' future actions could have stimulated more sophisticated reasoning including strategic anticipation.

[5] Some subjects had not taken the SAT, but had ACT scores. For those subjects, we converted their score using the SAT-ACT Concordance Chart provided on http://www.collegeboard.org. Reflecting local IRB policies, subjects were given the option of skipping questions to protect their privacy. There were few missing values for other questions, but a substantial fraction of the subjects (31%) chose not to report an SAT/ACT score. We discuss below how the analysis reflects these missing values.

After all of the individual measures were gathered, subjects played Block 1 (Rounds 1 – 10) of the turnaround game. Subjects were randomly assigned to groups of four, which remained fixed for the block. All subjects in all sessions were assigned to Contract 1 in Block 1. Given the weak incentives to raise effort above 0, we anticipated that play would converge to low effort levels. The goal was to generate a productivity trap that subjects would need to escape from in Block 2. Achieving efficient coordination in a weak-link game is a challenging task, and having to overcome a history of low effort makes it even harder.

Prior to the start of Block 2 (Rounds 11 – 20) of the turnaround game, subjects were assigned to either Contract 1 or Contract 2, new groups were formed with subjects assigned to the same contract, and W was set for Contract 2. Only one version of Contract 2 was used in any session, and the number of subjects in each session was a multiple of eight, so an equal number of groups could be assigned to each contract. The four treatments varied how contract assignment was done.

In the **Auction** treatment, the subjects' assignment to contracts and the base wage W of Contract 2 was based on a reverse English auction, to be explained below.[6] While auctions are used in many real-world markets, including some labor markets, they are not an especially common market mechanism for labor markets.[7] Nevertheless, clock auctions have a number of advantages for our experiments. They are easy to implement in a lab setting, quickly equilibrate, and make the selection process transparent for subjects (giving strategic anticipation a better chance of having an effect). In a broader sense, the auction has two critical features: workers must sacrifice fixed wages (W) to obtain higher performance pay (B) *and* workers are aware that they and others make a choice whether or not to accept lower fixed wages in exchange for higher performance pay. Any reasonable mechanism for a labor market should share these features, so our qualitative conclusions should not strongly depend on our choice of a market mechanism.

At the beginning of the auction, W was set to 400 ECUs with 400 seconds on the clock. The clock then ticked down towards zero. The base wage was reduced by 5 ECUs every five seconds (400, 395, 390, etc.). We used discrete changes to give subjects time to react before W dropped

[6]This clock auction is strategically equivalent to a sealed bid uniform price auction and shares the property of having a dominant strategy (bid or drop out at your reservation value). Experimental evidence (Kagel and Levin, 1993) shows that subjects tend to use the dominant strategy in English auctions but not in the equivalent Vickrey auction. This implies that the clock auction is less prone to mispricing than the sealed bid auction.

[7]A number of online labor markets use auctions to set wages. An example is http://www.freelancer.com, the largest outsourcing marketplace with 15.6 million freelancers and almost 8 million projects posted. An employer posts a project and freelancers make bids. The employer decides on a freelancer based on the price and ratings.

again. At any given time, subjects could press a button labeled "Contract 1." As soon as a subject clicked on this button, he/she was immediately assigned to Contract 1. If there were N subjects in the session, the auction continued until $N/2$ subjects had dropped out by pressing the Contract 1 button. The remaining $N/2$ subjects were assigned to Contract 2 with the value of W based on the time at which the last dropout took place. Subjects saw the payoff tables under Contracts 1 and 2 throughout the auction, with the Contract 2 table adjusting to reflect changing values of W. At no point during or after the auction were subjects given any information about dropouts prior to the final dropout. After the auction was completed, subjects assigned to Contract 2 only knew the value of W at which the final dropout took place. Once subjects were assigned to contracts, they were randomly placed into four-person groups with others assigned to the same contract.

Depending on the show-up rate of subjects, we had either 16 or 24 subjects in each session of the Auction treatment. The realized base wages of Contract 2 are summarized in Table 2, broken down by session size.

Table 2: Realized Base Wages in Contract 2

	16-Subject Sessions	24-Subject Sessions
Base Wages (in ECUs)	180, 210	190, 195, 225, 235

Prior to the auction, our detailed instructions stressed the relationship between W and the payoff table for Contract 2 and provided comparisons between the payoff tables of the two contracts. The goal was to have subjects understand the relationship between dropping out or not and their payoff table for Block 2. Subjects also watched a short movie demonstrating how the auction mechanism worked and participated in an unpaid practice round played against computerized opponents with randomly determined dropout times.[8]

In Block 2 of the Auction treatment, we attribute differences in behavior between Contracts 1 and 2 to the following three sources.

1). Direct Incentive Effect: Subjects have higher incentives to coordinate at the efficient outcome under Contract 2 than Contract 1.

[8]Dropout times were distributed uniformly over the range [0, 400]. Subjects knew that the dropouts in the practice round were randomly determined and contained no useful information.

2). Selection Effect: Subjects assigned to Contract 2 by the auction may have inherently different characteristics than those assigned to Contract 1. Selection can occur along many dimensions (e.g. risk preferences, personality traits) but initial beliefs are particularly germane given the importance of beliefs in coordination games. Contract 2 is relatively attractive to "optimists", defined as subjects who are optimistic about the chance of efficient coordination *independent of the auction outcome*, implying that optimists are relatively more willing to accept a lower base wage in order to obtain Contract 2. It follows that the auction tends to sort optimists into Contract 2 and pessimists (defined as subjects who are pessimistic about the chance of efficient coordination independent of the auction outcome) into Contract 1.

3). Effect of Strategic Anticipation: Subjects who anticipate that individuals assigned to Contract 2 by the auction are likely to be optimists become more optimistic when assigned to Contract 2 by the auction and more pessimistic when assigned to Contract 1. This optimism/pessimism *depends on the auction outcome*, because an otherwise identical individual with strategic anticipation will have different beliefs depending on which contract they are assigned by the auction. Strategic anticipation can only have an effect when subjects are given information that makes it possible to anticipate selection due to the auction.

In the **Random Assignment** treatment, subjects were randomly assigned to new groups for Block 2. The groups were then randomly assigned to either Contract 1 or Contract 2, with a 50/50 split between contracts imposed in each session. To maintain parallelism between the Auction and Random Assignment treatments, we matched the session sizes and base wages of Contract 2 in the Random Assignment sessions to those used in the Auction treatment sessions.

The instructions gave subjects no indication there was anything systematic about how groups or contracts were assigned. Subjects were told that the groups had changed, and the instructions stressed that they were almost certainly not with the same people as in Block 1. As for the contract, subjects were told, "You will be assigned to a new contract in Block 2. This may be a different contract than the one you had in Block 1, or it may be the same." Subjects had no reason to believe that their contract assignment told them anything about the type of people in their group. Thus, selection cannot be present in the Random Assignment treatment. *Differences between Contracts 1 and 2 in the Random Assignment treatment can be attributed solely to the direct incentive effect.*

Differences between the Random Assignment and Auction treatments capture the difference between exogenous and endogenous assignment to contracts. The **Sort** treatment separates this difference into selection and strategic anticipation effects by assigning subjects to contracts through a mechanism designed to mimic (imperfectly) the selection occurring in the Auction treatment. Specifically, we used data from the Auction treatment to generate a model predicting each subject's dropout time as a function of his/her individual characteristics and his/her choice in Round 1 of Block 1. Specifics of this model are given below in Section 3.4. After Block 1, subjects in the Sort treatment were ranked from the highest to the lowest predicted dropout time. The clock in the auction showed the time *remaining*, so subjects with high dropout times dropped out and took Contract 1 while those with low dropout times stayed in and were assigned to Contract 2. In the Sort treatment, subjects ranked first, second, third and fourth were assigned to Group 1, subjects ranked fifth, sixth, seventh and eighth were assigned to Group 2, and so on. Groups 1, 2, and 3 were assigned to Contract 1 (Groups 1 and 2 for sessions with only 16 subjects) and Groups 4, 5, and 6 were assigned to Contract 2 (Groups 3 and 4 for sessions with only 16 subjects). In other words, subjects who were predicted to have chosen Contract 1 if they had participated in the Auction treatment were assigned to Contract 1 in the Sort treatment, and subjects who were predicted to have chosen Contract 2 in the Auction treatment were assigned to Contract 2.[9]

Subjects in the Sort treatment received the same information as subjects in the Random Assignment treatment about how groups and contracts were assigned. This information was intentionally vague, giving subjects no indication that there was anything systematic about how groups or contracts were assigned. The goal was to mimic (imperfectly) the selection present in the Auction treatment while disabling strategic anticipation. *The difference between the Sort and Random Assignment treatment provides a lower bound on the size of the selection effect.*[10]

The Auction and Sort treatments use different methods of assigning subjects to groups conditional on their assigned contract. By sorting subjects into groups within contracts, we can use the differences between groups within contracts as an additional measure of the impact of selection on

[9]See Rietz, Sheremeta, Shields, and Smith (2013) for an example of using an initial phase to type individuals. Rietz *et al.* use this typing exercise to help understand individual behavior in a repeated three-player trust game.

[10]Dal Bó, Foster, and Putterman (2010) faced a similar methodological challenge of separating selection effects from other effects of endogenous assignment in their study of the effects of democracy. We could have used a similar mechanism to separate selection effects from the effects of strategic anticipation, but this would have had major drawbacks in our experiment. The largest problem is that their method would have caused an enormous loss of power given the structure of our experiment, necessitating a vastly larger sample size. For a detailed discussion of this issue, see Appendix A.

