Experimental Organizational Economics*

Colin F. Camerer

Div HSS 228-77 Caltech, Pasadena CA 91125 USA

camerer@hss.caltech.edu

Roberto A. Weber

Social & Decision Sciences, Carnegie Mellon University, Pittsburgh PA 15213

rweber@andrew.cmu.edu

I. Introduction

This chapter is about experiments that study aspects of organizational structure and economic performance. Relative to field studies using empirical data, experiments often have obvious advantages, especially the value of control and randomized assignment to implement theoretical assumptions that can only be imperfectly measured or controlled econometrically using field data. Despite these advantages, the range of organizational hypotheses studied in experiments is small, although it is growing rapidly.

What makes an experiment organizational? The working definition we used to define the scope of this chapter is this:

> In an organizational experiment, two or more subjects engage in a productive enterprise (usually a highly reduced-form model of a more complex organization) where the impact of the structure of the organization and incentives on productivity, through the characteristics and choices of subjects, are the variables of central interest.

This definition is crafted to exclude (for brevity) consideration of many experiments relevant to organizations but that are not *special* to organizations and do not vary structural variables of most interest in the economic study of organizations. One prominent class of such experiments is those that use simple interactions to estimate the effect of social preferences on behavior (e.g., public goods, ultimatum, and trust games). While valuable, this research is not special to organizations, since the games also apply to market bargaining, groups, households, and other contexts (see Ledyard, 1995; Camerer, 2003 for reviews). Another class are early experiments on principal-agent contracts which emphasize risk-incentive tradeoffs but have very simple choices (e.g., Berg, Dickhaut Another prominent class is experiments on negotiations studied by scientists interested in organizational behavior. We exclude these important experiments because they do not typically alter incentives or hierarchical structure to test particular ideas in organizational economics, and their results are well-reviewed elsewhere (e.g., Bazerman, Curhan, Moore and Valley, 2000 Annual Review of Psychology).

An obvious concern about using experiments to study organizations is that the organizations that are the typical focus of theorizing – such as large firms – are complicated. Experiments are necessarily much simpler, and usually constrained in the numbers of agents and length of time that can be studied in a way that is both inexpensive and well-controlled. This mismatch in scope leads to heightened concerns about generalizability (which we discuss further below).

The chapter is organized in seven sections. Section II presents an introduction to experimental principles and some early history of organizational experiments in psychology and sociology. The central sections, III through VI, discuss experimental research in four important organizational areas: (1) basic theories of incentives and worker effort; (2) voluntary effort and reciprocity; (3) coordination; and (4) leadership and hierarchy. Section VIII concludes and discusses some types of organizational experiments that could provide valuable insights into organizational economics, but have not yet been conducted.

II. Experimental principles and some history

A. Basic principles of experimentation in economics

Experiments are useful in all sciences. The experimental environment is one carefully designed and controlled by the researcher to provide the cleanest test of a theory or hypotheses.

For instance, an experiment can allow researchers to measure the outcomes of both an event and its counterfactual (such as an institutional change or a merger), when the counterfactual is typically not observed outside the lab. An experiment also allows a researcher to avoid the endogeneity biases present in much empirical field research, by providing the researcher control over what is exogenous. Experiments also allow the researcher to distinguish and test between different explanations for an observed phenomenon (for instance, by controlling for transactions costs, risk aversion, etc) that are hard to separate parametrically in field data.

The key feature of any experiment is *control* – over factors that are important in the theory being tested and over factors that need to be excluded from the environment in order to properly test theory.[1]

B. What makes for a good experiment?

Good experiments typically satisfy four basic criteria: control over all relevant factors in the experimental environment; internal validity (the treatments work as intended and for the reasons claimed by the researcher); statistical integrity (using standards applicable to other empirical work[2]); and generalizability.[3] It is usually easy to evaluate these criteria. But even when it is not, the threat of direct replication disciplines researchers and gradually weeds out unreliable effects that might result from weaknesses in the basic criteria.

Really good experiments have other properties that are more difficult to define than the basic ones but are just as important. For example, an "interesting" experiment grabs attention by producing a result that is both surprising – Davis (1971) calls it 'denying assumptions of the audience' – but is also comprehensible after the surprise wears off. A "fruitful" experiment creates a paradigm (and sometimes tools, like software) and a result that piques curiosity and leads to natural extensions of the paradigm. One component of fruitfulness is the capacity of an experiment that rejects a theory to suggest an alternative theory and new experiments worth doing.

C. Organizational experiments and generalizability

The canonical organizations of most interest in economics – large firms – consist of agents who are both self-selected and selected by organizational promotion tracks, highly motivated by financial incentives, often highly experienced in their role, and familiar with other organizational members. The typical laboratory experiment uses random assignment of inexperienced subjects who are unfamiliar with one another to simple organizational settings with relatively modest incentives. Despite the obvious challenge of generalizing from such simple experiments, there are three arguments in favor of doing simple laboratory experiments on organization.

First, in science it is usually surprisingly useful to start with the simplest cases, then generalize (much as in producing formal theory). Basic genetics were first understood by studying fruit flies, which are much simpler than humans. Earthquakes are understood by using “slider block” models that study wooden blocks connected by metal springs (representing tectonic plates). At this point, organizational experiments are similar in complexity to fruit fly studies of genetics and slider block models of earthquakes. Fortunately, there is plenty of future opportunity to create experimental organizations that are larger in scale, complexity, and temporal scope than the simple ones we describe.

Second, some organizations *are* simple, so simple experiments are good analogies. A startup with a few inexperienced employees or a project team within a firm that involves a small group of strangers who interact in an unfamiliar production task for a limited period might be well-modelled by simple experiments like those described later.

Third, the presumption in economic theory is that basic principles apply widely unless the theory explicitly contains variables that parameterize boundaries of likely empirical applicability. Nothing in principal-agent theory, for example, says the theory should apply to CEO compensation but not to a three-person law firm with one lawyer and two paralegals, or to a three-person lab experiment with an economic structure of the law firm. If obvious distinctions between large firms and small experiments, like incentives or experience and selection of agents, are expected to make a difference in behavior, then the way in which those distinctions matter should be part of the theory, If they are not, experimental findings can suggest how those distinctions should be included.

When judging the generalizability of organizational experiments it is also important to distinguish the power of the *experimental method* with the progress made from *experiments thus far*. If there are doubts about whether worker reciprocity persists over long time periods, for example, this is a criticism about the generalizability of particular short-run experiments to particular long-run settings, but is not a criticism of the experimental method – because, in principle, experiments can be conducted over the long-run. Moreover, field experiments in markets and firms with some degree of control can be powerful complements to laboratory experiments.

Thus, readers should think of the summary of what has been learned so far as lessons about interesting effects. Where there is a doubt that a particular experimental finding will generalize, those doubts are most useful if expressed as a conjecture about how extending the criticized experiment in some feasible way would change the results.

D. Early (non-economic) organizational experiments

The earliest organizational experiments, as we define them, were simple social psychological experiments aimed at understanding how performance and productivity change in the presence of others. Early experiments on "social facilitation" (Triplett, 1898), measured how the presence of other individuals can lead to increased effort and performance in simple tasks (such as pedaling a bicycle). Later experiments by management researchers and industrial psychologists (Taylor, 1911; Mayo, 1933) studied factors that influence productivity and performance. Most of this research took place in real firms. Researchers exogenously varied factors such as incentives, task design, and environmental influences, and then measured performance under varying factors. This early research demonstrated the value of experimentation for understanding organizations, and influenced the development of management and organizations as areas of scholarly inquiry (March and Simon, 1958 Organizations).

Starting in the 1950's organizational psychologists started using simple laboratory experiments to study social factors that might influence efficiency and productivity in organizations. For example, several experiments explored the impact of group structure or "communication nets" on productivity in simple tasks (e.g., Bavelas & Barrett, 1951; Leavitt, 1951; Guetzkow & Simon, 1955 Management Science). These experiments provided valuable insights into how communication and authority structures in organizations might influence performance, and also demonstrated the potential value of simple laboratory experiments for understanding organizational phenomena.

In subsequent decades, managerial and organizational researchers – usually trained as psychologists – continued to use experiments to understand organizations. In 1969, a special issue of a leading management journal was dedicated to experiments on organizations (see Weick, 1969 Administrative Science Quarterly). This special issue addressed topics such as performance in hierarchical vs. flat organizations (Carzo and Yanouzas, 1969), internal resource allocation (Pondy and Birnberg, 1969), and employee performance in response to varying levels of payment (Evan and Simmons, 1969) – an experiment similar to recent "gift exchange" experiments in economics (see section III.B). The experiments varied in whether they took place in a laboratory or in the field, and as with many subsequent non-economic organizational experiments they also frequently involved deception of subjects by the researchers (a practice frowned upon by experimental economists – see Ortmann and Hertwig, 2002 Experimental Economics).

Following the above research, organizational and managerial researchers have since continued to rely on laboratory and field experiments to address similar questions (e.g., Argote, Turner & Fichman, 1989 OBHDP; Greenberg, 1990 Journal of Applied Psychology; Bottom et al 2006 ASQ). While we omit this research to focus on experiments in organizational *economics*, this work frequently explores similar questions as the work we review below.[4]

We now review experimental research in four areas of interest to organizational economists: (1) theories of incentives and worker effort; (2) voluntary effort and reciprocity; (3) coordination; and (4) leadership and hierarchy. Each of the next four sections reviews experimental research on one of these topics, focusing on a few key studies and cumulation of regularity across studies.

III. Incentives and worker effort

The first wave of experiments we describe study the most basic propositions of incentive theory. The relationships between different kinds of incentives, effort, and productivity are central to organizational economics. These experiments exploit the sharp precision of predictions that can be derived when cost of effort is induced experimentally and comparative statics make clear predictions about which types of contracts will work well or poorly.

A. Basic incentives: Piece-rates and teams

Nalbantian and Schotter (NS, 1997 AER) studied a variety of incentive schemes in simple lab experiments. In their setup, teams of six agents choose efforts e_i in the interval [0, 100]. Group output is a simple sum of effort plus a random term, $Y=1.5(\Sigma_i e_i+\varepsilon)$, where ε is uniform in [-40,+40]. Revenue is equal to group output times a constant ($R = 1.5Y$). Agents' effort costs are $(e_i)^2/100$.

Optimal net revenue for the group (less effort costs) occurs at e_i*=75. NS consider a variety of contracts that either induce this effort (in theory) or that theoretically induce lower effort due to shirking. These contracts and some basic results are summarized in Table 1.

Phase I experimental sessions begin with a 25-period revenue-sharing segment, followed by a switch to another scheme. Phase II experiments used the opposite order. Two experimental sessions used monitoring schemes, reversing the order of two different probabilities of being monitored.

The baseline contract is simple revenue-sharing (RS), in which players earn R/6 and pay their own effort costs.[5] This contract should induce shirking and does so reliably; mean efforts start around 35 and drift down to 18, toward the equilibrium of 12.5. A related contract, profit-sharing (PS), involves players sharing revenue only if average output exceeds the equilibrium quantity. This treatment also works roughly as predicted and generates little effort.

Table 1: Equilibrium and actual effort in various incentive schemes (Nalbantian and Schotter, 1997)

Contract	Individual payoff to i	Equil. effort e*	Equilibrium group profit	Mean effort
Revenue-sharing	$Y/6 - c(e_i)$	e*=12.5	103.1	19
Forcing contract (75)	$Y/6 - c(e_i)$ if $Y \geq R^*=450$; B=1.125otherwise	e*=75	337.5	16
Profit sharing	$(Y-Y^*)/6 - c(e_i)$ if $Y \geq Y^*=75$; B=0 otherwise	e*=19.1	150	22
Team competition	$(Y_1+T)/6-c(e_i)$ if $Y_1>Y_2$ $(Y_1-T)/6-c(e_i)$ if $Y_1<Y_2$ T=360	e*=75	337.5	35
Monitoring (p=.7)	monitoring p=.7, earn W=112.5 if e_i =75, w=18.75 if e<75	e*=75	337.5	68
Monitoring (p=.3)	monitoring p=.3, earn W=112.5 if e_i =75, w=18.75 if e<75	e*=0	-506.25	4

Note: Mean effort is from the last 5 periods (out of 25) phase I only. Means are estimated from plots. [a]Mean from phase II, estimated from Figure 10 (corresponding mean for forcing contract with e*=75 is effort of 19).

In a Hölmstrom-type (1982 Bell Journal of Economics) forcing contract, players share revenue only if total revenue is above the efficient level of 675 (corresponding to e_i*=75); otherwise, all players earn a low penalty wage $B \leq 1.125$. This forcing contract does badly because it is so fragile: Mean efforts start around 45, but all groups fall below the threshold output requirement, everyone earns the penalty wage, and efforts then fall. A forcing contract that is inefficient, inducing effort of 40 in theory rather than 75, actually works much better in practice than the efficient forcing contract.

The contracts that work best, both at producing high effort and in the accuracy of theoretical predictions, involve team competition and monitoring. In the competitive teams (CT) treatment, firms consist of two six-person teams and efforts create team output and revenue as in treatment RS. However, the lower-output team in each pair of teams pays a transfer of T=360 (shared equally among team members) to the team with higher team output. In theory this contract should induce equilibrium effort of 75. Actual mean efforts are about half that level (with low variance across subjects).

Monitoring contracts work best of all. Players choose effort and are monitored with probability p. If they are monitored and chose effort of 75 or more, or are not monitored, they earn a high wage W=112.5. If they are monitored and chose $e<75$, they earn 18.75. When the monitoring probability is p=.70, they should choose $e^*=75$, and monitoring does work incredibly well – median effort is 75 in about 90% of the experimental periods. When p=.30 they should shirk and choose $e^*=0$, and median efforts are indeed very low.

B. Tournaments and sabotage

Lazear and Rosen (1981 JPE) note that many labor markets, particularly internal competition for promotion in firms, are organized as "tournaments." In the simplest tournament model, individual workers exert costly effort e_i, their individual output (y_i) is a combination of (unobserved) effort and a measurement error term θ_i, and workers are ranked by total output $y_i=e_i+\theta_i$. Workers with rank r earn a pre-announced prize M(r). Ranking workers is often cognitively easier than measuring their absolute outputs. Lazear and Rosen show conditions under which tournaments elicit maximum effort, and are better incentive schemes than piece-rates and other individual-specific wage packages.

Explicit tournaments are often used in sports (e.g. Ehrenberg and Bognanno, 1990 Industrial and Labor Relations Review) and some areas of the economy (e.g. Knoeber, 1989; Lambert, Larcker, and Weigelt (1993, Admin Science Quarterly). Generally, however, it is difficult to test the sharpest predictions of tournament theory with field data because it is difficult to measure or influence key variables.

Bull, Schotter and Weigelt (1987 JPE) did the first such tournament experiment. Subjects chose effort ($e_i \in [0,100]$) in two-player tournaments, with output $y_i = e_i + \theta_i$ determined by effort and an i.i.d. stochastic term θ_i drawn from uniform interval [-a,+a], with a=40, and convex effort costs $c(ei) = e_i^2/K$, where K varied from 10,000 to 25,000. The winner prize M(1) varied from $1.02 to $1.45, while the second (loser) prize M(2) varied from $0.43 to $0.86.

Table 2 gives results from the last half of the 12-period experimental session. The piece rate incentive produced outcomes very close to the theoretical prediction under risk neutrality. While aggregate behavior in the first three tournament sessions (baseline, high-effort, and wide random number) converges towards the respective theoretical predictions, the variance in effort is quite large (much higher than for the piece rate) and does not decline across the session.

The results of providing additional information are a little puzzling. Providing ranks (baseline), or ranks and opponent effort (high information) produce efforts close to the equilibrium prediction, but supplying only rank and outputs (medium information) produces efforts that are much higher. Subjects appear to overreact to learning that their opponent output is high, even though they realize that high output could be due to luck.

Table 2 : Predicted and actual behavior in two-player tournaments (Bull, Schotter and Weigelt, 1987)

	equilibrium prediction	rounds 7-12 effort	
Treatment		mean	variance
piece rate	37	38.91	87
baseline	37	38.75	499
high effort equilibrium	74	69.91	892
wide random number (a=80)	37	33.41	286
asymmetric costs	70,35	73.60, 56.91	708, 905
medium information (output)	37	48.91	442
high information (output, e_i)	37	33.48	552
automaton opponent	37	44.62	276
25-round experiment	37	48.00*	362*

*Data from the last half of the experiment (rounds 13-25).

Sessions with automated opponents choosing Nash effort levels generally produce efforts close to equilibrium, and lower variance than sessions with human opponents, which shows that strategic uncertainty when playing other subjects may increase variance.

