NBER WORKING PAPER SERIES

LAB LABOR: WHAT CAN LABOR ECONOMISTS LEARN FROM THE LAB?

Gary Charness
Peter J. Kuhn

Working Paper 15913
http://www.nber.org/papers/w15913

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2010

Manuscript prepared for Handbook of Labor Economics, volume 4, Orley Ashenfelter and David Card, editors. We thank Michael Kuhn and Loni Spilberg for excellent research assistance. We have done our utmost to identify as many lab experimental papers in labor economics as possible, and to represent the results of these papers accurately. Given the volume of recent research in this area and the deadline we have faced in writing this review, it is inevitable that we have missed some important work. We also sincerely hope that we have not mischaracterized any paper's conclusions. The authors apologize for these limitations in advance and nevertheless hope that our review provides a helpful guide to the 'lab labor' literature to date The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Lab Labor: What Can Labor Economists Learn from the Lab?
Gary Charness and Peter J. Kuhn
NBER Working Paper No. 15913
April 2010
JEL No. C9,J0

ABSTRACT

This chapter surveys the contributions of laboratory experiments to labor economics. We begin with a discussion of methodological issues: why (and when) is a lab experiment the best approach; how do laboratory experiments compare to field experiments; and what are the main design issues? We then summarize the substantive contributions of laboratory experiments to our understanding of principal-agent interactions, social preferences, union-firm bargaining, arbitration, gender differentials, discrimination, job search, and labor markets more generally.

Gary Charness
Department of Economics
University of California, Santa Barbara
2127 North Hall
Santa Barbara, CA 93106
charness@econ.ucsb.edu

Peter J. Kuhn
Department of Economics
University of California, Santa Barbara
2127 North Hall
Santa Barbara, CA 93106
and NBER
pjkuhn@econ.ucsb.edu

The economics literature has witnessed an explosion of laboratory experiments in the past 20 years. Many of these experiments have focused on topics that are central to the field of labor economics, including how workers respond to various forms of compensation, and the economics of discrimination, arbitration, bargaining, and matching. In this chapter we survey the contributions of laboratory experiments to our understanding of these questions.

We begin our review with a discussion of methodological issues: First, we pose the general question of why (and more importantly when) a labor economist might want to conduct a laboratory experiment: What types of questions, if any, are laboratory experiments best suited to answer? How do laboratory experiments compare to field experiments? Next, once one has decided to conduct a laboratory experiment, how should it be designed? Here we review the main methodological decisions an experimenter typically needs to make, and the advantages and disadvantages of the various choices.

The second half of our review turns its attention to the substantive issues in labor economics that have been addressed using laboratory experiments. While these are wide-ranging, we focus our review on the set of issues that have generated probably the largest volume of experimental papers in labor economics: the effects of compensation policies on the supply of effort by workers. We do this in two parts. The first uses 'traditional' principal-agent theory as a theoretical lens to derive predictions regarding the effects of incentives, and tests these predictions in the lab. Many are confirmed; at the same time a number of robust 'anomalies', such as apparent gift exchange, also appear. The second part focuses specifically on the use of experiments and the development of new theoretical models of social preferences to understand these anomalies. Finally, we also provide brief guides to the laboratory literature on a number of other labor-related topics, including union-firm bargaining, arbitration, gender differentials, discrimination, and job search.

I. Why Laboratory Experiments?

Why should labor economists care about laboratory experiments? After all, there are plenty of field data available for empirical tests. In addition, there have been a number of objections to lab experiments concerning issues such as a lack of realism (external validity), demand effects, and selection effects. Indeed, the laboratory is an artificial environment. On the other hand, lab experiments have some important advantages over other approaches; we begin this section with a discussion of these advantages, then move on to critiques and responses.

1. Advantages of Laboratory Experiments

Most practitioners of lab experiments would probably agree that a key advantage is the ability to control conditions more tightly than in any other context. For example, testing theory is a basic component of both the physical and social sciences, and the scientific method relies upon explicit tests of theory. While empirical data are indeed rich and abundant, they reflect a variety of environmental factors; disentangling these factors is difficult if not impossible.

Falk and Fehr (2003) provide the concrete example of testing tournament theory (e.g., Lazear and Rosen 1981), where contestants should in equilibrium choose effort levels to equate marginal effort with marginal gain. Since a direct empirical test of this theory must take into account "the number of workers who compete for the prize, the effort cost functions of the workers, the exact level of the prize, and the production level including the nature of the error term" (p. 400), such a test seems impossible with traditional empirical data.[1] However, all of these factors can be controlled (and systematically varied) in a laboratory experiment. In this manner, it is also feasible to study the impact of specific institutional arrangements on behavior by systematically varying them.

Of course, the same argument can be made regarding the effects of workers' outside options, minimum wages, sick pay, discrimination, etc. As another example, the gift-exchange game (an experimental model of Akerlof 1982) tests for a positive relationship between wages and effort; as this is a critical assumption of efficiency-wage theories (as discussed for example in Akerlof 1984 and Akerlof and Yellen 1986), it is useful to test it; however, it is quite difficult to do so with standard field data. In general, the laboratory offers superior control, which make it possible to identify causal relationships. In field data, variables are often determined endogenously and usually it is only possible to identify correlation. Finally, if one has doubts concerning data reported in a laboratory experiment, one can readily replicate the experiment (particularly when a standard subject pool is used).[2]

One of the greatest strengths of lab experiments is the ability to take a specific theoretical model (say of behavior under a specific group incentive scheme, with no communication between players and one-shot interaction), where theory says exactly what, say, the Perfect Bayesian equilibrium should be, and have real agents play exactly that game with real monetary consequences. One can then compare the predictions of the model to what happens; if the theory is rejected it is then relatively easy to test competing explanations (e.g. inequity aversion, 'reciprocity', loss aversion, framing) for the rejection. One way to think about a lab experiment is as a first link in a longer chain running from theory to actual interactions in real firms. A distinct role of lab experiments

[1] Furthermore, Falk and Fehr (2003) point out that other factors such as communication, peer pressure, and whether the interaction is one-shot or repeated may well affect behavior and are unlikely to be known to a researcher using field data.

[2] See Charness (forthcoming) for more discussion regarding replicability.

applies to situations where 'standard' game theory doesn't give us crisp predictions, as in games with multiple equilibria. Important examples include production in teams with complementarities among the agents' efforts (Brandts and Cooper 2007), where the multiple equilibria are due to coordination problems, and repeated principal-agent interactions (e.g. Brown, Falk and Fehr 2004) where the folk theorem can generate an infinity of equilibria. In these cases, experiments provide information on how people behave in situations where existing theory provides little or no guide to what should happen.

The theory of mechanism design often provides suggestions for a number of mechanisms that are predicted to yield socially-efficient outcomes, or sometimes for a number of alternative mechanisms that are all predicted to be equivalent (for example, it is well known that a properly-designed tournament should be able to exactly mimic an efficient individual piece rate). Lab experiments however can be used to show that mechanisms/institutions that work in theory don't always do so in practice, *even under conditions designed to be ideal* for the institution. Also, institutions that are predicted to yield identical results may not do so. In fact, the use of lab experiments to 'pre-test' proposed allocation mechanisms before implementing them in the real world already has an established history (see Plott 1987 for some examples). The same is true of lab experiments in at least two labor contexts: the design of matching mechanisms in professional labor markets (e.g. McKinney, Niederle and Roth 2005) and the design of arbitration mechanisms for public-sector union bargaining (e.g. Deck, Farmer and Zeng 2007).

A sometimes-overlooked advantage of lab experiments is their low cost, especially compared to field experiments and survey data collection: competing explanations can often be tested or distinguished quickly and inexpensively with a modest number of sessions. In this sense, dropping laboratory experiments from our toolkit would be a little like dropping animal studies from cancer research: while results from animal studies do not always apply to humans, the ability to test many hypotheses cheaply under carefully controlled conditions provides an indispensable tool for the development of models that work in the real world.

Another comparative advantage of lab experiments is in the study of phenomena that are hard to measure in the field because they are illegal or face disapproval, such as acts of sabotage, discrimination, and spite. It is also relatively easy to measure agents' *beliefs* in the lab, using monetary incentives. This is important in view of the role played by beliefs in many game-theoretic models. Belief data from experiments has been central in the development of new behavioral theory such as guilt aversion (see Charness and Dufwenberg 2006, Battigali and Dufwenberg 2007, 2009).

Finally, lab experiments offer unique opportunities to researchers who are interested in the *form of strategies* used by agents in solving dynamic problems (or playing dynamic games). An illustrative example here is the classic problem of search from a fixed wage distribution, which is often used by labor economists to model individual workers' unemployment spells. Theory has strong predictions here –that the

optimal strategy has a reservation wage property--, but it is difficult to test this prediction from field data (whether experimental or not) because strategies need to be inferred from choice histories. In contrast, the lab makes it easier to elicit subjects' strategies more directly in a number of ways, including asking subjects to describe their strategy (Hey 1982), observing subjects' use of information boards (Sonnemans 1998), and forcing subjects to play the game using the strategy method (Sonnemans 1998, Brown, Flinn and Schotter forthcoming). This approach has identified some interesting deviations between actual and predicted strategies (for example, subjects seem to condition their acceptance behavior on factors like their total earnings to date, which is not optimal) that researchers are now attempting to understand using a variety of behavioral approaches.

2. Objections to laboratory experiments and some responses

The most common objection to the data obtained through laboratory experiments is that they have nothing to do with the field environment (no external validity). In principle, this is a serious objection, of course. It comprises a number of facets, such as the fact the participants are usually undergraduate students, who typically have little experience in labor markets (particularly as firms), the issue that the stakes are low, and the fact that "labor task" is often simple the choice of how much money to assign to another party. One might also be concerned that participants are affected by the mere fact that they are being observed (Hawthorne effects).

There are at least two main responses to these objections. First, as pointed out by Falk and Heckman (2009), "for the purpose of testing theories, [representative evidence] is not a problem because most economic models derive predictions that are independent of assumptions concerning participant pools (p. 537). Of course, it seems better to have a richer variation than is provided by undergraduate students, the most convenient source of participants for experiments conducted by academic researchers.[3] The second response involves the use of more "real-world" participants, allowing for agent self-selection *in the lab*, less artificial tasks, and higher stakes.

Laboratory experiments have been conducted on soldiers (Fehr, Kirchsteiger, Weichbold, and Gächter 1998), Costa Rican coffee-plantation CEOs (Fehr and List 2003), Chinese central planners (Cooper, Kagel, Lo and Gu 1999), professional arbitrators (Farber and Bazerman 1986), Ghanaian manufacturing workers (Barr and Serneels 2007), Japanese shrimp fishermen (Carpenter and Seki forthcoming), and employees at large French firms (Charness and Villeval 2009), among others.[4] In fact, the performance by student participants is often fairly closely matched in such experiments. In addition, many if not most field experiments on incentive effects focus

[3] Note that it is also inexpensive to use undergraduate participants, who also tend to be fairly intelligent (in fact, evidence suggests that the undergraduates who participate in experiments are more capable than the average); this is a nice combination for the experimenter.

[4] In the latter study, a mobile laboratory was taken to the workplaces of two French firms, where older (over the age of 50) and younger workers (under the age of 30) were tested in three experimental environments.

on highly specific industries or occupations, such as windshield repairers (Lazear 2000), tree planters (Shearer 2004), fruit pickers (Bandiera, Barankay and Rasul 2005), and bicycle messengers (Fehr and Goette 2007).[5] Virtually all of these seem likely to be a more highly-selected population than college students as a group, who can reasonably be considered to be representative of the college-educated labor force. Therefore, if the goal is to identify general principles that apply broadly to a large population of workers, college students might be a more attractive choice than workers in a single, narrowly-defined occupation or industry.

A second approach to the 'representativeness' issue is to mimic, in the lab, the same sorts of self-selection that generate different subpopulations in the real world. Clearly, such selection can be important, for example, if altruistic workers tend to self-select into cooperative work environments (such as teams), or risk-loving (or overconfident) workers self-select into highly competitive work environments. Persons in jobs that frequently require them to make 'tough' decisions such as cutting workers' pay or firing them (i.e. managers) might have highly selected social preferences indeed, so that laboratory experiments that randomly assign college students to represent 'principals' may provide a particularly poor guide to the decisions of real managers. To some extent, however, laboratory experiments can allow for such self selection and even shed important light on how it works. Interesting examples of this approach in a non-labor-market context include Lazear, Malmendier and Weber (2006), and Della Vigna, List and Malmendier (2009), who allow experimental subjects to self-select out of a situation where they are 'expected' to be altruistic. In a labor market context, Eriksson, Teyssier and Villeval (2009) show that allowing risk-averse subjects to self-select out of tournaments improves tournament performance. A series of interesting experiments beginning with Gneezy, Niederle and Vesterlund (2003) show that women tend to self-select out of tournaments. A key advantage of the lab in addressing these self-selection questions is the opportunity to directly measure, and control for, confounding factors such as the agent's ability at the task, her perception of her own and others' abilities, loss aversion, and risk aversion.

Concerning the objection that the labor task is abstract and artificial, there has been an increasing trend in "real-effort" experiments, in which tasks have included proofreading (Frohlich and Oppenheimer 1992), solving puzzles (Rütstrom and Williams 2000), mazes (e.g., Gneezy, Niederle, and Rustichini (2003), anagrams (Charness and Villeval 2009), complex optimization problems (van Dijk, Sonnemans, and van Winden 2001), simple clerical tasks (Falk and Ichino 2006, Carpenter et al., forthcoming) and cracking walnuts (Fahr and Irlenbusch 2002).

Regarding the issue of small stakes, laboratory experiments have been conducted in locations where the stakes translated into more than a month's earnings (e.g., Fehr and Tougareva 1995, Slonim and Roth 1998), with evidence that fairness considerations still

[5] A similar comment applies to 'natural experiment' studies of the effects of changes in compensation policy, which have focused on stadium vendors (Oettinger 1999), Continental Airlines employees (Knez and Simester 2001), textile workers (Hamilton, Nickerson and Owan 2003), steel minimill workers (Boning, Ichniowski, and Shaw 2007), and taxi drivers (Crawford and Meng 2008).

seem to apply.[6] Furthermore, it is not obvious whether stakes involving larger sums of money or the small stakes that apply to decisions people make on a daily basis are more relevant for economic purposes. It is also the case that large stakes do not necessarily lead to fewer mistakes, as in shown by Ariely, Gneezy, Loewentein, and Mazar (2008). Finally, it is possible that participants behave differently due to scrutiny (Levitt and List 2007). As discussed by Falk and Heckman (2009), in many laboratory experiments involving more complex decisions, this is likely to be only a minor problem.[7] And of course scrutiny can be present in the field as well, as workers are often monitored. In any event, scrutiny can be eliminated (or systematically varied); some experimenters use double-blind techniques, where payments are placed in envelopes by monitors who have not observed the experiment, so that participants understand that the experimenter cannot know their choices.

Thus, many of the objections raised against laboratory experiments are either red herrings or can be met by taking the laboratory to the field, using 'real' people (of course, students are real people as well, and they respond to the financial incentives provided), real-effort tasks, and varying the stakes.[8] So, while there are certainly issues in taking the results of laboratory experiments to the field environment, these can be ameliorated. The real value of laboratory experiments is in the enhanced opportunities for, and lower cost of carefully-controlled variation, as is required for causal knowledge rather than simple correlation. This control extends to environmental features such as institutions, payoff parameters, participant pools, the nature of the interaction among the participants (e.g., anonymity versus face-to-face; one-shot versus repeated), and even the level of scrutiny. To quote Falk and Heckman (2009): "Laboratory experiments are very powerful whenever tight control … is essential. … Tight control … also allows replicability of results, which is generally more difficult with field data" (p. 537). If one wishes to perform careful tests of theory, laboratory experiments are particularly useful.
Of course, none of the comments above should be taken to imply that laboratory techniques are intrinsically superior to standard empirical data or field experiments. We discuss this issue in some detail in the next subsection.

3. Laboratory experiments and field experiments

Is a hammer a better tool than a screwdriver (or *vice versa*)? Sometimes one needs a hammer and sometimes one needs a screwdriver. They are different tools, suited for different purposes. Claiming superiority for one tool over the other seems misplaced. This principle also applies to research methods, as each method has its own strengths and weaknesses. The idea is not a new one; Runkel and McGrath (1972) identify eight research strategies (including field studies, field experiments, and laboratory

[6] It is also worth mentioning that Bewley (1999) finds evidence of fairness considerations in his interviews of business people and that Krueger and Mas (2004) find evidence of negative reciprocity amongst Firestone workers and Mas (2008) shows that police performance suffers after unfavorable arbitration decisions.

[7] They also observe that "reanalysis of the original Hawthorne data [see Jones 1992] shows that no Hawthorne effect was present in the Hawthorne study" (p. 537).

[8] A comprehensive study by Camerer and Hogarth (1999) indicates that in many cases whether or not a laboratory participant is given financial incentives has little effect on behavior.

experiments), which they categorize along the two dimensions of obtrusive-unobtrusive research operations and universal-particular behavior. They state: "We cannot emphasize too strongly our belief that none of these strategies has any natural or scientific claim to greater respect from researchers than any other" (p. 89).[9]

This brings us to the current debate about the value of field experiments compared to the value of laboratory experiments. List (2008) and Levitt and List (2009) extol the value of field experiments, often mentioning the notion that these are a useful bridge between naturally-occurring environments and laboratory experiments. Levitt and List (2007) provide a criticism of laboratory experiments, pointing out factors that are beneficial in field experiments, while pointing out a number of factors that make the interpretation of data in laboratory experiments problematic. Their main issue is the degree to which "the insights gained in the lab can be extrapolated to the world beyond" (p. 153); this is also known as external validity. They mention five factors (p. 154) that can influence behavior in the lab; in our view, the three most relevant of these involve the nature and extent of scrutiny (on which the greatest emphasis is placed in the paper), the context in which the decision is embedded, and the stakes of the game.

Falk and Heckman (2009) provide a response to these comments, and strongly emphasize the value of laboratory experiments. The thrust of their argument is that the controlled variation possible in laboratory experiments facilitates tests of theory, causal effects, and treatment effects.[10] To a certain extent, field experiments also can provide fairly good control of the environment, although rarely to the level attainable with laboratory experiments.[11] Falk and Heckman address the notion that the conditions in field experiments are more "realistic"; for example, they point out how it is unclear whether undergraduate students are less representative of the overall population than sports-card traders in their natural setting.[12]

A number of arguments regarding scrutiny are mentioned above. Indeed, there is no doubt that the mere act of scrutiny can affect behavior.[13] However, since there is also scrutiny in field environments and since the sense of scrutiny and the associated possibility of 'demand effects' (where the participant acts in a manner that he or she

---

[9] We are indebted to Keith Murnighan for this information. For more detail, see the discussion in Murnighan (2010).

[10] They state: "This control allows for the testing of precise predictions derived from game-theoretic models" (p. 636).

[11] Indeed, Levitt and List (2009) discuss some of the limitations of field experiments, such as the difficulties with some forms of replication, and the fact that "they sometimes cannot be used to distinguish between alternative theories because the experimenter exerts less control than in the lab" (p. 1414).

[12] Holt (2007) points out that field experiments are particularly valuable when social context is critically important. However, he cautions: Although field experiments can induce a more realistic social context and environment, the cost is often a partial loss of control over incentives, over measurement of behavior, or over the ability to replicate under identical conditions" (p. 14). Regarding the independence of observations, he writes: "To the extent that social context and target demographics are important in a field experiment, each field experiment is in some sense like a data point that is specific to that combination of subjects and context unless appropriate random selection of subjects is employed" (p. 14).

[13] Charness, Rigotti, and Rustichini (2007) provide one example, as they find that the presence of an audience of peers with a common interest can lead to more aggressive behavior in experimental games)

believes reflects the experimenter's desired outcome) can be nearly eliminated,[14] this concern seems somewhat overstated. The notion that social preferences can be crowded out by large financial incentives is hardly new, as it is a feature of the Rabin (1993) model. Nevertheless, as Falk and Heckman point out, many real-life decisions involve small stakes, so that it is not clear that one requires larger stakes in order to provide incentives that are meaningful enough to match the relevant field environment.

Despite the divergent views expressed in these articles and others, it is worth noting that there is indeed common ground. For example, in both Levitt and List (2007) and Falk and Heckman (2009), the authors discuss how one needs a model or theory to transport findings to new populations or environments, whether these data originate in laboratory experiments or field experiments. These articles also appear to agree that the controlled variation in the laboratory is better for careful tests of theory. Both state that there are shortcomings in both laboratory and field experiments, but that each can provide useful insights. In fact, both camps apparently agree that both forms of experimentation (as well as hybrids) can be combined to yield a better understand of the phenomena involved. One interesting point is that two of the key players in the debate (Falk and List) have used and continue to use both laboratory and field experiments in their research.[15] In a certain sense, one wonders what the shouting is about.

Our own view is that laboratory experiments are best at testing theory and identifying treatment effects, and they can also provide useful qualitative insights. However, any assumption that the quantitative levels of behavior observed in the laboratory apply to naturally-occurring settings must be carefully considered, as the laboratory is only a model of the field environment and cannot include many details that may influence behavior.[16] Field experiments, for their part, offer promise in areas that are not readily susceptible to laboratory experimentation and generally involve a greater range of personal and demographic characteristics. Field experiments are especially valuable to the extent that they can capture more realistic behavior (particularly in settings where the participants are unaware that there is an ongoing experiment). That said, a similar level of care needs to be taken in applying quantitative estimates from

---

[14] Echoing the discussion on p. 4, the standard policy in the experimental laboratory at the University of Amsterdam is to separate the payment of experimental participants from any observation of behavior, by having non-experimenters place payments in envelopes that are then passed out by the experimenter. Given that this is standard procedure, experienced participants know to expect this and are not particularly suspicious about it being utilized in any specific experiment. List, Berrens, Bohara, and Kerkvleit (2004) use another approach, in which a "randomized response" means that the experimenter cannot ascertain whether a participant made a pro-social choice.

[15] Both authors of this chapter have also used both methods in their research.

[16] An example from the games in Charness and Rabin (2002) may provide a useful illustration. In one treatment, a participant unilaterally made a choice between (Other, Own) payoffs of (750,375) or (400,400); in a second treatment, the other paired participant first faced a choice between payoffs of (550,550) or passing the choice to the second participant, who would once again face a choice between (Other, Own) payoffs of (750,375) or (400,400). In the first treatment, roughly half of the population chose to sacrifice 25 units, selecting (750,375); however, only 10% did so in the second treatment. The comparison of the two treatments qualitatively shows a strong effect, which indicates that a form of negative reciprocity is present. Nevertheless one would be naive to conclude that these quantitative levels would also be observed in the outside world.

some highly selected field populations (fruit pickers, bicycle messengers, tree planters, school children) to other field populations. These research methodologies are complements, not substitutes. One should use the most appropriate tool or tools for the job at hand.

II: Issues in Designing Laboratory Experiments

Suppose you have decided that a laboratory experiment is a fruitful way to address a research question. This section reviews some of the main design questions the investigator typically needs to address. We do this in two stages: first, we consider general issues that arise in almost all laboratory experiments, not necessarily restricted to questions in labor economics. Second, we focus specifically on the design of 'supply of effort' experiments, which constitute the main focus of our review of the substantive research.

1. General Design Questions

The first and most basic question is how closely to try to match the field environment. This will depend to a substantial degree on whether one is testing theory, one is trying to isolate a treatment effect, or one is trying for realism in an effort to draw conclusions about the effect of policy changes in a specific environment.[17] While one should be fairly insistent that the details of the design deal correctly with the issues involved with a test of theory, one cannot expect the experimental design to precisely match the field environment. There are typically trade-offs between parsimony and richness. A general rule is to err on the side of simplicity, but to include the central elements of the question at issue. It is fundamental that the participants understand the task at hand, and this varies inversely with the degree of complexity.

Regarding the issue of comprehension, a choice variable is the degree of examples (or even actual coaching) that will be provided. Experimental practice has to some extent evolved over the years. In earlier times, it was customary to provide neither examples nor test questions, out of concern that giving examples could introduce bias or demand effects. However, this policy runs the serious risk that participants will fail to understand some important aspects of the task. It has become nearly standard practice to at least ask participants questions about what outcome would prevail in the event of various combinations of choices; it is also customary to provide examples in the instructions. Of course there is always the possibility that some bias may be introduced due to this process. Nevertheless, one can minimize this possibility by going over every contingency; if this is not feasible due to the presence of a large number of contingencies,

[17] Regarding the latter case, one example comes from industry. Hewlett-Packard was interested in the consequences of changing their minimum-advertised-price policy. Rather than use a costly test market, laboratory experiments were conducted; it this situation, it was critical to match their retailer environment as closely as possible. As another example, if firms in the experiment are losing money, it is difficult to argue that the observed behavior will persist over time in the field.

one can select 'representative' contingencies. Whether examples are needed will depend on the complexity present in the experiment.

A closely-related issue is whether to use an abstract context in the instructions or to provide a richer context that points to the field environment in question. In some experiments (e.g., Charness and Rabin 2002), the researchers are careful to choose completely neutral terms. The main advantage of this policy is that it may well limit bias.[18] On the other hand, many laboratory experiments explicitly label the subjects' roles as 'firms' or 'workers' (or even more specifically as 'high-ability workers', etc.) and to label the choices as 'wages' or 'effort', etc., even though (strictly speaking) the experiment is simply a game with no actual work performed. As with providing examples, the main advantage of a richer context is that it makes it easier for subjects to understand the game, while the primary disadvantage is that it might bring in established behavior patterns/expectations from those environments.

The question of context, framing and reference points is not innocuous. Sometimes the details of instructions given to the subjects can unwittingly cause their behavior to focus on certain outcomes. One nice example of framing comes from Liberman, Samuels, and Ross (2004). There are two treatments, both of which feature the identical prisoner's-dilemma game. However, in one treatment, the game is labeled "The Wall St. Game", while the game is labeled "The Community Game" in the second treatment. The rate of cooperation was less than 30% in The Wall St. Game, but was over 70% in The Community Game. A second example is provided by Cooper, Kagel, Lo and Gu (1999), who find that providing context for Chinese central planners in a lab experiment improved their understanding of the game; however this had no effect on the students in other sessions of the experiment. More recently, Levitt, List and Reilly (2009) present evidence suggesting that even professionals (such as world-class poker players who are skilled randomizers in the field) have difficulty transferring those skills to the unfamiliar context of the laboratory.[19]

Another important design question concerns whether an experiment features multiple periods or not. In our view, the first question the investigator needs to ask here is whether the real-world situation they are interested in understanding most closely resembles (a) repeated interactions between the same decision-makers over a long and indefinite horizon, (b) repeated interactions with a clear end date, or (c) one-shot interactions. Case (a) can be mimicked in the lab by having the same subjects interact repeatedly with the last period of the experiment unknown to the subjects[20]; case (b) is straightforward to implement; case (c) can be implemented either by having a single period, or by randomly re-matching subjects between multiple periods (i.e. a "strangers"

[18] However, it is quite possible that people bring their own (uncontrolled) personal experience or habits with them to the experiment, so that a neutral context may not achieved the desired effect.

[19] However, note that Palacios-Huertas and Volij (2009) find that professional soccer players (at least in the aggregate) play a mixed strategy that corresponds remarkably closely to the equilibrium predictions in the O'Neill game, where there are four choices for each of two players, but there is no pure-strategy equilibrium.

[20] Specific approaches include rolling of a die after each period with some continuation rule or having a pre-set ending that is not divulged to the participants.

design). If the experimenter is interested in one-shot interactions, having only one period is in a sense the cleanest design, but it is also the most expensive approach to gathering experimental data. It can also be problematic if agents need some experience to actually understand the game they are playing. This leads most investigators interested in one-shot interactions (where the predictions of theory are usually the sharpest) to implement multiple periods with re-matching. Interestingly, even though 'standard' game theory predicts no repeated-game effects under these conditions[21], behavior sometimes resembles the predictions of one-shot models more closely in the last few rounds.

"Partners" designs where agents are matched for the duration of the session are, of course, expected to yield repeated-game effects; the *predicted* effects of repeated interaction typically differ dramatically between finitely- and infinitely-repeated games (with the "folk theorem" applying to the latter case). In the latter case, experiments are less useful in testing theory than in providing some idea of what tends to happen when 'standard' theory has little predictive power. A related design question is whether people are always in the same role or whether this can change from period to period. There is disagreement concerning which approach facilitates learning, but role change permits the experimenter to compare an individual's behavior across the various roles.

This leads us to the question of whether to use a 'within-subjects' design or a 'between-subjects' design. Labor economists' experience with field data where there is typically substantial nonrandom heterogeneity disposes them towards research designs with subject fixed effects; in the laboratory, this requires administering both the treatment(s) and the control situation to the same subjects. Experimenters are accustomed to having (both observable and unobserved) heterogeneity handled by randomization, but are highly sensitive to framing and sequencing effects. Thus one's behavior under one condition may be influenced by his exposure to other conditions (something that is usually ruled out by assumption in fixed-effects econometric models). Thus, many experimenters tend to prefer 'between-subjects' designs, where each subject is exposed to one and only one treatment.

Labor economists need to be aware of this motivation for the between-subject approach. In some ways, a between-subjects design is cleaner and avoids sequencing effects (although sometimes these are a main topic of interest),[22] but it is typically less powerful and costlier to implement; on the other hand, a within-subjects approach tends to be more powerful in statistical tests, but can lead to spurious correlations. An advantage of a within-subjects design is that one can control for individual differences by letting each person serve as their own control. Some experiments combine both. One way to incorporate both approaches is (a) to vary the order of treatments in a within-

[21] A detail that is sometimes overlooked in designs with re-matching is that, in practice, many re-matching schemes incorporate a positive probability that an agent will be re-matched with the same person at a later point in the session. This can be important in calculating expected equilibrium behavior. Sometimes investigators choose to eliminate this possibility (thus reducing the number of rounds of data that can be collected), or to minimize it by informing subjects that they will never be matched with the same partner in the next period.

[22] One can systematically vary the sequence to test for sequencing effects.

subject design, then (b) use only the cross-sectional data from the first treatment as a between-subjects experiment to test the robustness of the within-subject approach.

A truly crucial issue in experimental design is the calibration of the parameters, as results can be very sensitive to parameter values and functional forms. Calibration typically involves establishing a baseline for comparisons. There is little 'science' to guide one to choosing parameter values; instead, this is an art that is informed by the experimenter's experience. However, one tip is to find a calibration for which the baseline treatment's results leave room to move in either direction (a calibration that leads to a very low or very high rate in the baseline permits movement in only one direction). The researcher must also consider how to justify the choice of parameter values and functional forms.

The choice of payoff method is intimately connected to the issue of calibration. Incentives should be large enough to induce thoughtful and motivated behavior by the participants. An additional consideration is whether to pay for each period or to pay the participant for only one (or several) periods randomly-chosen at the end of the session. It is more traditional to pay for each period, but there is a definite trend towards paying for only some random subset of all periods. The latter approach avoids wealth effects (the amount already earned in a session: participants often have some form of income targets), mitigates boredom in later rounds, and avoids issues of people taking chances because they know that they have negative earnings (bankruptcy) at some point in the session and that negative earnings are uncollectible.