Block 2 of the Sort treatment. It is not possible to do a similar sorting for the Auction treatment as we do not observe dropout times for half of the subjects, but we feel the additional measure is worth a departure from parallelism.[11]

The **Reverse Sort** treatment is designed to illustrate the power of selection. This treatment was identical to the Sort treatment, except subjects were assigned to the contract they were *not* predicted to be assigned in the Auction treatment. In other words, subjects predicted to be assigned Contract 1 in the Auction treatment were assigned to Contract 2 in the Reverse Sort treatment and subjects who were predicted to be assigned Contract 2 in the Auction treatment were assigned to Contract 1 in the Reverse Sort treatment. Within contracts, subjects were sorted into groups by their predicted dropout times. Strategic anticipation should play no role in the Reverse Sort treatment, and the selection effect should increase effort for Contract 1 and reduce effort for Contract 2. Since the selection effect counteracts the direct incentive effect, the difference between Contracts 1 and 2 is expected to be narrowed in Block 2 relative to the Random Assignment treatment, or possibly even reversed.

3.3 Procedures: All experimental sessions were conducted in the XS/FS computer lab of Florida State University using the software z-Tree (Fischbacher, 2007). All FSU undergraduates were eligible to participate, although subjects were drawn primarily from students taking social science classes (economics, political science, and sociology). Subjects' recruitment was done using the software ORSEE (Greiner, 2015), and subjects were allowed to participate in *only* one session. Subjects were guaranteed \$10 for arriving on time. Average earnings per participant were \$17.32 including the show-up fee, and sessions typically lasted 45 – 60 minutes. Each treatment had six sessions and 128 subjects (two sessions with 16 subjects and four with 24 subjects).

At the beginning of each session, subjects were randomly seated. Instructions (see Appendix B) were read aloud by the experimenter prior to each stage of the experiment, and subjects were given a short comprehension quiz.

At the beginning of each round, subjects were shown the base wage, the bonus rate, and the resulting payoff matrix. It was common knowledge that all subjects of a given group faced the *same* contract, and that the group and contract were fixed throughout the ten-round block. While

[11] In a companion paper (Cooper, Ioannou, and Qi, 2018), we use a structural model to study the effect of the assignment method (random or sorted within contract). Simulation exercises indicate that the assignment method has little effect on our results (average effort by contract).

viewing this information, subjects were asked to pick an effort level for the round. These choices were made simultaneously, so subjects did not know the effort choices of the others in their group before making their own choices. After all group members had made decisions, subjects were shown a summary of the round's results including the minimum effort, their payoff for the round, and their cumulative earnings. They were also shown, sorted from low to high, the effort levels that all the subjects of the group had chosen. Subjects were provided with a summary of results from the previous rounds of the block.

At the end of the session, each subject was paid via check the earnings for all rounds played plus the $10 show-up fee. Payment was done on an individual and private basis.

3.4 Predicting Dropout Times: The Sort and Reverse Sort treatments required us to predict when subjects would have dropped out of the auction if they had been in the Auction treatment. This was done in two steps. First, we ran a regression, fitting the times subjects dropped out (i.e. pressed the "Contract 1" button) in the Auction treatment as a function of their individual characteristics and their first choice in the turnaround game (i.e. before they interacted with other subjects). A Tobit model was used due to censoring of dropout times. Technical details of the model are given in Appendix A. In the second step, the fitted parameters for this Tobit model were used to generate predicted dropout times for subjects in the Sort and Reverse Sort treatments. After the first round of play in the turnaround game (Block 1) was completed, a research assistant entered data into a program that generated predictions, and subjects were sorted into groups for Block 2 using these predicted dropout times. Each stage of the experiment used a separate z-Tree program, so an experimenter could enter information about groupings into the Block 2 program while Block 1 was running. The data entry and calculations were sufficiently rapid that subjects observed no delays prior to the start of Block 2.[12]

[12]This was the main reason for not including the Round 10 choice in the Tobit model. Waiting until the end of Block 1 to start entering information would have caused a long delay before the beginning of Block 2. If the Round 10 choice is added to the Tobit model along with the Round 1 choice, then both variables are statistically significant but the ability to predict which contract a subject is assigned only improves slightly. A secondary concern was that endogeneity of the Round 10 choice could bias our estimates.

Table 3: Tobit Regression Independent Variables

Variables	Explanation	Data Source
$RISK$	Choice of option in the Risk Aversion Test	Risk Aversion Test
$MATH$	SAT Math Score if reported	Questionnaire
I_{MATH}	Indicator function: "=1" if Math Score reported & "=0" otherwise	Questionnaire
$SCORE$	SAT Comprehensive Score if reported	Questionnaire
I_{SCORE}	Indicator function: "=1" if Comp. Score reported & "=0" otherwise	Questionnaire
$EXTROVERT$	Big 5 Personality Test: Extroversion	Questionnaire
$AGREEABLE$	Big 5 Personality Test: Agreeableness	Questionnaire
$CONSCIENTIOUS$	Big 5 Personality Test: Conscientiousness	Questionnaire
$EMOTIONSTABLE$	Big 5 Personality Test: Emotional Stability	Questionnaire
$OPENNESS$	Big 5 Personality Test: Openness	Questionnaire
$FAIR$	Trust Test: "Would most people try to be fair?"	Questionnaire
$HELP$	Trust Test: "Would most people try to be helpful?"	Questionnaire
$TRUST$	Trust Test: "Can most people be trusted?"	Questionnaire
$TRANSFER$	Trust Test: "Should personal income be determined by work?"	Questionnaire
$STRANGER$	Trust Test: "Can you count on strangers?"	Questionnaire
$TRUSTME$	Trust Test: "Am I trustworthy?"	Questionnaire
$EFFORT_l$	Effort Choice Dummies in round 1 of Block 1 & $l \in \{10, 20, 30, 40\}$	Block 1
$CONS$	Constant	

Table 4: Regression Models to Predict Dropout Times

Dependant Variable	Model 1 Dropout Times	Model 2 Probability of Choosing 40
$RISK$	2.802	0.055
	(4.779)	(0.075)
$MATH$	0.147	0.008^{***}
	(0.167)	(0.003)
I_{MATH}	-30.279^{*}	0.253
	(16.972)	(0.303)
$SCORE$	0.041	-0.002^{*}
	(0.051)	(0.001)
I_{SCORE}	−5.000	−0.038
	(19.168)	(0.326)
$EXTROVERT$	-13.668^{*}	-0.212^{*}
	(7.469)	(0.112)
$AGREEABLE$	4.210	−0.084
	(7.050)	(0.098)
$CONSCIENTIOUS$	−6.422	0.067
	(8.099)	(0.117)
$EMOTIONSTABLE$	1.315	0.097
	(8.224)	(0.114)
$OPENNESS$	−1.718	0.036
	(8.023)	(0.114)
$FAIR$	−10.029	-0.252^{*}
	(9.034)	(0.140)
$HELP$	6.356	0.178
	(8.241)	(0.142)
$TRUST$	−12.462	0.092
	(8.499)	(0.143)
$TRANSFER$	−6.977	0.101
	(9.509)	(0.155)
$STRANGER$	−2.737	0.040
	(9.765)	(0.164)
$TRUSTME$	18.686	−0.095
	(12.067)	(0.145)
$EFFORT_{10}$	−18.824	
	(35.701)	
$EFFORT_{20}$	−23.875	
	(30.918)	
$EFFORT_{30}$	−24.710	
	(28.812)	
$EFFORT_{40}$	-101.631^{***}	
	(27.084)	
$CONS$	141.473	−2.435
	(135.634)	(1.897)
# of Observations	128	128

Notes: Model 1 uses a Tobit regression to predict dropout times in the Auction Treatment. Model 2 uses a probit model to predict the probability of choosing effort 40 in the 1^{st} round of Block 1. Standard errors are given in parentheses. Three (***), two (**), and one (*) stars indicate statistical significance at the 1%, 5%, and 10% respectively.

Table 4 reports the results of two regressions, with Table 3 providing a brief description of the independent variables. Model 1 gives the results of the Tobit regression used to predict dropout times. By far the most important variable is the dummy for choice of effort level 40 in the first round of the turnaround game. The parameter estimate is large and significant at the 1% level. Consistent with the Tobit results, subjects who chose 40 in the first round were more than twice as likely to be assigned to Contract 2 versus Contract 1 (69% vs. 31%). The extroversion component of the Big Five and the dummy variable for subjects who reported a math score also have significant effects, but only at the 10% level.[13]

The measures gathered prior to Block 1 have little predictive power in Model 1 for dropout times. One possible explanation is that these measures act on dropout times through their effect on the probability of choosing 40 in the first round. Model 2 supports this argument. This is a Probit model where the dependent variable is the dummy for choice of 40 in the first round of Block 1 and the independent variables are the individual measures gathered before Block 1. There is a strong positive relationship between math scores and choosing 40, and a number of the other measures also have weakly significant relationships with the likelihood of choosing 40. Somewhat to our surprise, risk aversion has minimal effect in Models 1 and 2. There is no indication that the auction sorts individuals based on their risk preferences.[14] A number of the survey questions measure subjects' social preferences, particularly whether they are trusting and/or trustworthy. These measures have little impact on the likelihood of initially choosing 40, but it is possible that incentivized measures or measures designed to measure different attributes such as altruism would have yielded stronger results.