BSW also conducted one session with asymmetric costs, in which one player's cost function was twice that of the other player. Both players chose effort levels above equilibrium, which is important since most naturally-occurring tournaments do have cost asymmetries.

Schotter and Weigelt (1992 QJE) studied asymmetric two-player tournaments with two kinds of handicaps for player 2. In "uneven" tournaments the disadvantaged player has effort costs that are $\alpha>1$ times larger than the advantaged player. In "unfair" tournaments the disadvantaged player must have output that is $k>0$ units higher than the advantaged player in order to win.

With a cost handicap α and the evaluation bias k, the equilibrium efforts for players 1 and 2 are $e_1^*=\alpha e_2^*$ and $e_2^*= \{[(1/2a)-(k/4a^2)]c(M-m)/2\alpha\}/\{1+[(1-\alpha)/4a^2]c(M-m)/2\alpha\}$. These tournaments have interesting properties. If there is no cost handicap ($\alpha=1$) then even when there is a bias k, both players should exert the same effort ($e_1^*=e_2^*$). When there is a bias k, both players' efforts are decreasing in k. Intuitively, when there is a bias the advantaged agent doesn't have to work too hard (since the marginal impact of effort on winning probability is lower) and the disadvantaged agent shouldn't bother to work too hard (because extra effort probably won't overcome the bias). Second, in the presence of a cost handicap (α), a bias (k) in evaluation can raise total effort. This provides a possible rationale for "affirmative action", adding a handicap to even a tournament where one player has a cost advantage.

Table 3 : Predicted efforts and actual efforts (periods 11-20) in asymmetric two-player tournaments (Schotter and Weigelt, 1992)

Type of tournament	Cost handi-cap α	Effort advantage k	Advantaged player effort		Disadvantaged player effort	
			equil.	Mean (s.d.)	equil.	Mean (s.d.)
1. Baseline	None	None	73.8	77.9 (24.8)	n.a.	n.a.
2. unfair	None	k=25	58.4	74.5 (19.9)	58.4	58.7 (37.9)
4. uneven	α=2	None	74.5	78.8 (21.8)	37.3	37.1 (28.9)
6. Affirm. Action	α=2	k=25	59.0	64.2 (32.3)	29.5	36.4 (27.7)
7. affirm. Action	α=4	k=25	60.2	85.8 (19.6)	15.1	32.4 (25.7)

Table 3 summarizes predicted efforts and mean efforts in the last 10 periods of Schotter and Weigelt's 20-period sessions. With no asymmetry ($\alpha=1$, $k=0$) average efforts of 77.9 are close to the predicted effort of 73.8, replicating the BSW result. When asymmetries are introduced, all the comparative statics move in the right direction. However, both a cost handicap and evaluation bias lead to more effort by both advantaged and disadvantaged players than is predicted by theory.

The experiments described above all use two-player tournaments. However, organizational tournaments typically involve more than two workers. Some recent experiments study multi-person (n>2) tournaments.

Orrison, Schotter and Weigelt (2004 Management Science) used the paradigm and parameter values in the experiments above. They varied the size of the tournament (k=2, 4, or 6 players), the proportion of players to be awarded the high prize, and the symmetry of the competition (by varying the handicap k between 0 and 25).

In their symmetric treatments, predicted effort is always the same (73.75). As predicted, OSW find that behavior is insensitive to tournament size when half the players win the high prize. However, the proportion of large prizes influences behavior in a way that is not predicted by theory. In six-player tournaments, when two or three of the players earn the highest prize effort is close that predicted, but when the large prize is awarded to four players, mean effort drops to 59. In the asymmetric treatments, predicted effort is higher in six-person groups than in four-person groups, as predicted. But contrary to the theoretical predictions, two-person groups do not produce lower effort – and often even have higher effort than the larger groups.

Müller and Schotter (2007, Unpublished) study a more complicated tournament in which four workers have private costs parameters and there are two prizes V1 and V2, with a fixed total V1+V2=1 (based on Moldovanu and Sela, 2001 AER). Theory predictions are sharply confirmed with two interesting exceptions. Low-cost workers should win all the time, but the predicted winners win only about half the time. And when costs are low, subjects typically exert too much effort ("workaholics"); when costs are high, subjects often "drop out" and choose $e_i=0$.[6]

While tournaments have many advantages – chiefly, they require only relative comparisons of output – they have a major drawback: workers may have an incentive to spend effort "sabotaging" perceptions of output of other workers. If sabotage is cheap and effective, heightened incentives to work hard from tournament bonuses may be undermined by both the diversion of effort into sabotage, and by the withdrawal of effort in anticipation of sabotage (cf. Milgrom and Roberts, 1988).

Carpenter, Matthews and Schirm (2007) did a clever experiment on sabotage with actual effort. Groups of eight subjects privately fill out forms and hand-address envelopes for 30 minutes. This simple activity permits differentiation of the quality of output because envelopes can be either clearly or sloppily addressed. The quality of each envelope was rated on a 0-1 scale of deliverability by an actual letter carrier from the U.S. Postal Service. Measured output is the sum of these quality ratings across all the addressed letters. In addition, after finishing the task all the workers looked at the output of other workers, selected one envelope at random to inspect, and rated its quality.

There are three experimental conditions: Piece-rate ($1 times output); tournament ($25 bonus for highest output); and tournament with sabotage. In the tournament with sabotage, the quality-adjusted output of a worker is determined by the average of the other worker's quality ratings (like a "360-degree evaluation" system). Opportunistic workers should work hard, but should also sabotage their peers by underrating the quality of their peers' work. Hiring the postal carrier to rate objective quality provides a clear way in which to address whether peer-rated quality is lower than objective quality (i.e., whether there is sabotage).

Table 4: Average raw output and quality-adjusted output (peer and expert ratings) in piece rate and tournament conditions (Carpenter, Matthews, Schirm 2007)

		Tournament output minus piece rate output	
quality adjustment?	piece rate	no sabotage	Sabotage
None (total letters)	13.6	+1.1	-1.3
peer-rated	11.1	-.4	-4.1
expert-rated	11.7	+1.2	-1.5

Note: Outputs are envelopes stuffed and addressed. Quality adjustment is output times a 0-1 measure assessed by peer ratings or expert (letter-carrier) ratings.

Table 4 summarizes raw and quality-adjusted output in the three experimental treatments. In the piece-rate condition, workers address 13.6 envelopes but both peers and the expert adjust the output downward about 20% for quality of handwritten envelope addresses. In the tournament with no sabotage (where peer ratings don't matter for payoffs), output goes up in raw terms, and as judged by the expert, compared to the piece-rate, but peers still rate quality as lower (even though underrating their peers does not benefit them). Carpenter et al interpret this as an automatic affective response to competition. In the sabotage condition, where workers might benefit from underrating peer quality, peer-rated quality is about 40% lower than in the piece-rate baseline – there is strong sabotage – and expert-rated quality falls as well. That is, when workers can have a direct influence on how their peers' work is perceived, there is both substantial sabotage of high-achieving peers, and a drop in effort (judged by expert quality ratings) in anticipation of sabotage. The drop in effort is so strong that output is higher in the piece-rate condition than in the tournament with sabotage.

C. Anomalous response to incentives

Several experiments show responses to incentives which are theoretically surprising and counterintuitive. These studies are important for understanding when incentives in organizations are likely to have unintended effects.

1. Incentives do not necessarily produce more effort (crowding out)

One of the simplest predictions of labor economics is that the supply of labor should increase in response to a transitory increase in wage. Studies of temporary wage increases (e.g., Mulligan 1995) suggest this law of supply is often true. However, using field data on cab driver hours, Camerer et al (1997 QJE) found that labor supply of inexperienced drivers slopes downward, because inexperienced drivers act as if they establish a daily income target and quit when they reach the target.[7] In a field experiment with bicycle messengers, Fehr and Goette,(2007 AER) found that hourly effort of messengers decreased with temporary wage increases (though they also worked more hours, which led to an increase in the amount of labor supplied under the wage increase)..The lesson of these studies is that labor supply and effort are sometimes sensitive to a point of reference in income.

Another kind of reference point is a standard for a fair or reasonable wage. Gneezy and Rustichini (2000) investigated effort in response to no incentive, very low incentives, and higher incentives, in performance on a 50-question IQ test. When pay is only 0.1 New Israeli Shekels (at the time of the experiment, about $.03) per correct answer, the average number of questions answered correctly is lower than when there is no pay, mostly because several subjects "quit" and do poorly on the test (like a labor strike) (see also Frey & Jegen, 2001; Heyman & Ariely; 2004).

Such "crowding out" of intrinsic incentives can also occur when people exert effort out of moral obligation, and pay drives out moral obligation. Gneezy and Rustichini (GR, 2000) studied this effect in a field experiment, at 10 private day-care centers in Israel. Before the experiment, parents were told to pick up their children on time. During the experiment, a modest fine was instituted for late pick-ups and then rescinded. When the fine was introduced, late pickups actually *increased* (compared to control groups with no new fine). When the fine was rescinded late pickups dropped but settled down to about twice the rate of late pickups before the fine. GR suggest the "fine is a price" – the moral obligation of picking up your children on time is driven out by a new low price you can comfortably pay for being late; and when the price is removed the moral obligation is not restored (e.g. Lepper, Greene and Nisbett, 1973).

2. Higher effort does not necessarily improve performance

The standard incentive model assumes that higher incentives lead to higher effort, which improves performance. Psychologists, however, have long suspected that at very high levels of arousal, performance can suffer because of arousal and distraction (such as "choking"). Yerkes & Dodson (1908), initially demonstrated this effect by testing the ability of rats to discriminate between safe and dangerous (i.e., shock-inducing) areas in a cage. The rats learned to do so most quickly when the shocks were at an intermediate level of intensity, and actually learned more slowly when shocks were stronger. Camerer (1998). noted this effect and reported data from NBA playoff free-throw shooting. Players make significantly fewer shots in the playoffs than in the regular season (about 75% vs 73%), presumably because of the heightened pressure during the playoffs.

Ariely, Gneezy, Loewenstein and Mazar (2005) demonstrated the Yerkes-Dodson pattern with a high-stakes experiment in a poor town in India, using six games that require concentration and physical skill. The highest performance level in each game earned 400 Indian Rupees (Rs), about a month's consumer spending. Subjects reached the highest performance in only 6% of games with high stakes, compared to 22% and 26% for low and medium stakes. The highest stakes seems to create nervousness or "choking" which actually lowers short-run performance.

D. Field experiments on incentives and monitoring

Some interesting experiments on the effect of incentives on employee effort and productivity have been conducted in the field. These studies serve as a valuable complement to laboratory experiments.

Nagin et al (2002) report an interesting field experiment that tests the effect of monitoring on employee moral hazard. Their experiment was conducted by a company that hires people to solicit donations to non-profit organizations over the phone. The experiment varies the likelihood of detecting low effort on the part of employees (as in the monitoring treatments of Nalbantian and Schotter, 1997).

The workers earn a piece-rate wage that increases with the number of successful solicitations. As a result, there is a moral hazard incentive to claim that people who were called had pledged donations over the phone, when those people actually had not. The employer tried to limit moral hazard by independently calling back a fraction of those people who had allegedly pledged to give money and identifying "suspicious bad calls" (SBCs), or cases indicative of misrepresentation by the employees. SBCs were deducted from the original piece-rate pay but were not directly linked to firing or other sanctions.

Because the callbacks are expensive, the company was interested in whether lowering the rate of calling back would lead to many more SBCs. They designed an experiment in which the monitoring (call-back) rates were varied across 16 separate call centers and across time, from 0 to 25%. In four sites, the actual rate of calling back was 25% during the experimental period, but only a fraction of those monitored call results were reported to the employees. For example, in one site 25% of the successful calls were called back, but only 8% of those results were reported back to the employees, so the employees were getting feedback as if they faced a 2% callback rate. This design enabled the firm to measure whether employees who were actually making a lot of SBCs but thinking that they were "getting away with it" most of the time would increase their SBC moral hazard in the future. (The design was double-blind so that both supervisors and employees did not know the percentage of monitored calls.)

Table 5 shows the marginal effect on the number of per-week SBCs for week t, based on the monitoring percentage observed by employees in the previous week t-1. For example, employees who were subjected to 0% reported monitoring in the previous week had .031 SBCs (per employee, per week) more than the baseline category (25% monitoring). Very low rates of monitoring (0-5%) produce substantial jumps in SBCs, although the numbers are still rather low.

Most of the aggregate effect is due to a jump in later SBCs among employees who had earlier made an SBC that was not reported back to them. For example, in the four 0-10% groups, about a third of employees made an SBC in week 61 that was not reported back to them. The SBC rate for those employees in the subsequent week 62, across the four experimental sites, was 5.29, 1.33, 2.73, and 0.47.

Table 5: Effect of low or high call monitoring in week t-1 on suspicious bad calls (SBCs) in week t (Nagin et al, 2002)

Previous monitoring % week t-1	Change in SBCs per week, week t	
	Coefficient	t-statistic
0 (no monitoring)	.031	2.69
2	.006	2.81
5	.019	6.29
10	.002	1.36
15 (high monitoring)	.002	1.20
25 (constant term)	.002	.44

Note: Taken from OLS column (1), Table 2, Nagin et al (2002). Constant, additional hours variables, and call center fixed effects included in the regression but omitted from this table.

There is a remarkable amount of heterogeneity in moral hazard. In the weeks when reported monitoring was 0%, the median number of SBCs per employee is .25. The 80th, 90th, and 95th percentiles of the cross-employee distribution, however, are 1.99, 3.80, and 6.45. The top tenth of cheaters are cheating 15 times as much as the median. In a distribution with such a long right tail, the basic managerial challenge is to prevent moral hazard *among that minority who are inclined to cheat a lot, if they can.*[8]

Several other field experiments have studied simple changes in incentives. Bandiera, Barankay and Rasul (2007 QJE) report the results of a natural field experiment on incentive change in a UK farm on which workers either pick fruit or do other tasks. Increasing bonus pay for managers who supervise workers increased fruit-picking productivity of workers, due to both closer monitoring and managerial selection of more productive workers to pick fruit, and less productive workers to do other tasks.

E. Summary

The experiments on basic incentives described in this section are mostly supportive of basic theory, with interesting exceptions. Effort in piece-rate and experimental tournaments are close to those predicted. However, when there is a deliberate tournament bias all workers work too hard, and when workers can sabotage others (through peer-rating of quality) they do so. Field experiments that increase high-powered incentives among managers increase productivity, due to more work effort and selection of harder workers by managers. In experimental team production (with no communication), forcing contracts – which penalize all workers if a team goal is not reached – work very poorly, and team competition does not work much better. The most robust experimental institution is the threat of monitoring. A field experiment with phone callers also shows that when monitoring drops to zero shirking rises sharply, but mostly among a small number of workers, and the effect is smaller if employees like their employer.

IV. Voluntary effort and reciprocity in organizations

Employment contracts typically have some degree of contractual incompleteness. Incompleteness raises the prospect of moral hazard (hidden action) – that is, when workers are expected to exert effort that is costly for them, but valuable to the firm, firms should not expect much effort (absent reputational concerns and contractual requirements or penalties). Several experiments explore the conditions under which organizational members voluntarily exert effort absent any extrinsic incentives for doing so.

A. Simple worker-firm gift exchange

A potential constraint on moral hazard is reciprocity: If A feels B treated her well (badly), then A may sacrifice to help (hurt) B. Moral hazard by reciprocity-minded employees may be limited, leading employees to exert costly effort that is valuable to firms if employees feel firms have treated them well. Akerlof (1982) called this phenomenon “gift-exchange” and used it to explain persistence of above-market wages and associated unemployment.

Gift-exchange is an ideal topic to explore in experiments because the crucial economic variable that is typically unobservable in field data – worker effort – can be observed by the experimenter, even if subjects acting as employers cannot contract on effort. It is hard to imagine how one could possibly collect these data in a typical firm, since any component of effort that cannot be measured (and contracted upon) by firms would typically be unmeasurable by an economist too.

The gift-exchange labor market paradigm was pioneered by Fehr, Kirschteiger and Reidl (1993). Their study was modified and expanded by Fehr, Kirchler, Weichbold and Gachter (FKWG 1998). Since the FKWG design became standard in many subsequent studies we will describe it in detail.

In FKWG's paradigm, firms offer a wage w. Workers who accept a wage then choose an effort in increments of .1 ($e \in \{.1,.2,\ldots 9, 1\}$). Firms earn (120-w)e and workers earn w-20-c(e), where c(e) is a cost-of-effort function that is convex (with c(.1)=0 and c(1)=18). Note that the gains from gift-exchange are large. In markets where effort is low (e=.1), the market-clearing wage is 21 so firms earn 9.9 and workers earn 1, a total surplus of 10.9. But if e=1 could be achieved, firms would earn 120-w and workers would earn w-20-18, for a total surplus of 118. The gains of obtaining high effort are therefore very large.