Finally, while it is traditional to tell a responder the choice of the paired first mover before the response, a more economical approach involves contingent payments. In this 'strategy method' (Selten 1967), the responder states an action at each and every information set. This permits the researcher to obtain an observation at every node of the game, which is particularly valuable when a node is reached rather infrequently. However, while the quantity of data is maximized, there remains the issue of the quality of the data. The strategy method is quite popular, but remains controversial. The most exhaustive study to date (Brandts and Charness 2009) examines many comparisons of results with the two methods, finding that there is generally no qualitative difference; we are unaware of any experiment in which a treatment effect is found using the strategy method that vanishes when the game is played through ('direct response'). In any case, this is an arrow in the experimentalist's quiver and is something to consider.

We close this section with a list of "fatal errors" mentioned on p. 14 of Holt (2007):

1. Inadequate or inappropriate incentives
2. Non-standardized instructions and procedures
3. Inappropriate context
4. Uncontrolled effects of psychological biases
5. An insufficient number of independent observations
6. Loss of control due to deception or biased terminology
7. The failure to provide a calibrated baseline treatment.

8. The change in more than one design factor at a time.

Needless to say, one should endeavor to avoid these pitfalls.

2. Design Questions in principal-agent/effort experiments:

A first question in these experiments is whether there will be a market for contracts. Most experiments simply start with firm-worker pairs that can realize some rents if they make an exchange and have a fixed outside option if they do not exchange. For many questions this is perfectly fine. But this leaves no room for labor markets, which can affect and be affected by the nature of principal-agent interactions. Early principal-agent experiments (e.g. Fehr, Kirchsteiger, and Riedl (FKR) 1993) incorporated an *ex ante* market for labor contracts, and showed that fairness considerations in the principal-agent interactions caused that market to fail to clear. More recently, Charness, Kuhn and Villeval (2009) have shown how causation can run the other way: introducing *ex post* labor markets can eliminate the well-known ratchet effect in the repeated principal-agent problem.

If one chooses to implement a market in the laboratory, how can this be achieved? One approach involves a form of an auction, as in FKR, who set up a two-stage game in their experiment. The first stage was a one-sided oral auction in which firms made wage proposals, but could not choose any individual worker, as every worker could accept every offer. If a worker accepted an offered wage, a binding contract ensued; people who were not paired at the end of three minutes received zero profits for this period. In the second stage, workers chose effort anonymously (only the paired firm learned the chosen effort). Other, simpler approaches to modelling agents' outside options include simply manipulating the agent's compensation if he/she chooses not to work for the principal to whom he/she has been assigned, or allowing an agent to receive simultaneous offers from more than one principal (see for example Charness, Kuhn and Villeval 2009).

As mentioned earlier, some labor experiments use some form of real effort, while others use a stated effort level that is simply a transfer (at some rate of exchange) from the agent to the principal. An advantage of stated effort is that we know the disutility-of-effort function and can therefore calculate exactly what the equilibrium effort levels should be, according to different theories. This approach also allows the investigator to induce, and manipulate, differences in ability/cost of effort, separately from other personal characteristics (e.g. risk aversion, competitiveness, reciprocity) that might be correlated with it in a sample of persons. Of course, the advantage of using real effort is that the task is more in line with what most people consider labor, and so might be considered to be a better match to the field environment.[23]

[23] A related issue is the distinction between hours of work and effort. While 'traditional' labor supply theory is framed as a worker's choice of hours at a fixed hourly wage, principal-agent theory and virtually all labor experiments are framed as a choice of effort. While in many ways the choices are isomorphic, there are some important distinctions; caution is required when generalizing 'effort' results to choices of hours worked. See for example Dickinson (1999).

A possible tactic for gathering more data is to place all participants in the role of agents, with the agent's compensation scheme manipulated by the experimenter. If the researcher's only interest is in the response of agents to different compensation schemes, one might argue that this is the simplest and and most economical design: all subjects are agents, essentially working 'for' the experimenter by performing either a real task or selecting a level of 'chosen effort'.[24] An alternative design assigns some subjects to the role of principals, who choose compensation policies to which agent-subjects respond. Some arguments in favor of the former approach are that (a) subjects might be more disposed to treat the experimenter (as opposed to a fellow student in the lab) like a 'real' employer, and (b) the behavior of college students acting as firms provides little insight into the behavior of 'real employers'. Also, when participants choose the pay scheme, it is *not* randomly assigned. On the other hand, the latter approach (with subjects as principals) may have advantages if one is interested in social preferences towards persons other than the experimenter, or in the behavior of principals *per se* (for example if the subjects are experienced managers).

Additional considerations in the design of principal-agent experiments include whether workers can self-select among reward schemes, such as a tournament or a piece-rate scheme (Niederle and Vesterlund 2007); whether the experimenter induces reference points (Abeler et al. 2009); allows communication (Charness and Dufwenberg 2006, Brandts and Cooper 2007); allows for a monitoring/fines technology (Fehr, Klein and Schmidt 2009); or for some coercion of agents (Falk and Kosfeld 2006). In the case of multiple agents per principal, the experimenter needs to decide whether agents can observe each other's actions (Falk and Ichino 2006) or wages (Charness and Kuhn 2007), whether pay is based on relative performance, or whether subjects interact in teams. All of these, and related questions, constitute the fabric of an extensive research agenda on principal-agent interactions in the lab, which we review in detail in Sections III and IV below.

3. Reading papers involving laboratory experiments

Labor economists may be at a loss in reading papers that report the results of laboratory experiments. The format may well be unfamiliar, the design mysterious, and the statistical methods foreign. In addition, many experimental papers seem written for experimental audiences, rather than the general population of economists. Nevertheless, there are some pointers that can be provided for labor economists interested in gleaning the substance and details of experimental papers.

Perhaps the most important factor in reading an experimental paper is to understand the experimental design. This is not always as clear as it should be in the text; often, experimental referees first read the experimental instructions. These should be consulted if there is any doubt concerning the exact procedures. It is critical for the reader to understand the flow of information; this means knowing what the participants

[24] Another approach is to use automated agents when there is really only one sensible response. See for example Charness, Kuhn, and Villeval (2009).

knew and when they knew it, in terms of the stages of the experimental game or task. When the design is complex, there is also a concern that participants may not have understood the game or task involved.

Once the instructions are understood, the reader should consider how well the experimental design constitutes an appropriate test of theory or matches the 'ideal' field environment of interest. While one should be fairly insistent that the details of the design deal correctly with the issues involved with a test of theory, one cannot expect the experimental design to precisely match the field environment. Still, it is important that the reader is persuaded of the relevance of the experiment to the field or to the theoretical environment. The reader should also be alert to the issue of framing, given the substantial possibility that this affects behavior. To a certain degree, framing effects may wash out when one compares across treatments, but this can be a delicate issue.

An important issue when reading an experimental paper is the presentation and analysis of the data. If one is concerned with "where the bodies buried" some degree of caution may be sensible. Authors have been known to put the best face on the data (for example, empirical researchers may tend to report the more useful regressions), so readers should keep this in mind. For example, sometimes articles emphasize (or only report) data from a subset of the periods; at times this can be justified and at times it is convenient. Sometimes authors will pool data from treatments; this increases the number of observations and makes statistical tests more powerful, but this pooling must be justified. In general, it has happened that experimental papers (and others) have interpreted their results in a favorable light. One should consider whether these interpretations are justified and whether there are alternative interpretations.

Regarding the issue of statistical and econometric tests, since most labor experiments in the laboratory feature multiple periods and interaction amongst the participants, one must have some approach towards determining how to treat multiple observations for the same individual. Labor economists are very familiar and comfortable with panel-data techniques, but experimenters are less so. Some feel that each session can only be considered to present one independent observation. A less strict approach is to collapse each individual's choices to an average, eliminating the issue of multiple observations (but not eliminating the issue of interactions during the session). In either of these cases, it is common for experimenters to report non-parametric tests, and sometimes no regressions are reported; labor economists may be unfamiliar with these tests. One workhorse is the Wilcoxon rank-sum test, which ranks the behavior of individual participants of individual sessions in each treatment and then compares the sums. When within-subject data are available, the binomial test is often used; here one can compare changes for each individual across tasks. If these changes go predominantly in one direction or the other, one can conclude statistically that the behavior is significant. The reader should understand how these tests (and the ones reported in the article) work.

III: Testing 'Traditional' Principal-Agent Theory in the Lab

The question of how workers' choices of effort and work hours respond to financial incentives is among the oldest questions in labor economics. In this section we consider how these questions have been addressed in the laboratory; our treatment roughly follows the literature on principal-agent models and in personnel economics by beginning with the simplest forms of work incentives (a wage per hour worked or an individual piece rate), moving on to incentives based on relative performance (tournaments), incentives for teams, multitask settings, and multi-period principal-agent settings. Not only does 'traditional' principal-agent theory serve as a useful organizing device for our discussion, many of its predictions are confirmed in the lab.[25]

That said, the experiments summarized in this section yield a number of robust results that are inconsistent with standard principal-agent models, including for example a strong apparent tendency by workers to 'reciprocate' generous wage offers from firms, even when such reciprocal behavior is costly to workers. In Section IV we focus specifically on the use of experiments and the development of new theoretical models of social preferences to understand these 'anomalies', with the ultimate goal of developing a more general class of models that is more firmly grounded in empirical fact and might be dubbed behavioral principal-agent theory.

1. The Basic Principal-Agent Problem: One Principal, One Agent, One Task, and One Interaction.

a) Animal Labor Supply Experiments

To the best of our knowledge, the earliest economic studies of the effects of material incentives on labor supply in the laboratory were the animal experiments of the early 1980s (Battalio, Green and Kagel 1981; Battalio and Kagel 1985).[26] Much of this work is summarized in Kagel, Battalio and Green 1995; see also Kagel 1987 for a general discussion of the contribution of animal experiments to economics. A key objective of these studies was to test the classic, static economic model of labor supply in which an agent chooses consumption (C) and leisure (L) to maximize a quasiconcave utility function $U(C,L)$, subject to the constraint $C = wL + G$ where w is the wage rate and G is unearned income. In these experiments, hungry animals expend real effort (key pecking for pigeons, lever presses for rats) to obtain income; the experimenters then vary both parameters of the budget constraint (w and G) exogenously and study the animals' reactions.

---

[25] For an alternative and complementary review of some of these issues, see Camerer and Weber, forthcoming).

[26] Battalio et al.'s work, in turn, has roots in an extensive literature in experimental psychology on the effects of reward structures and amounts on animal behavior. See for example Ferster and Skinner (1957), Barofsky and Hurwicz (1968) and Kelsey and Allison (1976). These studies commonly find backward-bending labor supply curves as the reinforcement rate is increased, but do not relate their results to a theory of utility-maximization, or consider the effects of income-compensated changes in 'wages'.

The key prediction tested by the authors is the labor supply response to an income-compensated wage decrease. As predicted by the standard model, both pigeons and rats reduce their labor supply and consumption (Battalio, Green and Kagel 1981; Battalio and Kagel 1985). The authors also study the pure income effects of declines in nonlabor income (G): in virtually all cases these raised labor supply, indicating that leisure is a normal good. The normality of leisure means that it is common to observe backward-bending labor supply curves among animal workers (see for example Table 3 in Battalio, Green and Kagel 1981).[27]

Another interesting feature of the animal studies that generalizes to the plethora of human studies is the presence of large subject effects: while most subjects respond to changes in incentives in the direction predicted by simple utility maximizing models, both the level of effort at any given reward *and* its responsiveness to incentives vary widely across subjects.

b) Piece Rates and Effort

To the best of our knowledge, the first laboratory experiment to examine labor supply responses to wage changes among humans that is couched in economic theory appeared in an accounting journal (Swenson 1988). Swenson's subjects supplied 'real' effort (repeatedly typing "!" then "enter" on a computer keyboard—this requires two hands and does not allow for continuous cursor movement).[28] Wages per character typed were fixed, but 'taxed' (this language was used in the subjects' instructions) at rates ranging from 12 to 87 percent. Total tax proceeds from the previous session were randomly distributed to the subjects in the following period, mimicking a balanced government budget but breaking most of the connection between current individual effort and future lump-sum income. The primary questions addressed were how labor supply and total tax revenues respond to the tax rate. Both curves were backward-bending, with tax revenues (i.e. the Laffer curve) peaking at the 73 percent tax rate.[29]

A decade later, economists Sillamaa (1999a, 1999b) and Dickinson (1999) conducted similar real-effort experiments.[30] Like Swenson's, Sillamaa's experiments

---

[27] Other labor economics questions addressed by these researchers in the lab include the "welfare trap" (do high levels of unearned income in past treatments reduce current labor supply?), and the "cycle of poverty": does deprivation raise animal subjects' discount rates (i.e. their preference for a small immediate reward over a larger delayed one), thereby leading to more deprivation in the future? Interestingly, little evidence of either effect was found (Kagel, Battalio, and Green 1995).

[28] Laboratory principal-agent experiments with human subjects can be divided into those where subjects are paid to perform an actual task ('real-effort') and those where effort decisions are represented by the choice of a decision number that imposes increasing marginal financial costs on the agent ('chosen effort'). Bruggen and Strobel (2007) find little difference between the two methods in a simple gift-exchange labor market game.

[29] It may be worth noting that backward-bending labor supply in the context of a laboratory experiment is actually somewhat of a puzzle for humans, since it is extremely unlikely that earnings in such an experiment could generate income effects.

[30] In Sillamaa's experiments, workers decoded numerical codes into letters; in Dickinson's workers repeatedly typed paragraphs, with a penalty for mistakes. Sprinkle (2000) considers the effect of

were motivated by questions about the impact of taxation (in one case, the impact of tax progressivity, in the other the effect of a zero top marginal rate), though in her case taxes were never mentioned in the subjects' instructions. Sillamaa found that (a) work effort responds more (positively) to real wage increases in the presence of an (equivalent) linear than a progressive income tax, and (b) introducing a zero top marginal tax rate also increased effort.

Like Sillamaa, Dickinson (1999) paid his subjects a piece rate, but in some treatments allowed his subjects to choose between two types of leisure: on- versus off-the-job. This modification is noteworthy because it provides one of the few empirical links between the types of work decisions that are usually studied in lab (and field) experiments (effort) and the traditional application of labor supply theory (to hours worked). Specifically, in the baseline ("intensity") treatments, subjects were required to stay for the entire two-hour experimental period; thus any time not working was spent in the lab. In the "combined" treatment, subjects could leave at any time during the experimental period. Consistent with theory and with previous research, subjects increased their output in the baseline treatment, substituting on-the-job leisure for effort when incentives were strengthened. In the combined treatment, many subjects responded to higher wages by working more quickly, but reducing their total work time by leaving the experiment early. This substitution of off-the-job for on-the-job leisure is offered as a possible explanation for why econometric estimates of labor supply elasticities (which use hours worked, not effort as their measure of labor supply) are often close to zero. Dickinson's analysis also points out that care must be taken in relating the results of laboratory labor supply experiments (where workers' *effort* during a fixed work period is the outcome of interest) to econometric studies of labor supply (where *hours worked* is the outcome).

Gneezy and Rustichini (2000) also studied the response of work effort to financial incentives; they conducted real-effort experiments in both the lab and the field, with similar results: the relationship between the piece rate and effort was U-shaped, with low piece rates eliciting less effort than a zero piece rate. (It may be interesting to note that this is exactly the opposite of the backward-bending labor supply curve in the one-period neoclassical model, which yields an inverted U). The authors hypothesize that small levels of financial compensation (explicit incentives) may 'crowd out' workers' intrinsic motivation to perform these tasks.[31] While this explanation may be more relevant to their field experiment (where the workers solicited charitable contributions) than their lab experiment (which had no charitable component) the phenomenon was observed in both settings. Since earlier studies did not, to our knowledge, implement treatments with a zero piece rate, Gneezy and Rustichini's results do not necessarily conflict with those findings, whether on human or animal subjects.

---

strengthening incentives for a considerably more complex task, spanning multiple periods and requiring belief revision and judgment calls. Here, incentives again increase effort, but only after the subjects had been exposed to considerable feedback and experience.

[31] Psychologists provided compelling evidence that monetary incentives can crowd out intrinsic motivation long before Gneezy and Rustichini's paper (see Deci et al 1999 for a meta-analysis). See also Frey and Oberholzer-Gee (1997) for an earlier economic analysis, though not in the domain of labor economics.

In more recent real-effort experiment, Cadsby, Song and Tapon (2009) show that the effect of performance incentives varies with agents' risk aversion. In their experiment, 25% of subjects actually perform worse when incentives are intensified; further the probability of such deterioration increases with risk aversion and with measures of stress. A similar result is obtained by Ariely, Gneezy, Loewenstein and Mazar (2008), who exposed subjects in the U.S. and India to incentives ranging from small to very large (relative to their typical levels of pay). In many cases, very high rewards had a detrimental effect on performance. Combining these results with the nonmonotonicity identified by Gneezy and Rustichini (2000) above suggests that the effect of stronger incentives on performance, predicted to be monotonic by basic labor supply theory (at least when income effects are unimportant, which is expected for laboratory experiments), may in fact be highly non-monotonic.

c) Selection into piece rate compensation

Since Lazear's famous Safelite study (2000) economists have realized that a significant share of the productivity improvements associated with piece rates can take the form of voluntary self-selection of higher-productivity workers into piece rate schemes, rather than changes in the work effort of existing workers. Laboratory studies that allow for self-selection into different pay schemes abound, though many of these focus specifically on selection into tournaments and teams. These studies are discussed later in this section. A recent study that considers the self-selection that occurs when a simple piece rate is introduced is Cadsby, Song and Tapon (2007). As in most studies, pay-for-performance raises productivity. Like Lazear (2000), they also find that sorting enhances this effect: more productive employees are more likely to choose pay-for-performance schemes.[32]

A somewhat different perspective on selection into pay-for-performance schemes is provided by two recent papers by Bandiera, Barankay and Rasul (2007, 2009) who conducted field experiments in a fruit-picking firm. When a pay-for-performance element (based on their unit's output) is added to *managers'* compensation schedules, Bandiera et al (2007) find that managers are more likely to select able workers into the units they manage. In their 2009 paper, the same authors show that this shift towards abler workers came at the expense of workers who were socially connected to the manager. In both papers, the shift away from friends increased the work group's total

[32] Dohmen and Falk (2006) examine laboratory subjects' voluntary self selection into several types of contracts (fixed pay, piece rate, tournament or revenue-sharing scheme) according to subjects' risk attitudes, overconfidence, social preferences, gender and personality. They find that most of the extra output generated by all three variable pay schemes is due to the selection of abler workers into them. Burks, Carpenter and Goette (2009) provide additional evidence on the importance of selection into pay-for-performance by conducting a context-rich prisoner's dilemma lab experiment on bicycle messengers who are employed in three types of firms: firms using pay for performance, those using hourly wages, or workers' cooperatives. Workers in the first type of firms were the least cooperative in prisoners' dilemma games. Using data from cities where bicyle messengers were not able to choose among these firm types, the authors argue, however, that these differences are not due to selection, but due to workers' (rapidly) adopting the norms governing cooperation in their workplace.

output and the manager's compensation. In contrast, Belot and van de Ven (2009) find in a field experiment with children that agents who are selected because they are friends increase their subsequent performance, presumably to reciprocate the favor of being selected. In such cases, favoring one's friends can be costless, or even beneficial to the manager and the firm. To our knowledge, neither the role of managers nor favoritism on selection into pay-for-performance has been studied in the lab.

d) Reciprocal Behavior

Evidence on the apparent presence of reciprocal behavior in workers' effort choices in economics goes back at least to the pure gift-exchange labor markets implemented by Fehr, Kirchsteiger and Riedl (1993). Labor contracts in these settings contained no explicit incentives; despite this, workers supplied costly effort, and supplied more effort the higher the (lump sum) wage the principal paid them. We treat this 'pure' gift exchange literature in another section; here we provide one or two examples of how workers' apparent concerns for reciprocity in the laboratory affect the performance of standard incentive contracts, such as piece rates.[33]

An illustrative paper in this regard is Anderhub, Gächter, and Königstein (2002), who study the behavior of both principals *and* agents where the contract specifies the agent's pay as a linear function of his/her output. (All the papers considered thus far study agents' reactions to reward schedules set by the experimenter.) Because there is no uncertainty, the efficient linear contract has a piece rate of 100%. Further, because principals make take-it-or-leave-it contract offers to agents in this experiment, the intercept term of the equilibrium linear contract is predicted to extract all of the agent's surplus if social preferences are absent. As one might expect, principals and agents behave relatively efficiently with respect to the slope of the contract (principals choose a 100% piece rate 30% of the time and a positive piece rate 98% of the time; agents optimized against this, choosing conditionally rational effort levels 87% of the time). Social preferences, however, clearly affected the both the principals' choice of the intercept and agents' responses to it: agents rejected offers that split the surplus too unevenly, and principals made few such offers. There was also some tendency for generous offers to lead to higher effort levels, though as already noted the vast majority of effort decisions were egoistically rational given the piece rate.

While effects such as those reported above are both dramatic and common in lab experiments, we note that more recently, Gneezy and List (2006) have argued that positive reciprocity effects detected in lab experiments can wear off very quickly in the field; Kube, Puppe and Marechal (2006a, 2006b) in turn generate longer-term effects of reciprocity in the field, especially for negative reciprocity. We discuss these questions further in the section on reciprocity and social preferences. Finally, we note that, in addition to modifying the nature of principal-agent interactions, social preferences may also explain why some principal-agent relationships exist in the first place. For example,

[33] Our section on the role of reciprocity and social preferences in principal-agent interactions also considers the closely-related role of communication, especially its ability to foster trust, cooperation and guilt. See for example Charness and Dufwenberg (2006, 2009).

Hamman, Loewenstein and Weber (2008) report on an experiment in which principals can hire agents to behave selfishly on the principal's behalf. Delegation of decisions that would otherwise make the principal act directly in a selfish manner appears to yield more lucrative outcomes for principals.[34]

e) Reference points

An emerging issue in the study of one-on-one principal-agent relationships is the effect of reference points on effort provision. Part of the inspiration for this is a lively debate in the non-experimental literature on the presence of reference points in daily labor supply decisions by agents (in particular, taxi drivers and bicycle messengers) who can make daily hours and effort decisions (see for example Camerer et al. 1997, Farber 2005, Fehr and Goette 2007, Farber 2008, Crawford and Meng 2008). One advantage of addressing this issue in the lab is that some possible reference points —for example, expected earnings in a round or session-- can not only be observed, but manipulated by the experimenter. This is the approach taken by Abeler et al (2009).

In Abeler et al's experiment, subjects are paid a piece rate to perform a tedious task. At the end of the period, with 50% probability they are paid their accumulated piece rate; otherwise they receive a fixed payment that is known in advance. Subjects decide how much to work before they know whether they will receive the fixed payment or their accumulated piece-rate earnings. Abeler at al find significant bunching of piece-rate earnings at the level of the fixed payment. Further, this spike in the earnings distribution moves when the fixed payment is changed. Neither of these is consistent with the 'standard' effort-leisure choice model (unless one were to introduce fairly unusual forms of non-separability between income and leisure). Instead, the authors argue that their results are consistent with Koszegi and Rabin's (2006) model of reference-dependent preferences, where the reference point is the subject's expected earnings for the experimental session (which is manipulated by the experimenter). The authors take some care to ensure that subjects' choices of 'target' earnings are not driven purely by the salience of those particular numbers in the instructions and experimental environment.[35]

f) Motivational 'Crowding Out'

Another question addressed in the experimental literature on worker-firm interactions is the effect of certain 'coercive' features of contracts, such as minimum effort requirements or employee monitoring, on agents' effort levels (and more broadly, on contract efficiency.). In this regard, Frey (1993) proposed that, especially in environments where the principal and agent know one another personally, the principal's decision to monitor the agent may be interpreted as a signal of distrust, and may reduce effort as a result despite obvious direct 'disciplining' effects of monitoring. Falk and

[34] See also Bartling and Fischbacher (2008) and Coffman (2009).
[35] Another possible reference point in a bilateral relationship is the terms of an ex-ante competitively negotiated contract between the parties. Experimental evidence of these effects is provided in Fehr, Hart and Zehnder (2008), though not in a labor or principal-agent context.

Kosfeld (2006) test a closely related idea in the lab using a very simple game where agents choose effort (which costs them less than it benefits the principal), and principals' *only* decision is whether to impose a minimum effort level on the agent. (This is essentially a gift-exchange game without an initial "gift"—agents' endowments are positive and principals' are zero). If principals' decisions to impose a minimum effort level had no effects on agents' behavior, the truncated distributions of agents' effort (above the imposed minimum) should be the same whether the minimum is imposed or not. This is not the case: Falk and Kosfeld find "hidden costs of control" in the sense that the majority of agents reduce effort when firms attempt to 'control' their actions (though the effort levels of a smaller number of 'opportunistic' agents' efforts were mechanically increased by the effort minimum). In most treatments, these net reductions in effort were so substantial that principals who 'controlled' earned lower payoffs then those who did not. In a follow-up survey, the authors asked agents the free-form question "What do you feel if [the principal] forces you to transfer at least [x] points?" The most common response was 'distrust', especially among agents who reacted negatively to control. The authors' results suggest that, at least in a laboratory environment, rigid attempts to control agents' behavior can 'backfire'; the authors also provide some support for the external validity of their results by administering a survey eliciting students' self-reported 'work motivation' in a variety of hypothetical work situations involving different degrees of employer control or trust.

In a clever variation on Falk and Kosfeld's design, Schnedler and Vadovic (2007) show that control by principals does not elicit negative reactions from agents when the principal's control is legitimized in two alternative ways. In one of these, the principal must set a common control policy that applies not only to the agent, but also to a computerized 'automaton' agent who supplies minimum effort whenever this is allowed. Perhaps not surprisingly, agents 'understand' the principal's decision in this case and do not reduce their effort when controls are imposed. In the other, the principal is given a small endowment (in contrast to zero in Falk-Kosfeld), and agents are allowed to take from this endowment by choosing a very low effort level. Here as well, experimental subjects treat control decisions that simply protect the principal's endowment from agent 'pilfering' as legitimate.

Irlenbusch and Sliwka (2005b) suggest an intriguing explanation of the negative incentive effects of paying for performance in a simple experiment where principals and agents first play a pure gift-exchange game, followed by a game in which principals had the option of offering a piece rate in addition to the fixed payment (gift). Consistent with Gneezy and Rustichini, agents' effort actually fell after a low piece rate was introduced. This is particularly interesting since the task performed by the agents was chosen to yield little or no intrinsic reward. More importantly, effort fell even further when, in a third treatment, piece rates were once again disallowed. A possible explanation is that agents' perception of the implicit contract offered by principals is changed by the introduction of piece rates: the presence of piece rates signals that agents are expected to behave egoistically; while the offer of a fixed wage signals that, as in many real-world employment relationships, a reasonable amount of effort is simply expected in return for a wage. Certainly, Irlenbusch and Sliwka's results suggest that *studies of what appears*

to be intrinsic motivation should pay close attention to subjects' interpretation of the implicit contractual understandings that may be signaled by different pay schemes.

Fehr, Klein and Schmidt (2007) find a similar pattern when they compare the performance of three types of contracts in a simple laboratory experiment. In an "incentive" contract, the principal stipulates a wage w, a required effort level e^*, and a fine f. If the agent accepts the contract, he is 'audited' with exogenous probability p, and is forced to pay the fine if the effort he has chosen falls short of e^*. In a "bonus" contract, the principal announces a wage, a *desired* effort e^* and her (unenforceable) intention to pay a bonus b if $e \geq e^*$. Finally, a "trust" contract is pure gift exchange in which the principal offers a wage and simply requests effort in return).[36] For their parameterization, FKS find that, when principals must choose between trust and incentive contracts, incentive contracts performed better: they yielded higher effort levels, higher payoffs for both principals and agents, and were increasingly selected by principals over the course of the experiment. These results are consistent with findings of Lazear (2000) and others that incentives increase effort. In contrast, however, when principals must choose between incentive contracts and bonus contracts[37], bonus contracts now dominate incentive contracts: they constitute the overwhelming majority of contracts offered, and yield higher levels of effort and payoffs to principals; this result contradicts the predictions of contract theory with egoistic agents. The authors explain these contrasting results by parameterizing Fehr and Schmidt's (1999) model of inequity aversion. Essentially, if the 'fair-minded' share of the population is neither too high nor too low, there are too few fair-minded persons to make trust contracts perform best, and two many fair-minded persons to make incentive contracts work best. That said, the authors recognize that inequity-aversion is not the only possible explanation for their results; indeed, the likelihood that incentive contracts signal distrust (plus the fact that the authors constrain the enforcement technology in these contracts to make a first-best allocation infeasible) may also help explain this pattern of outcomes.

Dickinson and Villeval (2008) also consider the effect of monitoring (the key element of FKR's "incentive contract") on work effort; their setting is a real effort laboratory experiment, where the task was designed to contain an element of intrinsic motivation.[38] Principals choose monitoring intensity, which raises the probability the agent is audited (and penalized via a 'fine' paid to the principal if his output did not exceed the target). Dickinson and Villeval vary two main aspects of the environment: in the "variable" treatment, the principal's profit, as usual, depends directly on the effort chosen by the agent. In the 'fixed' treatment it does not. The other aspect that is varied is the degree of anonymity; interaction is either anonymous or preceded by five minutes of face-to-face interaction. Dickinson and Villeval find that monitoring raises agents' effort in the anonymous setting, as predicted in the standard agency model. Motivational crowding-out is observed *only* when interactions are not anonymous *and* when the

[36] Note that the agent's actual effort, e, is observed by the principal *ex post* in all contracts, whether the agent is 'audited' or not; auditing simply means that a punishment can be enforced if effort is short of the requirement.

[37] Note that the class of bonus contracts includes the 'trust' contract as a special case, with bonus of zero.