To check how well the model replicates contract assignment by the auction, we compared the contract assignments predicted by the Tobit model with the actual contract assignments in the Auction treatment. The model correctly predicts the contract assignment for 69% of the subjects (88/128). Among subjects for whom our model incorrectly predicts the contract assignments, most are close to indifferent with 65% having predicted dropout times within 30 seconds of the

[13]As noted previously, about a third of subjects chose not to report scores. The regressions include a dummy for subjects who did report a score as well as a slope parameter. The missing scores were set equal to the median reported score. Subjects who did not report a math score were less likely to be assigned to Contract 2 in the Auction treatment (36% vs. 58%).

[14]We anticipated that subjects who were risk averse would avoid strategic risk by making low initial effort choices, but find no evidence to support this conjecture. This could be because our subjects are largely risk neutral given the low stakes involved, or because the relevant risk preferences are domain specific and not captured by the Eckel and Grossman measure that we employ.

cutoff to switch to the other (correct) contract. As an alternative method of checking how well the model predicts contract assignment by the auction out of sample, we performed a Monte Carlo exercise where we randomly drew 75% of the observations, fit Model 1, and then predicted the contract assignments in the remaining observations. Averaging over multiple draws, 63% of the predictions are correct. Overall, the model does a reasonable job of replicating the contract assignment that occurred in the Auction treatment.

For the six questions relating to fairness and trust, we made an error in the coding when subjects chose not to answer.[15] This caused 6% of the subjects to be flipped between contracts (6 in the Sort treatment and 10 in the Reverse Sort treatment). The regressions reported in Table 6 (Model 3 and 4) include controls for the sorting mistakes. The sorting mistakes have little impact on the results since affected subjects tend to be close to the cutoff between contracts.

3.5 Hypotheses: In the Random Assignment treatment, differences in effort between Contracts 1 and 2 reflect only the direct incentive effect. The Sort treatment adds in the selection effect and the Auction treatment adds in selection and strategic anticipation effects (as defined previously). All three effects should have a positive effect in Contract 2 relative to Contract 1, leading to higher effort under Contract 2 than Contract 1.

H1: Average effort in Block 2 will be higher for Contract 2 than Contract 1 in the Random Assignment, Sort, and Auction treatments.

Along similar lines, effort under Contract 2 should be increasing from Random Assignment to Sort to Auction. The Reverse Sort treatment is different since the selection effect is reversed. Given that pessimists tend to get Contract 1 rather than Contract 2 in the Auction treatment, the assignment process in the Reverse Sort treatment should tend to assign pessimists to Contract 2 rather than Contract 1. Instead of getting the direct incentive effect *plus* the selection effect, as in the Sort treatment, effort in the Reverse Sort treatment is affected by the direct incentive effect *minus* the selection effect. This implies that effort under Contract 2 should decrease between the Random Assignment and Reverse Sort treatments.

H2: Average effort in Block 2 for Contract 2 will be increasing in the following order across

[15] All six questions used a Likert scale over how much subjects disagreed or agreed with a statement. They also had an option labeled "prefer not to answer." Averaging across the six questions, 2.2% of the subjects responded "prefer not to answer." These responses should have been coded as neutral responses, but instead were coded as strong disagreement due to our error. The error was only noticed after the fact.

treatments: Reverse Sort, Random Assignment, Sort, and Auction.

Our predictions about differences between treatments under Contract 1 are a flipped version of our hypotheses for Contract 2, given that the direct incentive effect, selection effect, and strategic anticipation effect are all expected to *reduce* effort under Contract 1.

H3: Average effort in Block 2 for Contract 1 will be decreasing in the following order across treatments: Reverse Sort, Random Assignment, Sort, and Auction.

H1, *H2*, and *H3* are all ordinal in nature. We have no *ex ante* hypotheses about the size of the treatment effects as they depend on the fraction of optimists, pessimists, and individuals with strategic anticipation in the population. These are purely empirical matters.

Likewise, *H2* and *H3* jointly imply that the difference between Contracts 1 and 2 in Block 2 should be smaller in the Reverse Sort treatment than in the Random Assignment treatment. The selection effect may be sufficiently strong that the difference will flip, with Contract 1 yielding *higher* effort than Contract 2, but we have no reason to formally hypothesize this outcome in the absence of data or a firm theoretical basis for our conjecture.

4 Results

Our discussion of the experimental results focuses on individual effort as a measure of performance. Individual effort directly measures subjects' choices (unlike payoffs) and has the advantage of using all subjects' choices, making it less susceptible than minimum effort to a single outlier within a group. We also discuss other performance measures such as group minimum effort, waste (individual effort – minimum effort), and payoffs.

At various points as we introduce the main results, we refer to the significance of various treatment effects. Unless we state otherwise, the tests being reported are Wilcoxon matched-pairs signed-rank tests and the reported results are p-values. An observation is the average value of individual effort (unless another variable is indicated) for a session in Block 2. Sessions are matched by the session size and base wage for Contract 2.[16] These are conservative tests, which correct for potential session effects in the most extreme way possible. As such, we are more likely to make Type II errors (false negatives) than Type I errors (false positives). Later in the results

[16]Recall that sessions in the Random Assignment, Sort, and Reverse Sort treatments are matched to sessions in the Auction treatment, using the same session size and base wage for Contract 2.

section we present Tobit models that explore difference methods of correcting for session effects.

All groups play Block 1 using Contract 1. The goal is to establish a history of low effort that must be overcome in Block 2. We largely achieve this goal as 73% of groups have minimum effort of 0 in Round 10 and 61% of individuals choose an effort level of 0. Subjects are randomly assigned to treatments, and there are no significant differences between treatments in Block 1.[17]

In Block 2, subjects are assigned to either Contract 1 or Contract 2 in the Random Assignment, Auction, Sort, and Reverse Sort treatments. Table 5 gives information about individual effort and a number of alternative performance measures in Block 2. Data are broken down by treatment. Within each treatment, data are provided by contract ("Contract 1" and "Contract 2") as well as the average across both of the contracts ("Both Contracts").

Table 5: Summary of Block 2 Outcomes

	Individual Effort	Minimum Effort	Waste	Variable Payoff	Total Payoff
Random, Contract 1	21.17	16.63	4.55	−6.11	293.89
Random, Contract 2	31.34	27.56	3.78	118.91	326.09
Random, Both Contracts	26.26	22.09	4.16	56.40	309.99
Auction, Contract 1	9.64	3.69	5.95	−26.08	273.92
Auction, Contract 2	39.81	39.25	0.56	193.44	400.63
Auction, Both Contracts	24.73	21.47	3.26	83.68	337.27
Sort, Contract 1	13.45	9.88	3.58	−8.02	291.98
Sort, Contract 2	37.48	35.69	1.80	169.45	376.64
Sort, Both Contracts	25.47	22.78	2.69	80.72	334.31
Reverse, Contract 1	30.03	26.00	4.03	5.84	305.84
Reverse, Contract 2	16.06	10.94	5.13	29.06	236.25
Reverse, Both Contracts	23.05	18.47	4.58	17.45	271.05

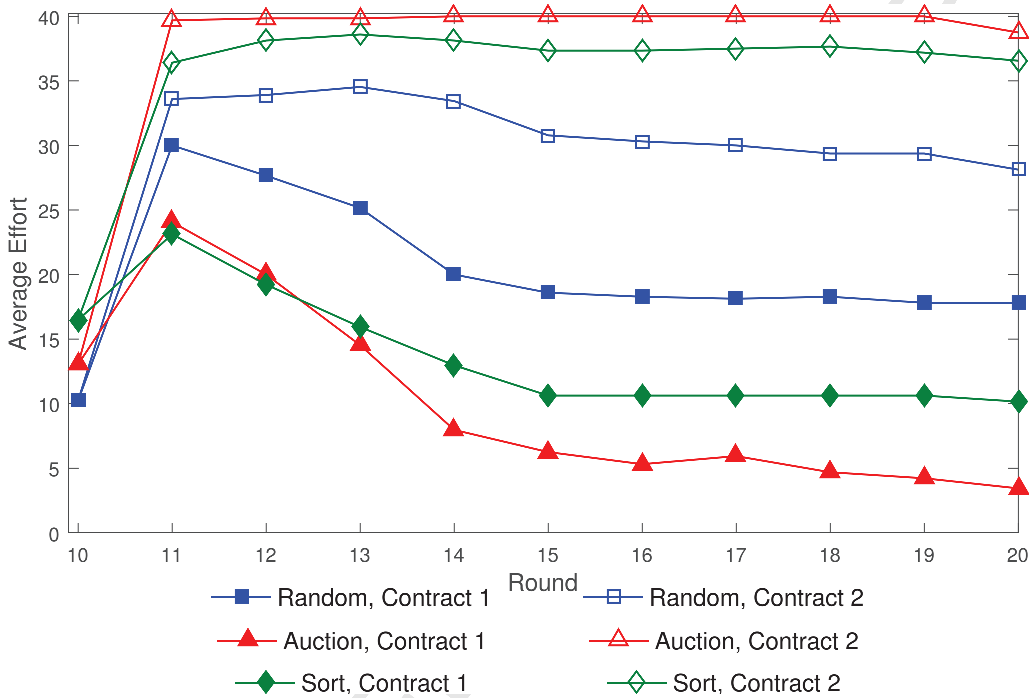
Notes: Each observation is the Block 2 average for a group. "Waste" is defined as the difference between individual effort and minimum effort. "Variable Payoff" does *not* include fixed wages (W) while "Total Payoff" is the sum of fixed wages and variable payoffs.

Figure 1 compares average effort in Block 2 for the Random Assignment, Auction, and Sort treatments, broken down by contract. Data from the final round of Block 1 (Round 10) are included to give a better sense of how the incentive contracts change behavior.