FKWG use explicit labor market language to describe the experiment,[9] and compare three market institutions. One institution is one-sided oral bidding by firms with an excess supply of labor (gift-exchange markets, GEM), another is a bilateral market (BGE, workers and firms are matched individually, firms offer a wage to their worker and their worker takes it or leave it); the third institution is a complete contract condition in which effort is fixed at e*=1 and workers' cost c(e)=0.

In the standard gift-exchange markets and bilateral matching markets, wages and efforts are persistently above the competitive level, do not decline over time, and efforts are strongly correlated with wages. In the complete contract condition, at the mandated effort e*=1, firms earn 120-w and workers earn w, so a wage offer is essentially an ultimatum offer of how the surplus of 120 is divided. In simple ultimatum games average offers are around 40%, which would predict a wage offer of 48. Wages in these experiments are close to that figure although clearly lower. In this complete contract condition, wage offers of soldier subjects are higher (average 44) than those of students (average 32), and the student offers decline over time more strongly than soldiers.

Fehr and Falk (1999) ran experiments with double-auction institutions, in which firms and workers both offer wages (subject to an improvement rule—firms must outbid and workers must underbid standing offers on their side of the market). There is a large amount of underbidding by workers but firms often take the highest wage bids and reject lower ones (as occurs in the field, too; see Agell and Lundberg, 1995). This turns out to be profitable for firms—profits increase in higher wages, because higher wages elicit higher effort from most subjects.

B. Replication and robustness of simple gift exchange

The apparent strength of gift exchange surprised experimental economists and inspired many replications and extensions designed to establish the boundaries and robustness of the phenomenon. Many studies are summarized in Table 6. Three questions that immediately arise are the effects of experimental procedures and subject demographics, the influence of trading institutions and other components of economic structure, and the persistence over time of reciprocity. Estimates with individual fixed effects show that there certainly is heterogeneity although most individual workers' efforts respond positively to higher accepted wages.

Charness, Frechette and Kagel (2004 ExEc) find that when the possible payoffs from various (w,e) combinations are presented in a table form, wages and efforts are lower, although the effort-wage relation is still strong. Hannan, Kagel and Moser (2002) explored differences in subject pools, using University of Pittsburgh undergraduates and MBA students.[10] The mean wages and efforts of MBA's are close to those observed in FKWG but the undergraduate wages and efforts are substantially lower. However, the effort response to wages is comparable in magnitude and significance across both groups.

They also conducted sessions in which firms both offer wages and request effort levels; this change seems to raise effort. Interestingly, regressions of actual effort choices which include both wages and effort requests show that MBAs respond to requests but not to wages, and undergraduates respond to wages but not to requests. This difference may reflect some socialization in workplace norms (or other differences in MBA and undergraduate groups).

Fehr and Tougareva (1995) did experiments in Russia with very high stakes (about one week's income per trade) and found similar results to FWKG.

Brandts and Charness experimented with four changes in structure: First, they varied the degree of excess labor supply, from four extra workers to four extra firms. Wages and efforts are lower, but insignificantly so, when there is excess labor, which suggests that excess labor supply is not that important.[11] Second, they had sessions with one-shot matches in the bilateral BGE institution; in these sessions, wages and effort are substantially lower, which shows a role for repeated-market effects (even though matches are anonymous). Third, in one condition they required a minimum wage equal to the midpoint between the highest and lowest wage. Accepted wages are very similar when the minimum wage is instituted, and efforts are a little lower. There is also a sharp downturn in wages, a drop of about half, in the final period 10 (except in the minimum wage condition).

The reciprocity account of these patterns depends on workers knowing that firms *intentionally* shared surplus with them. Charness (2004 J Labor Econ explored this by comparing three treatments: Firms intentionally choose wages; and wages were either clearly drawn randomly from a bingo cage or "preselected by an experimenter." In the latter two treatments, wages were actually drawn from a distribution matched to the wages chosen by firms in the first condition (although subjects did not know this). The hypothesis is that effort response to wages will be higher when firms intentionally chose wages. Overall effort is slightly lower in the intentional-wage condition but the response of efforts to wages is significantly higher (.0120 vs. .0088).

Hennig-Schmidt, Rockenbach and Sadrieh (2003) explored the effect of incomplete information on the profitability of worker effort for firms. Incomplete information could undermine reciprocity because it limits workers' confidence that firms are sharing surplus generously. They used a firm profit function (120-w)ef, where f is equally likely to be .5 or 1.5. In one condition neither side knows the value of f; in the more interesting condition firms know f but workers don't (and the asymmetry is commonly-known). In the firm-knows-f condition wages are about 15 points lower but do respond to the profitability factor f (increasing f from .5 to 1.5 raise average wages by 6.67), as if firms are willingly sharing surplus. Efforts are modestly lower when f is unknown. Overall, the information asymmetry has a small corrosive effect on wages and effort (and hence, on efficiency).

Healy (in press) explores the possibility that even with random rematching and excess labor supply, reputational considerations might generate cooperative behavior. He derives conditions under which a "full reputational equilibrium" (FRE) exists in which

Table 6: Results of several worker-firm gift-exchange experiments

Authors (year)	Experimental treatment	# firms, workers	Wage range	Worker utility	Firm profit	mean wage	Mean effort	Effort-wage coeff. (t))	Comments
Fehr, Kirchler, Weichbold, Gachter (98)	Bilateral (BGE)	10,10	0-120	w-c(e)-20	(120-w)e	57.7[a]	.34[a]	.0057 (t=8.0)	Ss are soldiers
	Cometitive (GEM)	6,9				56.4[f]	.395[f]	.0088 (t=8.8)	
	Complete contract			c(e)=0	120-w	44,32	e*=1 (fixed)	n.a.	Soldiers,students differ
Fehr, Falk (99)	GEM double auction	7,11	FKWG	FKWG	FKWG	62.5	.348	.0064 (t=6.4)	
Charness, Frechette, Kagel (2004)	BGE	10,10	0-100	w-c(e)	(100-w)e	39.8	.315	.014 (t=10.0)	
	with payoff table					33.5	.227	.011 (t=10.7)	
Hannan, Kagel, Moser (2002)	Posted offer GEM MBAs	10,12	FKWG	FKWG	FKWG	59.4	.34[c]	.0214 (t=2.6)	High productivity firms (v=120) only
	Undergraduates					45.0	.21[c]	.0280 (t=3.9)	
Brandts, Charness (03)	Excess labor	12,8	0-10	5w-e	5e-w	74.5[b]	.384[b]	.384 (t=10.4)	
	Excess demand	8,12				83.6	.440	.354	
	BGE one-shot	16,16				48.5	.242	n.a.	incentive higher
	Minimum wage	12,8	w$\geq$5			78.2	.293	.451	
Charness (2004)	BGE firms choose wages	10,10	FKWG	FKWG	FKWG	54.9	.313	.0120 (t=11.2)	Workers must take offer
	random wages					same	.331	.0088 (t=14.4)	
Henning-Schmidt, Rockenbach, Sadrieh (2003)	Baseline BGE	10,10	FKWG	FKWG	(120-w)ef, f=.5 or 1.5	59.9	.32[d]	.0054	
	unknown f					55.5	.27	n.a.	
	firm knows f					46.1	.27	n.a.	
Rigdon (2002)	linear v(e), double-blind	6,8	10-35	w-(2e-2)	27+8e-w	17.0	.129[e]	n.a.	P(pay)=.5 last period
Engelmann, Ortmann (2002)	interior high m=80	6,6	n.a.	w*min(1. 5.5+.5e)-c(e)	em-w*min(1.5.5+.5e	44.4	1.77	.007[g] (t=5.6)	e in [1-3], equil. w*=20, e*=1.8
	interior low m=50					31.8	1.75		
	corner low m=50			wn-c(e)	50e-wn	22.9	1.33		equil. w*=2-3, e*=1.2
				n=1 for e=1, n=1.5 for e>1					

Empty cell entries are the same as the entries in the cell just above. [a]Figures reported in Charness, Frechette, Kagel. [b]Normalized to 0-1 for effort, 0-100 for wages. [c]Mea effort estimated from effort at mean wage, Figure 1. [d]Mean efforts estimated by multiplying average wage and period in effort-wage regression Table 4. [e]Effort normalized by subtracting minimum (1) and dividing by range (6-1)=5. Unnormalized mean effort is 1.64. [f]Estimated from binned histogram Figure 1. [g] Correlation between wages and effort minus best reply effort (Table 9) excluding corner-equilibrium treatment.

firms pay the highest wage and workers reciprocate with the highest effort (until the last period, when workers shirk by giving low effort). He derives a threshold probability that workers are reciprocal which supports FRE. Parametrically, this probability is often rather low when the marginal profit from hiring shirking workers is low relative to the gain to firms from high effort at high wages. Importantly, Healy notes that if workers believe that firms think reciprocal types are correlated, then workers are more inclined to reciprocate. Intuitively, when types are uncorrelated then a worker who is lucky enough to get a high wage in one period can shirk and not suffer direct consequences from low wages in the future. But when types are correlated, the worker's shirking affects firms' perceptions about *all workers*, which may affect the shirking worker's wages in the future. Incorporating "stereotyping" of this sort expands the range of beliefs about the proportion of reciprocal workers which supports FRE.

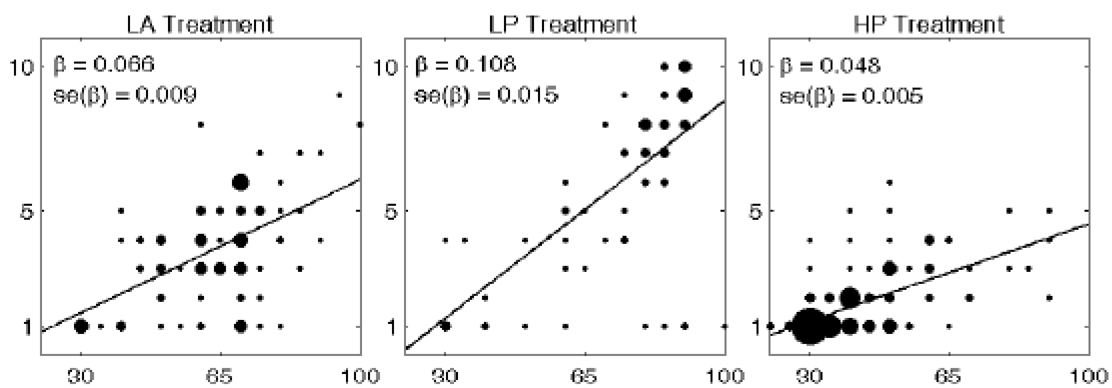


Figure 2: Effort (y-axis) and wage (x-axis) bubble plots (Healy, in press). Treatment LA treatment is Fehr, Kirschsteiger, Reidl (1993) replication. Treatment LP is the same economic structure but with posted worker ID's. Treatment HP uses firm profit that is quasi-linear in worker effort and where reputational equilibria are more difficult to sustain. Efforts and wages are clustered around the self-interested equilibrium in HP, but the wage-effort relation is still strongly positive.

Healy then conducts experiments in which the threshold is low – replicating the original Fehr, Kirchsteiger and Reidl (1993) design (Healy treatment LA), and changing the design so that worker ID numbers are known (LP). In these designs there is substantial gift exchange and a strong correlation between wages and effort until a sharp dropoff in the last period (see Figure 2). In another treatment (HP), firm profits are quasi-linear in effort, approximately .275+.0725e, rather than the multiplicative (v-w)e as in earlier designs. He shows that the threshold probability of reciprocal workers necessary to sustain FRE with these parameters is much higher, which implies that unless firms are very sure lots of workers are reciprocal gift exchange will not emerge. Indeed, wages and efforts are much lower in this treatment where the value of effort to firms has much lower marginal product (see Figure 2), although there is still a strong effort-wage relation.

Rigdon's (2002) parameters and design provide a boundary case in which gift exchange is likely to disappear, and does. She uses linear effort costs (efforts 1-6 have cost c(e)=2e-2) and firm profit is linear in effort, v(e)=27+8e-w (with a minimum wage of 10). Because v(e) is linear in effort, the gains from gift exchange are much smaller than in other designs (25 at the self-interest equilibrium vs. a maximum of 65). She also uses a double-blind procedure (subjects receive keys to a box where their payoffs are deposited). There are T=16 periods and only one is chosen for payment. The probability that the last period is chosen for payment is .5 (all earlier periods are equally likely), so there is a strong economic emphasis on the last period. The trading institution is a sealed-bid offer market and employers request effort levels. Employers request 3.5 units of effort and receive 1.64 on average (see Table 6). Wages and efforts decline over time until they are very close to the competitive equilibrium in the last period. This example shows that gift exchange does not survive a combination of procedural factors (double-blindness), and design parameters which make the gains from gift exchange small and the economic value of the last period large.

Engelmann and Ortmann (2002) varied the efficiency gains from gift exchange and implemented a design with an interior wage-effort equilibrium (assuming self-interest). Creating an interior equilibrium is important because when a predicted equilibrium is at the boundary (like zero effort), any confusion or deviations from equilibrium create positive effort, which could look like reciprocity. (This concern is allayed by the fact that every study reports that wage-effort relations show a strong response of effort to wages, which suggests high effort choices are typically deliberate.) To create an interior equilibrium, they used efforts in [1,3] and a worker utility function which is 1.5w-c(e) for $e \leq 2$ and (e+1)w/2-c(e) for $e \geq 2$. Profit functions are em-1.5w for $e \leq 2$ and em-(e+1)w/2-c(e) for $e \geq 2$. Values of m=50 and m=80 create high and low efficiency gains from effort with an equilibrium of w*=21 and e*=1.8. They also compare abstract and worker-firm experimental framing (which has only a small positive effect when m=80 and the efficiency gains are large). little effect). Efforts are close to those predicted under self-interest, although wages are above the equilibrium level and there is a strong effort-wage relationship.

C. A second wave of gift exchange experiments

Other experiments explore how institutional changes designed to enhance the power of reciprocity affect behavior.

Fehr, Gachter and Kirchsteiger (1997) study the influence of types of reciprocity on wages and effort. In their "weak reciprocity" treatment, firms offer a wage w, a requested effort ẽ, and a fine f. A worker who accepts the offer chooses a costly effort e. If e<ẽ the worker is fined with probability f, and earns w-f-c(e) rather than w-c(e). In a "strong" reciprocity treatment the firm can also punish a worker who shirks and is fined, at a cost k(p), by multiplying the worker's total payoff b a shrinkage factor p<1. The firm can also reward a worker who chooses effort $e \geq ẽ$, at cost k(p) by multiplying the worker's payoffs by a multiple $1 \leq p \leq 2$.

The results are simple: In the weak treatment efforts are low (around .1) even though firms offer substantial job rents (w-c(ẽ)) and efforts do respond to offered rents. However, strong reciprocity works very well. Efforts are around .8 and workers only shirk around 20% of the time. Shirkers are typically punished by the firm and workers who choose the requested effort are rewarded about half the time.

Brown, Falk and Fehr (2004 Econometrica) examined the effect of allowing implicit long-term contracting by announcing ID numbers of workers, and allowing firms to offer contracts to specific (ID'd) workers. Their session had L=7 firms and N=10 employees and T=15 periods.

Firms pay wages $w \in \{1,2,\ldots,100\}$ and workers choose efforts $e \in \{1,2,\ldots,10\}$. Firms earn $10e - w$ and workers earn w-c(e) (where c(e) is convex as in FWKG and other studies). In the specific-contract condition, about 2/3 of the offers are earmarked for a specific employee, wages are higher, and wage offers depend on previous effort. For example, if the worker provided the maximum effort in period $t - 1$, the probability of getting a new contract in t is close to 1. However, the probability of renewal falls to 0.2 for effort levels below 7.As a result, efforts are generally higher when offers are specified for particular workers than when worker IDs are not provided. Both employees and firms in the specified-offer condition earn more per round the longer they have been in a relationship. Firms initiate long-term relationships by initially paying relatively high wages, and employees can signal their trustworthiness by meeting or exceeding the firm's effort expectations.

D. Field experiments

Three field experiments on gift exchange shed light on how long reciprocity persists. Gneezy and List (2006 Econometrica) recruited a small small sample of 19 undergraduate students (unaware it was an experiment) for a one-time six-hour job entering information from library books into a computer. Ten of the subjects were paid the $12/hour wage offered during recruitment ("no gift" condition) and nine were paid $20/hour instead ("gift" condition), with no explanation given for the increased wage.