[38] Agents try to attain a maximum value of a non-decreasing function by costly sequential search.

principal's payoff depends directly on the agent's effort (their 'variable' treatment). This suggests that the *motivational 'crowding out' by monitoring is not driven primarily by a reduction in the intrinsic rewards derived from the task, but from a form of negative reciprocity (punishing the principal for a lack of trust).*[39]

g) Nonlinearities: Targets, Fines and Bonuses

While most of the experimental literature on piece rates considers simple linear reward schedules, and while linear contracts are theoretically sufficient to achieve efficiency when agents are risk neutral, it is interesting to consider the effects of nonlinearities in individual piece rate contracts, since these do occur in the real world.[40] The one experimental paper we know of that focuses on this topic is Cadsby, Song and Tapon (2008). The distinguishing feature of their experiment is that the task allowed agents to misrepresent their own performance (the number of words created in an anagram game). While actual output was similar under target-based pay schemes versus a continuous (linear) reward scheme, the former produced significantly more cheating. Further, cheating is more likely under a target-based scheme the closer a participant's actual production is to the target. Since the agent's rewards to cheating are also greatest in these situations, Cadsby et al.'s results are both consistent with theory and indicative of a possible drawback with sharp discontinuities in reward schedules.

h) Peer Effects and Wage Comparisons

Although 'peer effects' on effort can only exist when a firm employs multiple workers, we consider peer effects in this section because 'pure' peer effects refer to a situation where workers work, side by side, for the same firm but do not interact in any way (except that they observe each others' work activity). For example, suppose that two workers are each paid an individual piece rate, and there are no substitutabilities or complementarities in production, but can observe each other's effort or output. Does anything change? Perhaps surprisingly, it does. In a real-effort experiment, Falk and Ichino (2006) find that average output is higher. Further, the standard deviation of output is lower within worker pairs than between pairs. Essentially, low-productivity workers raise their output towards that of their co-worker when a co-worker is present. One can imagine a number of possible explanations for this behavior, including subject uncertainty about the 'true' compensation schedule. Similar results were found in a field experiment by Bandiera, Barankay and Rasul (2009), but *only* when the co-workers were friends. Specifically, in a situation where workers received individual piece rates and no appreciable production externalities existed, workers who were less able than a co-worker with whom they are friends increased their effort (and hence income) by 10 percent. In

[39] Ellingsen and Johannesson (2008) provide a theoretical model of crowding out in which some audiences are more worth impressing than others. In their model, the principal's choice of monetary incentives signals that he/she is not work impressing. See Bénabou and Tirole (2003, 2006) for closely related models.

[40] Of course, the voluntary bonuses studied by Fehr, Klein and Schmidt (2007a) are nonlinear *ex post*, but here we focus on enforceable reward schedules, announced *ex ante*, that contain discrete jumps. Incidentally, FKS 2007b adds an enforceable fine to FKS 2007a, but finds it is rarely selected by principals and has little effect on agents' effort.

contrast to Falk and Ichino, however, Bandiera et al. also found that workers who are more able than their co-worker friends reduce their effort and forgo 10% of their earnings.[41]

A final, related question is how workers' effort changes when they can see each others' *wages*. Charness and Kuhn (2007) pose this question in a pure gift-exchange game in which workers knew that their productivity was different from their co-worker's, but did not know the size or direction of this difference. If between-worker equity concerns are important determinants of effort, we might expect that low-productivity workers (who tend to receive lower wages) would reduce their effort in treatments where they observe their co-worker's wage than when they do not (indeed this is suggested by Akerlof and Yellen 1990). Perhaps surprisingly, they do not. Agents' primary concern seems to be to reciprocate generous wage offers from the firm; the authors speculate that responding to wage offers made to their co-workers would likely muddy this 'signal' and is therefore avoided by workers.[42]

In sum, laboratory tests of the "single task/single agent" principal-agent model have identified the following broad empirical regularities:

a) Compensated wage cuts reduce effort in animal labor supply studies. In addition, leisure is normal, and uncompensated wage changes generate backward-bending labor supply curves (Battalio, Green and Kagel 1981; Battalio and Kagel 1985). For both animals and humans, there are large individual subject effects, both in the level of effort supplied for a given level of incentives, and in the *responsiveness* of effort to incentives.

b) For humans, higher piece rates raise effort (Swenson 1988, Sillamaa 1999a,b, Dickinson 1999). Agent self-selection into pay-for-performance schemes reinforces these effects (Lazear 2000; Dohmen and Falk 2006, Cadsby et al. 2007).

c) Not paying at all can yield higher effort than low pay (Gneezy and Rustichini 2000).

d) Effort also responds to the *intercept* of the worker's compensation schedule, at least when generosity is seen as intentional (Fehr, Kirchsteiger and Riedl 1993; Charness 2004).

e) Effort decisions can be affected by at least one type of reference point that can be manipulated in the lab: the subject's expected earnings for the session (Abeler et al, 2009).

---

[41] Mas and Moretti (2009) also detect small effects on productivity of being observed by a more productive worker, in a setting where workers were not paid for performance.

[42] Maximiano, Sloof, and Sonnemans (2006) also examine gift exchange in multi-worker firms, but in their treatment all workers are equally productive and receive the same wage by design. Güth et al (2001) consider a situation in which the principal can offer a menu of contracts to two independent agents with different productivities. As in Charness and Kuhn, public observability of the co-worker's contract induces the employer to compress compensation schemes.

f) Forcible restrictions on agents' choice sets can reduce the efforts of agents on whom they are not binding (Falk and Kosfeld 2006), but not when the restrictions are seen as 'legitimate' (Schnedler and Vadovic 2007).

g) A decision by a principal to use piece rates can also reduce agents' efforts; a likely explanation is that the introduction of piece rates changes the agents' interpretation of the implicit contract for labor services (Irlenbusch and Sliwka 2005).

h) A principal's decision to monitor the agent can also reduce agents' efforts, but only when the agents 'know' the principal (Dickinson and Villeval 2008).

i) Unenforceable promises by principals to pay bonuses for 'satisfactory' worker performance can elicit surprising amounts of effort, and can outperform more objective mechanisms such as random monitoring combined with punishment (Fehr, Klein and Schmidt 2007).

j) Sharp discontinuities in reward schedules induce workers to misrepresent their output (Cadsby et al., 2008).

k) Even when there is no strategic independence between workers, workers' efforts may depend on their co-workers' efforts (Falk and Ichino 2006). Co-workers' *wages* do not appear to affect effort, at least in our earlier work (Charness and Kuhn 2007).

2. Tournaments

a) Theory

Consider now a situation where a firm employs multiple workers, who still do not interact in production. However, because the firm bases rewards, at least in part, on agents' performance relative to each other, workers' effort decisions, pay levels, and utilities are interdependent. Relative performance plays a key role in a number of features of real-world compensation schemes, including promotions and bonuses for top-performing workers. Since at least Lazear and Rosen's seminal 1981 paper, economists have understood that, if workers are risk neutral, pay structures which award prizes based only on workers' relative performance can generate identical allocations in Nash equilibrium as would be achieved by optimal individual piece rates. This may be useful if rank order is easier to measure than cardinal performance, or if rewards are inherently indivisible (some promotions might be an example). Further, tournaments can be more efficient than piece rates if workers are risk averse and if their outputs are affected by a common shock. These results are robust to ability differences between workers if ability is public information: in that case, efficient tournaments typically include handicaps for the abler agents. As Carmichael (1983), among others, has pointed out, payment by relative performance also mitigates an incentive problem affecting the principal, namely the incentive to understate workers' true outputs, or to provide suboptimal levels of complementary inputs after the contract has been signed. Finally, as O'Keefe, Viscusi

and Zeckhauser (1984) have pointed out, contests may have an efficiency advantage if workers derive direct utility from competition itself.

Counterbalancing the above advantages, tournaments may be less efficient than individual piece rates when workers' abilities are hidden information, especially if workers can self-select into tournaments (Lazear and Rosen, 1981). Also, in contrast to individual piece rates, tournament games in general require agents to think strategically about their co-workers' effort levels to find a Nash equilibrium; this may make them less robust as incentive schemes. It is also worth noting that, in general, optimal contest design will be different if (part of) the contest's objective is not simply to induce effort, but to identify the most talented contestant (e.g. for promotion). In this section we examine how tournament reward schemes work, not just theoretically, but when the games are played by human subjects in what has become a sizable experimental literature.

b) Early Experiments

To our knowledge the first laboratory experiment on tournament-based incentives was by Bull, Schotter and Weigelt (1987). Bull et al. implemented tournaments between pairs of experimental subjects whose output was subject to independent, uniformly distributed productivity shocks (this guarantees a unique, pure Nash strategy equilibrium if the spread of the distribution is high enough), parameterized to yield identical effort levels to a simple piece rate. As predicted, average effort levels were similar between the tournament and the piece rate, but effort variance across subject pairs was much greater under the tournament.[43] This suggests that, while equivalent in principle, tournaments may be a much less robust incentive scheme than piece rates in practice because agents have difficulty finding a Nash equilibrium in strategies. Bull et al also studied tournaments between players with different abilities, and found that less-able agents systematically exerted more effort than the Nash equilibrium.

Schotter and Weigelt (1992) implement a very similar laboratory protocol but focus in more detail on 'uneven' tournaments (where the participants' abilities differ), as well as on 'unfair' tournaments (where the rules favor one identical agent over another). They are also interested in the effects of policies that (a) restore fairness in unfair tournaments (termed 'equal opportunity laws'), or (b) give handicaps to less-able agents in uneven tournaments (termed 'affirmative action laws'). Schotter and Weigelt find (a) again, that mean effort levels in fair, symmetric tournaments match the Nash equilibrium; (b) disadvantaged contestants in unfair tournaments supply more than Nash equilibrium effort; (c) mean effort in uneven tournaments matched theoretical predictions, though largely because the less-able agents either worked too much or chose zero effort, neither of which was Nash behavior; (d) symmetrizing previously-unfair tournaments raised both agents' effort levels; and (e) handicapping abler contestants raised total worker output and the principal's profit when the ability difference between contestants was large. It did so largely by eliminating drop-out behavior among the less able contestants. The

[43] This variance result appears quite robust. See, for example, van Dijk, Sonnemans and van Winden (2001) for a replication.

authors suggest that these results might have some relevance to the effects of affirmative-action programs in the real world.

c) Selection into tournaments:

The early theoretical literature on tournaments considered the question of selection into tournaments largely from an adverse-selection perspective. For example, Lazear and Rosen (1981) predicted that, if workers have private information about their own ability, less-able workers will 'contaminate' tournaments designed for abler workers. Whether or not firms modify the structure of their tournaments to address this adverse selection problem, the resulting effort allocations will no longer be as efficient as individual piece rates.

Experimental studies of self-selection into tournaments tend to focus on different questions. For example, building on earlier market-entry experiments in Industrial Organization (e.g. Rapaport 1995), which tend to find surprising levels of co-ordination on the efficient outcome in entry decisions despite the absence of communication among subjects, Camerer and Lovallo (1989) designed a game in which MBA students chose whether to enter a 'market' where their success depended on performance relative to other entrants, and on the subjects' own skill level (on a sample of logic puzzles or trivia questions about sports or current events). Camerer and Lovallo found excessive entry, which appear to stem not from inaccurate forecasts of the number of entrants or other factors, but from the subjects' substantial overestimates of their own ability. Vandegrift et al. (2007) conduct a similar experiment, where they allow a fixed population of agents to choose whether to be paid an individual piece rate or to enter a pool where they receive a prize of fixed value for the best performance. The task performed by workers (forecasting the price of a fictitious stock based on cues that correlated with the true price) is deliberately chosen to allow for 'winning' to have some intrinsic or signaling value. These authors do not detect significant levels of excess entry. In this same vein, Niederle and Vesterlund (2007) also allow subjects to self-select between a piece rate and a tournament; their main interest is gender differences. They find that, at given levels of ability, men exhibit significantly more overconfidence in their tournament entry decisions than women. We discuss this and related articles in more depth in the section on discrimination.

More recently, Eriksson, Teyssier and Villeval (2009) replicate Bull et al.'s (1987) original experimental design as closely as possible, with the exception that subjects are allowed to choose between a tournament (where they are randomly matched with another player) and an individual piece rate, calibrated to yield the same levels of optimal effort and expected utility. Thus, effort is a decision number and all agents are equally able. Eriksson et al., however, elicit subjects' risk aversion after the experiment. They find that risk-averse subjects are less likely to enter the tournament. This has the additional effect of reducing the high between-subject variance of tournaments, which was cited by Bull et al. as a possible disadvantage of tournaments. Mean effort was about one third higher in the tournament scheme; as in Lazear (2000) half of this was due

not to incentive effects but due to selection. This is of particular interest here, since agents' abilities in this experiment were equal by construction.

d) Tournaments and Risk-Taking:

Another prediction of tournament theory is that, in a tournament setting, agents have incentives to increase the spread of their output distribution, i.e. that increase risk. To our knowledge, Bronars (1986) was the first to discuss risk taking as a choice variable in tournaments; he argued that leading agents in sequential tournaments prefer a low risk strategy (to 'lock in' their gains, whereas their opponents choose higher risk. Hvide (2002) shows that, even in a one-stage tournament, if there are no limits on risk taking, tournaments 'collapse' in the sense that, for any given prize spread, agents choose infinite risk and zero effort in the Nash equilibrium. When there are limits on risk-taking, the Nash equilibrium is at the maximum level of risk; furthermore an exogenous reduction in the maximum permissible risk level raises both effort and welfare. Finally, Hvide shows that contests in which agents' performance is ranked according to its absolute distance from a target level, k (thus outputs in excess of k are *punished*) have superior efficiency properties in this environment.

Agents' tournament-induced preferences for risk have been studied in various field contexts, including stock car racing --Becker and Huselid (1992) show that drivers take more risks when the prize spread is large--, and investment fund managers (Brown et al, 1996); Chevalier and Ellison (1997). Brown et al. find, as predicted by Bronars, that expected losers prefer high risks while expected winners prefer low risks. In the lab, Vandegrift and Brown (2003) find that high-variance strategies are indeed attractive in tournaments, but primarily to agents with low capabilities performing a simple task. Nieken and Sliwka (2008) extend the theory of agent risk selection in tournaments to cases where the agents face correlated risks. They argue that --in contrast to Bronars' prediction-- leading agents, rather than 'playing it safe', may be forced to imitate their opponent's risky strategy. This prediction is confirmed experimentally. A possible application is to the case of mutual fund managers investing in the same, or similar, risky assets.

e) Sabotage

Since Lazear's important article (1989), economists have recognized that any compensation system based on relative worker performance rewards workers who take actions that reduce the measured performance of their peers, i.e. to engage in sabotage. Of course, sabotage is inherently difficult to study in the field, because workers may go to considerable lengths to conceal their acts of sabotage. This consideration has led a number of authors to study sabotage in the laboratory.

The first published laboratory experiment on sabotage appears to be Harbring and Irlenbusch (2005) who investigate sabotage in both a baseline treatment where the prizes are exogenously manipulated by the experimenter, and in a setting where principals in the experiment can choose the prize structure. Four agents compete against each other with the top two receiving a 'winner' prize. Investing in sabotage reduces the output of all the

other agents. When the prize spread is exogenously set by the experimenter, Harbring and Irlenbusch find, as predicted in simple tournament models, higher prize spreads encourage both greater effort and more sabotage; interestingly the latter effect dominates, suggesting that pay compression may be an optimal strategy. This finding does not, however, generalize to the case where prize spreads are selected by participants in the experiment, perhaps because in the authors' design higher spreads imply a higher expected prize. Now, agents appear to reciprocate more generous compensation packages by refraining from sabotage. While it is not clear if this result would persist in designs that held the expected prize constant, the result does remind us that agents' intentions to reward or punish the *principal* may also play a role in real-world sabotage decisions.[44]

In a 2007 article, Harbring, Irlenbusch, Kräkel, and Selten consider sabotage in a contest where players are heterogeneous in ability. Three contestants play a two-stage "Tullock" contest where each agent selects *targeted* levels of sabotage aimed at each of the two other players. (Sabotaging other players makes it more costly for them to exert effort.) All sabotage levels are then revealed, and effort choices made in the second stage. In addition, players can be of two types—those with *ex ante* high effort costs ("favorites") or *ex ante* low effort costs ("underdogs"); this is publicly known in advance. Three treatments are implemented: one with homogeneous contestants, one with two underdogs and one favorite, and one with one underdog and two favorites. Contrary to expectations, when there are two underdogs, they do not 'conspire' against favorites by directing their sabotage against the favorite. In line with expectations, an underdog engages in less sabotage when she is playing against two favorites than one; this resembles the dropout behavior of less-skilled agents in Schotter and Weigelt (1992). Harbring et al also examine some treatments in which a saboteur's identity is revealed to the other contestants; they find that retaliation occurs in future rounds and that overall sabotage is less common.

Harbring and Irlenbusch (2008) implement tournaments with two, four, or eight contestants, and with the share of contestants who receive the winning prize equal to 1/4, 1/2, or 3/4. Agents choose both their own effort and a level of sabotage that affects all other agents equally. Parameters are chosen to yield identical Nash equibrium effort levels in all treatments. Consistent with Nash behavior, neither tournament size nor the share of winning prizes has strong effects on effort, or on sabotage, though there is some tendency for effort to be higher when the share of winning prizes is 1/2, compared to 1/4 or 3/4. In a more recent paper (2009) the same authors introduce communication among the principal and agents. This turns out to curb sabotage via agreements on flat prize structures and increased output.

To our knowledge, the only paper to study sabotage in a real-effort experiment is Carpenter et al (forthcoming). Sabotage in Carpenter et al.'s context takes the form of peers' subjective evaluation of the quality of each others' output. (The task is printing

[44] Falk, Fehr and Huffman (2008) also focus on the behavior of principals in a setting where agents can invest in sabotage. They find that both sabotage *and* loss-aversion among agents compromise the ability of large prize differentials to increase effort levels; principals respond to this by choosing wage compression.

letters, placing them into envelopes, and handwriting addresses on envelopes. Subjects also had an opportunity to miscount the number of envelopes produced by their co-workers). Carpenter et al. found that subjects responded to the possibility of sabotage by their co-workers by producing less output than when peer review was not possible. Piece rate compensation performed much better than the tournament with peer-review.

f) Collusion

In all tournaments, agents have an incentive to collude against the principal: if they can agree to both exert zero effort, the prize in any fair tournament will be randomly assigned to one of them, and they will both be better off than if they had truly 'tried' to win the prize. Despite this feature, collusion rarely appears to occur in tournament experiments.[45] This may be because various features of the design, including anonymity and re-matching, are deliberately chosen to make collusion difficult. To our knowledge, Harbring and Irlenbusch (2003) is the only laboratory experiment that addresses the issue of collusion. They find (as one might expect) that the smallest (two-person) tournaments are the most conducive to collusion. It would be interesting to see if greater amounts of collusion are observed when there is less anonymity and more opportunity for communication and repeated interaction among agents.[46]

g) Feedback

Suppose that the production process takes time during which partial information about the agents' relative performance might become available. How will this information affect agents' subsequent effort decisions? Theoretically, even with standard preferences, this depends on both the distribution of the randomness in the effort-output relationship and the shape of agents' disutility-of-effort functions; thus Eriksson, Poulsen and Villeval's (2008) recent finding that feedback has no effect is not necessarily surprising.[47] More recently, however, Gill and Prowse (2009) consider a case --where the probability of winning is linear in the difference in the agents' efforts-- where such information should have no effect under standard preferences; despite this, they find that agents who are behind exert less effort.[48] They interpret this 'discouragement effect' as a consequence of disappointment aversion, and estimate both the level and heterogeneity of disappointment aversion in their sample using structural methods.

---

[45] Of course, while zero effort by all agents is probably a good sign of collusion, the question of how one would identify collusive behavior in a tournament is an interesting one. Clearly, individual effort can be below the privately-optimal level for reasons other than collusion.

[46] In an interesting field experiment, Bandiera, Imran and Rasul (2005) found that, when engaged in a tournament with their friends, fruit pickers moderated their output (relative to a piece rate) in apparent response to the negative externality their effort imposed on those co-workers. Since this behavior occurs only when workers can monitor others and be monitored, it seems more likely to be motivated by collusion than altruism.

[47] Eriksson et al.'s finding that feedback tends to reduce the *quality* of the low-performers' work is however evocative of the greater risks taken by agents who find themselves running behind, discussed earlier in this section.

[48] In Gill and Prowse's context (where the agents select their outputs sequentially, and the interim feedback consists of informing the second agent about the first agent's performance), being 'behind' simply corresponds to a high level of performance by the first agent.

Using field data from over 60,000 basketball games, however, Berger and Pope (2009) find the opposite: being slightly behind at halftime leads to a discontinuous *increase* in a team's winning percentage; this apparent psychological effect is roughly half the size of the home-team advantage. This field data is corroborated with experimental evidence; the experiment sheds important additional light on the field data by helping to rule out alternative explanations. The contrasting results of these three studies on feedback suggest there is much we still do not understand about the effects of providing interim performance feedback on effort decisions.[49]

In sum, laboratory tests of the tournament models have identified the following empirical regularities:

a) A properly-designed tournament can replicate the results of an efficient piece rate in expectation, but generally yields greater variance in mean output across agent groups (Bull, Schotter and Weigelt, 1987).

b) Handicaps, or 'affirmative action' tend to improve the performance of tournaments between unequal agents. Part of this effect, however, is due to non-Nash choices by less-able agents in the absence of handicaps (sometimes working too hard, sometimes dropping out completely) (Schotter and Weigelt 1992).

c) Under some conditions, decisions to enter into tournaments are surprisingly close to optimal levels (Rapaport 1995). For certain populations, however, entry can be excessive due in part to overconfidence. See Camerer and Lovallo (1989) for MBAs and Niederle and Vesterlund (2007) for men. Allowing risk-averse agents to self-select out of tournaments reduces the between-group variance in output (Eriksson, Teyssier and Villeval 2009).

d) Tournaments can increase risk-taking (Vandegrift and Brown 2003); this effect is not necessarily confined to agents with a low probability of winning (Nieken and Sliwka 2008).

e) Increases in tournament prize spreads can raise sabotage as well as effort; this effect can be strong enough to reduce total output (Harbring and Irlenbusch 2005, Carpenter et al forthcoming). Agents do not always 'target' their sabotage in the expected direction (Irlenbusch et al. 2007).

f) Collusion is rare in anonymous tournaments with more than two contestants (Harbring and Irlenbusch 2008).

---

[49] An additional question concerning interim feedback involves the *principal's ex post* incentives to reveal this information honestly; for example, it may be in the principal's interest to report that the race is closer than it really is. Both Gürtler and Harbring (2007) and Ederer and Fehr (2007) consider this question; they do find evidence that the gap is underreported, though the level of underreporting is lower than predicted by a model without aversion to lying.

g) The effects of interim performance information on agents' subsequent efforts remain poorly understood. Recent studies have found intriguing effects in opposite directions, with Gill and Prowse (2009) finding that being behind reduces effort (when standard preferences predict no effect) and Berger and Pope (2009) finding the opposite.

We conclude this section by noting one aspect of tournaments that has so far received relatively little attention in the lab. This is the use and effectiveness of tournaments as a tool, not for eliciting efficient amounts of effort, but for *identifying the more able player*. (A recent exception in the field is Calsamiglia et al., 2009). If –as Gibbons and Waldman (1999) argue-- it is efficient for organizations to promote abler persons into higher-level positions, then the almost-exclusive focus of the experimental literature on tournaments' consequences for effort levels (as opposed to efficient inference of the agents' underlying abilities) may be missing a key function of relative performance evaluation schemes in real organizations.

3. Teams

a) Holmstrom's Model

In the classic model of agency in teams (Holmstrom 1982), a group's output, x, is a differentiable function of the effort levels of its n members, e_i. The principal's problem is to design a set of compensation functions, $w_i(x)$, that depend only on the group's total output and induce efficient effort choices by *all* agents. Holmstrom's well known result is that such a function cannot exist if $w_i(x)$ takes the form of a 'sharing rule'. A group compensation function is a sharing rule it satisfies $\sum_{i=1}^{n} w_i(x) = x, \ \forall x$, i.e. it balances the budget for all possible group output levels, not just the group's equilibrium output. Thus, free riding is inevitable unless the principal can commit to paying agents as a group more than their combined output for some out-of-equilibrium effort choices, and less in others.

To our knowledge, the first economists to study team production in the lab were Nalbantian and Schotter (1997).[50] In their experiments, Nalbantian and Schotter create two teams of six workers each whose group output is given by the technology $x = \sum_{i=1}^{6} x_i + \varepsilon$, where ε is a uniformly distributed random variable. Four compensation schemes are compared: revenue sharing (where each worker is paid 1/6 of the group's output), 'forcing' contracts or targets (where revenue is shared *if* group output exceeds a target; otherwise all agents receive a low payoff), "gainsharing" (where the target is a function of the team's past performance in the experiment), and a tournament between the firm's two teams (where all members of the team with the higher output receive a prize). Consistent with the 'classic' model's predictions, effort levels under the revenue

[50] The team production problem is, however, closely related to the problem of voluntary contributions to a public good, which has a longer experimental history.

sharing scheme converge towards individually rational levels, which entail a high degree of free riding (though they start out considerably higher in early rounds). Forcing contracts (including gainsharing) perform poorly, perhaps because of the multiple equilibria that are theoretically associated with the induced game among agents. Competition between teams generates the highest level of efficiency among all these treatments, at least for Nalbantian and Schotter's parameterization, which sets a prize high enough to (theoretically) induce efficient effort levels.[51]

More recently, Meidinger, Rulliere, and Villeval (2003) have introduced 'active' principals in to Nalbantian and Schotter's protocol: here, a subject in the role of the principal first makes an offer to two prospective team members; the offer is the share of the team's output that the team gets to keep— either one half or two thirds. Agents then decide independently on effort levels. In addition to the usual cooperation versus free-riding considerations, agents' effort decisions in this context also appear to be affected by a desire to reciprocate generous 'share' offers from the principal.

b) Team production and the voluntary contributions mechanism (VCM)

A second branch of the 'team production' literature traces its origins not to the principal-agent literature but to that on public goods. Team production in these studies is defined as contributors' behavior in a voluntary contributions mechanism (VCM) (see Isaac and Walker 1988 for an early example; Ledyard 1995 for a review of the experimental literature, and Chaudhury 2007 for a more recent review of specific issues). In the standard VCM, each member of a group, i, voluntarily contributes m_i 'tokens' out of his endowment, w_i to a common account. Each member's payoff is then just $(w_i - m_i) + a\sum_i m_i / N$, where N is group size and a is the efficiency gain from public provision. Thus, in the 'classic' VCM, the reward schedule for the agents is predetermined to be a sharing rule, with equal shares accruing to all agents. The typical experimental result for the basic VCM is that individual contributions start out above the individually-rational (but socially inefficient) level but converge to that level as agents gain experience with the game. Work within this tradition has examined the impact of factors like group size, communication, and group heterogeneity on voluntary contribution levels; see Chaudhuri, Graziano and Maitra (2006) for a recent example. Recent summaries of results in the VCM literature are available in chapters 82-90 of Plott and Smith (2008).

Beginning in the 1990s, some experimenters interested in team production in the workplace studied effort decisions using a VCM framework. The key modification to the VCM framework that was introduced by these researchers is reward schedules that do

[51] Nalbantian and Schotter also implement a scheme where the principal can observe an *individual's* effort at a cost. Not surprisingly, they find that "monitoring works, but is costly". Van Dijk, Sonnemans and van Winden (2001) implement individual, team, *and* tournament based compensation in a real-effort experiment. While some free riding occurred in teams, overall effort levels were the same under individual and team payment, since free-riding was compensated by many subjects providing more effort than in case of individual pay.

depend in some way on the effort contributions of individual team members.[52] Since we know (at least theoretically) that either individual piece rates or tournaments can induce efficient effort levels in this case, the main questions concern the efficacy of particular types of reward schedules in raising individual contributions toward efficient levels. Examples of this approach include Dickinson and Isaac (1998) who introduce prizes for the highest individual contribution into a standard VCM environment; not surprisingly, these prizes raise contribution levels. (Of course, to be effective, such prizes must be committed to regardless of the level of output that is attained by the group; thus they constitutes precisely the kind of 'budget-breaking' that Holmstrom (1982) showed is necessary to attain efficiency in a team production environment.) Other papers who have introduced different types of relative rewards into VCMs are Dickinson (2001), which considers monetary *fines* on the lowest contributors, and Irlenbusch and Ruchala (2008).

c) Complementarities in Production

An ironic feature of most models of team production that have been implemented in the lab (including the basic VCM) is that the assumed production technology –in which all agents' efforts are perfect substitutes for each other-- rules out one of the main reasons why teams exist: production complementarities among the members. An extreme form of complementarity that has, however, been studied in the lab is the "minimal effort" or 'weakest link' game (Van Huyck, Battalio, and Beil 1990) where group output depends only on the lowest effort supplied among the team members. Especially in the absence of communication among agents, deductive methods provide little guidance regarding what equilibria to expect in such co-ordination games; Van Huyck et al. find massive co-ordination failures in the lab as agents play 'safe' strategies that leave them relatively invulnerable to the hard-to-predict actions of their co-workers.[53]

More recently, Brandts and Cooper (2007) consider Leontief production in a team of four workers. With no communication or management, Brandts and Cooper show that – as in earlier studies-- such games almost always converge on co-ordination failure, with all agents supplying low effort levels. The authors then assign a manager to each such team, and give the manager two types of tools: (a) increasing the (common) rate of pay

[52] Eckel and Grossman (2005) consider a classic VCM problem, framed as a team production problem, with no modifications to the incentive structure at all. Instead, they manipulate the amount of anonymity, contact and 'identity' of team members, for example by having the members perform a cooperative task before the experiment. They find some positive effects of increased contact among team members, though the two interventions that yielded the greatest gains also offered extra financial incentives relative to the base case (wages for teamwork, and a monetary bonus to the team with the highest total output).

[53] Bornstein, Gneezy and Nagel (2002) find that this co-ordination is somewhat mitigated when intergroup competition is introduced—specifically, two groups compete for a prize received by the group with higher minimum. The members of the losing group were paid nothing. Sutter and Strassmair (2009) introduce communication into experimental tournaments between teams. In their experiment, communication within teams increases efforts (by facilitating coordination) while communication between teams reduces effort (by facilitating collusion).

received by each team member for an extra unit of team output –recall that both the manager and the workers see only the team's total output, which equals the lowest effort level chosen by any team member--, or (b) communicating with the workers by sending (and in one treatment also receiving) completely unstructured messages to the group.

Brandts and Cooper's provocative finding in this environment is that increased financial incentives are essentially powerless in raising group performance (small increases in group output can be achieved, but never enough to compensate the principal for the cost of the additional compensation). Simple messages, however, can be highly effective in overcoming co-ordination failures of the type modeled by these authors. The authors supplement their experimental analysis with an interesting econometric analysis of the effects of different types of messages exchanged between principals and agents.

Most recently in this context, Georg, Kube and Zultan (2009) test an intriguing theoretical result about incentives in teams with production complementarities due to Winter (2004). Winter's result refers to a group of n identical agents contributing effort to a project, each of whom receives a reward of b_i , $i= 1, ...n$ if the project succeeds and 0 otherwise. The production function is said to exhibit complementarity when $p(k+1) - p(k)$ increases in k, where k is the number of agents exerting positive effort and p is the probability of success. In this situation, Winter shows that the *only* reward scheme that induces efficient effort as a unique Nash equilibrium when workers' efforts are complementary is fully discriminating, in the sense that no two workers' rewards, b_i, are the same. Thus, efficiency *requires* inequity, in the sense of treating identical workers differently. Intuitively, it is easier (and cheaper) to make some agents work if they 'know' that working is a dominant strategy for some other (identical) agent because that agent will be extremely well rewarded if the entire group succeeds.[54]

Of course, given many authors' arguments that workers' effort decisions can be strongly, and negatively, influenced by 'unfair' wage differentials (e.g. Akerlof and Yellen 1990), it is not at all obvious that Winter's proposed mechanism –which *requires* arbitrary wage differentials—would work well in practice. Perhaps surprisingly, it does: When the production technology exhibits complementarity, higher efficiency is achieved under a discriminatory reward mechanism than under a cost-equivalent symmetric one. Further, despite concerns that fairness considerations might affect workers' behavior, "subjects' effort choices are highly sensitive to their own reward, but largely unresponsive to the rewards of the other …subjects in their group". This echoes Charness and Kuhn's (2007) result for individual labor contracts, which also found that workers' effort decisions in a multi-worker firm were insensitive to other workers' wages. Taken together, these two papers suggest that, in contrast to the role of social preferences in exchanges between individual 'workers' and 'firms', the importance of horizontal comparisons (*among* workers) within firms seems limited in the experimental literature.