[17] This is based on Wilcoxon rank-sum tests. There is a single observation per group giving the group's average effort across Rounds 1 – 10.

Groups assigned to Contract 2 perform better than those assigned to Contract 1 in all three treatments. This difference is weakly significant ($p = .075$) for the Random Assignment treatment and significant for the Auction ($p = .028$) and Sort ($p = .028$) treatments. The data support *H1*.

FIGURE 1: COMPARISON OF RANDOM ASSIGNMENT, AUCTION AND SORT TREATMENTS IN BLOCK 2



Individual effort increases sharply with endogenous assignment to incentive contracts in the Auction treatment. Groups assigned to Contract 2 in the Auction treatment achieve nearly perfect coordination at the efficient equilibrium, with all four group members choosing effort level 40 in 96% of the observations (compared with 54% in the Random Assignment treatment). On the flip side, groups assigned to Contract 1 in the Auction treatment perform poorly with a minimum effort of 0 in 87% of the observations (versus 57% in the Random Assignment treatment). Comparing the Auction and Random Assignment treatments, the difference in effort is significant under both Contract 1 ($p = .028$) and Contract 2 ($p = .035$).

The Sort treatment attempts to replicate the selection process of the Auction treatment while eliminating effects due to strategic anticipation. Even though this replication is obviously imperfect, the effect of the Sort treatment is similar to the effect of the Auction treatment. The efficient equilibrium is played in 78% of the observations for Contract 2, and there is a decline in average effort for Contract 1 relative to the Random Assignment treatment, albeit not as large as in the

Auction treatment. This does not reflect a big increase in groups with a minimum effort of 0 (61% vs. 57% in Random Assignment) but instead shows up as a large decrease in efficient coordination (16% vs. 40%). Comparing the Sort and Random Assignment treatments, the difference in effort is weakly significantly under Contract 2 ($p = .093$) but not Contract 1 ($p = .249$).[18]

The Reverse Sort treatment is intended to drive home the power of selection. We intentionally create a situation that presumably never occurs in field settings, assigning subjects who would most likely have been assigned to Contract 1 in the Auction treatment to Contract 2 and vice versa. The selection effect now works in the opposite direction of the direct incentive effect. If the selection effect is sufficiently strong, the Reverse Sort treatment should yield *lower* effort in Contract 2 than Contract 1.

FIGURE 2: COMPARISON OF REVERSE SORT AND RANDOM ASSIGNMENT TREATMENTS IN BLOCK 2

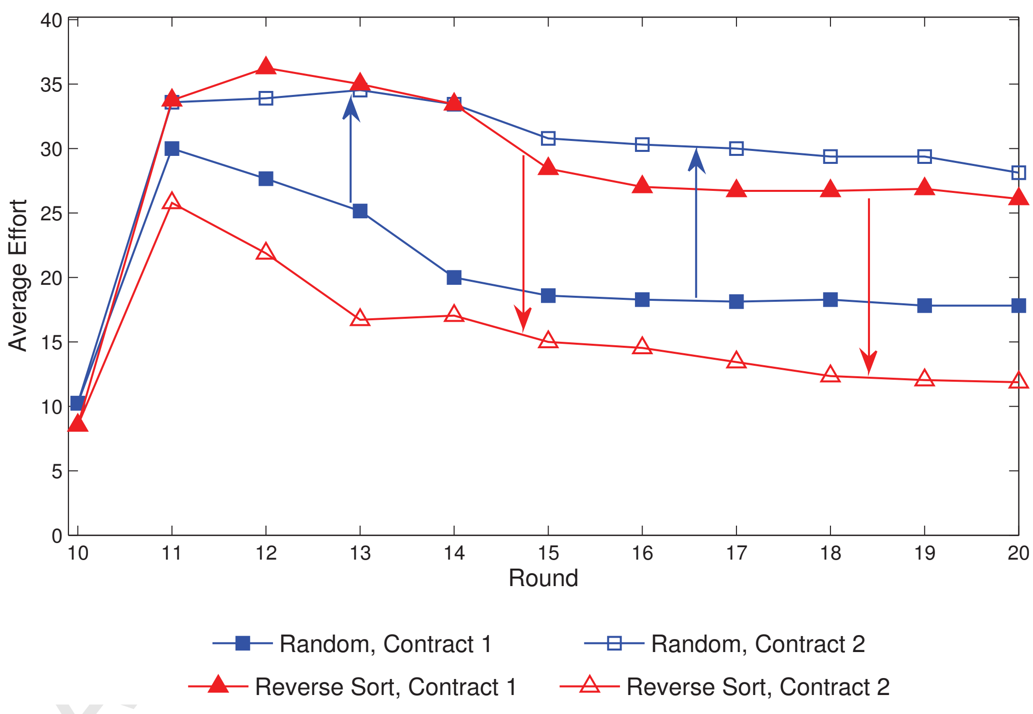


Figure 2 compares effort levels for the Random Assignment and Reverse Sort treatments. Colored arrows point from Contract 1 to Contract 2 to make it easier to see the difference between the two incentive contracts by treatment. In Random Assignment, where only the direct incentive

[18] As an alternative approach, we have run Wilcoxon matched-pairs signed-rank tests using the difference between Block 2 average effort and Block 1 effort (by session). This has relatively little effect on the results with one exception. Comparing the Sort and Random Assignment treatments, the difference is significant for Contract 1 but not Contract 2, rather than the other way around.

effect is present, effort is higher under Contract 2 than Contract 1. This is reversed in the Reverse Sort treatment. The selection effect is sufficiently strong to make effort higher under Contract 1 in spite of lower incentives to coordinate. This difference between contracts is significant ($p = .028$). Across Block 2, efficient coordination occurs in 54% of the observations under Contract 1 as opposed to only 18% for Contract 2. Likewise, the group's minimum effort is 0 in 61% of the observations under Contract 2 versus only 28% for Contract 1. Comparing the Reverse Sort and Random Assignment treatments, the difference in effort levels is significant for Contract 2 ($p = .028$) but not Contract 1 ($p = .249$).

H2 hypothesizes that average effort in Block 2 under Contract 2 would increase in the following order across treatments: Reverse Sort, Random Assignment, Sort, and Auction. The data largely support this hypothesis, with the exception that the difference between the Sort and Auction treatments is not significant ($p = .173$). *H3* flips the predictions of *H2* for Contract 1, with average effort predicted to decline across treatments in the following order: Reverse Sort, Random Assignment, Sort, and Auction. The differences between treatments all have the predicted signs, but fewer of the differences are statistically significant than for Contract 2 (albeit using a conservative statistical test).[19]

Comparing the performance under Contracts 1 and 2 in the Auction treatment illustrates an important point for interpreting our results. If we look solely at groups assigned to Contract 2, then endogenous assignment to contracts has a strong positive effect. Looking at *all* of the groups paints a different picture, as the positive effect for groups assigned to Contract 2 is almost perfectly offset by the negative effect for groups assigned to Contract 1. Averaging across both contracts, the average effort in Block 2 is almost identical in the Random Assignment and Auction treatments (26.26 vs. 24.73), and the difference is not statistically significant ($p = 0.600$). To appreciate the effect of endogenous assignment to incentive contracts, we have to understand the effect for groups that choose high performance pay *as well as the effect for groups that do not*. More generally, average individual effort across both contracts is roughly the same for all four treatments (25.47 and 23.05 for the Sort and Reverse Sort treatments, respectively). Not surprisingly, none of the differences between treatments are significant when we average across contracts (comparing the Sort and Reverse Sort treatments with Random Assignment, we get

[19]For differences between consecutive treatments, statistical significance for all pairs under Contract 1 are reported above except for the difference between the Sort and Auction treatments which is not significant ($p = .600$).

$p = .600$ and $p = .463$, respectively).

Conclusion 1. *Endogenous assignment to contracts increases effort under high performance pay (Contract 2) and decreases effort under low performance pay (Contract 1). The Sort treatment generates similar effects as the Auction treatment, albeit weaker under both contracts, while the Reverse Sort treatment flips the difference between high and low performance pay. Increased effort under high performance pay (Contract 2) is largely offset by reduced effort under low performance pay (Contract 1), making individual effort, averaging across both contracts, roughly the same in all four treatments.*

Thus far we have focused on individual effort, but Table 5 provides data on a number of measures other than individual effort. For the most part, the pattern across treatments is similar to what we have described for individual effort. This is unsurprising as minimum effort, variable payoffs, and total payoffs are strongly correlated with individual effort. Waste, defined as the difference between individual effort and minimum effort, has somewhat different treatment effects than the other variables. Waste decreases for Contract 2 in the Auction treatment relative to Random Assignment, reflecting the tendency of all group members to choose effort level 40. This difference is weakly significant ($p = .092$). A similar effect is observed for the Sort treatment; the effect is smaller but statistically significant ($p = .046$). There is no significant difference in waste between the Random Assignment and Reverse Sort treatments under Contract 2 ($p = .600$), and waste is similar (and not significantly different) across all four treatments under Contract 1 (comparing the Auction, Sort, and Reverse Sort treatments with the Random Assignment treatment yields $p = .600$, $p = .917$, and $p = .917$, respectively). To summarize, endogenous assignment to contracts reduces waste, but only under Contract 2. The Sort treatment replicates this effect. Coupled with increased effort, decreased waste reinforces the positive effect of endogenous assignment to high performance pay.