Figure 3 presents the mean number of books catalogued by subjects in the gift and no gift treatments averaged within 90-minute intervals (GL). There is a significant difference in output in the first 90-minute period, consistent with the gift exchange hypothesis, but the difference decreases over time. A second field experiment used 23 participants in a door-to-door fundraising task. In that study, the effect of the gift wage is sensitive to how effort and output are measured, and to whether demographic subject controls are included (which are known to have a large effect on output).

Kube, Marechal and Puppe (2006 Working Paper) ran a close replication of the GL field experiment on cataloging books. Their study consisted of 29 participants in three conditions who all expected initially to earn 15€/hour: neutral (paid 15€), kind (paid 20€), and unkind (paid 10€). The results are also shown in Figure 3. Here, there is evidence of positive gift exchange from the kind wage increase which persists across the six hours (although the statistical significance is weak, perhaps due to small samples as in the Gneezy-List experiment). In the unkind treatment, participants produced significantly lower output throughout the six hours and this difference was always statistically significant. Therefore, in Kube et al's experiment there is clear evidence of persistent reciprocity in worker output, with a stronger negative reaction to a below-expected "unkind" wage than the positive reaction to the "kind" wage.

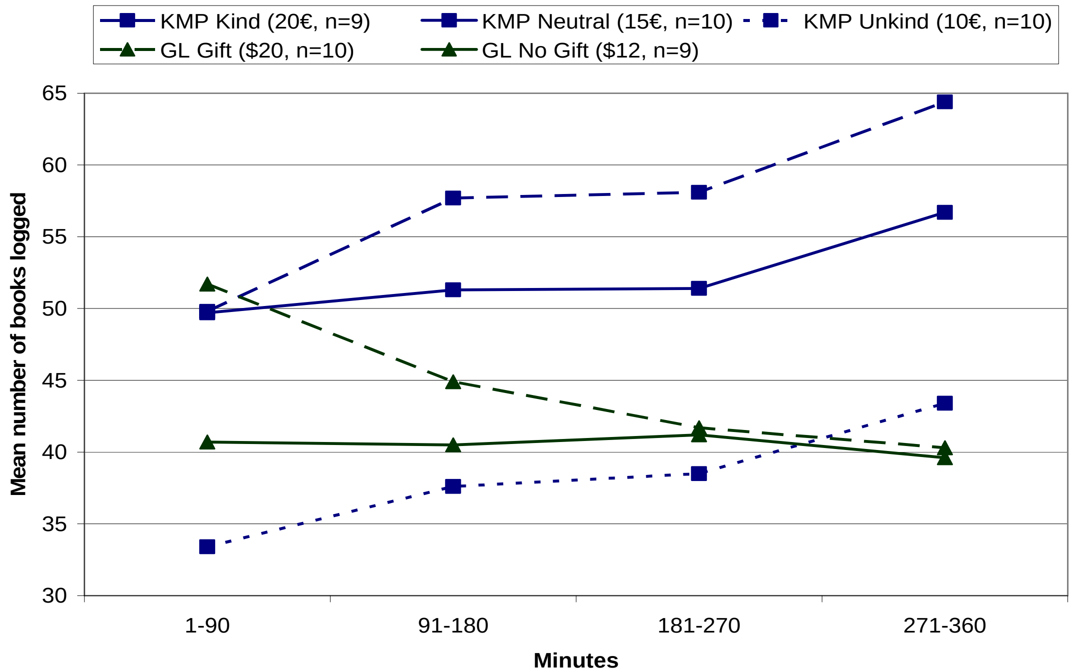


Figure 3. Mean output (books catalogued) in Gneezy and List (2006) and Kube et al (2006) experiments. Note: GL "gift" and "no gift" conditions are relabeled "kind" and "neutral" for comparability.

Taken together, the experiments show modest evidence of positive reciprocity (about 10 percent higher effort in the gift/kind treatments relative to the baseline), and stronger evidence of negative reciprocity. (Of course, the fact that subjects worked at all could be thought of as evidence in support of the gift exchange hypothesis). The field experiments therefore demonstrate that the main laboratory phenomenon does extend to the field, but also raise questions about when and precisely why they do so.

Bellemare and Shearer (2007) describe a gift-exchange experiment in which British Columbia tree-planters were awarded a one-day bonus, which was said to result from unpaid wages negotiated for a different purpose. Workers plant about 1000 trees a day and usually earn a piece rate of $.20/tree (about $200/day). The work is hard, and varies with terrain and weather. The one-day bonus raises productivity by about 100 trees per worker-day, and is highly significant. The effect is also stronger for longer-tenured workers, and for high-tenure workers there is a productivity boost in the Monday after the week in which there was a one-day gift.

The results from these three field experiments do not all agree. In one book-cataloguing study the effect of the wage gift appears to wear off after 90 minutes. In the other book-cataloguing study it does not. A one-day bonus for tree planters has a day-long effect.[12] In all cases there is at least some effect of the gift wage.

Reciprocity is probably most strongly activated if workers feel that employers are sharing surplus with them voluntarily (the book-cataloguing experiments said nothing about why the wage was higher than originally offered). This issue is addressed in a large literature on "procedural justice" and its effect on organizational outcomes (e.g. Lind and Tyler, 1992 Book chapter). One of the best early studies was done in two aerospace/car parts plants by Greenberg (1990). The company determined that it needed to cut wages for 10 weeks by 15%, to avoid laying off workers. It randomly assigned to each plant one of two methods of explaining the pay cuts to workers. In one plant the company's president spent 90 minutes explaining the decision and expressing remorse, then answered questions for 60 minutes. In the other plant a vice-president spent 15 minutes explaining the decision, and wrapped up the meeting by saying "I'll answer one or two questions, but then I have to catch a plane for another meeting." Compared to the short speech, the longer and more remorseful explanation lowered the number of resignations in the subsequent pay cut period from 27% to 2%, and lowered employee theft from 8.7% to 5.7% (though both theft rates were higher than the pre- and post-pay cut rates of 3%).

E. The impact of sorting on voluntary effort and reciprocity

An experimental topic with important potential implications for organizational research involves the study of how individuals sort among economic environments. Several experiments have studied endogenous sorting into market entry games but those experiments are of limited interest in the study of organizations.[13] Our focus here is on the kind of sorting that occurs *within* organizational contexts and deals with voluntary effort and reciprocity.

1. Voluntary cooperation and endogenous group formation

Public goods experiments and n-person prisoners' dilemmas are a useful model for some kinds of organizational processes, such as voluntary effort towards production with group incentives. Experiments using the voluntary contribution mechanism (VCM), in which individuals contribute linearly to a public good, show that contributions typically begin at about 50% and steadily decline close to the inefficient Nash equilibrium level, with a small percentage of "moral exemplars" who always contribute regardless of what others do (see Ledyard, 1995).

When the game is surprisingly or randomly "restarted" contributions tend to jump back up to high levels, then decline again over further periods (Andreoni, 1988; Ambrus and Pathak, 2007).

These experiments typically assign individuals to groups randomly. In naturally-occurring organizations people can usually form groups endogenously based on their preferences and what they know about potential group members. A handful of recent studies explore how contributions vary by allowing such endogenous group formation. Page, Putterman and Unel (2005 Econ J) explore the influence of endogenous selection of group members on cooperation in a linear VCM public goods game.[14] In one treatment, group members are allowed to rank others in all groups according to their match desirability every three periods, but ranking is costly. Most subjects (80%) choose to rank others and generally rank others who contributed most highly. Subjects are then regrouped in an assortative-matching scheme, creating groups of players who matched each other most highly. Regrouping based on costly ranking has a huge effect (compared to a no-rank, no-regrouping baseline), doubling overall contributions (70% versus 38%). In the most "selective" groups (i.e., among those who were most frequently chosen), contributions approach the maximum.

Ahn, Isaac and Salmon (in press JPET) show similar effects of endogeneous grouping in public goods games. Allowing subjects to determine who enters their group increases contribution rates (though only by about 10%). Charness and Yang (2006 working paper) also find a strong effect of endogenous group formation through voluntary mergers of existing groups.

2. Endogeneous sorting among organizations

A central question in organizational economics is how workers sort (or self-select) themselves into different economic environments. Lab experiments typically assign subjects to roles and environments, while most firms use various kinds of non-random sorting (including self-selection). However, it is easy to study sorting experimentally. Indeed, the ideal way to study sorting is to assemble a pool of people, measure their characteristics and expectations, then allow them to sort into economic environments and see what types of people sort into what opportunities. It is very difficult to infer this type of sorting from field data and extremely easy to do so in experiments.

Lazear, Malmendier and Weber (2006 WORKING PAPER) show the effects of sorting on the expression of social preferences: Subjects who can take a certain amount of money, or play a dictator game, often prefer to take the certain money rather than play.

An experiment on sorting that is central to organizational outcomes was conducted by Eriksson and Villeval (2004 WORKING PAPER). Incentive theory predicts that higher performance-based incentive will both raise effort of all workers, and also induce sorting of higher-productivity workers into such a scheme if workers could choose those incentives over a fixed wage (as in Bandiera et al, 2007, above; and see Lazear, 2000). It is also conceivable that when reciprocity plays a strong role, even highly productive employees might prefer high-fixed-wage contracts to performance-based incentives if the high wages indicate an employers' trust in the workers' willingness to work hard even without extrinsic incentive. It is therefore helpful to understand how sorting into different incentive systems works, and the lab provides a good measure of control.

In Eriksson and Villeval's experiment, there are eight firms and eight workers Effort is less costly for some (high-skill) workers and more costly for other (low-skill) workers. Firms either post contract offers to *all* employees each period (a market condition) or offer contracts to the *same* employee each period (a matching condition). In the first eight periods, firms can only offer a fixed wage and stipulate a desired (but non-contractible) level of effort. Employees either agree to a contract or reject all contracts (earning nothing in that period); those that agree to a contract then choose their level of effort. In the last eight periods, each firm offers a fixed wage and offers a piece-ratre wage, specifying a marginal wage for each unit of effort. In these periods, employees who agree to work with a particular firm also choose among the two offered fixed-wage and piece-rate contracts, and also choose an effort level.

In the market condition, firms initially offer wages above the minimum, but workers do not reciprocate with high effort, so wages and efforts rapidly disintegrate to a low-wage/low-effort equilibrium. However, when piece-rate contracts are allowed, average effort increases substantially, because high-skill (i.e., low-effort-cost) sort into the piece-rate contracts and low-skill workers choose fixed wages. This sorting parallels a field study in which an auto-glass installation firms switched to piece rates, increasing effort by all employees and also leading less-productive workers to leave the firm (Lazear, 2000).

In the matching condition, firms offer higher fixed wages than firms in the market condition, and employees reciprocate with high effort. When piece-rates are allowed, high-skill workers are more inclined to continue working for fixed wages than are high-skill employees in the market condition.

Together, these results show that high-productivity workers often sort into piece-rate work environments in a market condition. However, in the matching condition (also called relational contracting), sorting is less important because the long-term contract generates effective reciprocity and high effort even when wages are fixed. Sorting across piece rates, and long-term contracting, are therefore substitutes in creating high worker effort.

F. Summary

Many experiments explore conditions under which reciprocity and voluntary worker effort occur or do not occur. The existence of such reciprocity, in the context of labor relationships, seems to be robust to trading institutions and to variables which should matter in theory (such as excess supply of labor and time horizon), but results are sensitive to other procedural details of experiments.

The clearest result is that when the gain to gift exchange is largest, with multiplicative productivity (v-w)e, gift exchange is substantial and robust across time, but when productivity is linear in effort, ve-w, or has an interior equilibrium, there is much less gift exchange. However, *every* study reports strong response of efforts to wages, even when wages and efforts are low (e.g., Figure 2, HP)). Moreover, non-pecuniary factors such as how wages are communicated appear to be important as well. Field experiments involving temporary wage increases that last a day yields mixed results, with some disappearance of gift exchange and some persistence.

Finally, sorting works very well in creating cooperative groups in public goods games. In a labor-market experiment allowing workers to sort into high- or low-incentive contracts (piece-rates or fixed wages) does create strong sorting, but sorting is much weaker when there is long-term contracting (even with fixed wages), because of reciprocity.

V. Coordination

The need for coordination arises when interdependence between organizational agents requires them to act consistently with one another, but with strategic uncertainty about what others are likely to do (see March & Simon, 1958; Schelling, 1960 Strategy of Conflict; James Thompson, 1967 Organizations in Action). Coordination is central in games with multiple equilibria, creating strategic uncertainty about which equilibrium will be played (see Mehta, et al. 1994 AER; Bacharach and Bernasconi, 1997; Cooper, 1999 Coordination Games). From an organizational perspective, coordination problems are important because they create the possibility of inefficiencies, when agents behaviors are mismatched or when an organization is stuck at an inefficient set of practices that a small group of workers cannot change on their own.

Heath and Staudenmeyer (2000) highlight the importance and difficulty of solving coordination problems in organizations. They note that organizations often neglect the importance of coordination, and tend to focus excessively on partitioning tasks into specialized sub-task, without paying proper attention to the need to eventually integrate the specialized output in a coordinated manner.[15] As a result, they argue, coordination failure is ubiquitous in organizations and often quite hard to overcome.

The research below seeks to understand mechanisms and interventions that might help organizations overcome coordination problems.

A. Minimum-effort (weak-link) coordination

The coordination game applied most frequently to study organizational phenomena is the "minimum-effort" (or "weak-link") coordination game (see Van Huyck, Battalio, and Beil, 1990 AER).[16] In this game, $n \geq 2$ players simultaneously choose numbers $a_i \in \{1,2,\ldots,K\}$ and earn $\alpha + \beta * \text{minimum}_{j \neq i}(a_i, a_j) - \gamma a,$, with $\beta > \gamma$ (so that each player would prefer to choose a higher number only if it increases the minimum). The payoff is increasing in the minimum and decreasing in the player's own choice. A payoff function of this form used in many experiments is shown in Table 7 (corresponding to $\alpha=60$, $\beta=20$, $\gamma=10$).

In this game, all of the outcomes in which players select the same number are pure-strategy Nash equilibria. The equilibria are Pareto-ranked because equilibria in which all players choose higher numbers are better for everyone.

The game is appealing for organizational research because it captures situations in which strategic uncertainty creates a conflict between efficiency and security. Players would all prefer to coordinate on the equilibrium corresponding to a choice of 7, but uncertainty about what others are likely to do also creates an incentive to choose lower numbers. Situations like this arise frequently in organizations when the quality of a good depends on the lowest-quality input or when the completion of a joint activity is determined by the slowest completion of any component activity (see Camerer and Knez, 1994).

Table 7. Payoffs for weak-link coordination game

		Minimum choice of all players						
		7	6	5	4	3	2	1
Player's choice	7	130	110	90	70	50	30	10
	6		120	100	80	60	40	20
	5			110	90	70	50	30
	4				100	80	60	40
	3					90	70	50
	2						80	60
	1							70

Initial experiments on weak-link games demonstrated that when players choose simultaneously without preplay communication, coordination near the efficient equilibrium (in which everyone chooses 7) is almost impossible for groups larger than 3 (Van Huyck, et al., 1991; see also Weber, 2006 AER). However, groups of two often coordinate on the efficient equilibrium if one player chooses 7 and patiently "teaches" the other player to learn to choose 7. Thus, coordination failure in groups larger than 2 raises an important issue for organizational research: Given the prevalence of similar kinds of problems in real firms, how can we solve such coordination failure?

B. Communication

A natural candidate for increasing coordination is preplay communication about strategies. Cooper, DeJong, Forsythe and Ross (1992 QJE) investigated one- and two-way communication in a two-player version of the weak-link game with strategies 1 and 2, and payoffs α=600, β=1000, γ=800 (this two-player, two-action version of the game is frequently refered to as a "stag hunt," or assurance game). Subjects playing without communication overwhelmingly choose strategy 1 – the inefficient, but safe strategy. With one-way communication, however, 87% of subjects say they will choose 2 and, conditional on receiving a message indicating action 2, the other subject chooses 2 76% of the time. The efficient (2,2) equilibrium results roughly half the time. With two-way communication subjects always indicate the intention to play action 2, and the efficient equilibrium results in 91% of groups.

Blume and Ortmann (2007 JET) measured the effect of communication in 9-person weak link games with the payoffs in Table 7. In the baseline, subjects play eight times without any communication. Choices begin around 6-7 and drop to 1-2 within five periods (as is typical in many studies). In the pre-play communication treatment, every subject also sends a message (1, 2, . . .7) before playing, and observes the distribution of messages others sent. Preplay communication has a very strong positive effect on choices: The median message is 7 (the maximum choice) and the post-message choices have a median of 7 and an average around 6.