[54] In the opposite case of 'substitutability', the only reward scheme that induces efficiency as a unique Nash equilibrium is the symmetric one that treats all workers alike. Thus, Winter's model provides interesting predictions for the relationship between the group's production technology and the efficient compensation policy.

d) Selection into Teams

As noted, probably the most common team compensation policy is a sharing rule in which all members receive an equal share of the group's output. An immediate consequence of such a policy –which contrasts starkly with tournament-based compensation-- is that team members will prefer to have *abler* co-workers. Thus the issue of how teams are formed can be an interesting one, especially when complementarities exist among team members: what are the effects of different team formation mechanisms on the matches that are formed, and are these matches efficient?

We are aware of only a handful of laboratory studies of the team-formation process. Weber (2006) conducts experiments on the minimum-effort co-ordination game that start with small groups –who find it easier to co-ordinate—, then adds entrants who are aware of the group's history. Using this procedure, coordinated large groups can be created, as long as the rate of growth is not too large. In contrast to Brandts and Cooper (2007), no communication is involved. Charness and Yang (2008) evaluate a specific voting mechanism for group formation in a VCM where there are economies of scale. Societies of nine people are initially formed randomly into three groups of three people who play the game for three periods. Individuals then learn about the average contribution of each individual (by ID number) in their current own group, as well as the average contribution in other groups, and can decide whether to exit the group. Remaining group members choose whether to exclude any current members from the group; the new groups and 'free agents' then choose whether to merge with other existing groups and/or other free agents. A critical element is the role of efficiency in terms of group size – the multiplying factor for each token contributed to the group account increases with group size. Groups of nine are common, particularly after a 'restart' after 15 periods. They find considerable success for the mechanism, as the threat of ostracism seems to keep contribution rates quite high and efficiency is a driving force.

In an interesting field experiment, Bandiera, Barankay and Rasul (2009b) find that, when the incentives facing an entire team are strengthened (in their case by introducing a tournament between teams or performance feedback), assortative matching into teams by ability is increased. Further, workers become less likely to form teams with those they are socially connected to. In Hamilton, Nickerson and Owan's (2003) field study of a textile plant, strong assortative matching into teams did not occur. Further, equal sharing of production bonuses within teams seems to have stimulated cooperation, information sharing, monitoring and even mutual training within teams, generating a productivity increase (relative to piece rates) despite the expected free-rider problem.

Finally, in addition to complementarity between team members' effort levels, there may also exist complementarity of a different sort—actual gains from heterogeneity, for example if different types of group members possess complementary skills. Charness and Villeval (2009) cite some work that suggests this and also finds that

there is a preference for mixed ages (and higher efficiency) in a VCM played at two French firms' work sites.

In sum, laboratory tests of the effort decisions in teams have identified the following stylized facts:

a) In the absence of communication and/or repeated interaction, teams in which agents are paid equal shares of the team's output perform poorly, with agents' efforts converging to low, individually rational levels after a few rounds of play (Isaac and Walker 1988; Nalbantian and Schotter 1997).

b) The forcing contracts (essentially group bonuses) suggested by Holmstrom (1982) typically fail to improve outcomes in these environments due to co-ordination problems among agents (Nalbantian and Schotter 1997).

c) Team performance may also be affected by considerations of reciprocity towards the principal, if one exists (Meidinger, Rulliere and Villeval 2003).

d) Adding incentives based on the relative contributions of individual members to the team's output can improve teams' performance, if such measures are available (Dickinson and Isaac 1998; Dickinson 2001, Irlenbusch and Ruchala 2008).

e) Adding competition between teams can be more effective than any of the above strategies (Nalbantian and Schotter 1997). Given the tremendous popularity of team sports, both to participants and spectators, it is not at all implausible to us that humans are naturally attracted to such situations and perform well in them.

f) When there is complement4arity between the efforts of team members, loss of output due to co-ordination failures can be severe in the absence of communication among team members (Van Huyck, Battalio and Beil 1990). Adding communication in such situations can generate dramatic improvements, much more so than strengthening financial incentives (Brandts and Cooper 2007). Other mechanisms that have been observed to work include asymmetric incentives –that facilitate co-ordination by making high effort a dominant strategy for at least one player—(Georg, Kube and Zultan 2009), and slowly adding new members to smaller groups, which find it easier to co-ordinate (Weber 2005).

We conclude by noting one aspect of team production that, to our knowledge, has not been addressed in the lab: opportunistic behavior by principals. For example, principals who attempt to commit to incentives that 'break the budget' at out-of-equilibrium effort levels may face strong temptations to understate the team's total output (see for example Eswaran and Kotwal 1984).[55]

[55] Note that tournaments do not face this problem, since the principal's total compensation bill does not depend on which worker wins; indeed this has been suggested as a possible advantage of tournament-based compensation (Carmichael 1983).

4. Multi-task Principal-Agent Problems

Suppose the agent performs multiple tasks the principal cares about, but the principal is only able to base the agent's compensation on a subset of those tasks. At least since Farrell and Shapiro (1989) outlined their "Principle of Negative Protection", and Holmstrom and Milgrom (1991) introduced the 'multi-task principal agent problem', economists have recognized that, depending on circumstances, incentive systems based on a subset of the tasks performed by the agent may be less efficient than a compensation system with no incentives at all.

The first experimental implementation of a multi-task principal-agent model we are aware of is Fehr and Schmidt (2004).[56] In their experiment, an agent exerts two types of effort, both of which are observed by the principal and agent, but only one of these is contractible. The two types of effort are complements in the production of total output (specifically, output equals the product of e_1 and e_2). Disutility of effort, on the other hand, is given by $c(e_1 + e_2)$, where c is increasing and convex. Principals can choose between two types of contracts: A "piece rate contract" pays the agent a fixed base wage plus a linear piece rate per unit of task one performed. A "bonus contract" consists of a fixed base wage plus an unenforceable announcement that the agent might receive a bonus from the principal if his overall performance is "satisfactory". Principals and agents interact anonymously, and only once.

Absent concerns for reciprocity (in the sense of perceived intentions for offering a bonus rather than a fine), it is clear that both the above contracts should perform relatively poorly: the bonus contract should yield zero effort, and the piece rate should produce effort only on the first task, which leads to very low output due to the assumed complementarity in production. Allowing for reciprocity, it is not immediately clear which contract should perform better, since both contracts allow for principals to make generous fixed wage payments that agents could conceivably reciprocate. Empirically, however, --echoing results in Fehr, Klein and Schmidt (2007)--, bonus contracts perform much better than piece rates in this environment: many agents reward high effort levels on both tasks with generous bonuses, and agents seem to anticipate this. (Some apparent reciprocation of high fixed pay was evident in both treatments, but was not very effective, relative to the possibility of earning a bonus).

In some sense, Fehr and Schmidt's results support Holmstrom and Milgrom's prediction that powerful incentives may be a mistake when they are based on a strict subset of the agents' actions that affect the principal's welfare. Indeed in such situations, Fehr and Schmidt's results suggest that vague and completely and unenforceable subjective performance evaluations by the principal can outperform piece rates in this case.[57] Of course, the advantages of such 'bonus' contracts depend heavily on what Fehr

[56] There is, of course, a sense in which tournaments in which agents can allocate their effort between productive activities and sabotage are a multi-task situation. We discuss these in the section on tournaments.

[57] See for example Baker, Gibbons and Murphy (1994) for a more explicit argument in favor of subjective performance evaluation.

and Schmidt interpret as workers' concerns for fairness vis-à-vis the firm; if workers' perceptions of fairness are manipulable or highly context-dependent, the widespread use of vague expectations as a solution to multitask principal-agent problems may not be practical.

The only other multitask principal-agent experiment we know of is by Oosterbeek, Sloof and Sonnemans (2006). In contrast to Fehr and Schmidt, rather than choosing between two complementary, productive activities, Oosterbeek, Sloof and Sonnemans' agent chooses between a productive and a 'rent-seeking' activity. The latter activity is pure social waste, but increases the agent's bargaining power by improving his outside options. In the 'classic' model with no social preferences, effort devoted to rent seeking should increase when the experimenters raise the marginal efficacy of effort in that activity. In some behavioral models, however, the opposite could happen, because abstaining from rent-seeking activities becomes a more powerful signal of the agent's good intentions the more effective those activities are. The authors' experimental results are largely in line with the classic model.

In sum, economists are just beginning to study the multitask principal-agent problem in the lab.[58] Some key patterns that have been observed to date are:

a) As predicted by the 'standard' model (Farrell/Shapiro 1989; Holmstrom/Milgrom 1991), rewarding the observable task via a piece rate while not rewarding the other yields poor outcomes, especially if the tasks are complements (Fehr and Schmidt 2004).

b) Unenforceable promises by the principal to reward 'satisfactory' *overall* performance by the agent perform remarkably well (Fehr and Schmidt 2004). Perhaps this situation is familiar to subjects and they act according to norms that are highly effective in the real world.

c) When agents can choose to invest in an unproductive rent-diverting activity, raising the returns to that activity generates more of it. Agents seem neither to anticipate, nor to receive, increased rewards for refraining from such activity when it becomes more tempting for them to undertake it (Sloof and Sonnemans 2006).

5. Multi-period Principal-Agent Interactions

a) Ratchet effects

The 'ratchet effect' applies to a situation where a principal contracts with an agent more than once, the agent has some persistent private information (such as his ability or the productivity of the principal's technology), and binding multi-period contracts are not

[58] Field and econometric studies are also rare. Slade (1996) studies multitasking in contracts between oil companies and gasoline stations; more recently Griffith and Neely (2009) consider multitask incentives in a U.K. distribution firm.

enforceable. In such situations, actions taken by the agent early in the relationship reveal information to the principal, which can be used by the principal later on to the agent's disadvantage. The classic example in labor economics is in the context of piece rates, where an agent's choice of a high effort level in the first period reveals either that (a) the agent's effort costs are low, or (b) the firm's technology is more productive than expected, either of which leads the firm to reduce the generosity of the agent's compensation package in the future. Anticipating this, 'able' workers (or workers who have discovered that the firm's new technology is highly productive) will choose low effort levels (Gibbons 1987; Ickes and Samuelson 1987). This benefits those workers by preventing the firm from extracting their rents later in the relationship, but is socially inefficient.

Aside from some early ethnographic studies (see for example Mathewson 1931), the *only* empirical evidence of ratchet effects of which we are aware is experimental in nature.[59] Chaudhuri (1998) conducted a laboratory experiment in which principals and agents interacted for two periods, and agents were one of two types that were unobserved by the principal. There was little evidence of ratcheting: most agents played naively, revealing their type in the first period even when an informed principal would use this information to the agent's disadvantage, and principals often did not exploit agents' type revelation. Possible explanations for this result include the relative complexity of the game, and the lack of context provided to the subjects that might have impeded the learning process.

Cooper, Kagel, Lo and Gu (1999) frame their experiment in a context-rich way, as a game between central planners and firm managers, use both students and actual Chinese firm managers as subjects, and implement experimental payoffs with high stakes relative to the participants' real-world incomes. They also simplify the interactions between principals and agents, focusing the experiment only on the stages of the game where information revelation matters: the agent's effort choice in the first period, and the principal's choice of a payoff schedule in the second. Cooper et al. do find evidence of ratchet effects, though even in their context it took some time for the players to learn the consequences of type revelation.

Finally, Charness, Kuhn and Villeval (2008) experimentally test a prediction of Kanemoto and MacLeod (1992) that ex post competition for agents can eliminate ratchet effects and lead to first-best outcomes in equilibrium. Importantly, this prediction holds even when outside firms cannot observe the past performance of the agent. They also extend Kanemoto and MacLeod's theoretical analysis to show that ex post competition for principals has the same effect. They impose three conditions in their experiment: no ex post competition, competition with an excess supply of principals, or with an excess supply of agents. As predicted, both types of competition virtually eliminate ratchet effects, though of course their effects on the utilities of the agent differ dramatically.

[59] Carmichael and MacLeod (2000) interpret some historical evidence in light of the ratchet effect.

b) Career Concerns:

In ratchet effects models, high-ability agents exert low effort in early interactions to convince principals they have low ability; this prevents principals from reducing the agents' compensation in the future. In 'career concerns' models, such as Holmstrom (1982/99) agents exert high effort in early interactions, in order to convince principals they have high ability, and thereby high wage offers in the future. The key reasons for these dramatic differences are different assumptions about information and competition: In ratchet models, the principal and agent are in a bilateral monopoly situation, and the agent's performance is typically not seen by other firms; thus high effort by the agent signals that he has a high level of rents that can be extracted in the future. In career concerns models, there is a competitive market for agents in all periods and the agent's output is seen by other firms; thus the wage the agent can command in future periods increases with the market's assessment of his ability. In the literature, ratchet models have been used to describe long-term worker-firm interactions, or interactions between managers and planners in non-market economies. Career-concerns models have been applied mostly to the compensation of CEOs and other senior executives, beginning with Fama (1980).

We are aware of two experimental papers on career concerns; the more recent (Koch, Morgenstern and Raab 2009) is actually the simpler one. Consider a world with a finite number of agents and principals. In the first period, no agent's ability is known to anyone, and all principals and agents share the same prior for agents' abilities. In this period, an agent chooses an effort level e at a private cost, $c(e)$, which is common to all agents. Once effort has been chosen, the worker's output is publicly revealed to be $y = a + e$, where a is realized worker ability. A worker's ability is a permanent characteristic, thus in this simple version of the model a worker completely learns his ability at the end of the first period. At the end of period 1, all firms see every worker's first-period performance y, and will rationally attempt to infer each worker's ability from this information.

In period two, agents exert no effort but produce an output equal to their ability, a. Knowing this, principals engage in (Bertrand) competition for agents at the start of this period. (Specifically, four firms bid for three agents; a principal can employ more than one agent; principals simply offer lump-sum wages that agents accept or not.) In a Perfect Bayesian equilibrium, firms will therefore offer wages equal to agents' abilities in the second period; this gives agents incentives to take period-one actions that convince firms they have higher abilities. Subject to certain restrictions on out-of-equilibrium beliefs, Koch et al identify a unique Perfect Bayesian equilibrium in which all agents choose the same, strictly positive effort level in period one. In consequence, first-period outputs perfectly reveal agents' abilities, and agents are paid their abilities in period two. While all agents would be better off choosing zero effort in period one, the logic of 'signal-jamming' equilibria such as these compels them all to work harder, to prevent being misidentified as a lower-ability type.

In their experiment, Koch *et al* compare the above model it to a 'public ability' treatment, which is identical except that both the worker's output and effort (and therefore ability) are publicly revealed at the end of the first period. In this treatment agents have no reason to expend effort in the first period, so equilibrium effort is predicted to be zero. By and large, Koch et al's findings are consistent with the predictions of the career concerns model: effort is higher in the hidden-ability treatment, and subjects' first and second-order beliefs (which are elicited by the experimenters) are quite consistent with the model. That said, decision errors were high in early rounds (it apparently took some time for subjects to understand the game), and principals' offers were subject to a mild winner's curse.

Irlenbusch and Sliwka's (2006) model of career concerns is identical to Koch et al.'s, except that it contains two key features. First, firms bid for workers at the start of the first period as well as the second. Second, agents choose effort in the second period as well as the first. Since all workers are *ex ante* identical (so there is no reason to offer them different wages in the first period), and since all workers' privately optimal second-period effort is zero, neither of these differences changes the PBE of the game. Interestingly, however, the experimental outcome is dramatically different: now, contrary to predictions, first-period effort is much higher in the public ability treatment, precisely where there is no signal-jamming reason to exert effort. According to both Irlenbusch and Sliwka and Koch et al., the most likely reason is that Irlenbusch and Sliwka's model introduces opportunities for signaling of a different kind (that is not formally modeled in either paper): because effort is now revealed before period-two wage offers are made, agents now have an opportunity to signal that they are "high-effort" (or "reciprocal", or "fair") types by choosing high effort in the first period. Since effort is not directly observed in the hidden-ability treatment, such signaling is less effective there. It is noteworthy, however, that this apparent signaling behavior occurs even though, by construction, all agents have the same cost of effort function in the experiment. The effort costs, or 'willingness to work' that agents are apparently signaling, is some personal characteristic that is not induced by the experimental design. (Though it is clearly consistent with heterogeneous social preferences).

Of course, which of the two above designs is more representative of any particular 'real world' labor market is unclear. While it is clear that both bidding for workers and effort decisions are made at multiple points in any worker's career, Koch et al. argue that, to the extent that agents' actions in Irlenbusch and Sliwka's experiment are driven by reciprocity or fairness concerns, they may be irrelevant to the market for CEOs. On the other hand, if these actions are meant to signal a low cost of effort, or a high level of determination, drive and ambition, they may indeed be very relevant to the CEO market.[60]

---

[60] Akerlof (1976) presents a signaling model in which agents signal a low cost of effort by working harder; separating equilibria in which (almost) all workers provide socially excessive effort exist. For a more recent career concerns model with similar properties, see Acemoglu, Kremer,and Mian (2008). It may also be worth noting that labor economists have developed a variety of models in which workers' abilities are gradually revealed to markets over time (see for example Bernhardt and Scoones, 1993, and Altonji and Pierret 2001); we focus on career concerns models here because they are the only such models we are aware of that involve an effort choice by workers.

c) Investments and Hold-Up

The "hold-up" problem pertains to any multi-period relationship involving specific investments, where binding multi-period contracts are not enforceable. An early formal statement was in the context of a unionized firm's investments in plant and equipment, where market frictions create a gap between capital's purchase price and its resale value (Grout 1984). Grout showed that the firm, in general under-invests when union-firm bargaining occurs after the capital is in place, since the firm pays the full cost of the investment but (due to *ex post* surplus sharing) it only reaps a fraction of the returns.

Most applications of the hold-up problem in labor economics, however, refer to the problem of *workers'* investments in firm-specific skills. Long ago, Becker (1965) proposed that sharing both the costs and returns to firm-specific investments should achieve (presumably constrained) efficiency in separation decisions after the investment is made, though the precise efficiency properties of this arrangement were not specified, nor were its implications for the initial investment decision analyzed. Since then, Hashimoto and Yu (1980), Hall and Lazear (1984) and others have studied the problem more formally, and some ingenious institutional solutions (such as rigid wages in combination with a 'triggered' renegotiation (MacLeod and Malcomson 1993) have been proposed. Other institutional arrangements that have been argued to help solve the firm-specific training problem (in the sense of inducing both efficient investments and efficient *ex post* separations) include various types of promotion ladders (Carmichael 1983, Prendergast 1993), and multi-skilling policies (Carmichael and MacLeod 1993).[61] We note that a key feature of all these analyses is that principals (firms) cannot make binding commitments, for example to retain a worker if that is *ex pos*t unprofitable for the firm. If firms can make credible promises of this nature (perhaps because their interactions with previous cohorts of workers can be observed by current workers), the holdup problem can be substantially mitigated, or eliminated.

The first experimental study of the hold-up problem of which we are aware is Anderhub, Königstein and Kübler (2003). They consider a two-period worker-firm interaction in which workers choose whether to make a firm-specific investment that reduces their effort costs in the first period; the investment is profitable to the worker only if he is employed by the same principal in both periods. In perfect equilibrium, workers should make the investment whether or not firms cannot commit to do this; perhaps unsurprisingly, Anderhub et al however find that contractual form does matter, with workers being more willing to make specific investments when their re-employment is contractually guaranteed. More recent experimental papers have examined the effects on holdup of communication (Ellingson and Johannesson 2004); of which party (worker or firm) makes the investment and the nature of ex post wage bargaining (Oosterbeek, Sloof and Sonnemans (2007a); and of firms' promotion rules (Oosterbeek, Sloof and Sonnemans 2007b).

[61] For a review of the hold-up literature in the context of labor markets, see Malcomson (1997).

More specifically, the promotion rules compared in the latter paper, are 'up or stay' (Prendergast 1993), versus 'up or out' (Kahn and Huberman 1988); the setting is one in which workers' specific investments may affect their firm-specific productivity not only in their current job, but in an alternative job (in the same firm) to which they might be promoted. Promotion decisions are at the firm's discretion, and firms cannot commit to refrain from opportunism in those decisions (aside from being bound by an up-or-out or an up-or-stay rule). Workers' *only* incentives to acquire firm-specific skills in these models are to win promotions or to keep their jobs. In this context, up-or-stay promotion policies will induce workers to invest only if investments raise their productivity more in the job to which they would be promoted than in the current job. Thus investment incentives may be too weak, but specific investments are never wasted due to separations. Up-or-out policies can provide better investment incentives, but may waste investments. (See Gibbons 1988 for a more in-depth discussion). Sloof et al. implement these policies in the lab and find that workers' investment decisions are, at least on average, in line with theoretical predictions.[62]

In sum, as for multitask principal-agent problems, experiments on multi-period principal-agent problems remain few in number. Nevertheless a few interesting results can be identified. They are:

a) The early pooling equilibria at low effort levels predicted by ratchet effects models can be generated in the lab (Cooper, Kagel and Lo 1999).

b) Consistent with Perfect Bayesian equilibrium in the modified game, labor market competition essentially eliminates the ratchet effect (Charness, Kuhn and Villeval 2008), at least in the case where workers' private information is about their ability (as opposed to the firm's technology). To our knowledge, no lab experiments on the latter ("hidden technology") variant of the ratchet effects model exist.

c) The early signal-jamming equilibria at high effort levels predicted by career-concerns models can be generated in the lab (Irlenbusch and Sliwka 2006; Koch, Morgenstern and Raub 2009).

d) If agents choose efforts in both periods in a career-concerns game, the effects of making effort publicly observable contradict the career concerns model: Rather than reducing first-period effort (because signal jamming is no longer possible), making effort public actually raises first-period effort (Irlenbusch and Sliwka 2006). This suggests that agents are attempting to signal some personal characteristic that is outside the model, such as 'honesty' or a personal willingness to work hard.

[62] In yet another paper published in 2007, Sloof et al. (2007c) consider the effects of keeping the level of one's specific investment secret on the hold up problem: the simple theoretical intuition is that if the other party does not know how much one has invested, they will be at an informational disadvantage in bargaining over the *ex post* surplus. The framing (and likely relevance) however seems more appropriate non-labor market interactions, however.

e) Even in situations where short-term contracts should theoretically guarantee workers the same return to firm-specific investments as long-term contracts, enforceable long-term contracts induce more worker investments in firm-specific skills (Anderhub, Konigstein and Kubler 2003).

f) The nature of *ex post* wage bargaining (threat point versus outside offer), and promotion policies (up-or-stay versus up-or-out) affects investments in specific training (Oosterbeek, Sloof and Sonnemans 2007a, b).

IV: Towards Behavioral Principal-Agent Theory: Fairness, Social Preferences and Effort

Perhaps the main contribution of experiments to principal-agent theory is the cascade of papers demonstrating the presence of "social preferences" (where one takes into account the payoffs, actions, and/or beliefs of other parties) in the laboratory. The essential content of social preferences is that people will deliberately sacrifice money to help other people or hurt other people, or even to keep their promises and thereby avoid guilt or a "cost of lying." The fact that people do not simply maximize their earnings has far-reaching consequences for theory; some efforts have been made in this regard (for example, see von Siemens 2004 for a characterization of optimal contracts with social preferences and hidden action).

The earliest experimental paper to convincingly demonstrate the existence of social preferences is the article on the ultimatum game in Güth, Schmittberger, and Schwarze (1982). In this bargaining experiment (which has been replicated hundreds if not thousands of times), one person is provisionally allocated a sum of money and chooses how much to offer a paired participant. If the offer is accepted, it is implemented; if it is rejected, both parties receive nothing. The main result is that people reject lopsided offers, even though it is costly to do so.

This result and many others like it have led to a number of models of utility. These can be roughly classified as either distributional or reciprocal, with some hybrid models (e.g., Charness and Rabin 2002, Falk and Fischbacher 2006). Distributional models presume that one cares about the payoffs of others regardless of the choices and perceived intentions of these others. There are two primary types of distributional models: "Difference-aversion models" (e.g., Loewenstein, Bazerman, and Thompson 1989, Bolton 1991, Fehr and Schmidt 1999, and Bolton and Ockenfels 2000) assume that players are motivated to reduce differences between theirs and others' payoffs, while "social-welfare models" (e.g., Charness and Rabin 2002) assume that people like to increase social surplus, caring especially about helping those (themselves or others) with low payoffs. On the other hand, reciprocity models (e.g. Rabin 1993, Dufwenberg and Kirchsteiger 2004) assume that the desire to raise or lower others' payoffs depends on how fairly those others are behaving; in other words, how one perceives the intentions of other parties affects one's behavior.

In the field environment, it is difficult to rule out that behavior is driven by expectations of future material benefit, since there is typically repeated interaction in the field. In the laboratory, one can isolate social preferences by ruling out the possibility of future interaction, either by using a one-shot game or by anonymous re-matching. Of course, one must be concerned that the behavior found in the laboratory is specific to the laboratory. Regarding distributional preferences, charitable giving in the U.S. (a notoriously individualistic society) exceeds 2% of GDP, with 90% of people donating; thus, most people are willing to contribute materially to the well-being of even anonymous strangers in the field. Regarding negative reciprocity, workers have been known to engage in sabotage or increased theft rates after a pay cut or other actions perceived to be unfair (see for example Greenberg 1990, Shminki, Cropanzano, and Rupp 2002), particularly when procedural justice in the organization is low (Skarlicki and Folger 1997). And the studies by Krueger and Mas (2004) and Mas (2006, 2008) present results that show retribution with real firms and workers. The case for non-instrumental positive reciprocity in the field is weaker, but includes cases such as tipping when on the road or higher response rates to mailed surveys that include small gifts.

Nevertheless, institutional and contextual factors determine the extent that social preferences come into play. For example, it is conventional wisdom that inexorable forces drive out social preference in a market setting. Roth, Prasnikar, Okuno-Fujiwara, and Zamir (1991) find experimentally that competition essentially completely drives out fairness considerations when 10 people can potentially accept a demand (only one of the people who accept this demand is selected) for a share of the pie that is made by one paired person. Recent papers such as Gneezy and List (2006) and Fershtman, Gneezy, and List (2009) suggest possible limitations for the applicability of social preferences. On the other hand, even though a great deal of research (see Holt 1995 for a survey) has found a striking degree of convergence to the self-interested competitive equilibrium in experimental double auctions, Fehr and Falk (1999) demonstrate the presence of downward wage rigidity in an experimental labor market featuring the competitive double auction.

Fehr and Gächter (2000) suggest that a key determinant of whether social preferences come into play is whether contracts are complete and enforceable. They point out that incomplete contracts typically prevail in the labor market, where (for example) wages are often paid without any explicit performance incentives. They state: "In situations where contracts are reasonably complete, the underlying assumption of self-interest should continue to be especially important" (p. 178). However, based on evidence from Fehr, Klein, and Schmidt (2007), where an implicit and incomplete contract with an unenforceable bonus outperforms a complete contract with contracted effort enforced through partial monitoring and the incomplete contract is selected 88% of the time, they argue that reciprocal considerations are not only critical with incomplete contracts, but also: "The endogenous formation of incomplete contracts through reciprocal choices shows that reciprocity may not only cause substantial changes in the functioning of given economic institutions but that it also may have a powerful impact on the selection and formation of institutions" (p. 178).

Given the evidence presented above, it seems clear that there is scope for improving the principal-agent models through the incorporation of behavioral motivations such as social preferences. Perhaps the insights into the nature of non-self-interested behavior gleaned from experiments will eventually be applied to a variety of economic settings, including employee response to changes in wages and employment practices.

1. Models of Social Preferences

We begin the discussion of the relevant literature by summarizing some of the more prominent models of social preferences. We proceed historically, rather than by publication date. For detail of the models and their full functional forms, we refer the reader to the individual papers.

Bolton (1991) develops a model in which people care about both their own money and their relative position; this model is based on the Ochs and Roth (1989) finding that people frequently make disadvantageous counter-offers in a two-round ultimatum game.[63] One receives negative utility from receiving less than the other person, but is not bothered by receiving more. There is a trade-off between one's material payoff and one's relative standing, in that a person might prefer (Own, Other) material payoffs of (3,2) to material payoffs of (4,6). This model is entirely consequential, as intentions or previous history does not affect preferences.

Rabin (1993) was the first to provide a formal game-theoretic model incorporating reciprocity preferences. The central notion is that of *kindness*, defined in terms of the payoff options made available to the other player by one's own choice, with one's kindness increasing as the best available payoff for the other person increases. If you are unkind, the other person may sacrifice money to hurt you, while if you are kind, the other person may sacrifice money to help you. In the latter case, mutual cooperation in the prisoner's dilemma can be supported as an equilibrium. Dufwenberg and Kirchsteiger (2004) formally extend this model (still including kindness as the main concept) to sequential games, which facilitates application to the more standard experimental games with first movers and responders. These models concentrate on reciprocity and only employ simplistic notions of fairness and distributional preferences.[64]

The Bolton and Ockenfels (2000) and Fehr and Schmidt (1999) consequential models of relative position (or inequity or difference aversion) were developed approximately simultaneously. The heart of these models is that people may (depending

[63] In their experiment, if a proposal is rejected there is another round in which the pie is substantially reduced; the person who rejected the initial offer can then make a proposal to split the smaller pie. A disadvantageous counter-offer is a proposal that, if accepted, would lead to a smaller material payoff for the second-round proposer than he or she had rejected in the initial proposal.

[64] However, there are two models in the appendix of Rabin (1993) that do consider the distribution of material payoffs.

on parameter values) trade off money to make material payoffs closer together. The Bolton and Ockenfels model is more general and does not provide a specific functional form, while the Fehr and Schmidt model provides a simple and quite tractable functional form. The main difference between these models (in the two-player case) is that people are more bothered by being behind than by being ahead in the Fehr and Schmidt model, but that one's disutility from unequal payoffs is unaffected by whether one is ahead or behind.[65]

Falk and Fischbacher (2006) combine the Fehr and Schmidt notion of inequity aversion with reciprocity considerations in a complex hybrid model where a person is less bothered by another's refusal to come out on the short end of a split than by a refusal to share equally. Importantly, they assume that one does not resent harmful behavior by the other player if it seems to come only from the other player's unwillingness to come out behind rather than his or her selfishness when ahead. In other words, the prevailing social norm is one where it is bad form to be selfish when one has more than others, but is understandable when one has the short stick.