In addition to the non-parametric tests, we also use Tobit models to examine the statistical significance of differences between treatments. This analysis, reported in Table 6, serves two purposes. First, the non-parametric tests take a conservative approach to correcting for the session effects in the data. The Tobit analysis gives a sense of how robust our conclusions are to using different approaches to correct for session effects. Second, the non-parametric tests do not control for several factors that might affect the results. Models 3 and 4 include controls for the sorting errors described in Section 3.4. We also control for the average starting effort level of all subjects in a session. Using the terminology of Fréchette (2012), this is a direct control for the most obvious source of dynamic session effects.

Table 6: Tobit Analysis of Treatment Effects

	Model 1	Model 2	Model 3	Model 4
Dependant Variable	Group Average Effort in Block 2			
Correction for Session Effect	Clustering	Random Effects	Clustering	Random Effects
Contract 2	9.638	9.630**	9.695	9.683**
	(6.734)	(4.716)	(6.704)	(4.495)
Contract 1 × Auction	−13.102**	−13.132***	−12.037*	−12.145**
	(6.275)	(4.913)	(6.808)	(5.008)
Contract 2 × Auction	19.753***	19.611***	20.260***	19.974***
	(4.436)	(5.800)	(4.921)	(5.794)
Contract 1 × Sort	−9.290	−9.369*	−11.082*	−11.315**
	(6.278)	(4.915)	(6.670)	(4.961)
Contract 2 × Sort	8.534*	8.442*	10.354**	10.125**
	(4.474)	(5.052)	(5.037)	(5.109)
Contract 1 × Reverse Sort	8.805	8.815*	12.153	12.006**
	(7.000)	(4.967)	(7.714)	(5.393)
Contract 2 × Reverse Sort	−15.644***	−15.636***	−18.283***	−18.487***
	(4.995)	(4.917)	(5.940)	(5.369)
# of Mistaken Assignments to			−6.413*	−6.188
Contract 2 in Sort / Contract 1 in Reverse Sort			(3.354)	(4.575)
# of Mistaken Assignments to			12.722**	13.152**
Contract 1 in Sort / Contract 2 in Reverse Sort			(5.771)	(5.304)
Round 1 Effort by Other Group Members			0.377	0.366
			(0.656)	(0.554)
Other Differences Between Treatments				
Contract 1×Auction - Contract 1×Sort	−3.813	−3.763	−0.955	−0.829
	(2.682)	(4.861)	(3.094)	(4.924)
Contract 2×Auction - Contract 2×Sort	11.219***	11.169*	9.906***	9.849*
	(3.348)	(5.872)	(3.728)	(5.832)
Contract 2×Reverse Sort - Contract 1×Reverse Sort	−14.811***	−14.821***	−20.740***	−20.810***
	(2.584)	(4.704)	(1.989)	(4.978)
# of Observations	128	128	128	128

Notes: Standard errors are given in parentheses. Three (***), two (**), and one (*) stars indicate statistical significance at the 1%, 5%, and 10% respectively. The omitted category is the Random Assignment treatment under Contract 1

Effort choices from the same group are highly correlated across individuals and rounds, so we use a group's average individual effort across all 10 rounds in Block 2 as the dependent variable (a total of 128 observations). A Tobit model is used to correct for censoring since almost a quarter of the groups coordinate at the efficient equilibrium (all four subjects choose 40) in all ten rounds of Block 2.[20]

We report models using two different methods of correcting for session effects. Models 1 and 3 correct the standard errors for clustering at the session level while Models 2 and 4 include a session-level random effect. Both approaches assume that there is correlation between observations from different groups in the same session. The correction for clustering is agnostic about the specific form of that correlation (and generally relies on weaker assumptions about the error structure), while the random effects model imposes a specific functional form on session effects (session effects are constants drawn from a normal distribution). Clustering is generally a more conservative approach than using random effects. Rather than stating that one is the right approach, our goal is to get a sense of how much the choice of approach matters.

Turning to the explanatory variables, the omitted category is the Random Assignment treatment under Contract 1. The Contract 2 dummy captures the difference between Contracts 1 and 2 for the Random Assignment treatment. Interactions between a Contract 1 dummy and treatment dummies (Contract 1 × Auction, Contract 1 × Sort, and Contract 1 × Reverse Sort) capture treatment effects for Contract 1 using the Random Assignment treatment under Contract 1 as the base. Likewise, interactions between a Contract 2 dummy and treatment dummies (Contract 2 × Auction, Contract 2 × Sort, and Contract 2 × Reverse Sort) capture treatment effects for Contract 2 using the Random Assignment treatment under Contract 2 as the base. The top panel of Table 6 reports parameter estimates and the bottom reports additional estimated differences between treatments.[21]

Models 1 and 2 test for treatment effects with no additional controls. The results largely support our conclusions based on the non-parametric tests. Regardless of whether clustering or random effects are used to correct for session effects, the Auction treatment has a significant effect on effort under either Contract 1 or Contract 2. The Sort and Reverse Sort treatments

[20] There are no groups that had all individuals choose 0 in all ten rounds, so we only have right censoring.

[21] These differences and the associated standard errors are generated by changing the omitted category and rerunning the regressions.

have a significant effect on effort under Contract 2 relative to the Random Assignment treatment (weakly significant in the case of the Sort treatment). There are two noteworthy differences between the regression results and the results of the non-parametric tests. First, the difference between Contracts 1 and 2 in the Random Assignment treatment is *not* significant ($p = .155$) when clustering is used to correct for the session effects. Second, the difference between the Auction and Sort treatments is significant under Contract 2 in both Models 1 and 2, unlike the non-parametric test ($p = .173$).

Due to a mistake in coding missing values (see Section 3.4), a small number of subjects were accidentally assigned to the wrong contract in the Sort and Reverse Sort treatments. To control for these sorting errors, Models 3 and 4 include a variable for the number of group members accidentally assigned to Contract 2 (Contract 1) in the Sort (Reverse Sort) treatment as well as a variable for the number of group members accidentally assigned to Contract 1 (Contract 2) in the Sort (Reverse Sort) treatment. The parameter estimates for the two new variables in Model 2 have the expected signs and are jointly significant at the 5% level in both models ($p = .040$ and $p = .016$, respectively).[22] Subjects in different groups for Block 2 may have interacted in Block 1, creating links between different Block 2 groups in the same session. Models 3 and 4 control for the resulting session effects by including the average Round 1 effort, *by session*, as a new variable. This control has the expected positive sign but is not significant and has little impact on our conclusions.

Comparing Models 3 and 4 with Models 1 and 2, the estimated parameters and their significance do not differ much. The only difference worth noting is that the effect of the Sort treatment under Contract 1 is now weakly significant in Model 3, the model using clustering to correct for session effects.

Overall, the results of the Tobit analysis do *not* change our conclusions much from the simple non-parametric tests. Support for *H1* is weaker while support for *H2* and *H3* is stronger, but the overall picture remains much the same. Neither the use of different methods to account for session effect nor the inclusion of additional controls affect our main findings as stated in Conclusion 1.

[22] To understand the expected signs, consider a subject who should be assigned to Contract 1 but is accidentally assigned to Contract 2 in the Sort treatment. This subject should tend to choose low effort levels in Block 2, pulling down effort for his/her entire group. This implies that the expected sign for "# of Mistaken Assignments to Contract 2 in Sort / Contract 1 in Reverse Sort" is negative. The same logic implies a positive sign for the other control variable.

In Section 3.2, we attribute the total effect of high performance pay in the Auction treatment to three effects: the direct incentive effect, the selection effect, and the effect of strategic anticipation. The direct incentive effect is measured by the difference between average individual effort (shown in Table 5) under Contracts 1 and 2 in the Random Assignment treatment (31.3 – 21.2). The sum of the selection and strategic anticipation effects is measured by the difference between the Auction and Random Assignment treatments under Contract 2 (39.8 – 31.3) while the difference between the Sort and Random Assignment treatments (37.5 – 31.3) provides a lower bound on the selection effect. The ratio of the two preceding differences (73%) provides a lower bound on the proportion of the increased effect of high performance pay with *endogenous* assignment to contracts that can be attributed to selection.[23] The specific point estimate of 73% is not terribly important. It is only a lower bound and is based on behavior by a specific set of subject in a specific environment. The point is that selection has a powerful effect in our experiments, more so than strategic anticipation.[24]

The power of selection can also be seen in Figure 3. Recall that subjects in the Sort (Reverse Sort) treatment were placed into groups according to predicted dropout time. Within a contract, the groups only differed in how strongly we predicted that they would have gotten the same contract (the other contract) in the Auction treatment. Any differences between these groups were purely due to selection. Figure 3 reports Block 2 data from groups that were on the border between the two contracts and groups that were at the extremes (i.e. most likely to be assigned to Contract 1 or Contract 2).[25] There are separate panels for the Sort and Reverse Sort treatments. Groups assigned to Contract 1 are shown in red and groups assigned to Contract 2 are shown

[23]The difference between this ratio and 50% is weakly significant ($p = .086$).

[24]In a companion paper (Cooper, Ioannou, and Qi, 2018), we develop a structural model of learning and fit it to the experimental data. We then use the model to examine what the selection effect would look like if the Sort treatment *perfectly* replicated the assignment to incentive contracts from the Auction treatment. We estimate that the true proportion of the improvement with endogenous assignment to high performance pay that can be attributed to selection is 90%.