Other forms of communication are less successful. For example, in theory, players' bids to play a coordination game should signal intentions to choose high numbers (and hence expected profits) through a forward induction argument: Players who bid high expect others to choose high numbers, and since it is rational to reciprocate high number choices, inferring optimistic expectations of other players from their bids should improve coordination. Indeed, Van Huyck, Battalio and Beil (1993 GEB) find that preplay bidding works extremely well in coordinating activity in a game where the group payoff depends on the median choice rather than the minimum (a less "punishing" version of the game). In applying a similar procedure to the minimum (weak-link) game, however, Cachon and Camerer (1996 QJE) find that while players opting to pay a fixed amount to play the game coordinate on higher (more efficient) equilibria, they do not convergence to the efficient equilibrium in which all players select 7.

In contrast to the success of multilateral communication in Blume and Ortmann's experiment, Chaudhuri, Schotter and Sopher (2001 WORKING PAPER) find "inter-generational" communication much less effective. In their framework, players in one generation provide advice on which strategies to play to members of a subsequent generation, as when members of a firm train their successors. Players earn payoffs from their own play, and from the play of their successor. Advice rarely improves the efficiency of coordination, and often makes it worse. Advice only improves efficiency when it is public and commonly known (i.e., when the advice of all members of the preceding generation is read publicly to all members of the subsequent generation), a rather extreme form of inter-generational communication.

Taken together, the above results suggest that communication can help solve even the most difficult coordination problems, with relatively large numbers of players and where the minimum effort determines the entire group's output. However, the communication required to get large groups to efficiency is extreme—players must all send messages and have public knowledge of messages. Experimental research has therefore considered other possible mechanisms for obtaining efficient coordination.

C. Incentives

A natural candidate for improving coordination is incentives: If managers can modify the payoffs in Table 7 (while keeping the basic structure and equilibria intact) and obtain more efficient coordination, then similar incentives might produce efficient coordination in real organizations. Of particular interest is the possibility that a temporary incentive increase might have long-term positive effects on efficiency, by coordinating players on better equilibria which they stick to.

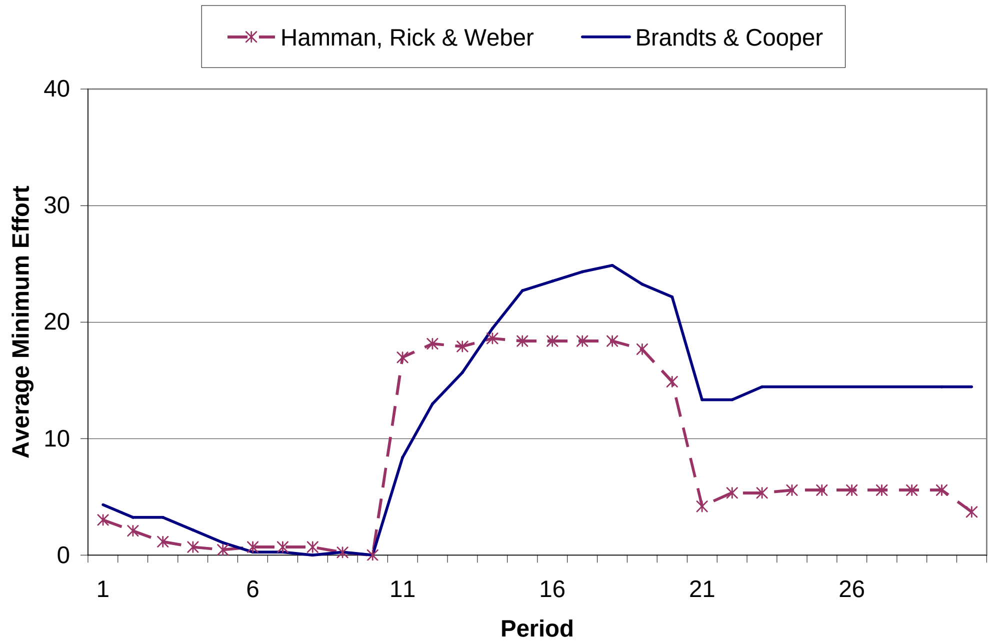
Goeree and Holt (2005 GEB) studied two-player version of the game with continuous strategy choices. Their experiment varied the cost of selecting high choices (i.e., varying γ in the equation above), keeping the equilibria unchanged. They find that this parameter has a strong effect on behavior. Subjects' choices are significantly lower when the cost of choosing too-high effort increases.

Bornstein, Gneezy and Nagel (2002 GEB) created team competition. Players in two seven-person teams earn their own outcome and a bonus if their group's minimum is higher than the other group's minimum. This manipulation produces a slight improvement, raising the average minimum by about 1.5.

Brandts and Cooper (2006) also show that incentives can improve coordination. They use a five-action "turnaround" game (with strategies {0, 10, 20, 30, 40}) in which coordination failure occurs reliably in the first 10 periods. The reward for efficient coordination (β) is then increased for 10 periods to induce a change to an efficient equilibrium.The solid line in Figure 4 shows the effect of increasing the bonus during periods 11-20 (for 37 groups), then reducing the bonus back to the original level (for 9 groups). The graph presents the average minima for groups in their experiment and also in a follow-up experiment by Hamman, Weber, and Rick (in press ExEc) that used a slightly different kind of incentive. Both lines report only groups that experienced coordination failure in the initial 10-period block.

In BC's data, the increase in incentive improves coordination reliably-- 70% of the groups move to a higher equilibrium and over a third coordinated on the efficient equilibrium. When the increased incentives were removed, half the groups were able to maintain coordination above the minimum. This suggests that even temporary incentive increases can permanently draw firms away from inefficient equilibria.

Figure 4. Average minima across periods in weak-link games (Hamman, Rick and Weber vs. Brandts and Cooper; initial failure only) For both experiments, Periods 1-10 involved baseline incentives, Periods 11-20 increased incentives, and Periods 21-30 a return to baseline incentives) From Hamman, Rick and Weber (in press).



An interesting contrast to the BC results is provided by Hamman, Rick and Weber (in press). This experiment uses an identical procedure to BC, in which groups are first induced to fall to the lowest minimum. Then "flat" all-or-none incentives of the kind regularly used in real organizations are introduced: Subjects receive a fixed bonus (or avoid a fixed penalty) if their group minimum reaches a certain threshold. Figure 4 shows that this flat bonus produces a large immediate improvement in group effort choices, as in the BC design, but the long-term improvement is significantly smaller, by about half. One interpretation of the difference in the two designs is that the all-or-nothing incentive either works immediately or doesn't work at all, but the linear bonus from increasing β allows groups to slowly "climb" out of the bad equilibrium and obtain more efficient coordination, The latter kind of incentive is more effective at aiding long-run "turnaround" in firms coordinated on inefficient equilibria.

D. Firm growth

Organizations might also achieve efficient coordination by obtaining it when they are small – when coordination is easier among a small number of members -- and then maintaining it by growing slowly and ensuring that new entrants are aware of the group's history of success. Weber (2006 AER) studied how a group's ability to maintain efficient coordination is influenced by the group's growth process and by the kind of entrants the group brings in. The experiment uses a 7-action game very similar to that in Table 7 (except with α=20).

The experiment compares groups that start off at a large size (12 players) with those that grow to that size, by starting small (2 players) and adding new entrants slowly. Replicating previous research (e.g., Van Huyck et al 1990 AER), groups that start off large converge to the inefficient outcome (minimum = 1) – all five 12-person groups converge to a minimum of 1. Also replicating previous work, 2-person groups all coordinate at higher levels of efficiency (minima of 6 or 7) than do large groups.

The two-person groups then grow slowly – usually by adding only one person at a time. Two treatments differ in the information provided to new entrants. In a "history" condition, new entrants observe full history of outcomes (minima) obtained by the group and this is common knowledge. In a "no history" condition, however, entrants are completely unaware of the minima obtained by the group prior to entry.

Groups without history all converge to the lowest minimum. However, while several groups in the history condition also fall to the lowest minimum as they grow, several groups do not. In fact, over half of the groups maintain coordination on minima greater than 1, and 2 of 9 groups remain coordinated on the efficient equilibrium (minimum = 7) throughout the entire growth process. This result presents one reason we might observe large efficiently coordinated organizations – they start off at a small size, when coordination problems are easier to solve, and then maintain efficiency by growing slowly and exposing new entrants to the group's history.

E. Coordination and organizational culture

Over the last decade, organizational economists have noted organizational culture as a potentially important concept for understanding behavior in firms. Much of this work focuses on the relationship between culture and equilibrium selection in coordination games (see, for example, Cremer 1993 Industrial and Corporate Change; Kreps, 1999 Book Chapter; Hermailin 2001 Book Chapter; see also Arrow 1974). Culture can serve as a means for resolving strategic uncertainty or selecting among several equilibrium behaviors by individual members of a firm. Thus a strong organizational culture might be one way to prevent coordination problems. (One way to interpret the above results on firm growth is that the efficient coordination obtained in small groups is similar to the strong culture that might develop in a small firm, and that such strong culture helps maintain efficient coordination as the firm grows.)

An important property of organizational culture is that it is sensitive to changes to an organizations' size and boundaries, such as with a merger. Given the difficulty in identifying organizational culture and its effects on performance in real firms, the laboratory presents a useful domain for isolating the effects of changes to an organization on its culture.

An experiment providing insights into what happens when firms when different "cultures" merge was provided by Knez and Camerer (1994 SMJ). They studied what happens when two groups playing the weak-link coordination game are combined into a single larger group. In their experiment three-person groups play the weak-link game in Table 7 for five periods. Three-person groups playing this game exhibit a large degree of heterogeneity in outcomes – some coordinate efficiently (4 of 20 groups obtain a minimum of 7 in the fifth period), others fall to the inefficient outcome (6 of 20 groups have a minimum of 1), and several end up in between (minima of 2, 3, or 4). Two three-person groups were then combined into six-person groups for another five-period block. The larger combined group fell to the minimum of its component three-person groups (or lower), and eventually to the minium of 1 in 80% of the cases. Maintaining efficient coordination (or a successful "culture") while merging small laboratory groups into larger ones appears to be difficult.

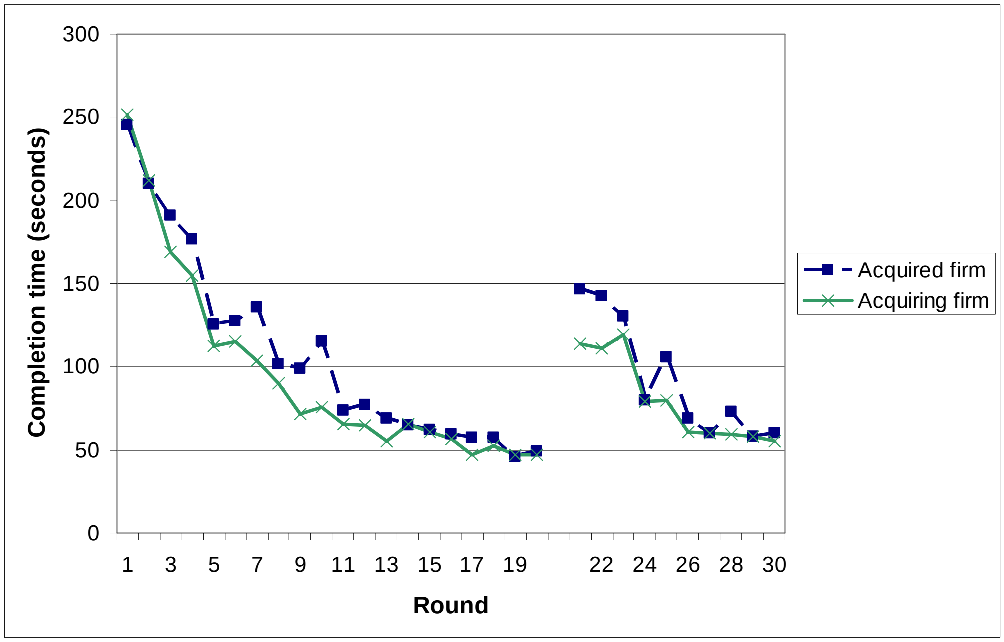
Weber and Camerer (2003 Management Science) further studied difficulties in merging firms with distinct cultures, and how such difficulties might be underestimated by firm members. The study introduced a paradigm for studying organizational culture using simple laboratory experiments. In the paradigm, groups of subjects develop a "code" in order to jointly identify sets of complex pictures as quickly as possible. More precisely, in every period a manager receives some information regarding an exogenous state of the world – which pictures the group must identify in that period – and must verbally communicate this information in order to get employees to respond appropriately by selecting the appropriate pictures. In order to do so, the manager and employees must develop a shared understanding of what words or phrases will be used to identify each picture, i.e., a code. This code is a shared understanding that allows firm members to solve coordination problems while minimizing the need for costly communication. As Cremer (1993 Industrial and Corporate Change) notes, such code shares many features with organizational culture. Much as with organizational culture, each firm's code is likely to be idiosyncratic, history dependent, and reflect the firm members' jointly-developed shared perspective. Moreover, each code is likely to be well-suited for that firm's performance, but different firms are likely to have codes that differ.

In each experimental session, two two-person firms develop code/culture independently over the first 20 periods. As the left part of Figure 5 indicates, firms initially lack a shared understanding, and it takes them a long time to complete the picture-identification task. However, with repetition, both two-person firms develop a code/culture that allows them to perform the task efficiently.

Following period 20, the two firms in each session merge, with one firm taking over the other, and eliminating the redundant manager. The new firm therefore comprises one manager, one familiar employee (who previously worked with the manager), and one new employee. The merged firm performs the same task as before (using the same set of pictures) for 10 additional periods. Prior to merging, members of both firms forecast the performance of the merged firm (with an incentive for accuracy).

As the right side of Figure 5 reveals, the merger has a detrimental effect on the performance of *both* employees. The manager and new employee find it very difficult to integrate their distinct codes/cultures, and this miscoordination negatively affects the performance of the familiar employee. Therefore, integrating two distinct codes/cultures – which had been equally efficient at solving the coordination problem pre-merger – proves quite difficult. Moreover, subjects' estimates significantly overestimate post-merger performance. Finally, questionnaires in which subjects rate the ability and competence of other subjects in the experiment, reveal that the post-merger integration difficulties are attributed to perceived incompetence on the part of the individual(s) from the other pre-merger firm. Taken together, this study provides an illustration of the pitfalls present in real-world mergers that are often discussed anecdotically after a highly-touted merger ends in failure, but that are difficult to demonstrate conclusively using non-laboratory data.

Fig 5. Average completion times (11 merger sessions) in Weber and Camerer (2003)



Feiler and Camerer (2007) replicated the Weber and Camerer using a computerized interface and endogenous mergers. Subjects in the separate groups (UCLA and Caltech students identifying pictures of their own campus) bid for payments demanded to join a mixed-picture (cross-campus) group (rather than staying in a same-campus group), in a first 'merger' and a second 'merger'. The central question is whether bids to join a merged group correctly price the economic loss from the merger (similarly to the pre-merger estimates in Weber and Camerer). In the first merger into mixed groups, the average bids to join the mixed groups are too low by $.32/period (using Vickrey auctions, which in theory are demand-revealing[17]). There seems to be an "organizational winner's curse" in which the successful bidders do not recognize how relatively optimistic they are about the profit that can be produced in the merger.

F. Summary

The experiments above analyze different ways in which activity in firms might be effectively coordinated. The results show that process of coordinating choices, even when there is a clearly payoff-dominant equilibrium, depends sensitively on communication, leadership, and how an organization grows. Communications must be impartial, highly participative, and commonly-known to work reliably. Incentives help, especially when complemented by communication of managers who create incentives. The studies on merging groups reveal that coordination among newly integrated units might be difficult, and the difficulty underestimated.

VI. Leadership and hierarchy

Organizations of almost any scale have some hierarchical structure in which authority relations substitute for price mechanisms and contract. Hierarchical superiors typically have residual control rights and authority over what subordinates do. A special aspect of hierarchy is leadership, in which values and broad strategic directions are espoused by a leader in an extremely implicit contract. These areas are also a natural meeting ground for rational theories based on information and contracting, and behavioral theories in which leaders inspire and serve symbolic functions.

The experiments below explore how hierarchy and leadership affect organizational behavior and productivity. Experimental control has a special power in these studies because experimenters can randomly assign firms to be either flat or hierarchical, can assign subjects to leadership roles, and can also allow self-selection to estimate the power of sorting and selection.