Charness and Rabin (2002) also combine distributional preferences with reciprocity.[66] The key innovation of this model is that people care about social efficiency, or the total payoffs for the reference group. Absent "misbehavior" (determined endogenously), one's utility is determined by a weighted average of one's material payoff and a social component, which is itself comprised of a weighted average of the total payoffs for the reference group and the lowest payoff for anyone in the reference group. With misbehavior, there is negative reciprocity: one withdraws one's willingness to sacrifice to help the miscreant, by diminishing or eliminating the weight put on their payoff; one may even be willing to sacrifice to hurt the other person. However, in line with considerable experimental evidence (see the discussion below), there is no positive reciprocity in this model.[67]

Cox, Friedman, and Gjerstad (2007) present a non-equilibrium approach that combines a form of distributional preferences with reciprocity considerations. In this approach, both status (relative position) and reciprocity affect one's emotional state, which in turn affects the choices that are made by a utility-maximizing agent. They introduce a parametric model of other-regarding preferences in which one's emotional state determines the marginal rate of substitution between own and others' payoffs, and thus my subsequent choices. In turn, one's emotional state responds to relative status and to the kindness or unkindness of others' choices. They find that structural estimations of

---

[65] In the *n*-player case, another difference is that Bolton and Ockenfels compare the ratio of one's material payoff to the total material payoffs, while Fehr and Schmidt consider the sum of pairwise comparisons.

[66] Note that the real model is in the appendix of the paper.

[67] One might wonder whether a complex model such as Charness and Rabin (2002) is needed, as the substantially more parsimonious and tractable Fehr and Schmidt (1999) model does a reasonably good job of organizing much of the observed experimental behavior. While this topic is still being debated, it seems fair to say that there are many papers that provide data that cannot be explained by distributional models without considering reciprocity. These include Kahneman, Knetsch, and Thaler (1986), Blount (1995), Charness (2004), Offerman (2002), Brandts and Charness [2003], Andreoni, Brown, and Vesterlund (2002), Falk, Fehr, and Fischbacher (2003), Charness and Rabin (2002), and Charness and Levine (2007).

this model with six existing data sets demonstrate that other-regarding preferences depend on status, reciprocity, and perceived property rights.

A final model involves a trade-off between one's material payoff and one's degree of guilt from violating the expectations of another person. This is called "disappointment aversion" in Dufwenberg and Gneezy (2000), but is probably better known as "guilt aversion" (Charness and Dufwenberg 2006). The idea is that the more one believes that a party with whom one is paired is expecting a favorable move, the more likely it is that one chooses the actual move. Charness and Dufwenberg (2006) find support for this notion, as there is a strong positive correlation between a responder's beliefs about the first-mover's beliefs and the responder's choice of the favorable action. Formal presentations of this concept ("simple guilt") and a more complex version ("guilt-from-blame") in which one feels guilt only to the extent that one believes that the other party blames one for an unfavorable outcome can be found in Battigalli and Dufwenberg (2007, 2009).[68]

In sum, a number of models of social preferences have been proposed, mainly classifiable into distributional models and reciprocity models and hybrids of these two approaches. Other motivations such as guilt aversion (and lying aversion, which has not generally been formalized; however, see Charness and Dufwenberg 2009 for one approach) are also being considered in recent models.

a) Distributional models posit that people are concerned with the payoffs of others in some fashion, with considerations of intention on the part of the others being irrelevant; in other words, only the consequences of one's choice matter. Bolton (1991) presumes that one cares about receiving less than another person in one's reference group, but doesn't mind receiving more than the other person. Bolton and Ockenfels (2000) instead presume that people care equally about coming out behind or coming out ahead of the average amount received by those people in the reference group. Fehr and Schmidt (1999) put forward a very tractable model that also presumes that people care about coming out behind or coming out ahead of others; however, one cares at least as much about coming out behind. In addition, one makes pairwise comparisons with the material payoffs of others, rather than comparing with the average of others (this only matters in environments with more than two people).

b) Reciprocity models instead posit that people respond to the intentions of others, with perceived kind intentions being met with kind responses and perceived unkind intentions being met with unkind responses. In this way, both mutual cooperation and mutual defection are potentially equilibria in a Prisoner's Dilemma. These models typically abstract away from distributional considerations. Rabin (1993) is the seminal paper in this area, incorporating reciprocity into simultaneous games. Dufwenberg and Kirchsteiger (2004) extend this approach to sequential games, typically better-suited to the laboratory.

[68] For another survey of social preferences, see Cooper and Kagel (forthcoming).

c) Since both distributional and reciprocity considerations appear to be relevant, some recent models have combined these factors. Falk and Fischbacher (2006) combine intentions with Fehr-Schmidt preferences. Charness and Rabin (2002) put forth the notion of social efficiency, whereby people are typically interested in improving the payoffs of others, unless said others have behaved badly. There is negative reciprocity in this model, but (based on the experimental evidence in that paper and elsewhere) there is no positive reciprocity. Cox, Friedman, and Gjerstad (2007) use a non-equilibrium approach that does not require knowledge of beliefs and is therefore considerably more tractable.

d) Other motivations such as guilt aversion have also been modeled (Battigali and Dufwenberg 2007, 2009). With guilt aversion, one trades off feelings of guilt against being selfish. The more one expects that another person expects her to behave favorably, the more likely one is to then behave favorably. Charness and Dufwenberg (2006) provide experimental support for guilt aversion, also finding that promises (statements of intent) are particularly useful in achieving optimal social outcomes.

2. The Gift-Exchange Game

Probably no experimental game in the area of labor economics has had as much impact as the gift-exchange game, which tests the notion (Akerlof 1982; Akerlof and Yellen 1988, 1990) that there is a positive relationship between wages and effort. This game was designed to mimic an employment relationship, in that labor contracts are typically incomplete and effort is not (fully) enforceable; thus a wage offer is binding, but effort is discretionary. In its basic form, the experimental participants are divided into "firms" and "workers", who interact anonymously either in some form of labor market or in one-to-one pairings; typically, the game is played for a number of periods. In this subsection we report results from the earliest gift-exchange experiments.

The first paper reporting results from the gift-exchange game is Fehr, Kirchsteiger, and Riedl (1993). They create a competitive labor market using a two-stage game. Workers and firms[69] (with more workers than firms) were separated into two rooms, and communication between the two rooms took place via telephone. The first stage was a one-sided oral auction with employers as bidders. Firms made wage proposals, which were posted in the room containing the workers. Once a worker accepted an offered wage, the first stage was concluded for both the firm and the worker. A firm could revise its (non-accepted) wage offer upward, so that it was higher than any existing posted wage offer. People who did not contract received zero earnings for the period. In the second stage, workers chose effort and this choice was only revealed to the paired firm. Firms chose wages (restricted to be multiples of five) from the interval [26, 126], earning (126-wage)*effort. Workers earned the wage less 26 less the cost of effort, shown below:

[69] The language used in this paper involves buyers and sellers who prices and quality; however, the paper emphasizes the labor interpretation, so we adopt the terminology used in almost all subsequent papers.

e	0.1	.0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
c(e)	0	1	2	4	6	8	10	12	15	18

If workers are purely self-interested, the prediction is of course that they will choose $e = 0.1$ regardless of the wage. Knowing this, firms will choose a wage of 30, the lowest wage ensuring participation. However, the main result is that both wages and effort levels far exceed the predictions with self-interested workers, as the average wage was 72 and the average effort chosen was 0.4. In addition, there is a very strong positive relationship between effort and wage.

The second paper in this series is Fehr, Kirchler, Weichbold, and Gächter (1998), who conduct both a one-sided oral auction treatment and a bilateral-gift-exchange treatment using Austrian soldiers to test the robustness of the Fehr et al. (1993) results and to determine the relative effect of competition on wages. An additional treatment involves complete contracts, in the sense that the experimenter enforced an effort level of 1.0 and no effort costs were subtracted from worker's earnings. The main finding is that gift exchange persists even in the absence of competition, when firms and workers are matched on a one-to-one basis; after a few periods wages in this treatment (and the ratio of effort to wage) coincide with wages in the one-sided oral auction without effort enforcement. Thus, the qualitative findings of Fehr et al. (1993) are replicated with non-student participants and even without competition. Wages were considerably lower when effort is enforced, although still substantially above the minimum needed to ensure participation (workers could reject offers).

Fehr, Gächter, and Kirchsteiger (1997) study the impact of reciprocity on contract terms and their enforcement with three experimental treatments involving competitive markets and more workers than firms. In each treatment, the firm specifies a wage, a desired effort level, and a fine imposed if the firm detects that the worker has shirked (provided less than the contracted level of effort. In the weak-reciprocity treatment, workers who have accepted posted contracts choose effort levels, and a random device determines whether shirking is verifiable (and fined at the specified level); the firm then learns the chosen effort level. The no-reciprocity treatment is identical, except that the experimenter exogenously fixes the effort level. Finally, in the strong-reciprocity treatment, there is an additional stage in which firms can reward or punish workers at a cost. The results show that firms demand and enforce much higher effort levels in the strong-reciprocity treatment than in the weak-reciprocity treatment; there is much less shirking in the strong-reciprocity treatment. In fact, both firms and workers earn more in the strong-reciprocity treatment, in large part because the higher effort levels lead to a larger pie. Nevertheless, firms' contract offers are much higher in the weak-reciprocity treatment than in the no-reciprocity treatment, and the offers increase with the desired effort level.

In sum, the gift-exchange game has been a very successful approach to modeling labor issues in the laboratory. The main finding is that higher wages lead to higher effort. This section reports only the earliest gift-exchange experiments.

a) Some early gift-exchange papers used a form of one-sided oral auction to create a market environment. Fehr, Kirchsteiger, and Riedl (1993) have more workers than firms, simulating unemployment conditions. Fehr, Kirchler, Weichbold, and Gächter (1998) find very similar results with Austrian soldiers and a bi-lateral design, where there are equal numbers of workers and firms.

b) Fehr, Gächter, and Kirchsteiger (1997) find that both firms and workers earn more when firms can punish or reward workers for their effort choices, since higher effort is socially efficient (there is a larger pie to divide). Even though contract offers are higher without this enforcement possibility, there is much more shirking (low effort). Nevertheless, offers increase with the desired effort level in both cases.

3. Multi-worker gift-exchange experiments

It seems substantially more realistic to consider an environment in which a firm has more than one employee. When there are multiple workers who receive wages and who can provide effort that benefits the firm, there is the possibility of dispersed responsibility for the firm's earnings (leading to possible free riding) and there are considerations of horizontal fairness (one's pay compared to the pay of other workers).

Maximiano, Sloof, and Sonnemans (2007) compare a bilateral gift-exchange game with one in which each firm has four workers; in the latter case the employer is likely to earn much more than any of her workers, thus reducing the need for any individual worker to sacrifice to help the relatively high-income firm. The authors did not expect gift exchange to survive in this environment.[70] In fact, effort levels in the latter treatment are only marginally lower than in the bilateral game, so that "the gift exchange relationship is quite robust to increases in the size of the workforce" (p. 1026).[71] Their results suggest that intentions-based reciprocity is a driving factor, although efficiency preferences (Charness and Rabin 2002) may also play a role in inducing this behavior.

Charness and Kuhn (2007) match two workers (with different productivity levels, although the precise levels are unknown to the workers) with one firm to investigate whether workers have concerns with pay inequality and whether pay secrecy and pay compression is therefore beneficial for a firm; in a within-subjects design, we varied

[70] They state on p. 1026: "Frankly, we did not believe that the gift exchange model would survive the more realistic multiple workers per employer design. We were wrong."

[71] We hasten to add that, despite the large number of papers that find a strong positive relationship between wage and effort, gift exchange is not robust to all experimental conditions. Hannan, Kagel, and Moser (2002) find little gift exchange with undergraduates; on the other hand, MBA students provide effort responses similar to those in most gift-exchange studies. Charness, Fréchette, and Kagel (2004) find that when a complete payoff table is provided in the experimental instructions, workers choose substantially less effort compared to a treatment in which participants had only the information needed to compute their payoffs.

whether a worker was aware of the other worker's wage. Under fairly general conditions, we demonstrated theoretically that workers' responsiveness to co-workers' wages should lead profit-maximizing firms to compress wages or maintain pay secrecy. And, as in other gift exchange experiments, we observed a strong positive empirical relationship between 'own' wages and effort. Surprisingly, however, the effort level provided by a worker was unaffected by the wage paid to his or her co-worker. Furthermore, although firms compress wages when the co-workers will know both wages, this did not raise profits. It seems that the relationship between a worker and the firm is much more salient than the relationship between the pay of the co-workers. Overall, our experimental evidence "casts doubt on the notion that workers' concerns with equity might explain pay policies such as wage compression or wage secrecy" (p. 693).

Gächter, Nosenzo and Sefton (2008) perform an experimental analysis of pay-comparison information and effort-comparison information in an environment in which firms are matched with two workers. While effort is highly sensitive to the worker's wage, co-worker wages *per se* have no effect. Further, when the firm pays different wages to the workers, the co-worker's effort decision is ignored. However, worker behavior is affected when both pieces of social information are provided: a generous wage generates higher effort when one's co-worker exerts high effort, but is ineffective when the co-worker contributes little or no effort. They suggest that group composition is a relevant factor for obtaining beneficial effects from social information.

Abeler, Altmann, Kube, and Wibral (2009) focus on a two-worker-one-firm environment in which a worker knows both the ability and effort level of the co-worker. However, they reverse the order of play, as workers first choose effort levels and the paired firm then chooses wages. In one treatment firms are constrained to pay equal wages, while in a second treatment there is no such constraint. Perhaps surprisingly, they find that there is lower effort when the firm is forced to pay equal wages, as a worker who chooses higher effort than the co-worker does not feel it is fair to receive the same wage as the co-worker, and subsequently reduces effort. Thus, it is important to pay attention to equity considerations rather than equality *per se*.[72]

In sum, there have been a modest number of gift-exchange experiments with multiple workers, a more realistic case than the standard game.

a) Maximiano, Sloof, and Sonnemans (2007) find that gift exchange is relatively undiminished even when each firm has four workers, with their results suggesting that intentions-based reciprocity is a driving factor in the effort choices.

b) Some recent papers have examined the effect of pay-comparison information and/or effort-comparison information when there are two workers per firm. Both Charness and Kuhn (2004) and Gächter, Nosenzo and Sefton (2008) find that pay-comparison information alone has little or no effect on the effort choices of workers. However, the

[72] In a related paper, Meidinger, Rullière, and Villeval (2003) examine team heterogeneity and productivity. There is considerable free riding as each agent is greatly influenced by his or her teammate's behavior. On the other hand, workers are better able to cooperate when the team is homogeneous.

latter study finds that worker behavior is affected when both pay-comparison information and/or effort-comparison information are provided and the co-worker provides high effort.

c) Abeler, Altmann, Kube, and Wibral (2009) examine the two-worker-per-firm environment in which a worker knows both the ability and effort level of the co-worker, with workers first choosing effort and the firm then choosing the wage. Effort is lower when firms are forced to pay the same wage, due to workers who choose higher effort objecting to being paid the same as a shirking co-worker

.

4. Positive and negative reciprocity

While the classic gift-exchange experiments provide strong evidence of reciprocal behavior, it is less clear that this represents reciprocity in a strict sense. The reason for this is that people may have distributional preferences that could lead to the same behavior that is observed in an environment where reciprocity is possible. Thus, positive or negative reciprocity reflect behavior that differs from what a responder would have done in the absence of a first-mover action that is perceived to be positive or negative. Some experimental games from Charness and Rabin (2002) illustrate this point. In the baseline case, a participant unilaterally chooses between (Other, Own) material payoffs of (400, 400) or (750, 375). Around fifty percent of the population chooses to sacrifice 25 units to give the other person an additional 350 units, even though this leads to a large difference in material payoffs.

In a second case, the other paired participant first faced a choice between payoffs of (750, 0) or passing the choice to the second participant, who would once again face a choice between (Other, Own) payoffs of (750,375) or (400,400). Positive reciprocity would imply that the rate of (750,375) choices should increase, as the first mover would clearly seem to be kind by allowing the responder to receive a positive payoff. However, in fact the rate goes down slightly (to 39%)! So positive reciprocity doesn't seem to be present here. In a third case, the other paired participant first faced a choice between payoffs of (550, 550) or passing the choice to the second participant, who would once again face a choice between (Other, Own) payoffs of (750,375) or (400,400). Negative reciprocity would imply that the rate of (750,375) choices should decrease, as the first mover would clearly seem to be unkind by forcing the responder to receive a smaller material payoff than was available with the outside option. In fact, the rate of (750, 375) choices does decrease sharply to 11%; thus, negative reciprocity is a factor.[73]

The first experimental paper to carefully test for positive and negative reciprocity was Charness (2004).[74] This paper considers gift-exchange in a bilateral setting, varying

[73] A further example from Charness and Rabin (2002) provides evidence of negative reciprocity. In ultimatum-game experiments, a high proportion of responders reject offers that would give them only 20% of the pie. However, when we offered 36 participants a choice between (Other, Own) payoffs of (800, 200) or (0,0), exactly zero chose (0,0). Thus, distributional considerations are insufficient to explain rejections in the ultimatum game, so that negative reciprocity would appear to be the driving force.

[74] Once again, we order papers by the date they were written (in this case, 1996), rather than the publication date.

whether the wage was determined by a self-interested firm or generated by an exogenous process (such as a draw from a bingo cage); in all cases, the firm benefits from the worker's chosen effort. There is a strong positive relationship between wage and effort in all treatments. However, the effort level with low wages is lower when the wage was chosen by a self-interested firm than when it was generated exogenously, suggesting the presence of negative reciprocity. On the other hand, there was virtually no difference across treatments in the effort level with high wages. Thus, this paper was the first to provide experimental evidence that positive reciprocity seems to be much weaker than negative reciprocity, while at the same time clearly identifying the effect of the distribution of payoffs *per se* on behavior.

A subsequent and related paper (in a non-labor setting) is Offerman (2002), who studies the effects of random choice mechanisms while allowing for both positive and negative reciprocity. He considers players' responses to a helpful or hurtful choice, as a function of whether the "choice" was made by an interested party or generated at random (responders could sacrifice one unit to either increase or decrease the first-mover's payoff by four). Following the helpful choice, responders never paid to lower the first-mover's payoff, but paid to help first movers more often when the first mover made the choice than when the choice was randomly-determined. This suggests some positive reciprocity may be present, although the effect is not significant. On the other hand, the effect on the response to the first-mover's perceived intentions was dramatic and significant following a hurtful choice, again suggesting that hurting hurts more than helping helps (the title of the paper).

Brandts and Charness (2003) test for punishment and reward in a cheap-talk game and find that intention is a critical issue, finding substantial negative reciprocity and limited positive reciprocity. One player sends a message about her intended play to another player; after play takes place, the other player is then given an opportunity to punish or reward the first player. The authors found that the responder was twice as likely to punish unfavorable play by the first player if that first player had lied about his play than if he had told the truth. A relatively small number of responders chose to reward a favorable play by the first mover.

On the other hand, Cox (2004) reports significant positive reciprocity in the investment game (Berg, Dickhaut, and McCabe 1995).[75] The triadic design compares behavior in the standard game with behavior when the first mover is a dictator and to behavior when the experimenter determines the amount received by the responder rather than this being determined by a self-interested first mover; this procedure should allow one to distinguish distributional preferences *per se* from reciprocal preferences. Cox, Sadiraj, and Sadiraj (2008) finds significant positive reciprocity, but does not find significant negative reciprocity in the "moonlighting game", where the first mover can

[75] In the standard version of this game, both a first mover and a responder are endowed with 10 units. The first mover can pass up to 10 units, with the amount passed tripled by the experimenter and then received by the responder. The responder can then pass back any number of units (not tripled).

take from the responder as well as pass to the responder. Finally, Cox and Deck (2005, 2006) report mixed findings on positive reciprocity; the results depend on whether or not the experimenter can observe the actions of the players (double blind or single blind), as well as on the sex of the subject.

In sum, the experimental evidence regarding intention-based positive and negative reciprocity is mixed, although the general result is that negative reciprocity is stronger than positive reciprocity. In a sense, this can be seen as reflecting expectations and violations thereof, as seen in a self-serving way. If one expects kind or favorable treatment and receives it, there is no strong emotional jolt; on the other hand, if one has this expectation and receives unkind or hurtful treatment, the emotional response is much stronger.

a) Negative reciprocity is highly pervasive in experiments, much as it is in the field. A number of papers, such as Charness (2004), Offerman (Charness and Rabin (2002) find little or no evidence of intentional positive reciprocity, in the sense that responses after a kind action do not receive a more favorable response than when no action has been taken.

b) On the other hand, a number of papers, such as Charness (2004), Offerman (Charness and Rabin (2002) find little or no evidence of intentional positive reciprocity, in the sense that responses after a kind action do not receive a more favorable response than when no action has been taken.

c) However, this is not a universal result, as papers such as Cox (2004) and Cox, Sadiraj, and Sadiraj (2008) report significant positive reciprocity Still, Cox and Deck (2005, 2006) report mixed findings on positive reciprocity. So this topic remains open.

5. Pay regulation

Governments often consider whether to regulate pay, by means such as mandating a minimum wage or a sick-pay rate. What are the effects of such policies? In the field, it is difficult to ascertain this, as there are many factors present and changing at the same time. Thus, experimental techniques are likely to be useful, as the effects of such policies can be successfully isolated.

Brandts and Charness (2004) were the first to study the effect of a minimum wage in a gift-exchange format. They used a more symmetric payoff design. Each person was endowed with 10 units, and the first mover could pass up to 10 units, and the responder received five times whatever was passed; the responder could then send back up to 10 units, with the original first mover receiving five times whatever was passed back. Brandts and Charness imposed a minimum wage of five in a condition with an excess supply of workers. The mandated minimum wage was counterproductive in the sense that the average effort was 30% lower than without it, even though the average wage was 5% higher. In addition, they found that the highest wage was chosen only half as

frequently in the minimum-wage condition, with effort also reduced by 30% at this top wage.

Fehr, Falk, and Zehnder (2006) test the effect of imposing a minimum wage in an environment in which each firm is matched with three workers and must choose the same wage for each of the workers. The workers decide on the minimum wage that they would conditionally accept (the strategy method) and the actual assigned wage is then accepted or rejected accordingly; however, workers do not choose effort levels, and a firm's profits only depend on the number of workers who accept the wage offer.[76] In some sessions, fifteen periods of a mandated minimum wage were followed by 15 periods with no minimum wage; the order was reversed in the other sessions. They find that there are lasting consequences of the mandated minimum wage even after it has been removed, in that firms must pay higher wages after the removal than before it was imposed. Thus, the authors conclude that policy can affect people's sense of what is a fair wage.

Owens and Kagel (2009) use a variant of the payoff design in Brants and Charness (2004). They find that the effect of an imposed minimum wage are a reduction in effort in the neighborhood of the minimum wage but no systematic effects on effort levels for higher wages. The minimum wages leads to improved incomes for both firms and workers (particularly the latter). Since there is little effect at higher wages, it appears that the minimum wage requirement is less salient at these higher wages and/or that employees recognized that wages set a good deal higher than the minimum represent just as large a monetary gift as without it.

Dürsch, Oechssler, and Vadovic (2008) report the results of a form of gift-exchange game that investigates the issue of sick pay in an experimental labor market. The main variation is whether there is one-to-one matching of firms and workers or whether there is a market setting in which firms post wage offers and workers select the ranking of their preferences regarding the possible offers; in the latter case, a firm may end up hiring more or less than one worker. In all cases, there is a one-third chance (chosen at random) that a worker will be "sick" and so be unable to complete his or her intended effort. Firms chose one of five possible contracts, each of which specified a wage to be paid in case of zero effort (not showing up for work, either because the computer made this choice or because the worker chose zero effort voluntarily; the firm cannot tell if the worker is actually sick or simply chose to stay home) and a wage to be paid in case of positive effort. The main findings are that higher wage offers significantly increase effort choices (both with respect to sick pay and non-sick pay), firms can attract more reciprocal workers by offering sick pay, but firms benefit from offering sick pay only in the market setting, where there is competition for workers.

Bauernschuster, Dürsch, Oechssler, and Vadovic (2009) use a similar design to consider the effect of sick pay with heterogeneous workers with different likelihoods of being "sick". They use a 2x2 experimental design, in which they vary whether there is a

[76] In this sense, the game resembles an ultimatum game. a monetary gift as without it.

minimum sick-pay rate (40% of the wage) and whether workers all have the same 20% rate of being sick or whether half of the workers have a 10% chance of being sick and the other workers have a 30% chance of being sick. Issues of interest include the degree of moral hazard (pretending to be sick), whether the adverse-selection problem is severe enough to lead to a collapse in the market for sick pay), and whether higher levels of sick pay lead workers to choose higher effort levels. The main results are higher-risk workers do indeed select into contracts with higher sick pay rates, and they pay for this by getting lower wages and an overall worse deal than the low risk workers, higher sick pay leads to increased effort, but only in the case where the sick pay rate is freely-chosen, and the sick-pay market does not break down due to the adverse-selection problem.

In sum, there are mixed results regarding the effect of imposing a minimum wage. Regarding sick pay, it appears it is possible (but perhaps not easy) to design systems where this is a useful policy intervention..

a) Brandts and Charness (2004) find that a mandated minimum wage was counter-productive in terms of effort provision. High wages were not offered nearly as frequently when this mandated minimum wage is present.

b) On the other hand, Fehr, Falk, and Zehnder (2006) find that a minimum wage is potentially beneficial. They find lasting positive consequences from the mandated minimum wage even after it has been removed, as firms higher wages after the removal than before it was imposed. Owens and Kagel (2009) also find some beneficial effects from imposing a minimum wage, with Pareto improvements resulting for firms and workers

c) Dürsch, Oechssler, and Vadovic (2008) and Bauernschuster, Dürsch, Oechssler, and Vadovic (2009) find results that on the whole show some promise for sick pay regulation. In the first paper, higher wage offers significantly increase effort choices (both with respect to sick pay and non-sick pay), but firms only benefit from offering sick pay when there is competition for workers. In the second paper, workers with a greater likelihood of being "sick" select into contracts with higher sick pay rates, but at a cost in terms of wages. Higher sick pay can result in increased effort in some circumstances, and the adverse selection problem does not destroy the sick-pay market.

6. Do gift-exchange and social preferences map into the field?

A major question that labor economists might have concerning laboratory gift-exchange results is the degree to which these mean anything in the field environment (external validity). Two recent papers call into question the degree to which the social preferences exhibited in these experiments are robust or in competitive environments. On the other hand, a number of studies provide evidence that the social preferences identified in the laboratory map well into the real world.

Gneezy and List (2006) investigate whether gift-exchange behavior persists over time in two real-effort tasks, one involving work in a library and the other involving door-to-door solicitation. The mechanism for testing for reciprocity consists of the experimenters telling people that they will be receiving a certain piece rate for their work, but then announcing at the time of the six-hour task that in fact they will be paid a substantially higher piece rate. The results are persuasive, particularly for the fundraising task: people do in fact work harder for the surprise pay, but this effect vanishes over the course of time. Thus, these results constitute a cautionary note against applying the results of these laboratory experiments to the field environment. However, one caveat is that these experiments only pertain to *positive* reciprocity, as even the lower advertised pay rate is above the alternative wage that the student workers would normally earn. Given the relative dearth of positive reciprocity in laboratory experiments, it may not be so surprising that positive reciprocity in these field experiments is fleeting.

Fershtman, Gneezy, and List (2009) present compelling evidence that social preferences are subject to framing effects and are not general to all environments. They show that introducing a competitive frame to the environment crowds out social preferences in mini-dictator and trust games. For example, 72.5% of dictators choose an (Own, Other) allocation of (8,8) instead of (11,2). This is followed by a real-effort competition between the dictator and the recipient, where the (8,8) outcome is implemented if the recipient wins but the (11,2) outcome is implemented otherwise. If the dictator really wanted the (8,8) split, the simple strategy would be to solve no problems; however, dictators contribute considerably more effort when solving more problems leads to higher own payoffs than when solving more problems doesn't affect these payoffs (in the baseline condition). Furthermore, in another treatment the (11,2) split is implemented if the dictator loses to the recipient, while otherwise the (8,8) split is implemented. In this case, 85% of the dictators do nothing. This result demonstrates that social preferences may not have much effect when competition is salient.

Two other experiments present a different view than Gneezy and List (2006). Bellemare and Shearer (2009) conduct a field experiment that investigates worker responses to a monetary gift from their tree-planting firm using incentive contracts. Workers were told that they would receive a pay raise for one day. Productivity on the day of the gift is compared with productivity on adjacent days, under similar planting conditions. They find direct evidence of a significant and positive effect on daily planter productivity on the day of the gift, controlling for a variety of other possible factors. Kube, Maréchal, and Puppe (2006) conduct a field experiment involving a six-hour task in which the participants were told that they would presumably receive 15 Euro per hour; in three treatments, they were then paid either 10, 15, or 20 Euro per hour. The main result is that there is little difference in performance in the treatments were people are paid either 15 or 20 Euro per hour, but that there is a strong, deleterious, and lasting effect on performance when people are paid less than the presumptive rate. These results suggest that there is an asymmetry between positive and negative reciprocity in the field,

very much in line with the evidence found in laboratory experiments. However, one caveat to these results is that the sample size is quite small.[77]

There are two main threads of evidence suggesting that social preferences are linked to the general population and the workforce. First, there are a number of non-laboratory experiments in European nations that show that the basic behavioral patterns observed in fairness-related experiments with students also prevail in representative samples of the general population, with older cohorts generally being more reciprocal than younger cohorts.[78] Holm and Nystedt (2004) analyze behavior in the investment game, with people selected from a public database in Sweden. While the average amount returned was similar for both cohorts, the proportions dispersed for the older responders, suggesting a greater degree of responsiveness to the environment. Other studies conducted with representative surveys in Germany (Fehr, Fischbacher, von Rosenbladt, Schupp, and Wagner 2002) and the Netherlands (Bellemare and Kröger 2007) conclude that older cohorts are more generous as responders. Falk and Zehnder (2006) find substantial evidence of social preferences in a sequential trust game conducted with 1,000 residents of Zurich, also finding evidence of discrimination and in-group favoritism. Sutter and Kocher (2007) find that the elderly are more reciprocal, in a study with participants ranging from 8-year-old children to people in their late sixties. In their study, trust increases from early childhood to early adulthood, but stays constant thereafter. Finally, Karlan (2005) finds that the amounts returned in a laboratory investment game conducted in Peru predict micro-finance loan payments a year after the experiment.

The second thread involves studies that show a relationship between worker productivity and social preferences. Barr and Serneels (2007) conducted a study with Ghanaian manufacturing workers. They find that a measure of trustworthiness (the ratio of the amount returned to the amount sent) has a positive relationship with the wages paid, with these wages being used as a proxy for productivity. This is particularly true with respect to the average trustworthiness in the workplace in question, although the direction of any causal relationship is unclear – are people more trustworthy because their wages are higher or *vice versa*? They conclude that behavioral characteristics and corporate culture are major determinants of the productivity of the firm and the operation of the labor market. Carpenter and Seki (forthcoming) conducted a series of experiments with local shrimp fisherman in Japan. They used a laboratory public-goods game to measure social preferences and also obtained measures of fishing productivity. The main finding is that there is a strong correlation between measured social preferences and productivity for the fisherman; a second finding is that social preferences grow with the degree to which team production is present in the fishing environment.