[25] Each treatment had four sessions with 24 subjects and two sessions with 16 subjects. In a 24-subject session of the Sort treatment, the groups range from Group 1, consisting of the individuals predicted with the *highest* probability to be assigned to Contract 2 by the auction, to Group 6 which contains the individuals predicted to have the *lowest* probability of assignment to Contract 2 by the auction. Groups 1, 2, 3 are assigned to Contract 2 and Groups 4, 5, 6 are assigned to Contract 1. Groups 1 and 6 are defined to be the ***extreme*** groups, the groups predicted to be *most* likely to be assigned to their respective contracts. Groups 3 and 4 are defined to be the ***border*** groups, the groups *least* likely to be assigned to their respective contracts. Similarly, in a 16-subject session of the Sort treatment, Groups 1 and 4 are the extreme groups, and Groups 2 and 3 are the border groups. Border and extreme groups are defined in an analogous fashion for the Reverse Sort treatment. There is one border group and one extreme group per contract in each session.

in blue. For the Reverse Sort treatment, recall that groups assigned to Contract 1 consisted of individuals predicted to be assigned to Contract 2 in the Auction treatment, and vice versa for Contract 2. Border groups are shown with solid markers and extreme groups with hollow markers.

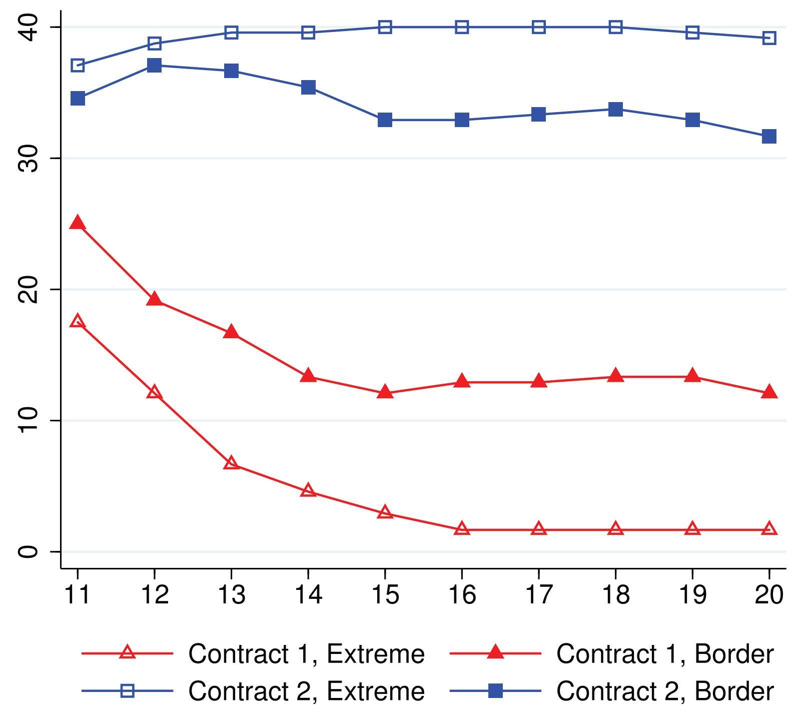
For both contracts in both treatments, the extreme groups differ from the border groups. For groups predicted to be assigned to Contract 1 (Contract 1 in Sort and Contract 2 in Reverse Sort), average effort is lower in Block 2 for the extreme groups. This flips for groups predicted to be assigned to Contract 2 (Contract 2 in Sort and Contract 1 in Reverse Sort). The difference between border and extreme groups is significant for groups predicted to be assigned to Contract 1 ($p = .010$) and groups predicted to be assigned to Contract 2 ($p = .003$). The power of selection can be seen along a number of other dimensions. In groups predicted to be assigned to Contract 2, the frequency of efficient coordination (all group members pick 40) rises from 51% for border groups to 84% for extreme groups. Likewise, in groups predicted to be assigned to Contract 1, the likelihood that the group's minimum effort equals 0 rises from 39% for border groups to 90% for extreme groups. This illustrates how powerful selection can be in isolation, and shows that our attempt to sort subjects into groups picks up something relevant for how the game was played.

Conclusion 2. *A number of features in our data point to the importance of selection. Comparing the effects of the Sort and Auction treatments suggests that the selection effect is stronger than the effect of strategic anticipation. The effect of selection is so strong that it can overcome the direct effect of lower incentives to yield coordination at high effort levels in the Reverse Sort treatment. In the Sort and Reverse Sort treatments, extreme groups (most likely to be selected into contracts) behave significantly differently than border groups, illustrating the power of selection.*

To understand why selection is so powerful, recall from Table 4 that Round 1 choices, particularly choice of 40, are by far the best predictor of dropout times. The tendency of the auction to assign subjects who chose 40 in Round 1 to Contract 2 is important because those subjects were extremely likely to also choose 40 in Round 11 regardless of their circumstances.

Figure 3: Comparison of Border and Extreme Groups in Block 2

Sort Treatment



Reverse Sort Treatment

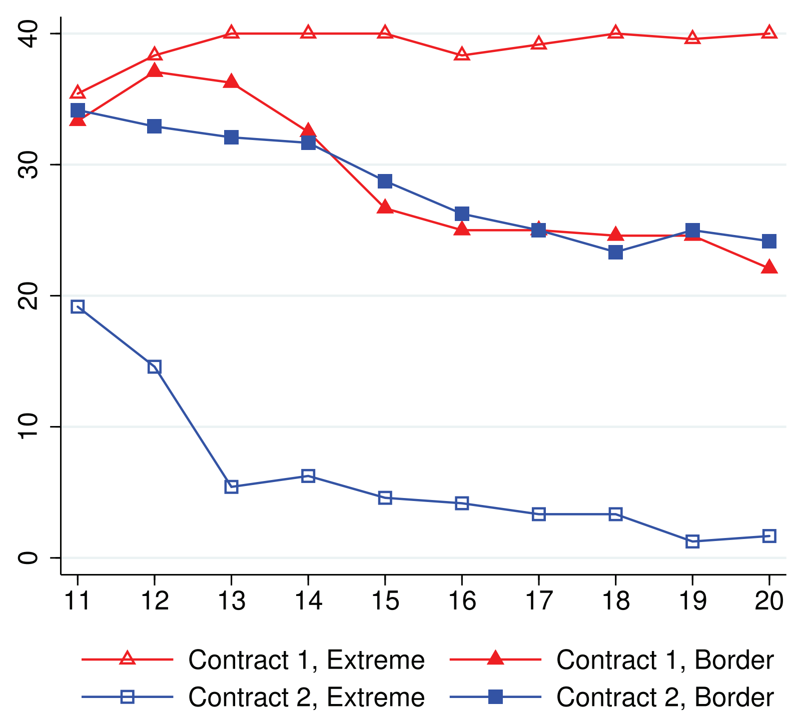
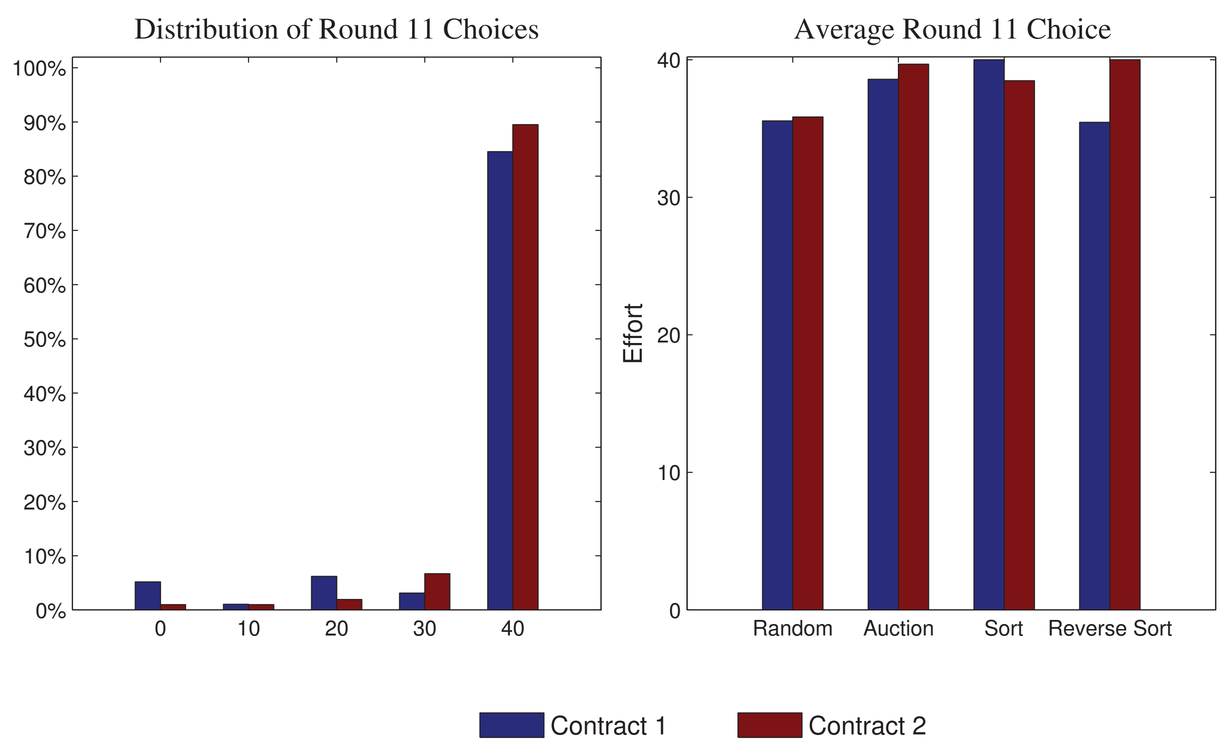


Figure 4 illustrates this point. The data for this figure come from all subjects who chose 40 in Round 1. The left hand panel displays the distribution of their choices in Round 11, broken down by the contract they were assigned in Block 2. For either contract, subjects who chose effort level 40 in Round 1 almost always chose 40 in Round 11. This is not because they failed to learn across Block 1. Most groups had low minimum efforts by the end of Block 1, and even the subset of initially optimistic subjects learned to adjust their choices during the course of Block 1. Only 36% chose 40 in Round 10, and only 10% chose something other than 0 in Round 10 if they were in a group with a minimum effort of 0. Rather, these subjects returned to their initial optimism when the game restarted with a new group and, possibly, a new contract. The right hand panel demonstrates how general this phenomenon was. It reports average effort in Round 11 broken down by treatment and contract. In all cases the average is close to 40.