A. Hierarchy and incentives

Alchian and Demsetz (1972) were among the first to recognize the advantage of an organization's members hiring a supervisor to monitor them. In an experimental analogue of this process, van der Heijden, Potters and Sefton (2006 Working Paper) studied the effect of supervisory power in a simple public goods game with predicted free-riding. Subjects in groups of four choose simultaneously to either shirk or work (a binary choice, $e_i \in \{0,1\}$). Shirking enables a worker to keep their endowment of 120; working means they lose the endowment but contribute to group output, determined by $Q = 60(\sum_i e_i)^2$, which is shared equally. If all four work, the net total payoff is $60(4)^2$=960, which is higher than the net total payoff when all shirk (480).

In a baseline ("revenue-sharing") condition, group output is shared equally. As is typical in repeated public goods experiments, and in the experiments by Nalbantian and Schotter (1007) that we reviewed previously, total effort begins at a high level (three workers choose to work) and shrinks to an average total effort of 0.5 in 15 periods. There is substantial variation across groups, showing that local cooperative norms can emerge.

In a hierarchy ("leader-determined sharing") condition, one person is randomly chosen to both choose effort as a worker in the group, and also to assign output to individual subjects after each round however he or she likes. While this hierarchy could backfire (the manager can grab all the output), the manager typically divides output evenly among those who exert effort, and gives nothing to those who shirked. Subjects learn this pattern and work, because free-riding will earn them nothing. As a result, the mean group effort is around three until the last round, when there is a sharp endgame drop in effort.

This experiment nicely shows that cooperation is sensitive to institutional rules and to the willingness of some players to punish defectors, as in many experiments in which costly punishment reliably raises contributions in public goods games (e.g., Fehr and Gaechter 2000 AER). Having a worker-manager who can allocate output to all workers coordinates punishment of *all* free riders, because of her ability to allocate group output to each worker, and leaders do not grab too much output (presumably because they know it would lower future effort). As a result, leadership works well enough to produce sustained cooperation close to full efficiency.

B. Leadership and public goods

An important aspect of leadership in organizations is signaling (e.g., Hermalin 1998 AER). Leaders signal private information about desirable collective outcomes through observable announcements or through their behavior, or by changing economic incentives. Recent experiments explore this phenomenon in public goods.

Güth, Levati, Sutter and van der Heijden (2004 Working paper) explored the effectiveness of leadership for improving cooperation in public goods games. In their experiment groups of k=4 subjects play a public goods game under the voluntary contribution mechanism (VCM). Endowments are 25 tokens, which have a private return of 1 and a public return of 1.6 (so the marginal per capita return is .4).

In the baseline condition, contributions begin at around half of the 25 tokens and decline to about 6 after T=16 periods, a typical pattern (e.g. Ledyard, 1995). In the weak-leader condition, one player moves first and her contribution is visible to others. In this condition average contributions are about 3 units higher and exhibit the same typical decline over time. In the strong-leader condition the leader can also penalize free riders by excluding one player from the group in the following round (the excluded player earns nothing). In this condition, average contributions start at 18 tokens and rise slightly (until a sharp drop in the final period 16). The strong and weak leader conditions together show that the ability of one player to establish a norm by moving first has a small effect, but combining that with the ability to 'lay off' a free rider increases contributions substantially.[18]

Potters, Sefton and Vesterlund (2007 Economic Theory) examined contributions to a public good when the value of the good is only known by an early-moving leader. In this case, the leader's action can signal information about the value of the public good.

Subjects play a public goods game in pairs, composed of players 1 and 2. Each player can choose one of two options: A or B. Choosing A gives the individual a private return of .4 (in £). Choosing B gives each of the players a return of 0, .3 or .6. The exact return from choosing B was randomly determined. For the small return (0) choosing A maximizes both individual and collective payoffs. The medium return (.3) is a prisoner's dilemma (A is privately better but two choices of B are best for both). For the highest return (.6) choosing B is both individually and collectively better.

The design is fixed-role, random rematching with T=18 periods. Choices are either made simultaneously or sequentially. Additionally, either both players observe the return rate prior to making a choice (full information) or only player 1 is able to do so (asymmetric information). Under full information, the order of play should not matter and leadership should have no effect. However under asymmetric information (when the leader observes private information), sequential play should implement efficient behavior (B choices by both players) when the return rate is 0.30 or 0.60, which is not predicted under simultaneous play. The interesting case that might show an effect of leadership is the sequential leader information case with return of .3. Suppose the leader knows the return is .3. If she chooses the cooperative B in those cases, a follower should infer that the leader's observed B choice either means the return is .3 or .6; the mean inferred return is therefore .45, which makes the marginal public return above the private return, so the follower should contribute. A leader contributing when she observed a return of .3 therefore coaxes followers, in theory, into contributing, which is mutually beneficial.

As predicted, under full information both subjects rarely contribute when the return is 0 (less than 5%) and almost always contribute when the return is .6 (over 90%). When the return is .3, leaders contribute a little more when playing sequentially (27%) than simultaneously (16%). However, followers in the sequential treatment mimic leaders only one-third of the time (and therefore contribute 9% of the time).

In the asymmetric information treatment, when the return is .3 player 1's do lead by choosing B 75% of the time when choosing sequentially, as theory predicts, compared to only 15% in the simultaneous condition. Followers contribute 81% of the time when leaders contribute, so the efficient outcome occurs over half the time.

The experiments above highlight the significance of leadership in organizations. In both cases, leadership has beneficial effects, inducing higher contributions and generating greater earnings for the group as a whole (see also Moxnes and van der Heijden, 2004 Journal of Conflict Resolution; Meidinger and Villeval, 2002 working paper; Gacter and Renner, 2004 working paper). However, the experiments also shed light on when leadership is likely to work and what characteristics make it effective

C. Leadership in coordination

Another way in which leadership might be valuable for organizations is in coordinating members' activity. When interdependent agents are faced with the problems of strategic uncertainty and equilibrium selection that we described in Section V, having a leader direct behavior toward an efficient equilibrium through public statements might help produce efficient coordination (Foss 2001 Industrial and Corporate Change). Understanding when public recommendations influences behavior is a baby step toward understanding leadership.

Van Huyck, Gillette and Battalio (1992 GEB) studied the effectiveness of public statements directing players towards one equilibrium in coordination games (see also Brandts and MacLeod, 1995 GEB). Pairs of subjects play three-action games with multiple equilibria as in the game in the top panel of Table 8, which has three pure-strategy equilibria (the outcomes along the diagonal). The experiment varies whether there is a recommendation from the experimenter to play a particular strategy.

Advice is strongly followed when it coincides with efficiency, but recommendation to play a Pareto-inferior equilibrium produces mixed behavior.

Table 8. Coordination-game payoffs and influence of recommendations.

		Column's choice		
		A	*B*	*C*
Row's Choice	*A*	5+X,5+X	0,0	0,0
	B	0,0	5,5	0,0
	C	0,0	0,0	5-X, 5-X

Recommendation	A	B	C
Van Huyck et al (X=4)			
None	97	2	1
A	99	1	0
B	51	48	1
C	62	0	38
Kuang, Weber & Dana (X=1)			
None	100	0	0
B	50	50	0
KWD (X=1; 3rd subject receives $5 for B)			
None	39	61	0
B (from interested 3rd subject)	37	63	0
B (from uninterested 3rd party)	19	81	0

This suggests that public statements by leaders have strong effects on coordinating behavior on a desired equilibrium, but can even coordinate behavior on an inefficient equilibrium, such as when employees are asked to adopt a new technology or practice that may be less desirable to them than their current practice, but which may benefit the firm.

Kuang, Weber and Dana (2007) used payoffs similar to those of Van Huyck et al (1992), and compared recommendations from the experimenter, from an uninterested third party, and from a self-serving third "player" who earns $5 if they follow her recommendation to play B (although the third player does not play). The third player's recommendation of B is ineffective (compared to no recommendation), but the recommendation of the independent third part party Recommendations are more likely to be followed when they are impartial.

In the last section, we discussed weak-link coordination problems. Given the difficult of obtaining efficient coordination in such games, understanding the effectiveness of leaders' recommendations towards the efficient equilibrium is important. Can a leader urging players to move away from the secure but inefficient equilibrium towards the efficient but risky equilibrium get them to do so?

Brandts and Cooper (2005 Working Paper) explored this question. In their experiment, firms of four subjects play a weak-link coordination game. Worker subjects choose among five possible effort levels $e_i \in \{0,10,20,30,40\}$. Worker i's payoff $u_i(e_i)$ is $200\text{-}5e_i\text{+}(B \min_{j=1 \text{ to } 4} (e_j))$; that is, each unit of effort privately cost 5 units, and the group payoff is B times the lowest effort, where $B \in \{6,7,8,\ldots,15\}$. Each firm also consists of a fifth subject, in the role of manager. In the first 10 periods, the manager is a passive observer (B is fixed at 6). In these first 10 periods, firms typically sink to the inefficient equilibrium in which efforts are close to zero.

In the remaining periods of the experiment (11-20), the manager takes control of the firm and attempts to induce coordination on the efficient equilibrium. The manager sets the bonus and receives the payoff: 100+(60–40B) min(e). This payoff function means managers benefit if the firm coordinates on higher minimum effort, but also prefer to pay the lowest B that increases effort. In different treatments, managers had the ability to send natural-language messages to employees (one-way communication) and receive messages back from employees (two-way communication).

Fig 6. Average minimum effort (group output) across conditions (Brandts and Cooper, 2005)

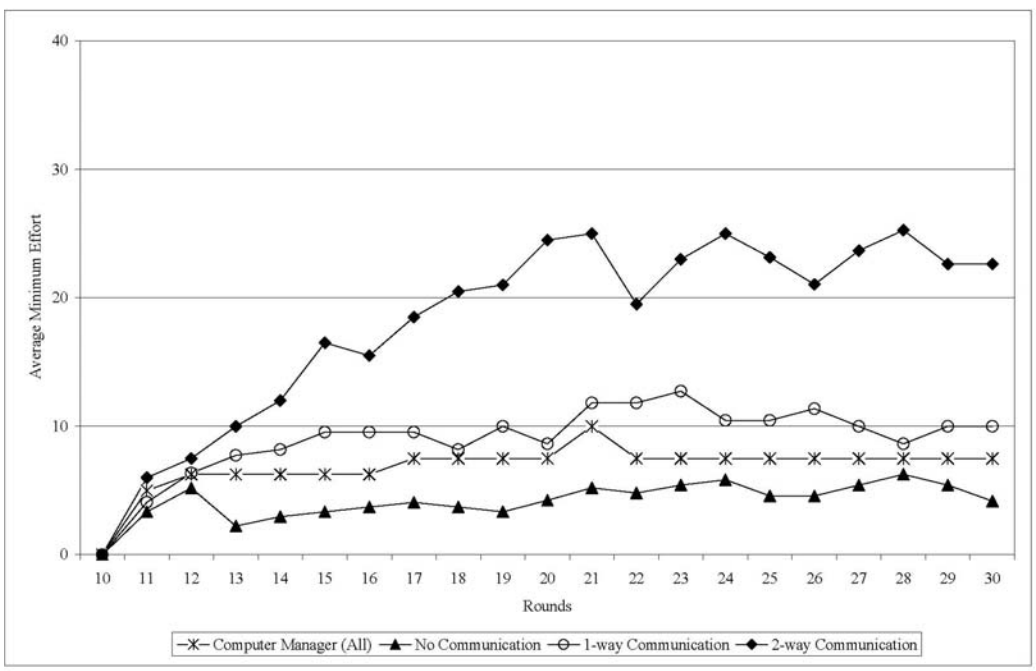


Figure 6 shows the average minimum effort across the manager-run periods, using only those firms that experience initial coordination failure (i.e., firms in which the manager could "turn things around"). "Computer manager" refers to a treatment in which the role of manager is played by a computer who always selects B = 10. (This bonus is very close to the average bonus set by human managers across all treatments.)

Managers who are unable to communicate (but who set the bonus) produce very little improvement in efficiency, as do computer managers. However, communication improves the firm's performance, and two-way communication produces the greatest improvement. Thus, allowing greater channels of communication between managers and employees improves the efficiency of coordination.

Brandts and Cooper also explored the content of messages from managers to employees. Managers who ask employees for greater effort, who draw their attention to the bonus, and who describe a long-term plan for improving efficiency obtain the greatest improvements in efficiency. Communication is a complement to incentive change.

In a follow-up paper, Cooper (2006 Advances in Economic Analysis & Policy) used the one-way communication treatment above to explore differences in the effectiveness of student subjects and actual managers (executive MBA students). Actual managers are slightly more effective than student managers, particularly at obtaining a rapid improvement in group outcomes, because they are more likely to employ the communication strategies found to be effective by Brandts and Cooper (2005).

D. The illusion of leadership

The above studies show that leaders, by recommending particular equilibria to organizational members, can improve the efficiency of coordination. Another study, however, highlights that such leadership may have limits to its effectiveness, and that the perceived quality of leadership may be affected by factors beyond leaders' control.

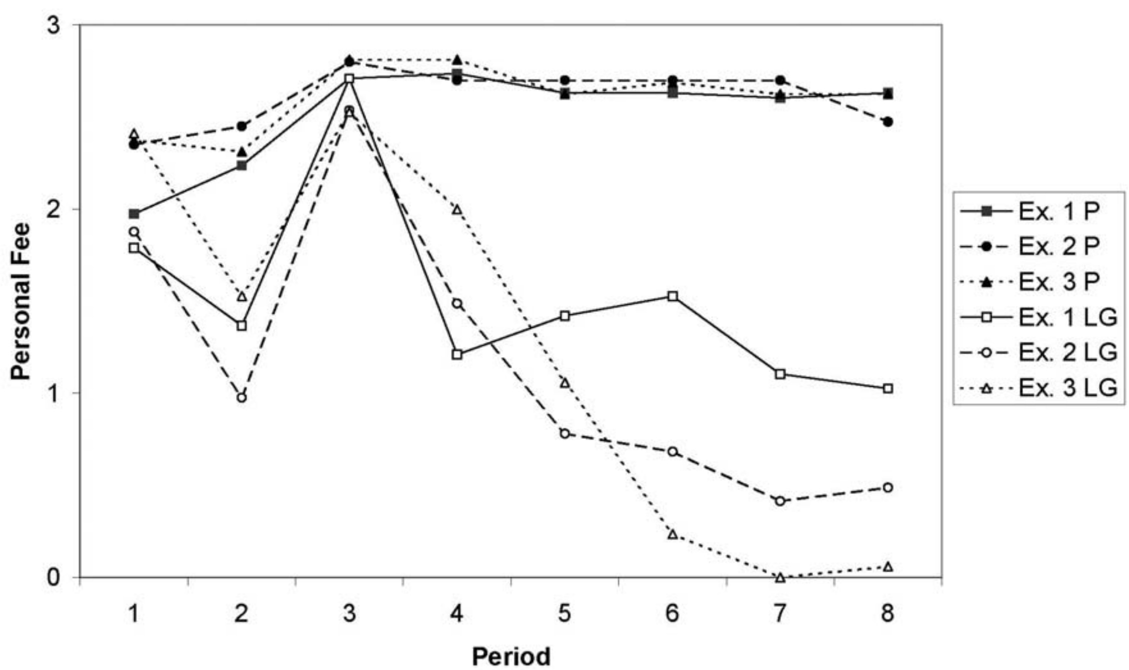
Weber, Camerer, Rottenstreich and Knez (2001 Organization Science) examined the ability of large and small groups to coordinate efficiently in the weak-link coordination game under a simple form of leadership. As we mention in Section V, previous studies using this game show a consistent group size effect: small groups almost always coordinate successfully while large groups never do (Van Huyck, Battalio & Beil, 1990 AER). As in Brandts and Cooper's experiment, Weber et al (2001) examined whether simple leadership in the form of one subject addressing the group and urging them to coordinate efficiently would improve the ability of large groups to coordinate.

Sessions consisting of 9-10 subjects play in one of two conditions. In the "large group" condition, all nine or ten subjects play a four-action version of the weak-link game together. In the "pairs" condition, subjects are randomly and anonymously paired with one of the other nine people in the room and play the same game with this person only. The game is played eight times.

Prior to the first round, a leader is randomly selected and told that he or she will, after the second round, address the entire group (or all pairs, in the pairs condition) in order to "organize" and "prepare" the players for the remaining rounds. Leaders are given a simple message to guide their statement – highlighting the efficient equilibrium and the mutual gains from selecting it

Figure 7 shows the average effort choices (which were framed in the context of a "personal fee") for three sessions each of large groups and pairs. As expected, the pairs coordinate efficiently in the first two periods while the large groups begin to fall towards the inefficient equilibrium.