---

[77] In a related study, Kube, Maréchal, and Puppe (2008) conduct a controlled field experiment that tests the extent to which cash and non-monetary gifts affect workers' productivity. The main result is that a non-monetary gift leads to a significant and substantial increase in workers' productivity, while a cash gift of the same value is ineffective.

[78] Charness and Villeval (2009) also find that students are less cooperative than young workers, older workers, and retirees.

In sum, there is mixed evidence regarding the extent to which social preferences manifest in the field. It is clear that these are not present in all environments, but there are a number of papers providing evidence from laboratory and field experiments, as well as work environments and surveys, that suggest that social preferences can be a factor in the field.

a) Gneezy and List (2008) find positive reciprocity in the first three hours of field experiments, but not in the last three hours. They interpret their results as indicating that social preferences are ephemeral. Similarly, Fershtman, Gneezy, and List (2009) use a clever design to show that social preferences are crowded out when a competitive game is played after an initial choice. Clearly social preferences are not ubiquitous.

b) However, other studies find rather different results. Bellemare and Shearer (2009) present direct field-experimental evidence of a significant and positive effect on productivity on the day when a gift is given, controlling for a variety of other possible factors (yet it remains unclear how long such an effect will persist). Kube, Maréchal, and Puppe (2006) use a design in a field experiment that can test for both positive and negative reciprocity. In line with many experimental results, there is little evidence of positive reciprocity, but strong evidence of negative reciprocity when the pay rate is lower than what was anticipated.

c) Survey evidence indicates that social preferences are linked to the general population and the workforce. Holm and Nystedt (2004), Fehr, Fischbacher, von Rosenbladt, Schupp, and Wagner (2002), Bellemare and Kröger (2007), Falk and Zehnder (2006), Sutter and Kocher (2007), and Karlan (2005) provide such evidence from surveys in several European nations, as well as Peru. In general, it seems to be that older people are more reciprocal than younger people, suggesting that this is a learned trait.

d) There is also direct evidence from the work environment. Barr and Serneels (2007) find a relationship between worker productivity and social preferences for Ghanaian manufacturing workers. Carpenter and Seki (forthcoming) observe a definite positive relationship between social preferences and productivity for shrimp fishermen in Japan, with said social preferences increasing to the extent that team production is present in the fishing environment.

7. Communication

In this final subsection, we consider how communication can affect behavior in principal-agent relationships. While it is true that other forms of social considerations such as distributional preferences and negative reciprocity are likely to lessen the applicability of the equilibrium contracts derived from standard theory, some studies demonstrate that communication can lead to better social outcomes than can be achieved with more standard social preferences.

Charness and Dufwenberg (2006) examine experimentally the impact of communication on trust and cooperation with hidden-action (moral hazard). The principal has an outside option that (in the main treatment) gives (Principal, Agent) payoffs of (5,5); alternatively, the principal can leave matters up the agent, who chooses between (Principal, Agent) payoffs of (0, 14) and expected payoffs of (10, 10).[79] With standard preferences, the agent would choose (0, 14), so that the principal will choose (5,5). In fact, there are substantial rates of cooperation even in the absence of communication, but the likelihood of the (10, 10) expected outcome increases from 20% to 50%. Furthermore, this increase is almost entirely driven by free-form and endogenous *promises* (statements of intent) by the agent to behave cooperatively. The evidence is consistent with people striving to live up to others' expectations so as to avoid guilt. When players exhibit such guilt aversion, communication may influence motivation and behavior by influencing beliefs about beliefs. Charness and Dufwenberg argue that guilt aversion may be relevant for understanding strategic interaction in a variety of settings, and that it may shed light on the role of language and social norms in these contexts.

Brandts and Cooper (2007) study manager-worker interactions in an environment where payoffs depend on employees coordinating at high effort levels; the lowest effort level chosen by any worker determines the overall production. One treatment investigates the effect of increased incentives (the marginal benefit of coordinating on a higher effort level increases substantially) on worker behavior, while a second treatment allows the manager to communicate via chat with the workers. It turns out that communication is a more effective tool than incentive changes for improving coordination on high effort levels. An analysis of the content of communication indicates that the most effective communication strategy is to request high effort, pointing out the mutual benefits of high effort. Thus, direct financial incentives have some benefit in terms of coordination, but messages are the better strategy if one cannot choose both.

Charness and Dufwenberg (2009) investigate the impact of communication on trust and cooperation with hidden-information (adverse selection). There are two possible agent ability levels, with a 1/3 chance that an agent will have high ability. The principal has an outside option or can leave matters up the agent. There are two main cases. In the first, there are possible Pareto-improvements over the outside option for both types of agents. In the second case, there is no feasible Pareto-improvement for the low-ability agent. While the standard game-theoretic prediction is for the principal to choose the outside option in both cases, the authors find a dramatically different effect of free-form communication from the agent to the principal in these two cases. When a Pareto-improvement is feasible with a low-ability agent, communication doubles the rate at which the low-ability agent sacrifices money to help the principal. However, communication has no effect in the alternative case. The difference is driven by the fact that many messages in the first treatment confess that the agent has low ability, but that he or she will choose the Pareto-improvement over the selfish option. Charness and

[79] In fact, there is a 5/6 chance that these payoffs are (12, 10) and a 1/6 chance that they are (0, 10). In this way, the principal cannot be certain of the agent's choice when a zero payoff is received.

Dufwenberg conclude that is good policy to offer lower-ability workers an opportunity to participate in a socially-beneficial outcome, as they are likely to behave cooperatively.

In sum, while experiments on cheap talk have been conducted for over 20 years, some recent work has applied such nonbinding pre-pay communication in clear principal-agent environments. This is a very promising area for both future research and policy .

a) Charness and Dufwenberg (2006) find that promises (statements of intent) by the agent to behave cooperatively are very effective in achieving optimal social outcomes in a hidden-action (moral hazard) setting, where the standard prediction is a total lack of cooperation. The evidence supports the notion of guilt aversion, where people striving to avoid guilt stemming from disappointing the expectations of others.

b) Charness and Dufwenberg (2009) also find that promises are effective in a hidden-information (adverse selection) environment, but only when less-able agents can participate in a Pareto-improvement in material payoffs. A potential policy application is to offer lower-ability workers an opportunity to participate in a socially-beneficial outcome, as they are likely to behave cooperatively.

c) Brandts and Cooper (2007) find the striking result that coordination on higher effort is facilitated more by communication between a manger and the workers than by increased incentives. Communication about the mutual benefits of high effort is particularly valuable. Thus, while direct financial incentives may be somewhat useful, communication in a team-production environment is even more so.

V: More Lab Labor: Bargaining, Search, Markets, and Discrimination

1. Bargaining, Strikes and Arbitration

The bilateral bargaining problem --how a surplus is divided between two players-- has a long and rich history in economic theory, dating back at least to Nash's (1950, 1953) contributions to co-operative and non-cooperative bargaining, and important subsequent developments such as Rubinstein (1982).[80] Experimental studies of bargaining also have a long history, both in psychology (see for example Deutsch and Krauss 1960) and economics (Siegel and Fouraker 1960). Roth (1995) reviews the experimental economics literature on bargaining up to that point, much of which aims to test predictions of the Nash, Rubinstein and related models. As already noted, a well known and robust result of these tests is that bargaining outcomes in the lab are not well described by perfect equilibria of games between rational, selfish agents; behavior is often highly suggestive of social preferences among the bargainers.

Given these early laboratory findings, it is perhaps not surprising that much of the more recent experimental bargaining literature has attempted to understand the nature of

[80] Summaries of these theoretical developments are available in Binmore and Dasgupta (1987), and in Osborne and Rubinstein (1990), as well as many other sources.

apparent other-regarding preferences in bargaining situations; this is typically done using extremely simple bargaining games (such as the ultimatum and dictator games) where there are fewer confounding factors, including strategic considerations, that might also affect outcomes. Recent examples of this approach include Charness and Rabin 2002, Andreoni, Castillo and Petrie 2003; and Falk, Fehr, and Fischbacher 2003; for reviews see Hoffman, McCabe and Smith (2008) and Murnighan (2008). Other aspects of the bargaining problem that have been examined using laboratory experiments since Roth's review include the effects of gender and other demographic differences (Eckel and Grossman 1997, Holm and Engseld 2005); of cultural differences (Chen and Tang 2009); stakes (Slonim and Roth 1998, Cameron 1999, Munier and Costin 2002); risk aversion and beliefs (Dickinson 2009); deadlines (Sterbenz, Phillips and Owen 2001; Gneezy, Haruvy and Roth 2003; Guth, Levati and Maciejovsky 2005); incomplete information (Guth, Huck and Ockenfels 1996); face-to-face interaction (Ockenfels and Selten 2000); delegation (Schotter, Zheng and Snyder 2000); "hot" versus "cold" decisionmaking (Brosig, Weimann and Yang 2003); communication (Brosig, Weimann and Yang 2004; Croson, Boles and Murnighan 2003); self-serving biases (Babcock and Loewenstein 1997); outside options (Carpenter and Rudisill 2003) and situations involving bargaining among more than two players (Fréchette, Kagel and Morelle 2005; Charness, Corominas-Bosch and Frechette 2007).

While the above bargaining models and experiments were motivated and framed in a variety of ways (including relatively abstract frames, or as buyer-seller interactions), labor economists have found them of primary interest for the light they might shed on union-firm negotiations.[81] In particular, labor economists have used formal bargaining models and laboratory experiments to shed light on (a) the determination of disputes (strikes, lockouts, holdouts) and (b) on mechanisms –in particular, arbitration-- that are designed to reduce the costs of such disputes.[82] We consider these two applications of bargaining theory and experiments in turn below.

While economists have been discussing the determinants of strikes since Hicks (1932) stated his famous 'paradox' (if parties can forecast the bargaining outcome after a strike they can always do better to settle on that outcome immediately without a strike), theoretical and experimental analysis of strikes by economists did not begin in earnest until the advent of asymmetric information models of bargaining, which produced disputes such as strikes as equilibrium outcomes. Early theoretical contributions of this nature include Hayes (1984); Kennan (1987) provided an early review of the 'economics of strikes'. Since then, Sopher's (1990) laboratory experiment tested the simple "joint cost" theory that strikes will be less common when their social cost to the bargainers is higher in a shrinking pie bargaining game. Forsythe, Kennan and Sopher (1991) studied 'strikes' in similar games with one-sided private information. Consistent with other

[81] Another important and promising application of bargaining experiments of interest to labor economists is to gender differences in labor market outcomes (e.g. Eckel and Grossman 1997, Babcock and Laschever 2003).

[82] There is also a literature that models the wage and employment outcomes of union-firm bargaining, and attempts to test for the appropriate model of the bargaining process. For reviews, see Farber (1987) and Kuhn (1998); we are not aware of any laboratory experiments in this area.

laboratory bargaining experiments, they found that agents often forgo rewards in order to inflict punishments, and engage in other apparently fairness-driven behavior. In (1991b) the same authors describe a symmetric-information experiment. Kennan and Wilson (1993) review the literature on bargaining under asymmetric information up to that point, including the contributions of lab experiments. Since then, a few papers have used simple models of bargaining and laboratory experiments to address an old question in industrial relations: the effects of centralized versus decentralized bargaining on strikes and other outcomes (Berninghaus, Güth and Keser, 2001, 2003; Tournadre and Villeval 2004).[83]

The aspect of bargaining on which labor economists and others have conducted the largest number of laboratory experiments is not strikes *per se*, but the effectiveness of dispute resolution (i.e. arbitration) mechanisms in encouraging bargaining and reducing disputes. In at least one way this is not surprising, as arbitration is a classic example of mechanism design, and a laboratory experiment is a natural way to 'pre-test' a mechanism under ideal conditions to see if it has a reasonable chance of functioning as predicted in the field.[84] In addition, there is of course considerable policy interest in the design of arbitration mechanisms that encourage (efficient) settlements and minimize the use of the mechanism itself (especially in view of Crawford's 1979 dictum that the best arbitration system is one that is never used). Thus, the evaluation of arbitration mechanisms is one area where laboratory research by labor economists has already been successful in influencing economic policy.

While laboratory experiments testing different arbitration (and other forms of third-party intermediation) systems have a long history in psychology (see for example Johnson and Tullar 1972), the earliest experiment on arbitration by economists of which we are aware (Farber and Bazerman 1986) is noteworthy for its methodology in view of current debates in economics about the lab 'versus' the field: Farber and Bazerman essentially conduct what is now called an (unpaid) 'artefactual' field experiment by confronting professional labor arbitrators with hypothetical cases and asking them to choose a settlement under different arbitration schemes. Since then, economists have performed a multitude of experiments evaluating different arbitration schemes, most of them 'traditional' laboratory experiments, such as Ashenfelter, Currie, Farber and Spiegel (1992).

A key question addressed in early experimental studies was the differential effects of conventional arbitration (where the arbitrator can impose any settlement he feels is 'fair') versus final-offer arbitration (where the arbitrator is constrained to choose *either* the firm's or the union's final offer). In line with its proponents' expectations, early laboratory experiments (e.g. Notz and Starke 1978, Starke and Notz 1981, Neale and Bazerman 1983) tended to find that final-offer arbitration encouraged more pre-arbitration settlement. These results were strongly rejected by Ashenfelter, Currie, Farber

[83] Non-experimental studies of strikes motivated by asymmetric information bargaining models include Card (1990), Cramton and Tracy (1992), Gu and Kuhn (1998) and Kuhn and Gu (1999).
[84] Labor economists have also been active in the use of field data to study the effects of arbitration; see for example Currie and McConnell (1991), who studied dispute rates for public employees across Canadian provinces with different arbitration laws.

and Spiegel (1992), who –unlike earlier treatments—provide bargainers with information about past arbitrator decisions in the same scenario (in their experiment, a list of random draws from the same distribution). Since then, Ashenfelter et al.'s main experimental results have been confirmed by Dickinson (2004, 2005) and Deck and Farmer (2003). Deck & Farmer (2007) and Deck, Farmer & Zeng (2007) further show that this result is robust to uncertainty over the final value of the quantity bargained over. Additionally to the credit of conventional arbitration, Kritikos (2006) finds that final-offer arbitration pushes parties to an equal split of the pie 80% of the time, which he interprets as an unnatural distortion of bargaining.

In addition to conventional versus final-offer arbitration, laboratory experiments have been used to study the effects of the following dispute-resolution mechanisms: tri-offer arbitration (an amendment to final-offer arbitration where the arbitrator can also choose a third option provided by an outside 'fact finder'); combined arbitration (which imposes final offer arbitration when the arbitrator's 'fair' settlement lies between the parties' final offers, and conventional arbitration otherwise); and other schemes known as double-offer arbitration, amended-final-offer arbitration, and automated negotiation. Providing a full description of these results is beyond the scope of the current paper. A recent review, however, is available in Kuhn (2009). By pointing the reader towards this large literature, we hope instead simply to provide an example of a 'success story' for laboratory experiments in labor economics: The use of the lab as a testing ground for dispute-resolution mechanisms in labor markets has yielded results of interest not only to readers of *Econometrica*, but also to policymakers who are faced with the real problem of designing dispute-resolution mechanisms, especially in public-sector labor markets where strikes are prohibited by law. This is clearly a case where 'lab labor' has already passed the market test.

2. Search

Labor economists have been using sequential search models to interpret field data on unemployment durations since at least Lippman and McCall's classic survey (1976a, b). Search theory also plays an important motivational and interpretive role in field experiments on the determinants of unemployment, such as those conducted on unemployment insurance recipients in Illinois and other states. These studies randomly exposed job searchers to re-employment bonuses and to job search assistance programs, and assessed workers' responses in light of job search theory. Such field experiments have produced useful results regarding the likely effects on workers' job search behavior and other outcomes of making very specific policy changes to existing UI programs. Excellent reviews are available in Meyer (1995, 1996).[85]

This said, both the field experiments and econometric studies of job search and unemployment durations discussed above are probably better suited to evaluating the effects of specific policy changes in specific markets than to testing the most basic predictions of sequential job-search models; indeed many of the estimated effects of policy changes are just as consistent with simple labor-supply models (e.g. Moffitt and

[85] For a more recent example of field experiments on job search and UI, see Dolton and O'Neill (2002).

Nicholson 1982) as they are with the search paradigm. The history of laboratory experiments on search, however, focuses much more on the decision processes of individual searchers, and on assessing whether these are consistent with the rules derived from search-theoretic optimization. Given the significant role still played by search models in interpreting microeconomic studies of unemployment and other durations, and in general equilibrium and macroeconomic studies of the labor market, laboratory search experiments make an important, but distinct contribution to the economics of labor market search.

To our knowledge the first economists to conduct laboratory search experiments were Schotter and Braunstein (1981, 1982).[86] Subjects engaged in search for a wage drawn at random from a known distribution (either rectangular or triangular), where each draw from the distribution cost the subject *c*. In the baseline treatment, subjects could take as many draws as they wished, and could accept any of the offers they had received (i.e. there was unlimited recall of past offers). Optimal search theory has strong and well-known predictions in this situation –including the fact that the optimal strategy is a constant reservation wage-- and the authors were interested in comparing subjects' actual behavior to these predictions.

In many ways, the subjects' behavior conformed to theory quite closely. The authors elicited reservation wages by asking workers to bid for the right to search; these reservation wages were, on average, very close to the theoretical optimum for a risk-neutral searcher. Actual search behavior conformed to these elicited reservation wages. Subjects responded to increased search costs and to a mean-preserving spread of the wage distribution as predicted. This said, two interesting anomalies were identified. First, in violation of the theoretical predictions, reservation wages were not constant over time—they fell. (Since reservation wages were optimal at the start of a search spell, this meant that, on average, subjects did not search long enough). Second, when the subjects were asked to describe their search strategy, most subjects reported that their strategy had two components: (a) a reservation wage, and (b) a maximum number of searches they would conduct. It is tempting to conclude that the latter aspect of the subjects' self-reported strategies explains the falling reservation wage in Schotter and Braunstein's results.

Hey (1982) conducted search experiments similar to Schotter and Braunstein's, (infinite horizon, full recall, constant cost per offer) though in Hey's case the context was consumer search, and subjects did not know what distribution the offers were drawn from (it was normal in practice). Despite this, over 40% of subjects stopped searching with the first price below the theoretically optimal reservation price. Hey's study is however most interesting for the methodological innovation (in economics) of *protocol analysis* in which the subjects are asked to report their thinking during the session out loud, and their words are recorded. It is interesting to compare these self-reported strategies to the predictions of theory. Interestingly, while only a small minority reported a pure

[86] As in the case of bargaining, psychologists conducted laboratory experiments on search long before economists did. See for example Rapaport and Tversky (1966, 1970) and Kahan, Rapaport and Jones (1967).

reservation-wage strategy, many of the rule-of-thumb strategies described by the subjects performed well, both as (sometimes imperfectly) implemented by the agents themselves *and* as mechanically played by a computer against the price distribution in the experiment. Hey's paper thus illustrates another potential advantage of laboratory experiments—the ability to acquire information about strategies more directly than inferring them from choice behavior alone, which can often present serious econometric challenges in both the field and the lab.

During the next decade, a number of laboratory experiments explored the predictions of search theory under various scenarios, including a finite horizon (Cox and Oaxaca 1989, 1992), variable search intensity (Harrison and Morgan 1990), and search from unknown wage distributions (Cox and Oaxaca 2000). Like most of these studies, Sonnemans (1998) finds that, while overall search efficiency is high (subjects stopped optimally in more than 60 percent of cases), subjects stop searching too early on average; risk aversion cannot fully account for this. Sonnemans elicits information about subjects' strategies in two ways (apart from their search behavior): one is their use of an electronic information board, which gave subjects information about various aspects of their own past experience, including the number of draws, the highest bid to date, total costs and earnings to date, etc. It seems reasonable to infer that if a subject consulted this information, it was an input into their strategy. Second, after conducting a sequential search, subjects were asked to conduct another search using the strategy method: they chose what explicit instructions to give to the computer, which conducted the search on their behalf.[87] Interestingly, only 22% of such strategies were of the pure reservation wage form, and a minority of these picked an optimal wage close to the optimal level. A significant share of such strategies had a satisficing property of stopping when total earnings exceeded a threshold.

More recent laboratory experiments on search include Carbone and Hey (2004), who study the excess sensitivity of consumption to income often noted in macroeconomic data; they find that subjects also over-react when facing similar problems in the lab and argue that agents' limited ability to plan ahead may help explain this phenomenon. Gabaix, Laibson, Moloche, and Weinberg (2006) explore the ability of a specific model of limited rationality --directed cognition—to account for laboratory evidence in search problems, finding that the directed cognition model better matches the lab evidence whenever its predictions diverge from full rationality. Schunk (2009) considers instead the ability of reference point updating (in a model where subjects are assumed to care about the total net earnings –including those that are sunk-- from the search) and loss-aversion to explain behavior in a laboratory search problem. As in many previous studies, on average Schunk's subjects end their search too soon; he argues that a model in which a subset of agents use suboptimal strategies that depend in part on sunk costs accounts well for this phenomenon. Additional results in Schunk and Winter (2009) support the claim that 'early stopping' is related to loss aversion, rather than, for example, simple risk aversion.

[87] See also Sonnemans (2000) for a more direct comparison of search decisions under the strategy method versus decisions made 'by hand'.

The most recent paper to address the 'early stopping'/declining reservation wage phenomenon in the lab is Brown, Flinn and Schotter (forthcoming). Like other authors, they find sharply declining reservation wages over time.[88] By comparing subjects' behavior in a treatment where offers arrive at a fixed rate but with stochastic cost to one where the time between offers is random but there are no search costs, the authors are able to distinguish hypotheses (such as Schunk and Winter's loss-aversion model) in which searchers respond to accumulating search costs from a competing scenario in which the subjective costs of searching rise sharply over time. They find that rising subjective search costs play a more important role, and argue that the behavioral factors they identify in the lab may help explain the phenomenon of declining reservation wages and/or increasing re-employment hazards in field data. Alternatively, laboratory subjects' rapidly rising time costs may simply reflect some students' short-term scheduling commitments outside the economics laboratory that are not relevant to real-world job searches.

In sum, laboratory experiments on search have, to date, focused almost exclusively on the 'partial equilibrium' question of choices by individuals facing an exogenously given wage distribution.[89] In doing so, laboratory experiments have exploited the strengths of that methodology, which include the ability to compare behavior to the exact numerical predictions of a theoretical model, and the ability to study the searchers' *strategies* in a dynamic choice situation more directly than attempting to infer strategies from choice histories alone. The main findings of the literature are that, while (especially early in the search process) many subjects behave *as if* they pursued a theoretically optimal strategy, subjects tend to search too little overall, with falling reservation wages and rising acceptance hazards over time. In addition, a majority of subjects report that their strategies depended on factors other than the highest offer received (including total earnings in the search), and indeed selected such strategies when required to play the game using the strategy method. Playing such strategies did not cost subjects much in the search situations examined in the lab, but this does not imply that such strategies are not costly in other environments, including some that are encountered in the real world. Ongoing research continues to attempt to understand these 'anomalies' in the light of different behavioral models. What remains unclear is to what extent these anomalies help shed light on job search in real labor markets. Still, the notion that, even under ideal laboratory conditions, subjects systematically choose search strategies that are suboptimal in a particular way, raises interesting hypotheses for research using field data, and for structural models of job search.

3. Labor Markets

It is probably fair to say that the vast majority of existing labor market experiments focus on interactions between a single 'firm' and 'worker', whose payoffs if they choose not to interact are exogenously set by the experimenter. That said, as mentioned in

[88] Brown et al. elicit subjects' reservation wages directly in advance of *every* offer by asking the subjects to name such a wage; a computer then automatically accepts all offers above this and declines the rest.
[89] For an exception in the context of consumer search, see Abrams, Sefton and Yavas (2000).

Section II.2, experimenters have sometimes implemented various forms of markets in which workers compete for jobs (and/or firms compete for workers) in the lab. These papers fall into two main categories, the first of which –the 'market design' papers— studies the effects of different institutional mechanisms on labor market performance. The focus here is typically on thin markets where match quality matters a lot, as in the markets for professional workers. Worker effort decisions are typically not considered. The second group of papers is primarily interested in the effects of contractual incompleteness regarding worker effort on the functioning of labor markets.

a) Market design:

One highly successful set of laboratory experiments on labor markets forms part of a larger literature that is sometimes referred to as 'market design' (for a recent survey and assessment, see Roth 2008). This literature draws on a combination of economic theory, market case studies, computation, field experiments and lab experiments to compare the performance of different forms of decentralized market mechanisms and centralized clearinghouses (the latter often designed by the researcher) for specific goods. Markets that have been studied in this literature include airport landing rights (Grether, Isaac and Plott 1981; Rassenti, Smith and Bulfin 1982); college admissions (Roth and Sotomayor 1989); sorority rushes (Mongell and Roth 1991); macroeconomic risks (Shiller 1993); newly privatized firms in transition economies (Svejnar and Singer 1994); postseason college football bowls (Roth and Xing 1994); radio spectrum licenses (McMillan 1994); space shuttle payload priorities (Ledyard, Porter, and Wessen 2000); student housing (Chen and Sonmez 2002); electric power (Wilson 2002); internet auctions (Ariely, Ockenfels and Roth 2005); human kidneys (Roth, Sonmez and Unver 2007); and admissions to New York City high schools (Abdulkadiroglu, Pathak and Roth 2009). Laboratory experiments play an important role in this literature, both in testing hypotheses about particular market mechanisms and in comparing the mechanisms' performance; one lab experiment in this area that may be of particular interest to labor economists is Chen and Sonmez's (2006) study of school choice mechanisms.

Labor markets that have been studied in the 'market design' literature include those for American physicians (Roth 1984, 1986; Roth and Peranson 1999); baseball players (Nalbantian and Schotter 1995); clinical psychologists (Roth and Xing 1997); British physicians (Kagel and Roth 2000); gastroenterologists (Niederle and Roth 2003; McKinney, Niederle and Roth, 2005); and law clerks (Haruvy, Roth, and Unver 2006; Avery, Jolls, Posner and Roth 2007). Of these labor market studies, Nalbantian and Schotter (1995); Kagel and Roth (2000); McKinney, Niederle and Roth (2005) and Haruvy, Roth and Unver (2006) all use laboratory experiments as at least part of their research design.[90] Key issues studied in these investigations include the causes of 'unraveling' that is observed in some markets (e.g. for law clerks) where offers are made so early that little information is available about candidates, and devices that effectively ensure enough market thickness to allow for efficient matching.

[90] Recent examples of laboratory experiments with applications to two-sided matching markets include Niederle and Roth (2009) and Niederle, Roth, and Unver (2009).

Lab experiments play a number of roles in the above studies of labor markets, one of which is to isolate the effects of different institutional changes that cannot be identified in field data, because they co-vary in the field (for example, Niederle and Roth 2009 use the lab to disentangle the effects of exploding offers and binding acceptances). While this could in principle be done in a field experiment, the time, expense, and ethical barriers involved place strict limits on the amount of experimental manipulation that can be done. In contrast, after cross-validating the lab and field in a baseline case, one can quickly and cheaply compare many design permutations in the lab. Market designers are also interested in the robustness of market mechanisms to large potential swings in demand and supply (McKinney, Niederle, Muriel and Roth 2005); in most cases these cannot be practically (or ethically) manipulated in a field experiment. A final role of lab experiments is simply to provide a level of detail on how individual actions change, for example during transition between market regimes, that is unavailable in field data (Kagel and Roth 2000).

In sum, the literature on market design, like the arbitration literature, is another economics 'success story', where economic theory combined with a carefully chosen set of research methods *including* lab experiments has created institutions that improve the process of economic exchange in the real world.[91] This includes some labor markets, especially for professionals --where markets are thin and match quality is arguably highly heterogeneous-- and some allocation problems in education (such as school admissions) that are of considerable interest to labor economists.

b) Contractual Incompleteness and Labor Markets

While the vast majority of laboratory experiments on principal-agent interactions incorporate labor markets in only the most minimal of senses (as noted, the experimenter typically fixes the two parties' outside options, usually at zero), it is noteworthy that one of the earliest papers to demonstrate the importance of social preferences in principal-agent interactions (Fehr, Kirchsteiger and Riedl 1993) included an explicit labor market, and was explicitly interested in the effect of social preferences on the nature of labor market equilibrium. As already noted, this experiment modeled the labor market as a one-sided auction in which firms posted contract offers to all workers in the session. A main result was that these markets failed to clear when effort was not contractible: firms posted above-market-clearing wages because these wages elicited more effort than market-clearing wages. Firms that tried to underbid such wages did poorly due to workers' endogenous effort responses; equilibrium unemployment resulted. This result is evocative of Shapiro and Stiglitz's (1984) well known model of involuntary unemployment, though the mechanism (reciprocity by workers) is different from the disciplinary dismissals that are central to Shapiro and Stiglitz's model.

Since FKR 1993, the most influential paper on the effects of contractual incompleteness on the nature of market interactions is probably Brown, Falk and Fehr (2004). This experimental paper is framed as a set of repeated interactions between firms and workers and generates, in our view, a set of results that may significantly influence

[91] See Roth (2002) for a review of methodological issues in the market design literature.

how labor economists conceptualize labor markets. In BFF's experiments, ten workers interacted with seven firms over fifteen trading periods. Each trading period had two stages: a market for contracts, followed by the exchange of effort for pay in contracts that had been concluded. In all cases, the market for contracts was, again, a one-sided auction: firms posted offers (consisting of a wage, a desired effort level, and the firm's ID number), which could be accepted or rejected by workers. Once a firm's offer was accepted by a worker, both the firm and worker were removed from the market for that trading period. Importantly, firms could make *both private and public offers*: private offers were only conveyed to the worker with whom the firm wanted to trade.

Under the above conditions, BFF compared two main experimental conditions: under complete contracts (C), the firm's required effort level was automatically implemented (by the computer) if a worker accepted a contract. Under incomplete contracts (ICF), this third party enforcement of desired effort levels was absent.[92] In both treatments, firms only observe the (current and past) effort levels of their 'own' workers. Perhaps unsurprisingly, BFF find that under complete contracts, markets resemble the textbook case: contract offers are public, long-term relations between firms and workers are absent, trading parties seem to be indifferent to their partners' identities, and rent sharing is driven toward the competitive equilibrium (where firms retain all the surplus). Under incomplete contracts, successful exchange is usually initiated by a generous contract offer that a worker reciprocates with a high effort level. After that, firms repeatedly seek out the same worker with a private offer. Rents are shared in these long-term bilateral relationships, which are disciplined by the threat of non-renewal. The result, rather than a classically competitive market, is a market where bilateral monopoly emerges endogenously once the employment relationship has been established, a change described eloquently by Williamson (1985) as the "fundamental transformation".