Figure 4: Round 11 Behavior by Subjects Choosing 40 in Round 1



Subjects who chose 40 in Round 1 were far more likely to choose 40 in Round 11 than other subjects (85% vs. 39% in Contract 1 and 90% vs. 56% in Contract 2).[26] Initial optimism is not

[26]Based on Probit regressions, we find that the likelihood of choosing 40 in Round 11 does not vary significantly across choices 0 – 30 in Round 1, but jumps significantly if 40 was chosen in Round 1.

transient but instead seems to represent a subject's type. The auction tends to assign subjects who chose 40 in Round 1 to Contract 2, leaving others behind in Contract 1. The Sort treatment shares a tendency to assign subjects who chose 40 in Round 1 to Contract 2 and therefore yields similar outcomes to the Auction treatment.

5 Conclusion

The primary purpose of this paper was to investigate how endogenous assignment to incentive contracts affects the efficacy of high performance pay as a tool for escaping a productivity trap. The impact on groups that receive such performance pay is much greater when assignment is endogenous, with groups achieving almost perfect coordination at the efficient equilibrium. However, at a global level, this is entirely offset by the negative impact on groups that keep the original incentive contract with low performance pay. Our design makes it possible to identify a lower bound on how much of the increased effect of high performance pay with endogenous contract assignment is due to selection rather than strategic anticipation. We find that selection accounts for substantially more of the increase than strategic anticipation, even though the impact of selection is presumably underestimated due to the Sort treatment's imperfect replication of the selection process in the Auction treatment. The effect of selection is sufficiently strong that it can overcome the direct effect of incentives, as manipulating the assignment process in the Reverse Sort treatment generates higher productivity with *low* performance pay.

While a number of studies have found that selection accounts for a high proportion of the effect of performance pay, the flavor of our results is rather different. In the existing literature, selection has an effect by choosing people who are good at performing some task – solving puzzles, multiplying numbers, etc. In our experiment, selection chooses people who are optimistic about coordinating at efficient outcomes (or believe that others are optimistic). Selection is primarily based on beliefs rather than abilities.

Our finding that high performance pay increases productivity by matching optimists with each other is related to Becker's work (1973) on the marriage problem and Kremer's work (1993) on the assignment of workers to firms. Becker considers the allocation of men and women into marriages in which the "productivity" of each marriage is assumed to depend on the ability levels of each of the partners. Becker then provides conditions under which the efficient allocation of partners results in assortative matching.[27] Similarly, when the productivity of each worker in an organization depends positively on the productivity of co-workers, incentives exist for relatively high-skilled workers to form firms that exclude their lower-skilled counterparts (Kremer, 1993).

A surprising finding from our experiments is the minuscule total effect, averaging over both contracts, of endogenous assignment to incentive contracts. The underlying issue is that assigning more optimists to Contract 2 necessarily means assigning more pessimists to Contract 1. This implies that labor mobility plays a central role in determining the total effect of endogenous assignment to incentive contracts. In future work we hope to explore the implications of making labor markets open versus closed, in the sense defined previously, as well as allowing firms to choose an incentive contract rather than exogenously assigning incentive contracts to firms.

Our experiments take advantage of the control available in the laboratory. The insights we could draw from our data would be limited without our Sort and Reverse Sort treatments, but it is hard to imagine a real-world setting that would mimic these treatments. A natural question is how our results apply to labor markets in the field. Real labor markets are far more complex than our lab experiments. In the simple environment of our experiments, the data suggest that workers generally do not take into account the implications of endogenous assignment to incentive contracts and the resulting effects of selection. What are the odds that workers in more complex field settings understand these subtle issues? This implies that strategic anticipation plays an even smaller role in field settings than in the lab, a conjecture that we hope to verify in the future.

More generally, we hope that future research will establish a relationship in field settings between selection for optimists and organizational success. For example, there has been movement towards "agile" software development over the past two decades. Many specific processes for software development (e.g. Extreme Programming, Scrum, etc.) fall under the umbrella of agile software development. There are a number of common features associated with agile software development, but of particular interest to us is a focus on small self-organizing teams. These provide a very different work environment from more traditional top-down organizations and there exists evidence that these different organizational forms attract software developers with different personality traits (e.g. Yilmaz, O'Connor, Colomo-Palacios, and Clarke, 2017, employing the

[27] Specifically, Becker (1973) shows that in a transferable utility matching market, if there are complementarities in production, then any core allocation, and consequently, any competitive equilibrium, exhibits assortative matching. In other words, individuals sort themselves into matches with likely mates such that the ablest male is matched with the ablest female.

Big Five, find that extroversion is more prevalent among members of agile software development teams). It would be interesting to know whether, as our research suggests, there is selection based on beliefs about the likelihood of cooperation and/or efficient coordination within teams and whether this type of selection plays an important role in organizational success. We view this and similar issues as promising directions for future work.

Finding field evidence for *negative* effects due to selection is also of interest to us. The literature on entrepreneurship provides an interesting example of this. Less than 10% of new businesses are started by teams of non-relatives rather than individuals or relatives (Shane, 2008), but firms founded by teams tend to do better. This can be attributed in part to preferences for working alone rather than in a team and/or pessimism about the profitability of teams. Specifically, Cooper and Saral (2013) find that entrepreneurs are significantly less willing to join teams than otherwise similar business people. In other words, entrepreneurship tends to attract individuals whose preferences and beliefs make them *less* likely to succeed as entrepreneurs![28]

Our paper and much of the related literature stresses workers' choices of jobs as a source of selection, but firms can and do actively use applicants' characteristics when choosing who to hire. Personality testing, particularly use of the Big Five, became popular starting in the late 1980 and early 1990s. One recent study finds that about 20% of large firms use personality testing (Piotrowski and Armstrong, 2006). To the extent that they correctly identify the characteristics associated with success, this practice should have a positive effect like the Sort treatment. It remains controversial whether or not personality testing is useful for employers. While some argue that that personality tests are good predictors of "soft skills" like persistence and willingness to follow organizational norms (Barrick and Mount, 2012), others claim that the usefulness of personality testing is limited, partially because job candidates have a strong incentive to manipulate the test results (Stabile, 2001).[29] It remains to be seen if some form of behavioral testing including instruments more familiar to experimental economists will prove useful to firms when hiring new employees.[30]

[28]Cooper and Saral's experiment is designed to limit the possibility that beliefs can explain the relative disinterest of entrepreneurs in joining teams, stressing the role of preferences. We think it is likely that beliefs also play an important role, a conjecture that we hope to explore in the future.

[29]For other papers discussing the usefulness of personality testing, see Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt, 2007; Ones, Dilchert, Viswesvaran, and Judge, 2007.

[30]Unilever is an example of a firm using personality tests that mirror standard experimental tasks like the trust game and the balloon analogue risk task (Feloni, 2017).

References

AIMONE, J. A., IANNACCONE, L. R., MAKOWSKY, M. D., AND RUBIN, J. "Endogenous Group Formation via Unproductive Costs." *Review of Economic Studies* 80, no. 4 (2013): 1215–36.

BANDIERA, O., GUISO, L., PRAT, A., AND SADUN, R. "Matching Firms, Managers, and Incentives." *Journal of Labor Economics* 33, no. 3 (2015): 623–681.

BARRICK, M. R., AND MOUNT, M. K. "11 Nature and Use of Personality in Selection." *The Oxford Handbook of Personnel Assessment and Selection* 225.

BECKER, G. S. "A Theory of Marriage: Part I." *Journal of Political economy* 81, no. 4 (1973): 813–846.

BEN-PORATH, E., AND DEKEL, E. "Signaling Future Actions and the Potential for Sacrifice." *Journal of Economic Theory* 57, no. 1 (1992): 36–51.

BLUME, A., AND ORTMANN, A. "The Effects of Costless Pre-play Communication: Experimental Evidence from Games with Pareto-ranked Equilibria." *Journal of Economic Theory* 132 (2007): 274–90.

BOHNET, I., AND KÜBLER, D. "Compensating the Cooperators: Is Sorting in the Prisoner's Dilemma Possible?" *Journal of Economic Behavior & Organization* 56, no. 1 (2005): 61–76.

BORNSTEIN, G., GNEEZY, U., AND NAGEL, R. "The Effect of Intergroup Competition on Group Coordination: An Experimental Study." *Games and Economic Behavior* 41, no. 1 (2002): 1–25.

BRANDTS, J., AND COOPER, D. "A Change Would Do You Good... An Experimental Study on how to Overcome Coordination Failure in Organizations." *American Economic Review* 96 (2006): 669–93.

———. "It's What You Say, Not What You Pay: An Experimental Study Of Manager-Employee Relationships In Overcoming Coordination Failure." *Journal of the European Economic Association* 5, no. 6 (2007): 1223–68.

Brandts, J., Cooper, D. J., Fatas, E., and Qi, S. "Stand by Me: Help, Heterogeneity, and Commitment in Experimental Coordination Games." *Management Science,* 62, no. 10 (2016): 2916–36.