Figure 7 Average choices by treatment and session (Weber et al 2001)



After period 2, the leaders address the groups. The effectiveness of leaders' statements can be seen in period 3 – in both treatments average choices are close to the maximum. However, for the large groups this is not enough to overcome coordination failure, and these groups subsequently converge towards the lowest effort level. Thus, even though Brandts and Cooper showed that simple leadership in the form of public statements is effective at inducing "turnarounds" in groups of size 4, this is not the case here for larger groups (of size 9 or 10).

Immediately after leaders speak, subjects in both treatments rate the leaders' speeches and leadership ability equally highly (means ratings of 5.8 (pairs) and 5.9 (large groups), on a 9-point scale). However, when the same questions are asked again at the end of the experiment, subjects in the large group condition rate leaders as significantly less able (4.5) than subjects playing in pairs (6.2). Thus, even though the situation in which subjects were placed (i.e., the size of the group they were leading) determines their outcomes – and subjects even recognize that leading small groups is easier than leading large groups – "followers" mistakenly attribute good outcomes to good leadership and poor outcomes to poor leadership. A second experiment replicated the above result, and showed that subjects are even willing to act on their perceptions of leadership quality, by casting a costly vote to replace the leader.

E. Summary

The experiments in this section all study the extent to which hierarchy and leadership can improve organizational outcomes. In each case, there is some inefficiency that obtains in "flat" organizations – either because of moral hazard or coordination failure. The experiments explore the extent to which efficiency is improved by assigning differential roles and giving some roles greater information, power, or prominence. In most cases, there is some beneficial effect of hierarchy and leadership on efficiency. However, the success of these treatments is not always perfect, even when theory predicts it should be, and they may even produce perverse effects, as in the misattributions of leadership ability in weak-link games with small and large groups.

VII. Conclusions and Future Directions

Laboratory and field experiments are a useful method for addressing questions such as how worker effort will respond to different kinds of monetary and social incentives and how hierarchy and communication can help improve the efficiency of firm outcomes. In each of these cases, laboratory experiments provide control which is helpful for testing theoretical predictions and comparing institutions, and which is typically difficult to obtain in the field.

The results vary in what they tell us about organizations. In some cases, the results are highly supportive of theory: For example, worker efforts are close to those predicted in tournaments, and workers do respond to piece rates and monitoring. However, sabotage in tournaments with peer-rating of quality is strong and team forcing contracts don't work.

Many of the results do not refute or confirm theory per se, but instead illustrate empirical effects of variables that have ambiguous theoretical effects. For example, communication typically expands the set of equilibria, so data are helpful in determining whether communication leads to better outcomes or not. It appears that large-group coordination is difficult, but extensive types of communication and leadership can solve coordination problems.

An advantage of the lab is that it is often possible to clearly measure heterogeneity in responses and its influence. For example, the field experiment on monitoring (Nagin et al) showed that reducing monitoring increases moral hazard, but largely among a small set of workers (and is moderated by workers' attitudes toward their employer). Heterogeneity is also evident in experiments on gift exchange: While a minority of subjects are self-interested and shirk if they can, a majority of subjects' efforts reliably respond to wages (indicating positive reciprocity), although procedural details and economic structure matter too.

When some results are surprisingly consistent with theory, and others are not, one hopes for a generalization of standard theory that could explain both kinds of evidence. In game theory, models which assume equilibrium beliefs but relax best response (quantal response equilibrium), or assume a hierarchy of cognitive sophistication, have proved useful in explaining when standard theory works well and when it does not (e.g., Goeree and Holt, 2001, Camerer, Ho and Chong, 2004). Models like these, along with specifications of reciprocal preferences, might be able to unify a lot of disparate results and make interesting new predictions.

In reviewing this research, we were struck by how many interesting topics have been understudied or completely unstudied. Most of what is discussed in a modern graduate Ph.D. course has never been studied experimentally. Employment contracts are usually surprising simple and incomplete (and relative performance evaluation is rare for topic executives), job design is crucial when multi-tasking makes it difficult to incentivize the optimal mix of activity, and sorting is presumably important—but all these topics are rarely studied, even though it is arguably even easier to study in a lab experiment than with naturally-occurring data. And in many cases, there is a clear complementarity of theory and experimentation because an obvious experimental design walks right off the pages of a theory paper right into the lab (particularly in game theory).

There are also many opportunities for entirely new experimental directions. Web-based software and reasonable availability of funds for subject payments (and the capacity do to experiments with highly literate subjects in low-income countries) mean that experiments which are much more elaborate than most of those described above are feasible. Such experiments could explore how results scale up to larger groups and longer spans of time. Some topics that are ripe for experimentation include the following:

Multi-tasking. When workers perform multiple activities, it is generally impossible to find contracts in which performance-based incentives induce agents to provide the mix of activity levels across tasks that firms would like. This fact shifts attention from single-activity contracts trading off risk and incentive to an incentive *system* including job design and monitoring (Holmstrom and Milgrom 1994). Amazingly, there are *no* experimental studies of multi-tasking. This is a ripe empirical opportunity because it is conceivable that studies would find that heterogeneity, sorting, and reciprocity could limit problems that appear to rise in theoretical multi-tasking models, or could guide job design in an unexpected direction.

Social networks: Recently, economists have caught up to sociologists who have been taking empirical 'snapshots' of networks and studying their properties and influence on behavior (e.g., Matt Jackson review paper World Congress). Networks are clearly important inside organizations. And since naturally-occurring networks are both created by behavior and influence behavior, experiments which assign networks exogeneously are scientifically useful to establish causality. Experiments have studied the influence of network structure on solving combinatorially-explosive OR problems (Kearns et al Science 06 CITE), financial contagions (Corbae and Duffy, in press) and other topics (see Kosfeld's, 2004 review), but none of these experiments deal with the central questions in organizational economics.

Virtual worlds: Commercial firms have developed online gaming "virtual worlds" in which players inhabit personal roles, interact socially, and develop economic exchange for scarce resources which enable players to succeed (such as SimCity and Second Life). These games represent a mouth-watering empirical opportunity because they attract large numbers of players who spend hours playing. Experimental control could either come from collaborations with designers, or from running true experiments inside the virtual world (see Bloomfield, 2007).

References

Ahn, T.K., Mark Isaac and Timothy C. Salmon. (in press). Endogenous Group Formation. *Journal of Public Economic Theory*.

Akerlof, 1982 G.A. Akerlof, Labor contracts as partial gift exchange, *Quarterly Journal of Economics* **97** (1982), pp. 543–569.

Alchian, Armen A. and Demsetz, Harold (1972). Production, Information Costs, and Economic Organization. *American Economic Review*, 62(5), 777-795.

Ambrus, Attila and Pathak, Parag. (20xx). Cooperation over finite horizons: Theory and experiments. Unpublished

Andreoni, James (1988). Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics*, 37(3), 291-304.

Argote, Linda Turner, Marlene E. and Fichman, Mark (1989) To centralize or not to centralize: The effects of uncertainty and threat on group structure and performance. *Organizational Behavior and Human Decision Processes*, 43(1), 58-74.

Ariely, Dan, Gneezy, Uri, Loewenstein George and Mazar, Nina (2005). *Large stakes and big mistakes*. Federal Reserve Bank of Boston, Working Papers: 05-11.

Arrow, Kenneth J. (1974). The Limits of Organization. Norton, New York.

Bacharach, Michael and Bernasconi, Michele (1997). The variable frame theory of focal points: An experimental study. Games and Economic Behavior, 19(1), 1-45

Bandiera, Oriana, Barankay, Iwan, and Rasul, Imran. (2005). Social Preferences and the Response to Incentives: Evidence from Personnel Data. Quarterly Journal of Economics, 120(3): 917-962.

Bandiera, Oriana, Barankay, Iwan, and Rasul, Imran. (2006). *Incentives for managers and inequality among workers: Evidence from a firm level experiment*. C.E.P.R. Discussion Papers, CEPR Discussion Papers: 5649.

Bandiera, Oriana, Barankay, Iwan, and Rasul, Imran. (2007). Incentives for managers and inequality among workers: Evidence from a firm level experiment. *Quarterly Journal of Economics*, 122(2), 729-773.

Bavelas, Alex and Barrett, Dermot. (1951). An experimental approach to organizational communication. *Personnel* 27, 367-371.

Bazerman, Max, Curhan, Jared R., Moore, Dona A. and Valley, Kathleen L. (2000) Negotiation. *Annual Review of Psychology* 51, 279-314.

Becker, Gary S. and Murphy, Kevin M. (1992). The division of labor, coordination costs, and knowledge. *Quarterly Journal of Economics*, 107(4), 1137-1160.

Bull, Clive, Schotter, Andrew & Weigelt, Keith. (1987). Tournaments and Piece Rates: An Experimental Study. *The Journal of Political Economy*, Vol. 95, No. 1. pp. 1-33.

Bellemare, Charles and Shearer, Bruce (2007). *Gift exchange within a firm: Evidence from a field experiment.* IZA DP no. 2696, March.

Blinder (1990)

Bloomfield, Robert. Worlds for Study: Invitation - *Virtual Worlds for Studying Real-World Business (and Law, and Politics, and Sociology, and....).* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=988984

Bloomfield 2007

Blume, Andreas and Ortmann, Andreas (2007). The effects of costless pre-play communication: Experimental evidence from games with Pareto-ranked equilibria. *Journal of Economic Theory*, 132(1), 274-290.

Bornstein, Gary Gneezy, Uri and Nagel, Rosmarie (2002). The effect of intergroup competition on group coordination: an experimental study. *Games and Economic Behavior*, 41(1), 1-25.

Bottom, William P., Holloway, James, Miller, Gary J., Mislin, Alexandra, and Whitford, Andrew (2006). Building a pathway to cooperation: Negotiation and social exchange between principal and agent. *Administrative Science Quarterly*, 51(1), 29-58.

Brandts, Jordi and Charness, Gary (2003) *Do labour market conditions affect gift exchange? Some experimental evidence*. Working paper.

Brandts, Jordi and Cooper, David. (2005). "It's what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure." Working paper.

Brandts, Jordi and Cooper, David. (2006). "A Change Would Do You Good …. An Experimental Study on How to Overcome Coordination Failure in Organizations." The American Economic Review, Volume 96, Number 3, pp. 669-693.

Brandts, Jordi and Macleod, W. Bentley. (1995). "Equilibrium Selection in Experimental Games with Recommended Play." Games and Economic Behavior, Volume 11, Issue 1, pages 36-63.

Brown, Martin, Falk, Armin, and Fehr, Ernst (2004). Relational contracts and the nature of market interactions. *Econometrica*, 72(3), 747-780.

Cachon, Gerard P. and Camerer, Colin (1996). Loss-avoidance and forward induction in experimental coordination games. *Quarterly Journal of Economics*, 111(1), 165-194.

Camerer, Colin F. (1998). Behavioral economics and nonrational organizational decision making. In J. Halpern & R. Stern, *Debating rationality: Nonrational aspects of organizational decision making* (pp. XX-XX). Ithaca, NY: ILR Press

Camerer, C. (2003).

Camerer, Colin, Ho, Teck-Hua, and Chong, Juin-Kuan (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119(3), 861-898.

Camerer, Colin and Lovallo, Dan (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, 89(1), 306-318

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

Camerer, Colin F., Babcock, Linda, Loewenstein, George and Thaler, Richard (1997). Labor supply of New York City cab drivers: One day at a time. *Quarterly Journal of Economics*, 111, 408-441.

Carpenter, Jeffrey, Matthews, Peter Hans and Schirm, John (2007). Tournaments and office politics: Evidence from a real effort experiment. Bonn IZA DP no 2972.

Charness, Gary (2004). Attribution and reciprocity in an experimental labor market. *Journal of Labor Economics*. 22(3), 665-688.

Charness, Gary, Corominas-Bosch, Margarida and Frechette, G.R. (2007). Bargaining and network structure: An Experiment. *Journal of Economic Theory*, 36(1), 28-65

Charness, Gary, Frechette, Guillaume, and Kagel, John. (2004). “How Robust is Laboratory Gift Exchange.” Experimental Economics, Vol 7, Issue 2, , pp. 189-205.

Charness and Jackson (CITE working paper).

Charness, Gary and Yang, Chun-Lei. (2006). “Exit, Exclusion, and Mergers: Endogenous Group Formation and Public Goods Provision.” Working paper.

Chaudhuri, A. & Schotter, A. & Sopher, B., 2001. "Talking Ourselves to Efficiency: Coordination in Inter-Generational Minimum Games with Private, Almost Common and Common Knowledge of Advice," Working Papers 01-11, C.V. Starr Center for Applied Economics, New York University.

Cooper, Russell W. (1999). *Coordination games: Complementarities and macroeconomics*. Cambridge: Cambridge University Press.

Cooper, David J. (2006). Are experienced managers experts at overcoming coordination failure? *Advances in Economic Analysis and Policy*, 6(2), 1-50.

Cooper, DeJong, Forysthe & Ross 1992 QJE

Cooper, Russell; Douglas V. DeJong; Robert Forsythe; Thomas W. Ross (1992). “Communication in coordination games.” The Quarterly Journal of Economics, Vol. 107, No. 2., pp. 739-771.

Corbae, D. and Duffy, John (in press). Experiments with network formation. Games and Economic Behavior.

Crawford, Vincent P. (1995). Adaptive dynamics in coordination games. *Econometrica*, 63(6), 103-143.

Crémer Jacques (1993). Corporate culture and shared knowledge. *Industrial and Corporate Change*, 2, 351 - 386.

Davis, Murray. (1971). That’s interesting! *Philosophy and Social Science* 1, 309-344.

DeMarco, Tom (1995). Why does software cost so much? and other puzzles of the Information Age. New York, N.Y. : Dorset House Pub.

Duffy, John and Hopkins, Ed (2005). Learning, information and sorting in market entry games: Theory and evidence. *Games and Economic Behavior*, 51, 31—62.

Ehrenberg, Ronald G. and Bognanno, Michael L. (1990) The Incentive effects of tournaments revisited: Evidence from the European PGA Tour. Industrial and Labor Relations Review, 43(3), S74-S88.

Engelmann, Dirk and Ortmann, Andreas (2002). *The robustness of laboratory gift exchange: A reconsideration*. CERGE-EI Working Paper.

Eriksson, Tor and Marie-Claire Villeval (2004). Other-Regarding Preferences and Performance Pay - An Experiment on Incentives and Sorting. Working paper.

Evan, William M. and Simmons, Roberta G. (1969). Organizational effects of inequitable rewards: Two experiments in status inconsistency. *Administrative Science Quarterly,* 14(2), 224-237

Farber, Henry S. 2004. Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers. Working paper.

Farber, Henry S.. Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers, Journal of Political Economy, 2005, vol. 113, no. 1 pp 46-82.

Fehr, Ernst and Gächter, Simon (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4), 980-994.

Fehr, Ernst. Gachter, Simon, and Kirchsteiger, Georg (1997). Reciprocity as a Contract Enforcement Device: Experimental Evidence. *Econometrica*, 65(4), 833-860.

Fehr, Ernst and Goette, Lorenz (2007). Do workers work more if wages are high? Evidence from a randomized field experiment. American Economic Review. 97(1), 298-317.

Fehr, Ernst & Kirchsteiger, George & Riedl, Arno, 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation," The Quarterly Journal of Economics, MIT Press, vol. 108(2), pages 437-59, May.

Fehr, Ernst, Kirchler, Rich, Weichbold, Andreas and Gachter, Simon (1998). When social norms overpower competition: Gift exchange in experimental labor markets. *Journal of Labor Economics*. 16(2), 324-351.

Fehr, Tougareva 1995

Feiler & Camerer 2007

Foss, Kirsten (2001). Organizing technological interdependencies: A coordination perspective on the firm. *Industrial and Corporate Change*. 10, 151 - 178.

Frechette 200x

Frey, B.S. and Jegen, R. (2001). Motivation Crowding Theory. *Journal of Economic Surveys*. 15, Part 5, 589-612.

Gacter & Renner (2004). Leading by Example in the Presence of Free Rider Incentives. Working Paper.

Gneezy, Uri and List, John A (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica* 74(5), 1365-1384.

Gneezy, Uri and Rustichini, Aldo (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics*, 115(3), 791-810

Goeree, Jacob K. and Holt, Charles A. (2005). An experimental study of costly coordination. *Games and Economic Behavior*, 51(2), 349-364.

Greenberg, Jerald (1990). Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of Applied Psychology*, 75(5), 561-568.

(Matt Jackson review paper World Congress).'