In many ways, the labor markets that emerge under BFF's ICF condition are much more familiar to today's workers than a 'classical' labor market: Workers invest in their reputations early in their careers by working hard; this reputation is (rationally) rewarded by their employer by continuing employment at a 'good' wage; both parties' actions are disciplined by the (rational) fear that their contract will not be renewed if they do not continue to cooperate by providing 'fair' levels of wages and effort. For the latter reason, BFF's results are thus more directly supportive (than FKR 1993) of Shapiro and Stiglitz's (1984) model where disciplinary dismissals generate equilibrium unemployment. In sum, BFF (2004) show very elegantly that if contractual incompleteness plays an important role in firm-worker interactions, the correct theoretical model of labor market equilibrium may not be the competitive one we are used to, nor one of the several interesting equilibrium search models that have been developed in the past two decades **(**e.g., Mortensen and Pissaridies 1994), but one in which reciprocity and bilateral monopoly play important roles.

---

[92] This condition is labeled ICF because workers' ID numbers were fixed throughout the entire session (as they were in the C condition). This allowed firms to target their offers to specific workers (presumably to a worker who had provided high effort in the past), and distinguishes it from an "ICR" treatment where contracts were incomplete but workers' identities could not be tracked during a session.

A number of interesting variations on BFF's (2004) design have yielded additional insights. Perhaps the most obvious such variation is replacing the excess supply of workers by an excess supply of firms (Brown, Falk and Fehr 2008). Interestingly, long-term employment relationships still emerge, in which firms pay workers above the going market rate and workers reciprocate with higher effort. Market performance remains high, suggesting that unemployment is not required to enforce high effort levels, though long-term relationships are less frequent than in the excess-worker case. Falk, Huffman and Macleod (2008) consider the effects of two institutions -- dismissal barriers and bonus pay—in the BFF (2004) model. Perhaps unsurprisingly, given the important disciplining role of contract renewal risk, legislated dismissal barriers, lead to large reductions in equilibrium effort and market efficiency. Giving firms the option to reward high effort with bonus pay, however, offsets much of these inefficiencies; interestingly, the resulting equilibrium exhibits many fewer long-term relationships.[93]

So far, all the papers in this subsection have considered the effects of hidden *actions* (effort choices) on the functioning of labor markets. We therefore conclude by noting a number of papers that have considered the effects of hidden information (adverse selection) on market equilibrium in the laboratory as well. Interestingly, experimental studies of signaling and screening models have quite a long history, dating back at least to Miller and Plott (1985). These papers may not be familiar to labor economists since they are typically cast in the context of markets for insurance (especially in the screening case), or product quality (signaling). For a recent experimental study of both signaling and screening that is cast in a labor market context, and that reviews the earlier experimental literature, see Kübler, Müller, and Normann (2008). Other recent experiments that have considered the effects of hidden information (about worker ability) in the presence of labor markets are Cabrales, Charness and Villeval (2006) and Charness, Kuhn and Villeval (2008).

4. Gender, Race and Discrimination: Insights from the Lab

The existence of discrimination and its effect on employment and wages is a major topic in labor economics. Discrimination is typically considered with respect to gender or ethnic considerations, but may also be relevant in relation to attractiveness.[94] A

[93] Two other papers that insert new variations into repeated experimental labor markets similar to BFF(2004) are Healy (2007) and Schram, Brandts, and Gerxhani (2007). Healy considers the ability of group reputations to substitute for individual reputations; Schram, Brandts and Gerxhani ask what happens when firms can see workers' past effort levels at *other* firms when bilateral contracts are negotiated.

[94] Discrimination may be based on bias or due to accurate perceptions about differences in ability or other characteristics across groups. While we do not discuss this at length in this section, there is experimental evidence in this regard. For example, Fryer, Goeree & Holt (2005) have "green" and purple" workers with different "investment costs". Investment improves chance worker does well on pre-employment test; this is observed by firm, who can hire either green (lower cost) or purple. Green workers were hired substantially more often.

fundamental issue is how to ameliorate concerns such as the male-female wage gap and racial disadvantage. Can policies such as affirmative action be effective or are these likely to be counter-productive? Once again, it is possible to focus specifically on these issues using laboratory techniques. We review the relevant literature in turn.

a) Gender

One of the puzzles in labor economics is the gender-wage gap. Even though the provision of equal opportunities for men and women has been a priority for decades, large gender differences prevail in competitive high-ranking positions. Why do women receive less pay than men? Some of this may stem a difference in a taste for negotiation. For example, Babcock and Laschever (2003) suggest that women are more likely to shrink from negotiation, potentially costing themselves thousands of dollars in not asking for promotions or in choosing career paths that don't involve negotiation. Another factor may be a difference in risk aversion: Charness and Gneezy (2009) find a very consistent result across a number of experimental studies is that women are more financially risk averse than men. Related to this point, Niederle and Yestrumskas (2008) offer participants a choice of a hard or an easy task, finding gender differences in seeking challenges: Men choose the hard task about 50 percent more frequently than women, independent of performance level.

Stereotypical attitudes towards female workers may well be at the heart of discriminatory gender-based compensation. Schwieren (2003) considers the issue of why women receive lower pay for comparable work in a gift-exchange experiment in which each firm knows whether workers are male or female. She adapts the double-auction format from Fehr and Falk (1999), with 4 firms and 6 workers in each session. Each group of workers in a session was all-female or all-male, as was each group of firms in a session; she varied whether male workers were paired with male or female workers and *vice versa*. The results are striking, in that female workers receive significantly lower wages than male workers, *even when women are in the role of the firm*. But this doesn't pay for firms, as a high discrepancy between the wage requested and the wage offered leads to low effort. The results suggest that stereotypes are the driving force, rather than statistical discrimination. Women also learn that they cannot successfully ask for high wages and reduce their bids over the course of a session.

The main thrust of experimental research on gender differences in performance is based on the notion that women are less inclined to compete. The first experiment to demonstrate this point is Gneezy, Niederle, and Rustichini (2003), who find that women appear to be less effective than men in competitive environments, despite the fact that their performance is similar to that of males when the environment is competitive. The experimental task was solving mazes, with six people in each session. In one treatment, people were paid two units for each maze solved; in a second treatment, where there were three men and three women in a session, the person who solved the most mazes received 12 units for each solved maze; the final treatment was identical to the second, except that

there were either six men or six women in each session.[95] There is a definite performance gap across gender in tournaments, even though there is no significant difference in performance across gender in the non-competitive treatment. This is driven by an increase in men's performance in the tournament, as women's performance does not differ across the tournament and piece-rate cases. Furthermore, the effect is more pronounced when women compete against men than when women compete against women, as men's tournament performance does not differ according to gender composition, while women's tournament performance in the same-sex treatment is significantly higher than in the piece-rate condition.

This result is confirmed in Gneezy and Rustichini (2004), a field experiment with school children (aged 9-10) in Israel. The performance measure was the time needed to run 40 meters. Each child first ran alone; the teacher then matched the students in pairs, with the two fastest children paired, then the next two fastest children paired, etc. The children in each pair then ran along side each other. The times in the first task were very similar for boys and girls; however, there were significant gender differences in the change in times from the first run to the second. When girls ran with other girls, the time actually increased slightly; however, the opposite effect was observed when boys ran with other girls. In mixed pairs, the reduction in time for boys was much larger than the reduction in time for girls.

Niederle and Vesterlund (2007) provide evidence that women "shy away" from competition in a task involving adding up sets of five two-digit numbers. In stage 1, participants performed this task under a non-competitive piece-rate scheme, and then performed the task in a four-person tournament in stage 2. In stage 3, people then chose whether they wished to choose the piece-rate scheme or the tournament (in the latter case, they competed against the performance of the other group members in stage 2). There are no significant differences in performance across gender, yet men select the tournament in stage 3 more than twice as frequently (73% versus 35%) as do women. This difference in rates remains significant even when differences in risk attitudes are taken into account. The authors conclude: "…the tournament entry gap is driven by men being more overconfident and by gender differences in preferences for performing in a competition. The result is that women shy away from competition and men embrace it." (p. 1067).

Gneezy, Leonard, and List (2009) perform field experiments (involving tossing a tennis ball into a bucket) in Tanzania and India. People chose whether they wished to be paid by a piece-rate scheme or to participate in a two-person tournament. The Maasai society in Tanzania is a patriarchal society while the Khasi society in India is matrilineal. The results in the patriarchal society correspond closely to the results in Western cultures, as the Maasai men choose to compete about twice as frequently as do the Maasai women. However, it is quite interesting that the comparison across gender goes in the opposite direction in the matrilineal society. Their results strongly suggest that it may be

[95] An additional treatment in which one person was chosen at random to receive 12 units per solved maze led to results quite similar to that in the first treatment.

differences in culture, rather than inherent genetic traits, that drive the results observed in previous experiments.

Two other experiments consider the context of the environment in relation to gender differences in risk-taking and in choosing to compete. Booth and Nolen (2009a) conduct an experiment in which participants choose between a risky gamble and a certain one, with the expected value of the risky gamble being higher than that of the certain outcome. They vary whether girls attend coed or single-sex schools. The main finding is that girls from single-sex schools are substantially more likely to choose the gamble than girls from coed schools; in fact, girls from single-sex schools choose the gamble roughly as frequently as do boys. This indicates that social learning and nurture may well be the source of the gender differences found in previous studies. Booth and Nolen (2009b) consider whether girls from single-sex schools differ from girls from coed schools in terms of willingness to compete. They find substantial differences, with girls from single-sex schools more likely to choose competition and behaving more like boys with respect to competitiveness.

b) Minorities

Discrimination occurs not only with respect to gender, but also with respect to ethnic background. As it is considerably more difficult to arrange experiments in which ethnic background is systematically varied, there are fewer studies in this area.

Fershtman and Gneezy (2001) consider discrimination in Israeli society between Ashkenazic Jews and Eastern Jews; the former tend to be wealthier and better educated. Three different experimental games were used in their study. First, in the investment ('trust') game, three times as much (17.16 versus 5.62) was sent to Ashkenazic male responders than to Eastern male responders; this pattern was observed for both male Ashkenazic and Eastern first-movers (in fact, this discrimination was only observed for males). However, there was no significant difference in the amounts returned by Ashkenazic and Eastern responders. Since first-mover behavior could reflect preferences *per se* rather than trust, a dictator-game control was also conducted, with the amount transferred to the recipient tripled, as in the investment game. Here there is no difference (5.6 versus 5.1) in the amounts sent to male Ashkenazic and Eastern recipients by male dictators, indicating that there is no *per se* taste for discrimination, but rather that the investment game results were driven by (mistaken) beliefs about the expected returns from each ethnic group. Finally, an ultimatum game was conducted to test for the stereotype that Eastern Jews are driven more by a sense of "honor" and so would be more likely to reject small proposals. In fact, Eastern male responders received significantly higher proposals (8.4 versus 5.9) than did Ashkenazic male responders, even though rejection rates for lopsided proposals were nearly the same for both groups. Overall, the observed discrimination in this paper appears to reflect stereotypes rather than rational statistical discrimination.

Fershtman, Gneezy, and Verboven (2005) attempt to distinguish between discrimination against people and "nepotism" in favor of one's own group. At issue is

the notion of anonymity rules, which forbid disclosure of group affiliation or background. Using the investment game, they compare behavior when group identity is fully observable and when it is not. The paper considers two different societies: Belgian (Walloons versus Flemish) and Israeli (ultra-orthodox religious versus secular). Belgian society shows evidence of discrimination, as both Walloons and Flemish sent much more to the responder when he or she was of the same background as the first mover than when the responder was of the other background (1200 versus 745 and 1009 versus 536, respectively). On the other hand, transfers made to anonymous responders were about the same as transfers to own-group responders for both Walloon and Flemish first-movers. However, a very different pattern was observed in Israeli society. In this case, ultra-orthodox religious Jews favor members of their own group, but send similar amounts to both secular Jews and anonymous responders.

c) Beauty

Does one's level of attractiveness matter in terms of hiring practices and behavior in the workplace? Many people feel that good-looking people are highly advantaged and in some occupations (e.g., pharmaceutical representatives selling to physicians) looks are essential for being hired. Job applicants are routinely counseled to look their best for job interviews. Some recent experimental studies have tested this issue in the laboratory.

Mobius and Rosenblat (2006) investigate the "beauty premium" in experimental labor market in which firms choose wages for workers in a real-effort task involving solving mazes. Physical attractiveness is found to be uncorrelated with skill in this task. The experimental design varies the degree of oral and visual interaction between firms and workers (and also elicits worker confidence). In the baseline, the firm only observes the "resume" of the worker, which includes the labor-market characteristics of the worker; in the visual treatment, the firm also sees a photograph of the worker; in the oral treatment, the firm sees the resume and also has a telephone conversation with the worker; in the face-to-face treatment, this conversation is face-to-face. They find a substantial beauty premium, which they decompose into three transmission channels: Physically attractive workers are more confident and higher confidence increases wages, these workers are also (incorrectly) considered to be more capable by firms, and these workers also have better oral skills that raise their wages.

Wilson and Eckel (2006) also investigate the beauty premium, using the investment game and photographs. They find a modest beauty premium, in that the average amount sent to an "unattractive" ("attractive") counterpart was $4.64 ($4.98); attractive trustees appear to reciprocate this trust. On the other hand, there is also a "beauty penalty" for attractive first movers, as responders return 35% (30%) of the amounts sent by "unattractive" ("attractive") first movers. It appears that this is driven by disappointed expectations, as responders withhold repayment in this case. Furthermore, responders expect more from attractive first movers, so their expectations are unmet more frequently and so this withholding is exacerbated when the first mover who disappointed the responder is attractive.

Andreoni and Petrie (2008) study the beauty premium in a public-goods game. When people are not told the contributions of others, there is indeed a beauty premium: attractive people earn 7% more than people of intermediate attractiveness and 12% more than unattractive people, even though attractive people contribute approximately the same average amount as others. In another treatment, participants are told the individual contributions of each other player. In this case, the beauty premium becomes a beauty penalty, as it appears (as in Wilson and Eckel) that people expect (hope for?) more from attractive people. As these expectations are more likely to be unmet, attractive people seem relatively more selfish, leading to decreased contributions from other participants. There are also some gender effects, as men make larger contributions and this appears to "lead" other people to contribute more in subsequent periods.

d) Affirmative action

A thorny policy issue is that of affirmative action, whereby some group that is considered to be generally disadvantaged is given an immediate advantage. Emotions run high on this topic, and "reverse discrimination" lawsuits are not uncommon. What are the effects of affirmative action? Clearly this helps disadvantaged people (at least in the short run), but it may well be detrimental with respect to overall performance. Some recent laboratory experiments have investigated this issue in controlled settings.

Schotter and Weigelt (1992) find, in a two-person tournament setting, that the effects of affirmative action depend on the severity of the cost disadvantage for the non-minority group. When the tournament was "unfair", one of these people had to win by a certain amount and/or was assigned a cost function greater than that for the other person. Affirmative action is found to discourage disadvantaged agents from "dropping out", as was observed without affirmative action. Nevertheless, overall the effect of affirmative action on output appears to depend on the degree of disadvantage. When this is large, then affirmative-action programs did improve profits for the "tournament administrators"; however, this did not occur when the cost disadvantage was less severe.

Corns and Schotter (1999) provide an 'existence proof' that there exist parameters under which price-preference auctions (often used to promote minority representation in government contracting) can increase minority representation while being cost effective; however, there is a decrease in cost effectiveness if the degree of price preference is too large. The experimental design involves high-cost firms (simulating minority firms) that receive either a 0%, 5%, 10%, or 15% price preference. There were six participants in each session, with costs drawn from a uniform distribution of [100,200] for the four low-cost firms and [110, 220] for the two high-cost firms. The frequency of a high-cost firm winning the auction increased from 12% in the 0% treatment to 43% in the 15% treatment, largely in line with the theoretical predictions. The average observed price was 121.24 in the 0% condition, 119.29 in the 5% condition, 122.84 in the 10% condition, and 124.41 in the 15% condition. There is no significant difference in the distribution of drawn costs across any two price-preference treatments. Nevertheless, the

results indicate that it is possible to use price preferences to increase representation of "minority" firms without increasing the overall cost or the average observed price.

Niederle, Segal, and Vesterlund (2008) follow up the Niederle and Vesterlund (2007) study by examining how affirmative action affects competitive entry and performance. Given that women shy away from competition, it may be advantageous to institute an affirmative-action policy under which women are favored in competitions. Participants are formed into groups of six, including three men and three women, and add up a series of five two-digit numbers; in one task, they are paid $0.50 for each correct answer, while in another task the two best performers are paid $1.50 for each correct answer. In the third task, people choose either the piece-rate or the tournament payment scheme. Next there is an affirmative-action task, in which the two winners in a tournament are the highest-performing woman and the highest performer of the other people in the group. The results indicate that when women guaranteed equal representation among winners, more women and fewer men enter competitions. Results also suggest that affirmative action need not be costly, but may be sensitive to parameters, as in Corns and Schotter (1999). Measured beliefs about one's rank and attitudes concerning competition are affected by the affirmative-action policy. While affirmative action may result in reverse discrimination towards men, the authors claim: "…this need not be the outcome when competitive entry is not payoff maximizing. The response in entry implies that it may not be necessary to lower the performance requirement for women to achieve a more diverse set of winners." Thus, affirmative-action policies may potentially be beneficial.

Balafoutas and Sutter (2009) investigate three alternative policy interventions that shall promote women in competitive environments: *Preferential Treatment*, *Repetition of the Tournament*, and *Affirmative Action*. These environments are compared with respect to the willingness of men and women to enter a tournament (addition exercises) and the impact on both tournament efficiency (selecting the most qualified people to be the winners) and post-tournament efficiency when groups then participate in a minimum-effort coordination game (the minimum effort translates into the efficiency level). Affirmative-action policy interventions lead to significantly higher entrance rates of women, expect for in the *Repetition* treatment. Tournament efficiency does not appear to deteriorate when these policies are introduced; in fact, a "small push" (giving women credit for one additional solved exercise) leads to the highest efficiency. Furthermore, efficiency in a post-tournament teamwork-task is slightly higher in the treatments with successful policy interventions; thus, to the extent that this experimental game serves as a proxy for dynamic considerations, cooperation is not harmed by the introduction of affirmative action.

Calsamiglia, Franke, and Rey-Biel (2009) consider affirmative action in two-player tournaments involving solving mini-Sudoku puzzles. Elementary school children at two otherwise-identical private schools in Spain were paired, with the person solving more (net) puzzles winning the tournament. In one of these schools, the students had experience with these Sudoku puzzles. These policies included adding some number of points to the disadvantaged students total and multiplying the number of solved puzzles

by disadvantaged students by some factor greater than one. Their results indicate that performance was not reduced for either advantaged or disadvantaged subjects and that it was in fact enhanced. Additionally, while affirmative action balanced the proportion of disadvantaged individuals winning their respective tournament, the average performance of the pool of winners only decreased slightly.

VI. Conclusions

Laboratory experiments have now been used to study almost every aspect of labor economics, ranging from the effects of final-offer arbitration in public sector labor disputes to the causes of gender wage differentials. This vast literature includes a number of noteworthy 'success stories', including the use of laboratory experiments as testing grounds for proposed institutional innovations --in particular, matching mechanisms for professional workers, and arbitration systems in public sector union bargaining-- that are subsequently used to organize the actual pricing and exchange of labor.

We see laboratory experiments as a useful tool among several at the labor economist's disposal; their key strengths include the ability to control conditions and confounding factors affecting a possible causal relationship more tightly than any other method, the ability to replicate results quickly and easily with newly collected data, the unique ease with which they can test the exact quantitative predictions of simple game-theoretic models, the ability to cleanly study behavior in situations where precise theoretical predictions are absent (as in the case of multiple equilibria), their low cost, the ability to more readily study phenomena that are hidden in the field because they are illegal or disapproved of (such as sabotage, discrimination, and spite), and the ability to elicit data relatively easily on such central game-theoretic concepts as beliefs and strategies that are often difficult to infer from observed behavior alone.

Key objections that have been raised to lab experiments in economics include the artificiality of the experimental context in general and the labor 'task' specifically, low stakes, short duration, and the effects of experimenter scrutiny. All of these are important issues, which experimenters should (and typically do) consider seriously. As we have argued in detail in this review, however, none of these concerns –with the possible exception of duration-- are insurmountable in the lab, and in general these disadvantages need to be weighed against the lab's potential benefits, enumerated above. In many cases a combination of methods, or a hybrid method --such as the artefactual field experiments and framed field experiments described by List (2009)-- may be the best approach.

Perhaps the most widespread concern with lab experiments is the representativeness of the laboratory population relative to the field population of interest; this concern seems particularly acute when discussing lab results that are hard to explain without some recourse to 'non-standard' preferences. As we have noted, this issue can be addressed in a number of ways, including conducting lab experiments directly on the field population of interest (e.g. Chinese central planners), and studying the selection process itself (for example, selection into competitive versus co-operative work environments, or into situations where altruism might be expected or not) in the lab.

Another important option, of course, is to move the entire experiment into the field, using the lab only as a pre-test or not at all.

Greater clarity on 'representativeness' can, in our opinion, be also achieved if future investigators are more precise about *which* field environment their experiment is intended to represent. Clearly, if the intended field population is highly specific (for example, new registrants for unemployment insurance in Illinois), and the researcher's main interest is in the *magnitude* of the behavioral response to a contemplated policy change affecting this population, the ideal research design is almost certainly what List (2009) calls a 'natural field experiment' on that population. However, when searching for underlying principles of behavior, derived from a specific theoretical model of strategic interaction and/or social preferences that one hopes might apply to a broader field population (for example, all college-educated employees, or even all humans), a lab experiment on a random sample from a broader group may be more useful. In all cases, the research question should guide the choice of method, and the investigator should specify the population on which the experiment is intended to shed light.

Probably the most frequent use of lab experiments in labor economics has been to principal-agent interactions between workers and firms. While many predictions of what one might call 'traditional' principal-agent theory have been confirmed in the lab --for example, in most cases effort increases when marginal work incentives are raised, contributions to team production fall with the number of team members, and ratchet effects disappear when *ex post* labor market competition is intensified--, lab experiments have also identified a significant number of robust departures from rational, selfish behavior in principal-agent interactions. As already noted, a catch-all phrase for the modifications to traditional preference structures that have been proposed to explain these 'anomalies' is "social preferences"; the discovery and mapping of social preferences is probably the most important contribution of lab experiments to principal-agent analysis, and to labor economics more generally.[96]

That said, much remains unknown about social preferences; several competing theoretical models exist and little is yet known about which models are most representative. On the theoretical side, one area in which progress is needed is in the development of social preference models that apply to large, multi-worker workplaces. Most current social preference models have been developed to explain behavior in the lab, where (for example) one college student representing an employer gives a 'gift' to another college student. While inequity aversion might explain behavior in these situations, it is hard to see how it (and some other versions of social preferences) applies to interactions between workers and large multi-worker firms. On the empirical side, social preferences seem capable of changing their shape and form considerably between different contexts, as well as between the lab and the field. Empirical work that sheds

[96] Non-selfish preferences are of course not the only possible explanation of departures from predicted equilibria in effort-supply models. Others include simple limits to human cognition (especially in complex, multi-period games), and practical difficulties of coordinating on Nash equlibrium even when a unique one exists.

light on *which* social preferences (including, among others, inequity aversion, reciprocity, guilt, spite, jealousy and simple altruism) come into play *when* will be of great value.

While much remains to be learned about the effects of compensation systems on worker behavior within firms, it is perhaps worth gauging our progress by considering the state of knowledge on this topic some 15 years ago in the words of some of the main contributors (emphasis added):

> "There is a large and growing interest in the economic theory of the internal workings of firms. However, this literature is based on *very little data* and *limited stylized facts.*" (Baker, Gibbs and Holmstrom 1994a.)

> "For a time there was considerable excitement about implicit contract theory… That literature soon fell out of favor, but in its place came more refined information economic analyses that viewed wage contracts as optimal responses to asymmetries in information between employees and firms. With this, the logical possibilities for explaining wage behavior grew dramatically. Today's large variety of models and modeling options has put *theory well ahead of observation.*" (Baker, Gibbs and Holmstrom 1994b).

And finally, from an article titled "Internal Labor Markets: Too Many Theories, Too Few Facts":

> "With the advent of information economics and contract theory, models of internal labor markets --or at least selected features of these markets-- have begun to emerge. The objective of these theories is to show that internal-labor-market outcomes can be construed as second-best solutions to contracting problems under incomplete information. … At this point, *there is hardly any feature of internal labor markets that cannot be given some logical explanation using the right combination of uncertainty, asymmetric information and opportunism.* (Baker and Holmstrom 1995).

Thus, just 15 years ago, research on incentives in organizations consisted to a disappointingly large extent of (a) identifying a 'paradoxical' feature of some internal labor market (for example, academic tenure, or pay raises based purely on seniority), then (b) crafting a theoretical model showing how that feature was in fact an efficient response to some contracting problem. Most analyses stopped there.

Today, thanks both to a laboratory literature that began in the mid-1990s and a more recent surge in field experiments, we now face a cornucopia (some might say, an unmanageable deluge) of facts about agents' and principals' behavior in the above models. Indeed, after reviewing this literature, it is tempting to conclude that we now have "Too Many Facts, Too Few Theories". Certainly, an important part of the road ahead is the development of models of social preferences (or of other causes of the robust departures from the predictions of 'traditional' principal-agent models) that account well not only for those departures in the lab but apply well to the sorts of employment relationships we see most often in the field. In our view, a variety of empirical tools including both lab and field experiments, natural experiments, econometric studies, and calibration of structural models will provide the necessary empirical discipline in this next phase of research on one of the most basic issues in labor markets: the exchange of effort for pay within firms.

References

Abdulkadiroglu, Atila; Parag A. Pathak, and Alvin E. Roth. "Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match" *American Economic Review*. Vol. 99 (5). p 1954-78. December 2009

Abeler, Johannes, Armin Falk, Lorenz Goette and David Huffman (2009), "Reference Points and Effort Provision", IZA Discussion Paper No. 3939 (Institute for the Study of Labor). (January)

Abeler, Johannes, Steffen Altmann, Sebastian Kube and Matthias Wibral (2009), "Gift Exchange and Workers' Fairness Concerns: When Equality Is Unfair", *Journal of the European Economic Association* (Forthcoming).

Abrams, Eric, Martin Sefton and Abdullah Yavas. "An Experimental Comparison of Two Search Models" *Economic Theory.* Vol. 16 (3). p 735-49. November 2000

Acemoglu, Daron, Michael Kremer and Atif Mian. "Incentives in Markets, Firms, and Governments" *Journal of Law, Economics, and Organization*. Vol. 24 (2). p 273-306. October 2008.

Akerlof, George A. "The Economics of Caste and of the Rat Race and Other Woeful Tales". *Quarterly Journal of Economics* Vol. 90 (4). p 599-617. November 1976.

Akerlof, George A. (1982), "Labor Contracts as Partial Gift Exchange", The Quarterly Journal of Economics 97(4), 543-69.

Akerlof, Geoge A. and Janet L. Yellen (1988), "Fairness and Unemployment", *American Economic Review* 78(2), 44-49.

Akerlof, George and Janet Yellen. Akerlof, G. and J. Yellen (1990), "The Fair-Wage Effort Hypothesis and Unemployment," *Quarterly Journal of Economics*, 105(2): 255-284.

Altonji, J. and C. Pierret. "Employer Learning and Statistical Discrimination" *Quarterly Journal of Economics*, Vol. 116, No. 1. (Feb., 2001), pp. 313-350

Al-Ubaydli, Omar, Steffen Andersen, Uri Gneezy and John List (2006), "For Love or Money? Testing Non-pecuniary and Pecuniary Incentive Schemes in a Field Experiment", Unpublished Paper (George Mason University).

Anderhub, Vital, Manfred Königstein and Dorothea Kübler (2003), "Long-term Work Contracts versus Sequential Spot Markets: Experimental Evidence on Firm-specific Investment", *Labour Economics* 10(4): 407-425. (August)

Anderhub, Vital, Simon Gächter and Manfred Königstein (2002), "Efficient Contracting and Fair Play in a Simple Principal-Agent Experiment", *Experimental Economics* 5(1): 5-27. (June)

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau and Elisabet Rutström (2008), "Eliciting Risk and Time Preferences", *Econometrica* 76(3): 583-618. (May)

Andreoni, James, Paul Brown, and Lise Vesterlund (2002), "What Makes an Allocation Fair? Some Experimental Evidence", *Games and Economic Behavior* 40(1): 1-24. (July)

Andreoni, James; Marco Castillo and Ragan Petrie. "What Do Bargainers' Preferences Look Like? Experiments with a Convex Ultimatum Game" *American Economic Review*. Vol. 93 (3). p 672-85. June 2003.

Andreoni, James and Ragan Petrie (2008), "Beatuy, Gender and Stereotypes: Evidence from Laboratory Experiments", *Journal of Economic Psychology* 29(1): 73-93. (February)

Ariely, Dan; Axel Ockenfels and Alvin E. Roth. "An Experimental Analysis of Ending Rules in Internet Auctions" *RAND Journal of Economics*. Vol. 36 (4). p 890-907. Winter 2005.

Ariely, Dan, Uri Gneezy, George Loewenstein and Nina Mazar. "Large Stakes and Big Mistakes". Unpublished paper, Duke University, August 20, 2008.

Ashenfelter, Orley, Janet Currie, Henry S. Farber and Matthew Spiegel (1992), "An Experimental Comparison of Dispute Rates in Alternative Arbitration Systems", *Econometrica* 60(6): 1407-1433.

Avery, Christopher; Christine Jolls, Richard Posner and Alvin E. Roth. "The New Market for Federal Judicial Law Clerks" National Bureau of Economic Research, Inc, NBER Working Papers: 13213. 2007.

Babcock, Linda, and George Loewenstein. "Explaining Bargaining Impasse: The Role of Self-Serving Biases" *Journal of Economic Perspectives*. Vol. 11 (1). p 109-26. Winter 1997.

Babcock, Linda and Sara Laschever. *Women Don't Ask: Negotiation and the Gender Divide* Princeton: Princeton University Press, 2003.

Baker, George, Robert Gibbons, and Kevin J. Murphy. "Subjective Performance Measures in Optimal Incentive Contracts". *Quarterly Journal of Economics* 109 (1994): 1125-56.

Baker, George, Michael Gibbs, and Bengt Holmstrom. "The Internal Economics of the Firm: Evidence from Personnel Data" *Quarterly Journal of Economics* 109, No. 4 (Nov., 1994a), pp. 881-919.

Baker, George, Michael Gibbs, and Bengt Holmstrom. "The Wage Policy of a Firm" *Quarterly Journal of Economics*, Vol. 109, No. 4 (Nov., 1994b), pp. 921-955

Baker, George, and Bengt Holmstrom. "Internal Labor Markets: Too Many Theories, Too Few Facts" *American Economic Review*. Vol. 85 (2). p 255-59. May 1995.

Balafoutas, L. and M. Sutter (2009), "Gender, competition and the effects of policy interventions," mimeo.