Brandts, J., Cooper, D. J., and Weber, R. "Legitimacy, Communication and Leadership in the Turnaround Game." *Management Science,* 61, no. 11 (2014): 2627–45.

Brekke, K. A., Hauge, K. E., Lind, J. T., and Nyborg, K. "Playing with the Good Guys. A Public Good Game with Endogenous Group Formation." *Journal of Public Economics* 95 (2011): 1111–8.

Broseta, B., Fatas, E., and Neugebauer, T. "Asset Markets and Equilibrium Selection in Public Goods Games with Provision Points: An Experimental Study." *Economic Inquiry* 41, no. 4 (2003): 574–91.

Cabrales, A., Miniaci, R., Piovesan, M., and Ponti, G. "Social Preferences and Strategic Uncertainty: an Experiment on Markets and Contracts." *The American Economic Review* 100, no. 5 (2010): 2261–2278.

Cachon, G. P., and Camerer, C. F. "Loss-avoidance and Forward Induction in Experimental Coordination Games." *Quarterly Journal of Economics* 111 (1996): 165–194.

Cadsby, C. B., Song, F., and Tapon, F. "Sorting and Incentive Effects of Pay for Performance: An Experimental Investigation." *Academy of Management Journal* 50, no. 2 (2007): 387–405.

Cason, T. N., Sheremeta, R. M., and Zhang, J. "Communication and Efficiency in Competitive Coordination Games." *Games and Economic Behavior* 76, no. 1 (2012): 26–43.

Charness, G., Cobo-Reyes, R., and Jiménez, N. "Efficiency, Team building, and Identity in a Public-goods Game.", 2011. UCSB working paper.

Chaudhuri, A., Schotter, A., and Sopher, B. "Talking Ourselves to Efficiency: Coordination in Inter-Generational Minimum Effort Games with Private, Almost Common and Common Knowledge of Advice." *The Economic Journal* 119 (2009): 91–122.

Cooper, D., Ioannou, C., and Qi, S. "Coordination with Endogenous Contracts: A Structural Learning Model Approach.", 2018. Working Paper, Florida State University.

Cooper, D. J., and Saral, K. J. "Entrepreneurship and Team Participation: An Experimental Study." *European Economic Review* 59 (2013): 126–140.

Cooper, R. C., De Jong, D., Forsythe, R., and Ross, T. "Communication in Coordination Games." *Quarterly Journal of Economics* 107 (1992): 739–71.

Crawford, V., and Broseta, B. "What Price Coordination? The Efficiency-enhancing Effect of Auctioning the Right to Play." *American Economic Review* 88 (1998): 198–225.

Dal Bó, P., Foster, A., and Putterman, L. "Institutions and Behavior: Experimental Evidence on the Effects of Democracy." *American Economic Review* 100 (2010): 2205–29.

Dohmen, T., and Falk, A. "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender." *American Economic Review* 101 (2011): 556–90.

Duffy, J., and Feltovich, N. "Do Actions Speak Louder Than Words? An Experimental Comparison of Observation and Cheap Talk." *Games and Economic Behavior* 39 (2002): 1–27.

———. "Words, Deeds and Lies: Strategic Behavior in Games with Multiple Signals." *Review of Economic Studies* 73 (2006): 669–88.

Eckel, C., and Grossman, P. "Forecasting Risk Attitudes: An Experimental Study Using Actual and Forecast Gamble Choices." *Journal of Economic Behavior & Organization* 68, no. 1 (2008): 1–17.

Eriksson, T., and Villeval, M.-C. "Performance-Pay, Sorting and Social Motivation." *Journal of Economic Behavior & Organization* 47, no. 3 (2008): 530–48.

Feloni, R. "Consumer-goods Giant Unilever has been Hiring Employees Using Brain Games and Artificial Intelligence — and It's a Huge Success." Business Insider. http://http://www.businessinsider.com/unilever-artificial-intelligence-hiring-process-2017-6, 2017. Accessed: 2018-03-10.

Fischbacher, U. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics* 10, no. 2 (2007): 171–8.

Fréchette, G. R. "Session-effects in the Laboratory." *Experimental Economics* 15, no. 3 (2012): 485–498.

Frey, M. C., and Detterman, D. K. "Scholastic Assessment or G? The Relationship Between the Scholastic Assessment Test and General Cognitive Ability." *Psychological Science* 15, no. 6 (2004): 373–378.

Glaeser, E., Laibson, D., Scheinkman, J., and Soutter, C. "Measuring Trust." *Quarterly Journal of Economics* 115, no. 3 (2000): 811–846.

Gosling, S. D., Rentfrow, P. J., and Swann, W. B. "A Very Brief Measure of the Big-Five Personality Domains." *Journal of Research in Personality* 37, no. 6 (2003): 504–528.

Greiner, B. "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE." *Journal of Economics Science Association* 1, no. 1 (2015): 114–125.

Gunnthorsdottir, A., Houser, D., and McCabe, K. "Disposition, History and Contributions in Public Goods Experiments." *Journal of Economic Behavior & Organization* 62, no. 2 (2007): 304–315.

Hamman, J. R., Rick, S., and Weber, R. "Solving Coordination Failure with All-or-None Group Level Incentives." *Experimental Economics* 10, no. 3 (2007): 285–303.

Knez, M., and Simester, D. "Form-Wide Incentives and Mutual Monitoring At Continental Airlines." *Journal of Labor Economics* 19, no. 4 (2001): 743–72.

Kogan, S., Kwasnica, A. M., and Weber, R. "Coordination in the Presence of Asset Markets." *American Economic Review* 101, no. 2 (2011): 927–47.

Kosfeld, M., Okada, A., and Riedl, A. "Institution Formation in Public Goods Games." *American Economic Review* 99, no. 4 (2009): 1335–55.

Kosfeld, M., and von Siemens, F. A. "Worker Self-Selection and the Profits From Cooperation." *Journal of the European Economic Association* 7, no. 2-3 (2009): 573–582.

———. "Competition, Cooperation, and Corporate Culture." *The RAND Journal of Economics* 42, no. 1 (2011): 23–43.

Kremer, M. "The O-ring Theory of Economic Development." *The Quarterly Journal of Economics* 108, no. 3 (1993): 551–575.

Lazear, E. P. "Performance Pay and Productivity." *American Economic Review* 90, no. 5 (2000): 1346–61.

Lazear, E. P., Malmendier, U., and Weber, R. A. "Sorting in Experiments with Application to Social Preferences." *American Economic Journal: Applied Economics* 4, no. 1 (2012): 136–163.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., and Schmitt, N. "Reconsidering the Use of Personality Tests in Personnel Selection Contexts." *Personnel Psychology* 60, no. 3 (2007): 683–729.

Myung, N. "Improving Coordination and Cooperation through Competition.", 2012. Working Paper, California Institute of Technology.

Ones, D. S., Dilchert, S., Viswesvaran, C., and Judge, T. A. "In Support of Personality Assessment in Organizational Settings." *Personnel Psychology* 60, no. 4 (2007): 995–1027.

Ones, U., and Putterman, L. "The Ecology of Collective Action: A Public Goods and Sanctions Experiment with Controlled Group Formation." *Journal of Economic Behavior & Organization* 62, no. 4 (2007): 495–521.

Piotrowski, C., and Armstrong, T. "Current Recruitment and Selection Practices: A National Survey of Fortune 1000 Firms." *North American Journal of Psychology* 8, no. 3 (2006): 489–496.

Riedl, A., Rohde, I. M. T., and Strobel, M. "Efficient Coordination in Weakest-Link Games." *Review of Economic Studies* 83, no. 2 (2016): 737–767.

Rietz, T. A., Sheremeta, R. M., Shields, T. W., and Smith, V. L. "Transparency, Efficiency and the Distribution of Economic Welfare in Pass-through Investment Trust Games." *Journal of Economic Behavior & Organization* 94 (2013): 257–267.

Salmon, T. C., and Weber, R. A. "Maintaining Efficiency While Integrating Entrants From Lower Performing Groups: An Experimental Study." *The Economic Journal* 127 (2017): 417–444.

Shane, S. A. *The Illusions of Entrepreneurship: The Costly Myths that Entrepreneurs, Investors, and Policy Makers Live By*. Yale University Press, 2008.

Sherstyuk, K. V., Karmanskaya, N., and Teslia, P. "Bidding with Money or Action Plans? Resource Allocation under Strategic Uncertainty.", 2014. Working Paper, Chapman University.

Stabile, S. J. "The Use of Personality Tests as a Hiring Tool: Is the Benefit Worth the Cost?" *University of Pennsylvania Journal of Labor and Employment Law* 4 (2001): 279.

Van Huyck, J., Battalio, R., and Beil, R. "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure." *American Economic Review* 80 (1990): 234–48.

———. "Strategic Uncertainty, Equilibrium Selection, and Coordination Failure in Average Opinion Games." *Quarterly Journal of Economics* 106 (1991): 885–911.

———. "Asset Markets as an Equilibrium Selection Mechanism: Coordination Failure, Game Form Auctions, and Tacit Communication." *Games and Economic Behavior* 5 (1993): 485–504.

Weber, R. A. "Managing Growth to Achieve Efficient Coordination in Large Groups." *American Economic Review* 96, no. 1 (2006): 114–26.

Yilmaz, M., O'Connor, R. V., Colomo-Palacios, R., and Clarke, P. "An Examination of Personality Traits and How They Impact on Software Development Teams." *Information and Software Technology* 86 (2017): 101–122.