Guth, Levati, Sutter & van der Heijden 2004 CITE Working paper

Hamman, John, Scott Rick and Roberto A. Weber. (in press). "Solving coordination failure with 'all-or-none' group-level incentives." *Experimental Economics*.

Hannan, Lynn, Kagel, John H., and Moser, Donald V. (2002) . Partial gift exchange in an experimental labor market, impact of subject population differences, and effort requests on behavior. *Journal of Labor Economics*, 20(4), 923-951.

Healy, Paul J. (in press). Group reputations, stereotypes, and cooperation in a repeated labor market. *American Economic Review*.

Heath, Chip and Staudenmeyer, Nancy (2000). Coordination neglect: How lay theories of organizing complicate coordination in organizations. *Research in Organizational Behavior*, 22, 155-193.

Hennig-Schmidt, Helke, Rockenbach, Bettina, and Sadrieh, A., (2003). Incomplete and asymmetric surplus information in labor relations. Center Tilburg paper 2003-121, December 2003.

Hermalin, Benjamin E. (1998). Toward an economic theory of leadership: Leading by example. *American Economic Review*, 88(5), 1188-1206.

Hermailin, Benjamin E. (2001). "Economics and Corporate Culture" in The International Handbook of Organizational Culture and Climate, edited by S. Cartwright et al, Chichester, England: John Wiley & Sons.

Heyman, James and Ariely, Dan (2004). Effort for payment a tale of two markets. *Psychological Science*, 15(11), 787-793.

Hirshleifer, Jack (1983). From weskest-link to best-shot: The voluntary provision of public goods. *Public Choice*, 41(3), 371-386.

Hölmstrom, Bengt (1982). Moral hazard in teams. *Bell Journal of Economics*, 13(2), 324-340.

Hölmstrom, Bengt and Paul Milgrom (1994) "The Firm as an Incentive System" The American Economic Review, 84(4), 972-991.

Kachelmeieir, Steven J. and Shehat, Mohamed. (1992). Culture and competition: A laboratory market comparison between China and the West. *Journal of Economic Behavior and Organization*, 19, 145-168.

Kearns, Michael, Suri, Siddharth, and Montfort, Nick. (2006). An experimental study of the coloring problem on human subject networks *Science*. 313(5788), 824-826.

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

Knoeber, Charles R. (1989). "A real game of chicken: Contracts, tournaments, and the production of broilers." Journal of Law, Economics, & Organization, Vol. 5(2), pp. 271-292.

Kosfeld, M. (2004). Economic networks in the laboratory: A survey. *Review of Network Economics* 3, 20-41.

Kremer QJE

Kreps, David M. (1990). "Corporate Culture and Economic Theory." In Perspectives on Positive Political Economy, eds. James Alt and Kenneth Shepsle, Cambridge, UK: Cambridge University Press.

Krueger & Mas

Kuang, Xi (Jason), Roberto A. Weber and Jason Dana. (2007). "How effective is advice from interested parties?: An experimental test using a pure coordination game." Journal of Economic Behavior & Organization, Volume 62, Issue 4, pages 591-604.

Kube, S., M. Marechal and C. Puppe. Putting reciprocity to work: Positive versus negative responses in the field. University of St Gallen discussion paper 2006-27.

Lambert, Richard A., David F. Larcker, Keith Weigelt (1993). "The Structure of Organizational Incentives." Administrative Science Quarterly, Vol. 38, 1993.

Lazear, Edward P. (2000). “Performance Pay and Productivity.” The American Economic Review, Vol. 90, No. 5, pp. 1346-1361

Lazear, Malmendier & Weber (2006). Sorting in Experiments with Application to Social Preferences. Working paper.

Lazear, Edward P. and Rosen, Sherwin. (1981). Rank-Order Tournaments as Optimum Labor Contracts. *The Journal of Political Economy*, 89(5).. 841-864.

Leavitt, H. J. (1951). “Some effects of certain communication patterns on group performance.” Journal of Abnormal Psychology, 46(1):38-50.

Levitt, Steven D. and John A. List (2007). “What do Laboratory Experiments Measuring Social Preferences tell us about the Real World,” ***Journal of Economic Perspectives***, 21(2): 153-174.

Ledyard, John O. (1995). “Public Goods: A Survey of Experimental Research.” In In Handbook of Experimental Economics, edited by J. Kagel and A. Roth, Princeton University Press

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology*, *28*, 129-137.

Lind, E. A. and T. R. Tyler. 1992. “Procedural Justice in Organizations,” Chapter 8 in The Social Psychology of Procedural Justice. New York: Plenum, pp. 173-177, 188-191.

Uri Gneezy, John A List (2006). Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments. Econometrica 74 (5), 1365–1384.

March, James G. and Herbert A. Simon (1958). Organizations. New York: Wiley.

Maréchal, M.A. and Thöni, C. (2006). Do managers reciprocate? Field experimental evidence from a competitive market. Mimeo.

Mayo, E. (1933). The Human Problems of an Industrial Civilization. Manchester, NH, Ayer.

McKelvey, Richard and Thomas Palfrey (1998). "Quantal Response Equilibria for Extensive Form Games," Experimental Economics, Springer, vol. 1(1), pages 9-41

Mehta, Judith, Chris Starmer and Robert Sugden (1994). “The Nature of Salience: An Experimental Investigation of Pure Coordination Games.” The American Economic Review, Vol. 84, No. 3, pp. 658-673

Meidinger & Villeval 2002. Leadership in Teams: Signaling or Reciprocating. Working Paper

Milgrom, Paul and Roberts, John (1988). An economic approach to influence activities in organizations. *The American Journal of Sociology Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure*, 94, S154-S179.

Moldovanu, Benny and Aner Sela (2001). "The Optimal Allocation of Prizes in Contests." The American Economic Review, Vol. 91, No. 3, pp. 542-558

Moxnes, Erling and van der Heijden, Eline. (2004). The Effect of Leadership in a Public Bad Experiment. Journal of Conflict Resolution, Vol. 47, No. 6, 773-795

Mulligan, Casey. The intertemporal substitution of work-- What does the evidence say? University of Chicago Population Research Center, 1995

Müller and Schotter (2007) "Workaholics and drop outs in optimal organizations" Working paper.

Nagin, Daniel S.; James B. Rebitzer; Seth Sanders; and Lowell Taylor. (2002). Monitoring, motivation and management: The determinants of opportunistic behavior in a field experiment. AER vol 92 no 4 02 850-873.

Nalbantian, Haig, ed. 1987. *Incentives, Cooperation, and Risk Sharing.* Totowa, N.J.: Rowman and Littlefield.

Nalbantian, Haig R. and Andrew Schotter. (1997). "Productivity Under Group Incentives: An Experimental Study." The American Economic Review, Vol. 87, No. 3, pp. 314-341

Orbell, John M. and Robyn M. Dawes (1993). "Social Welfare, Cooperators' Advantage, and the Option of Not Playing the Game" American Sociological Review, Vol. 58, No. 6, pp. 787-800

Orrison, Schotter and Weigelt (2004). Multiperson tournaments: An experimental examination. Management Science, February, vol ?, pp 1-10? CHECK CITE

Ortmann, Andreas and Ralph Hertwig (2002). The Costs of Deception: Evidence from Psychology. Experimental Economics, 5(2): 111-131.

Paarsch Shearer 2007 wp Cahier de Recherce CIRpee 07 FIELD EXP

Page, Talbot, Louis Putterman and Bulent Unel. (2005). "Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry and Efficiency." The Economic Journal 115 (506), 1032–1053.

Pondy, Louis R. and Jacob G. Birnberg (1969). "An Experimental Study of the Allocation of Financial Resources Within Small, Hierarchical Task Groups." Administrative Science Quarterly, Vol. 14, No. 2, pp. 192-201.

Potters, Jan, Martin Sefton and Lise Versterlund (2007). "Leading-by-Example and Signaling in Voluntary Contribution Games: An Experimental Study" Economic Theory. 33, 169-182.

Radner and van Zandt (CITE)

Rapoport, Amnon, Darryl Seale & Yuval Winter (2002). "Coordination and Learning Behavior in Large Groups with Asymmetric Players. Games and Economic Behavior, 39: 111-136.

Rigdon, Mary. Efficiency wages in an experimental labor market. PNAS 13348-13351, vol 99, no 20, Oct 1, 2002.

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

Schotter, Andrew and Keith Weigelt (1992). Asymmetric tournaments, equal opportunity laws, and affirmative action: Some experimental results. *Quarterly Journal of Economics*, 107(2), 511-539.

Shearer, Bruce. (2004). Piece rates, fixed wages and incentive effects: Evidence from a field experiment. REStud

Vernon L. Smith (1976). "Experimental Economics: Induced Value Theory." The American Economic Review, Vol. 66, No. 2, pp. 274-279

Smith & Williams in Kagel Book

Taylor, Frederick W. (1911). Principles of Scientific Management. New York: Harper & Brothers.

Thompson, James D. (1967). Organizations in Action: Social Science Administration Theory. New York: McGraw-Hill.

Triplett, Norman. (1898). "The Dynamogenic Factors in Pacemaking and Competition." The American Journal of Psychology, Vol. 9, No. 4, pp. 507-533

Van der Heijden, Eline; Potters, Jan; and Sefton, Martin (2005). Hierarchy and opportunism in teams. August

van der Heijden, Eline; Potters, Jan; and Sefton, Martin (2006). Working paper

Van Huyck, John B., Raymond C. Battalio, Richard O. Beil (1990). “Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure.” The American Economic Review, Vol. 80, No. 1, pp. 234-248.

Van Huyck, John B., Raymond C. Battalio, Richard O. Beil (1991). “Strategic Uncertainty, Equilibrium Selection, and Coordination Failure in Average Opinion Games” The Quarterly Journal of Economics, Vol. 106, No. 3, pp. 885-910

Van Huyck, John B., Raymond C. Battalio, Richard O. Beil (1993). Asset markets as an euqilibrium selection mechanism. Games and Economic Behavior, 5, 485–504.

Van Huyck, John B., Ann B. Gillette and Raymond C. Battalio (1992). “Credible Assignments in Coordination Games.” Games and Economic Behavior, 4, 606-626.

Weber, Roberto A. (2006). “Managing growth to achieve efficient coordination in large groups.” *The American Economic Review* 96(1):114-126.

Weber, Roberto A. and Colin Camerer. (2003). “Cultural conflict and merger failure: An experimental approach.” *Management Science* 49(4):400-415.

Weber, Roberto A., Yuval Rottenstreich, Colin Camerer and Marc Knez. (2001). “The illusion of leadership: Misattribution of cause in coordination games.” *Organization Science* 12(5):582-598.

Weick, Karl E. (1969). “Laboratory Organizations and Unnoticed Causes.” Administrative Science Quarterly, 14 (2): 294-303.

Yerkes, R. M. and J. Dodson (1908). “The relation of strength of stimulus to rapidity of habit formation.” Journal of Comparative Neurology and Psychology, 18, 459-482.

---

[*] 10/11/2007 10:43 PM This chapter was prepared for the Handbook of Organizational Economics (Bob Gibbons and John Roberts, eds). Comments welcome. Thanks to Scott Rick for very active collaboration in the earliest draft and ? for comments etc.

[1] One powerful approach for obtaining control is the use of random assignment to treatment. The experimenter conducts two almost identical kinds of experimental sessions, and varies only the factor(s) of interest. Human participants are randomly assigned to such varying “treatments,” thus creating statistically identical conditions under which the influence of only the varying factor can be tested. Experimental economists also frequently devote considerable attention to designing procedures (e.g., for determining incentives, see Smith 1976 AER) that produce a laboratory environment as similar as possible to the theoretical environment that the experiment is intended to test.

[2] In economics experiments with interacting groups and repeated behavior, it is generally inappropriate to treat individual observations as i.i.d. draws. Instead, the appropriate tools are those used for time-series panel data with some specification of interaction effects among agents and over time. Controlling for various levels of clustering is one approach (see Frechette, 2006 Working paper). In practice, researchers often report tests using both incorrect i.i.d. assumptions and overly conservative assumptions (e.g., treating behavior averaged across an entire session as a single data point, and testing across sessions).

[3] However, it is important to note that generalizability is simply one criterion on which many experiments can be evaluated, and not a litmus test that must be applied to determine the usefulness of every experiment. Testing theories often requires creating environments that are different than what we typically observe in the natural world, but that contain all the key elements of a theory (In the physical sciences theoretically-motivated experiments similarly often use environments that are atypical, dissimilar from where the theory is usually applied, and do not exist naturally (for example, particle accelerators, zero-gravity, or a vacuum).
[4] For example, the recent paper by Bottom et al (2006 ASQ) examines how employees respond to both financial and social incentives. This similar in focus to the work reviewed in Section III.A.
[5] Revenue-sharing contracts are interesting because many firms use variants of group-based performance schemes (such as employee stock ownership (ESOP) plans). In theory, these schemes should not work well because in large teams, workers should free ride if their shirking cannot be monitored and punished. Yet group-based performance schemes often work surprisingly well in practice (Blinder 1990, Nalbantian 1987), which motivates an interest in understanding when teams do not fail as badly as they should in theory.
[6]These deviations could be explored with newer models of stochastic response (McKelvey and Palfrey, 1998 Exp'l Econ) or limited strategic thinking (Camerer and Ho, 2004 QJE); see Camerer (2003 book) for a review of these theories, and others which incorporate players' regard for the payoffs of others.
[7] Farber (2005) finds a stronger effect of hours than cumulated daily income, but does find evidence consistent with targeting in three of five drives on which he has many days of data. Using a different approach, Farber (2004) finds clearer evidence of targeting, although inferred targets are statistically imprecise.
[8] Nagin et al also used surveys to compare the influence of psychological and economic factors on generating SBC's. Measures of difficulty of finding a comparable job, and expected tenure with the firm (typical economic variables), were not correlated with SBC's, but employee attitudes toward their employer did have a significant effect, consistent with pure reciprocity..
[9] In Fehr, Kirchsteiger and Reidl (1993) the labor market was described in plain buyer-seller language. Switching to explicit labor market language does not seem to make much difference.
[10] They also varied the value of effort to firms, comparing profit functions (120-w)e with (90-w)e. There are not large differences, except that the low-value firms pay lower wages in the MBA sessions and request higher effort levels.
[11] The insensitivity to excess labor supply or demand is in contrast to earlier experiments on goods markets (with private information about values and costs). In those markets, even small proportional excess demand or supply drive prices to high or low levels at which consumer and producer surplus are highly asymmetric (see Smith and Williams, book chapter Kagel Green edited)..
[12] Field data from police-union arbitration (Mas, 2006 NBER working paper) and a long tire-company strike (Krueger and Mas, 2004 JPE) also suggest very long-term effects linked to perceptions of fairness and reciprocity.
[13] See Camerer and Lovallo, 1999 AER; Rapoport, Seale and Winter, 2002 GEB ; Duffy and Hopkins, 2005 GEB.
[14] Orbell and Dawes (cite)? Report a similar effect in PD games where the mutual defection outcome is negative. Subjects who opt to play tend to expect cooperation and cooperate themselves. The conditional cooperators sort into the PD and earn positive profits, while defectors expect others to defect and stay out.
[15] They describe an experiment in which groups of MBA students have to assemble a "Lego man" consisting of arms, legs, and a torso. A long period of untimed planning is followed by a timed period of assembly. The goal is to assemble the Lego man properly in the fastest assembly time. Planning time spent on partitioning (dividing up tasks) and time spent on integration (discussing how to assemble the parts later) are equally valuable in speeding up assembly time, but groups tend to do more partitioning and less integration.
[16] See Hirshleifer, 1983 for an early discussion of this game and Kremer (QJE?) and Becker and Murphy (1992) for applications to economic growth and firm output. This game falls in a more general class of games known as order-statistic games, in which each player's payoff is a function of her own choice and some order statistic (such as the minimum or median) of all players' choices (see Van Huyck, Battalio and Beil, 1991 QJE; Crawford, 1995 Econometrica).
[17] Camerer and Feiler also conducted regular auctions in which successful bidders pay their bids (in this case, since bids are demands for payment, successful bidders are paid what they demanded). Since these bids are not, in theory, the same as actual valuations, an inferential procedure is used to estimate true willingness-to-accept from bids. Both raw bids and inferred valuations suggest that winning bidders are too optimistic about the merger.
[18] In a second part of the experiment (periods 17-24), groups voted on whether to have a leader or not (with unanimity required). Only 40 percent of the subjects voted to have leaders, but groups with leaders obtained significantly higher contributions and earnings. This raises the interesting question of why a structural change that

benefits the group as a whole is not more popular. Subjects might not like ceding authority to a leader, or they may underestimate the leader's economic impact which helps everyone.