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

Bandiera, Oriana, Iwan Barankay and Imran Rasul. "Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment" *Quarterly Journal of Economics.* May 2007; 122(2): 729-73

Bandiera, Oriana, Iwan Barankay and Imran Rasul. "Social Connections and Incentives in the Workplace: Evidence from Personnel Data" *Econometrica*, Vol. 77, No. 4 (July 2009a), 1047–1094

Bandiera, Oriana, Iwan Barankay and Imran Rasul. "Team Incentives: Evidence from a Field Experiment" unpublished paper, University of Pennsylvania, October 2009b.

Barofsky, I. and Hurwitz, D. July 1968. Within ratio responding during fixed ratio performance. *Psychonomic Science* 11(7):263-264.

Barr, Abigail and Pieter Semeels (2009), "Reciprocity in the Workplace", *ExperimentalEconomics* 12(1): 99-112. (March)

Battalio, R.C., Green, L., and Kagel, J. H. ""Income-Leisure Tradeoffs of Animal Workers" *American Economic Review* 71 (1981): 621-32.

Battalio, R.C., and Kagel, J. H. "Consumption-Leisure Tradeoffs of Animal Workers: Effects of Increasing and Decreasing Marginal Wage Rates in a Closed Economy Experiment" in Vernon L. Smith, ed. *Research in Experimental Economics*, volume 3. Greenwich, Conn.: JAI Press (1985): pp. 1-30.

Battigalli, Pierpaolo and Martin Dufwenberg (2007), "Guilt in Games", *American Economic Review Papers & Proceedings* 97: 170-76.

Battigalli, Pierpaolo and Martin Dufwenberg (2009), "Dynamic Psychological Games", *Journal of Economic Theory* 144: 1-35.

Bauernschuster, Stefan, Peter Dürsch, Jörg Oechssler and Radovan Vadovic (2009), "Mandatory Sick Pay Provision: A Labor Market Experiment", Jena Economic Research Papers in Economics 2009-076, Friedrich-Schiller-University Jena, Max-Planck-Institute of Economics.

Becker, Gary. *Human Capital.* New York: National Bureau of Economic Research, 1964.

Belot, Michele, and Jeroen van de Ven. Friendships and Favoritism at School - Evidence from the field" unpublished paper, University of Amsterdam, June 15, 2009

Bellemare, Charles and Sabine Kröger (2007), "On Representative Social Capital", *European Economic Review* 51(1): 181-202

Bellemare, Charles, Sabine Kröger and Arthur van Soest (2008), "Measuring Inequity Aversion in a Heterogenous Population Using Experimental Decsisions and Subjective Probabilities", *Econometrica* 76(4): 815-839. (July)

Bellemare, Charles and Bruce Shearer(2009), "Gift Giving and Worker Productivity: Evidence from a Firm-Level Experiment", *Games and Economic Behavior* 67(1): 233-44. (September)

Bellemare, Charles, Patrick Lepage and Bruce Shearer (2009), "Peer Pressure, Incentives and Gender: An Experimental Analysis of Motivation in the Workplace", Discussion Paper No. 3948 (Institue for the Study of Labor). (January)

Berg, Joyce, John Dickhaut and Kevin McCabe (1995), "Trust, Reciprocity and Social History", *Games and Economic Behavior* 10: 122-42.

Bernhardt, D. and D. Scoones, "Promotion, Turnover, and Preemptive Wage Offers" *American Economic Review* 83 (4) (Sep., 1993): 771-791.

Berninghaus, S., et al. "Decentralized versus Collective Bargaining--An Experimental Study" *International Journal of Game Theory*. Vol. 30 (3). p 437-48. 2001.

Berninghaus, Siegfried; Guth, Werner; Keser, Claudia. "Unity Suggests Strength: An Experimental Study of Decentralized and Collective Bargaining" *Labour Economics.* Vol. 10 (4). p 465-79. August 2003.

Binmore, Ken and Partha Dasgupta, editors [1987], *The Economics of Bargaining,* Blackwell, Oxford.

Bolton, Gary E. and Elena Katok (1998), "Reinterpreting Arbitration's Narcotic Effect: An Experimental Study of Learning in Repeated Bargaining", *Games and Economic Behavior* 25: 1-33.

Bénabou, Roland, and Jean Tirole. 2003. "Intrinsic and Extrinsic Motivation." *Review of Economic Studies*,70(3): 489–520.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review*, 96(5): 1652–78

Berger, Jonah, and Devin G. Pope. "Can Losing Lead to Winning?" unpublished paper, The Wharton School, University of Pennsylvania, August 2009.

Blount, Sally (1995), "When social outcomes aren't fair: The effect of causal attributions on preferences", *Organizational Behavior and Human Decision Processes* 63: 131-144.

Bolton, Gary (1991), "A comparative model of bargaining: Theory and evidence", *American Economic Review* 81: 1096-1136.

Bolton, Gary and Axel Ockenfels (2000), "ERC: A theory of equity, reciprocity and competition," *American Economic Review* 90: 166-193.

Bolton, Gary E. and Elena Katok (1998), "Reinterpreting Arbitration's Narcotic Effect: An Experimental Study of Learning in Repeated Bargaining", *Games and Economic Behavior* 25: 1-33.

Boning, Brent, Casey Ichniowski, and Kathryn Shaw "Opportunity Counts: Teams and the Effectiveness of Production Incentives" *Journal of Labor Economics*. October 2007; 25(4): 613-50

Booth, Alison L. and Patrick J. Nolen (2009a), "Gender Differences in Risk Behavior: Does Nurture Matter?", Discussion Paper No. 4026 (Institue for the Study of Labor). (March)

Booth, Alison L. and Patrick J. Nolen (2009b), "Choosing to Compete: How Different Are Girls and Boys?", Disucssion Paper No. 4027 (Institute for the Study of Labor). (March)

Borghans, Lex, Bart H. H. Golsteyn, James H. Heckman and Huub Meijers (2009), "Gender Differences in Risk Aversion and Ambiguity Aversion", *Journal of the European Economic Association* 7(2-3): 649-658. (April)

Bornstein, Gary, Uri Gneezy and Rosmarie Nagel (2002), "The Effect of Intergroup Competition on Group Coordination: An Experimental Study", *Games and Economic Behavior* 41(1): 26-45. (October).

Brams, Steven J. and Samuel Merrill, III (1983), "Equilibrium Strategies for Final- Offer Arbitration: There Is No Median Convergence", *Management Science 2*9(8): 927-941.

Brams, Steven J. and Samuel Merrill, III (1986), "Binding Versus Final-Offer Arbitration: A Combination Is Best", *Management Science* 32(10): 1346-1355.

Brandts, Jordi and Gary Charness (2003), "Truth or Consequences: An Experiment", *Management Science* 49: 116-130. (January)

Brandts, Jordi and Gary Charness (2004), "Do Labour Market Conditions Affect Gift Exchange? Some Experimental Evidence", *Economic Journal* 114(497): 684- 708. (July)

Brandts, Jordi and David J. Cooper (2007), "It's What You Say, Not What You Pay: An Experimental Study Of Manager-employee Relationships in Overcoming Coordination Failure", *Journal of the European Economic Association* 5(6): 1223-1268. (October)

Braunstein, Yale. and Andrew Schotter (1982), "Labor Market Search: An Experimental Study", *Economic Inquiry* 20(1): 133-144. (January)

Bronars, S. (1986): "Strategic Behavior in Tournaments" Texas A&M University.

Brosig, Jeannette; Weimann, Joachim; Yang, Chun-Lei. "The Hot versus Cold Effect in a Simple Bargaining Experiment" *Experimental Economics.* Vol. 6 (1). p 75-90. June 2003.

Brosig, Jeannette, Weimann, Joachim, Yang, Chun-Lei Communication, Reputation, and Punishment in Sequential Bargaining Experiments. *Journal of Institutional and Theoretical Economics.* Vol. 160 (4). p 576-606. December 2004.

Brown, K., Harlow, W., Starks, L., 1996. Of tournaments and temptations: an analysis of managerial fund incentives in the Mutual Fund Industry. *Journal of Finance* 51, 85– 110

Brown, Martin, Armin Falk and Ernst Fehr (2004), "Relational Contracts and the Nature of Market Interactions", *Econometrica* 72(3): 747-780.

Brown, Martin, Armin Falk and Ernst Fehr (2008), "Competition and Relational Contracts: The Role of Unemployment As a Disciplinary Device", Working Paper No. 3345 (Institute for the Study of Labor). (May)

Brown, Meta, Chris Flinn and Andrew Schotter. "Real Time Search in the Laboratory and the Market" *American Economic Review*, forthcoming.

Bruggen, Alexander and Martin Strobel (2007), "Real Effort versus Chosen Effort in Experiments", *Economics-Letters* 96(2): 232-236. (August)

Bull, Clive, Andrew Schotter and Keith Weigelt (1987), "Tournaments and Piece Rates: An Experimental Study", *Journal of Political Economy* 95(1): 1-33. (February)

Burchett, Richard and John Willoughby (2004), "Work Productivity When Knowledge of Different Reward Systems Varies: Report from an Economic Experiment", *Journal of Economic Psychology* 25(5): 591-600. (October)

Burks, Stephen, Jeffrey Carpenter and Lorenz Goette (2009), "Performance Pay Worker Cooperation: Evidence From an Artefactual Field Experiment", *Journal of Economic Behavior and Organization* 70(3): 458-469. (June)

Burks, Stephen, Lorenz Goette and Aldo Rustichini (2009), "Cognitive Skills Explain Economic Preferences, Strategic Behavior and Job Attachment", *Proceedings of the National Academy of Sciences* 106(19): 7745-7750. (May)

Butler, Richard J. and Ronald G. Ehrenberg (1981), "Estimating the Narcotic Effect of Public Sector Impasse Procedures", *Industrial & Labor Relations Review* 35(1): 3-20.

Cabrales, Antonio, Gary Charness and Marie-Claire Villeval (2006), "Competition, Hidden Information and Efficiency: An Experiment", Discussion Paper No. 2296 (Institute for the Study of Labor). (September)

Cadsby, C. Bram, Fei Song and Francis Tapon (2007), "Sorting and Incentive Effects of Pay for Performance: An Experimental Investigation", *Academy of Management Journal* 50(2): 387-405.

Cadsby, C. Bram, Fei Song and Francis Tapon (2008), "Are You Paying Your Employess to Cheat? An Experimental Investigation", Working Paper No. 2008-10 (University of Guelph)

Cadsby, C. Bram, Fei Song and Francis Tapon (2009) "The Impact of Risk Aversion and Stress on the Incentive Effect of Performance Pay". unpublished paper, University of Guelph, 2009.

Calsamiglia, C., J. Franke, and P. Rey-Biel (2009), "The Incentive Effects of Affirmative Action in a Real-Effort Tournament," mimeo.

Camerer, C. and D. Lovallo. "Overconfidence and Excess Entry" American Economic Review 89 (1999): 306-318

Camerer, Colin, and Roberto Weber. "Experimental Organizational Economics." In *The Handbook of Organizational Economics*. Edited by R. Gibbons and J. Roberts. Princeton, NJ: Princeton University Press, forthcoming.

Cameron, Lisa A. [1999], Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia," *Economic Inquiry*, 37,1,January, 47-59.

Carbone, Enrica; Hey, John D. "The Effect of Unemployment on Consumption: An Experimental Analysis" *Economic Journal*. Vol. 114 (497). p 660-83. July 2004.

Card, David "Strikes and Wages: A Test of an Asymmetric Information Model" *Quarterly Journal of Economics*. Vol. 105 (3). p 625-59. August 1990.

Carmichael, H. Lorne. "The Agent-Agents Problem: Payment by Relative Output" *Journal of Labor Economics*. Vol. 1 (1). p 50-65. January 1983.

Carmichael, H. Lorne. "Firm-Specific Human Capital and Promotion Ladders" *Bell Journal of Economics*. Vol. 14 (1). p 251-58. Spring 1983.

Carmichael, L. and W. B. MacLeod, "Worker Cooperation and the Ratchet Effect". *Journal of Labor Economics* 18 (January 2000): 1-19.

Carmichael, H. Lorne and W. Bentley MacLeod. "Multiskilling, Technical Change and the Japanese Firm" *Economic Journal*. Vol. 103 (416). p 142-60. January 1993.

Carpenter, Jeffrey; Rudisill, McAndrew. "Fairness, Escalation, Deference, and Spite: Strategies Used in Labor-Management Bargaining Experiments with Outside Options" *Labour Economics*. Vol. 10 (4). p 427-42. August 2003.

Carpenter, Jeffrey, Peter Matthews and John Schirm. "Tournaments and Office Politics: Evidence from a real effort experiment" *American Economic Review*, forthcoming.

Carpenter, Jeffrey and Erica Seki (forthcoming), "Do Social Preferences Increase Productivity? Field Experimental Evidence from Fisherman in Toyama Bay", *Economic Inquiry*.

Cesarini, David, Christopher T. Dawes, Magnus Johannesson, Paul Lichtenstein and Björn Wallace (2009), "Genetic Variation in Preferences for Giving and Risk Taking", *Quarterly Journal of Economics* 124(2): 809-842. (May)

Charness, Gary, and Matthew Rabin. "Understanding Social Preferences with Simple Tests" *Quarterly Journal of Economics*. Vol. 117 (3). p 817-69. August 2002.

Charness, Gary (2004), "Attribution and Reciprocity in an Experimental Labor Market", *Journal of Labor Economics* 22(3): 665-688. (July)

Charness, Gary and Martin Dufwenberg (2006), "Promises and Partnership", *Econometrica* 74(6): 1579-1601. (November)

Charness, Gary; Corominas-Bosch, Margarida; Frechette, Guillaume R Bargaining and Network Structure: An Experiment. *Journal of Economic Theory*. Vol. 136 (1). p 28-65. September 2007.

Charness, Gary and Martin Dufwenberg (2008), "Contracts and Communication", Unpublished Paper (University of California, Santa Barbara).

Charness, G. and U. Gneezy (2009), "Strong Evidence for Gender Differences in Experimental Investing, mimeo.

Charness, Gary and Peter Kuhn (2007), "Does Pay Inequality Affect Worker Effort? Experimental Evidence", *Journal of Labor Economics* 25(4): 693-724. (October)

Charness, Gary and David I. Levine (2007), "Intention and Stochastic Outcomes: An Experimental Study", *Economic Journal* 117(522): 1051-1072. (July)

Charness, Gary and Marie-Claire Villeval (2009), "Cooperation and Competition in Intergenerational Experiments in the Field and Laboratory", *American Economic Review* 99(3): 956-978. (June)

Charness, Gary, Peter Kuhn and Marie-Claire Villeval (2008), "Competition and the Ratchet Effect", Discussion Paper No. 3784 (Institute for the Study of Labor). (November)

Charness, Gary, and Chun-Lei Yang. "Endogenous Group Formation and Public Goods Provision: Exclusion, Exit, Mergers, and Redemption" Department of Economics, UC Santa Barbara, working paper no. 13-08. 2008.

Charness, Gary and Martin Dufwenberg (2009), "Hidden Information, Participation, and Communication", Unpublished Paper (University of California, Santa Barbara).

Charness, Gary, Guillaume Fréchette, and John Kagel (2004), "How Robust is Laboratory Gift Exchange?", *Experimental Economics* 7: 189-205. (June)

Chaudhuri, Ananish (1998), "The Ratchet Principle in a Principal with Unknown Costs: An Experimental Analysis", *Journal of Economic Behavior and Organization* 37(3): 291-304. (November)

Chaudhuri, Ananish, Sara Graziano, and Pushkar Maitra, Sara Graziano, and Pushkar Maitra. ""Social Learning and Norms in an Experimental Public Goods Game with Inter-Generational Advice" *Review of Economic Studies*, April 2006, 73(2), 357-380

Chaudhuri, Ananish, "Conditional Cooperation and Social Norms in Public Goods Experiments: A Survey of the Literature", Unpublished Manuscript, University of Auckland, August 2007.

Chen, Kang, Tang, Fang-Fang. Cultural Differences between Tibetans and Ethnic Han Chinese in Ultimatum Bargaining Experiments. *European Journal of Political Economy*. Vol. 25 (1). p 78-84. March 2009.

Chen, Yan and Tayfun Sonmez. "School Choice: An Experimental Study". *Journal of Economic Theory*. Vol. 127 (1). p 202-31. March 2006.

Chen, Yan and Tayfun Sonmez.. "Improving Efficiency of On-Campus Housing: An Experimental Study." *American Economic Review*. Vol. 92 (5). p 1669-86. December 2002.

Chevalier, J. A. and Ellison, G. D. (1997): Risk Taking by Mutual Funds as a Response to Incentives. *Journal of Political Economy*, 105, pp. 1167—1200.

Cooper, David J., John H. Kagel, Wei Lo and Qing Liang Gu (1999), "Gaming Against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers", *American Economic Review* 89(4): 781-804. (September)

Corns, Allan and Andrew Schotter (1999), "Can Affirmative Action Be Cost Effective? An Experimental Examination of Price-Preference Auctions", *American Economic Review* 89(1): 291-305. (March)

Cox, James C., and Ronald L.Oaxaca. "Laboratory Experiments with a Finite-Horizon Job-Search Model" *Journal of Risk and Uncertainty*. Vol. 2 (3). p 301-29. September 1989.

Cox, James C., and Ronald L.Oaxaca. "Direct Tests of the Reservation Wage Property" *Economic Journal* Vol. 102 (415). p 1423-32. November 1992.

Cox, James C., and Ronald L.Oaxaca. "Good News and Bad News: Search from Unknown Wage Offer Distributions" *Experimental Economics*. Vol. 2 (3). p 197-225. March 2000.

Cox, James C. (2004), "How to Identify Trust and Reciprocity", *Games and Economic Behavior* 46: 260-281.

Cox, James C. and Cary Deck (2005), "On the Nature of Reciprocal Motives", *Economic Inquiry* 43(3): 623-635.

Cox, James C. and Cary Deck (2006), "When Are Women More Generous than Men?", *Economic Inquiry* 44: 587-598.

Cox, James C., Klarita Sadiraj, and Vjollca Sadiraj (2008), "Implications of Trust, Fear, and Reciprocity for Modeling Economic Behavior", *Experimental Economics* 11: 1-24.

Cox, James C., Daniel Friedman and Vjollca Sadiraj (2008), "Revealed Altruism", *Econometrica* 76(1): 31-69. (January)

Cramton, P.C. and Tracy, J.S., 1992. Strikes and holdout in wage bargaining: theory and data. *American Economic Review* 81, pp. 100–121.

Crawford, Vincent P. (1979), "On Compulsory-Arbitration Schemes", *Journal of Political Economy* 87(1): 131-159.

Crawford, V. & Meng, J. (2008), `New york city cabdrivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income', UCSD Discussion Paper

Croson, Rachel; Boles, Terry; Murnighan, J. Keith. Cheap Talk in Bargaining Experiments: Lying and Threats in Ultimatum Games. *Journal of Economic Behavior and Organization*. Vol. 51 (2). p 143-59. June 2003.

Cunyat, Antoni and Randolph Sloof (2008), "Employee Types ann Endogenous Organizational Design", Discussion Paper No. 08-019/1 (Tinbergen Institute). (August)

Currie, Janet. (1989), "Who Uses Arbitration? The Case of British Columbia's Teachers, 1947-1981", *Industrial & Labor Relations Review* 42(3): 363-379.

Currie, Janet and Sheena McConnell (1991) "Collective Bargaining in the Public Sector: The Effect of Legal Structure on Dispute Costs and Wages" *American Economic Review*. Vol. 81 (4). p 693-718. September 1991

Currie, Janet, and Henry S. Farber (1992), "Is Arbitration Addictive? Evidence From the Laboratory and the Field", Working Paper No. 3952 (National Bureau of Economic Research).

Davis, Douglas D. (1987), "Maximal Quality Selection and Discrimination in Employment", *Journal of Economic Behavior and Organization* 8(1): 97-112. (March)

Deci, E.; R. Koestner and R. Ryan. "A Meta-Analytic Review of experiments examining the effects of extrinsic rewards on intrinsic motivation". *Psych. Bull.* 125 (1999): 627-668.

Deck, Cary A. and Amy Farmer (2003), "Bargaining with Asymmetric Uncertainty: Arbitration Mechanisms Compared", Discussion Paper (University of Arkansas).

Deck, Cary A. and Amy Farmer (2007), "Bargaining over an Uncertain Value: Arbitration Mechanisms Compared", *Journal of Law, Economics & Organization* 23(3): 547-579.

Deck, Cary A. and Amy Farmer (2009), "Strategic Bidding and Investments in Final Offer Arbitration: Theory and Experimental Evidence", *Journal of Economic Behavior & Organization* 70(1-2): 361-373.

Deck, Cary A., Amy Farmer and Dao-Zhi Zeng (2007), "Amended Final-offer Arbitration over an Uncertain Value: A Comparison with CA and FOA", *Experimental Economics* 10(4): 439-454.

Deck, Cary A., Amy Farmer and Dao-Zhi Zeng (2009), "Arbitration and Bargaining Across the Pacific", *Southern Economic Journal* 76(1): 183-197.

Dee, Thomas S. (2009), "Stereotype Threat and the Student-Athlete", Working Paper No. w14705 (National Bureau of Economic Research). (February)

Della Vigna, Stefano, John A. List and Ulrike Malmendier. "Testing for Altruism and Social Pressure in Charitable Giving". NBER working paper no. 15629, December 2009)

Deutsch, M. and R. Krauss. "The effects of threat on interpersonal bargaining" *Journal of Abnormal and Social Psychology*, 1960, 61, 223-230.

Dickinson, David L. (1999), "An Experimental Examination of Labor Supply and Work Intensities", *Journal of Labor Economics* 17(4): 638-670.

Dickinson, David L. (2001), "The Carrot versus the Stick in Work Team Motivation", *Experimentl Economics* 4(1): 107-124.

Dickinson, David L. (2003), "An Illustrated Example of the Effects of Risk Preferences and Expectations on Bargaining Outcomes", *Journal of Economic Education* 33(2): 169-180.

Dickinson, David L. (2004), "A Comparison of Conventional, Final-offer, and "Combined' Arbitration for Dispute Resolution", *Industrial & Labor Relations Review* 57(2): 288-301.

Dickinson, David L. (2005), "Bargaining Outcomes with Double-offer Arbitration", *Experimental Economics* 8(2): 145-166.

Dickinson, David L. (2006), "The Chilling Effect of Optimism: The Case of Final-offer Arbitration", *Journal of Socio-economics* 35(1): 17-30.

Dickinson, David L. and R. Mark Issac (1998), "Absolute and Relative Awards for Individuals in Team Production", *Managerial and Decision Economics* 19(4-5): 299-310. (June – August)

Dickinson, David L. "The Effects of Beliefs versus Risk Attitude on Bargaining Outcomes." *Theory and Decision*. Vol. 66 (1). p 69-101. January 2009.

Dickinson, David L. and Marie-Calire Villeval (2008), "Does Monitoring Increase Work Effort? The Complimentarity Between Agency and Crowding-out Theories", *Games and Economic Behavior* 63(1): 56-76. (May)

Dickinson, David L. and Ronald L. Oaxaca (2009), "Statistical Discrimination in Labor Markets: An Experimental Analysis", *Southern Economic Journal* (Forthcoming).

Dittrich, Dennis and Martin G. Kocher (2006), "Monitoring and Pay: An Experiment on Employee Peformance under Endogenous Supervision", Discussion Paper No. 5962 (Centre for Economic Policy Research). (December)

Dohmen, Thomas J. and Armin Falk (2006), "Performance Pay and Multi-dimensional Sorting: Productivity, Preferences and Gender", Discussion Paper No. 2001 (Institute for the Study of Labor). (March)

Dolton, Peter, and Donal O'Neill. "The Long-Run Effects of Unemployment Monitoring and Work-Search Programs: Experimental Evidence from the United Kingdom." *Journal of Labor Economics*. Vol. 20 (2). p 381-403. Part 1 April 2002.

Dufwenberg, Martin and Uri Gneezy (2000), “Measuring Beliefs in an Experimental Lost Wallet Game”, *Games and Economic Behavior* 30: 163-82.

Dufwenberg, Martin and Georg Kirchsteiger (2004), “A Theory of Sequential Reciprocity”, *Games and Economic Behavior* 47: 268-98.

Dürsch, Peter, Jörg Oechssler and Radovan Vadovic (2008), “Sick Pay Provision in Experimental Labor Markets”, Discussion Paper No. 476 (University of Heidelberg). (October)

Eberlein, Marion and Judith Przemeck (2008), “Whom Will You Choose? Collaborator Selection and Selector’s Self-prediction”, Discussion Paper No. 12/2008 (University of Bonn). (June)

Eckel, Catherine C. and Philip Grossman [1997], "Are Women Less Selfish Than Men?: Evidence from Dictator Games," *Economic Journal*.

Eckel, Catherine C. and Philip J. Grossman (2005), “Managing Diversity By Creating Team Identity”, *Journal of Economic Behavior and Organization* 58(3): 371-392. (November)

Ederer, Florian P. and Ernst Fehr (2007), “Deception and Incentives: How Dishonesty Undermines Effort Provision”, Discussion Paper No. 3200 (Institute for the Study of Labor). (December)

Edwards, Wanye and Lee Huskey (2008), “Job Search with An External Opportunity: an Experimental Exploration of the Todaro Paradox”, *Annals of Regional Science* 42(4): 807-819.

Ellingsen, Tore and Magnus Johannesson (2004), “Promises, Threats and Fairness”, *Economic Journal* 114(495): 397-420. (April)

Elingsen, Tore and Magnus Joahnnesson (2008), Pride and Prejudice: The Human Side of Incentive Theory”, *American Economic Review* 98(3): 990-1008. (June)

Eriksson, Tor, Sabrina Teyssier and Marie-Claire Villeval (2008), “Self-selection and the Efficiency of Tournaments”, *Economic Inquiry* (47): 530-548.

Eriksson, Tor, Anders Poulsen and Marie-Claire Villeval (2008), “Feedback and Incentives: Experimental Evidence”, Working Paper No. 3440 (Institute for the Study of Labor). (April)

Eswaran, Mukesh, and Ashok Kotwal. The Moral Hazard of Budget-breaking. *RAND Journal of Economics*. Vol. 15 (4). p 578-81. Winter 1984.

Falk, Armin (2007), “Gift Exchange in the Field”, *Econometrica* 75(5): 1501-1511. (September)

Falk, Armin and Simon Gächter (2002), “Reputation and Reciprocity: Consequences for the Labour Relation”, *Scandanavian Journal of Economics* 104(1): 1-26. (March)

Falk, Armin, Ernst Fehr, and Urs Fischbacher (2003), “On the Nature of Fair Behavior”, Economic Inquiry 41: 20-26.

Falk, Armin and Andrea Ichino (2006), “Clean Evidence on Peer Effects”, *Journal of Labor Economics* 24(1): 39-57. (January)

Falk, Armin and Urs Fischbacher (2006), “A Theory of Reciprocity”, *Games and Economic Behavior* 54(2): 293-315.

Falk, Armin and Christian Zehnder (2006), “Discrimination and In-group Favoritism in a Citywide Trust Experiment", IEW-Working Paper No. 318.

Falk, Armin, Fehr, Ernst, Fischbacher, Urs. "Reasons for Conflict: Lessons from Bargaining Experiments" *Journal of Institutional and Theoretical Economics*. Vol. 159 (1). p 171-87. March 2003.

Falk, Armin and Michael Kosfeld (2006), "The Hidden Costs of Control", *American Economic Review* 96(5): 1611-1630. (December)

Falk, Armin, Ernst Fehr, and David Huffman "The Power a4nd Limits of Tournament Incentives" unpublished manuscript, University of Bonn, March 28th, 2008.

Falk, Armin, Ernst Fehr and Urs Fischbacher (2005), "Driving Forces Behind Informal Sanctions", *Econometrica* 73(6): 2017-2030. (November)

Falk, Armin, Ernst Fehr and Christian Zehnder (2006), "The Beahvioural Effects of Minimum Wages", *Quarterly Journal of Economics* 121(4): 1347-1381.

Falk, Armin, Simon Gächter and Judit Kovács (1999), "Intrinsic Motivation and Extrinsic Incentives in a Repeated Game with Incomplete Contracts", *Journal of Economic Psychology* 20(3): 251-284. (June)

Falk, Armin, David Huffman and W. Bentley MacLeod (2008), "Instituions and Contract Enforcement", Working Paper No. 13961 (National Bureau of Economic Research).

Fama, E. F. "Agency Problems and the Theory of the Firm". *Journal of Political Economy* 88 (1980): 288-307.

Farber, Henry S. (1980), "An Analysis of Final-offer Arbitration", Journal of Conflict Resolution 24(4): 683-705.

Farber, Henry S. (1987) "The analysis of union behavior" in Richard Layard and Orley Ashenfelter, eds., *Handbook of Labor Economics*, volume 2. Amsterdam: North-Holland, 1987.

Farber, Henry. S. (2005), `Is tomorrow another day? "The labor supply of New York City cabdrivers", *Journal of Political Economy* 113(1), 46{82.

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

Farber, Henry S. and Max H. Bazerman "The General Basis of Arbitrator Behavior: An Empirical Analysis of Conventional and Final-Offer Arbitration" Econometrica 54(4) (July 1986): 819-844.

Farber, Henry S. and Max H. Bazerman (1989), "Divergent Expectations as a Cause of Disagreement in Bargaining: Evidence from a Comparison of Arbitration Schemes", *Quarterly Journal of Economics* 104(1): 99-120.

Farber, Henry S., Margaret A. Neale and Max H. Bazerman (1990), "The Role of Arbitration Costs and Risk Aversion in Dispute Outcomes", *Industrial Relations* 29(3): 361-384.

Farmer, Amy and Paul Pecorino. (1998), "Bargaining with Informative Offers: An Analysis of Final-offer Arbitration", *Journal of Legal Studies* 27(2): 415-432.

Farrell, Joseph, and Carl Shapiro. "Optimal Contracts with Lock-In" *American Economic Review*, Vol. 79, No. 1 (Mar., 1989), pp. 51-68

Fehr, E. & Goette, L. (2007), `Do workers work more when wages are high? Evidence from a randomized field experiment', *American Economic Review* 97(1), 298-317.

Fehr, Ernst and Georg Kirchsteiger (1994), "Insider Power, Wage Discrimination and Fariness", *Economic Journal* 104(424): 571-583. (May)

Fehr, E. and K. M. Schmidt. "A Theory of Fairness, Competition and Co-operation" *Quarterly Journal of Economics* 114 (1999): 817-868.

Fehr, Ernst and Klaus M. Schmidt (2004), "Fairness and Incentives in a Multi-task Principal-agent Model", *Scandanavian Journal of Economics* 106(3): 453-474. (September)

Fehr, Ernst and Klaus M. Schmidt (2007), "Adding A Stick to the Carrot? The Interaction of Bonuses and Fines", *American Economic Review* 97(2): 177-181. (May)

Fehr, Ernst and Jean-Robert Tyran (2008), "Limited Rationality and Strategic Interaction: The Impact of the Strategic Environment on Nominal Inertia", *Econometrica* 76(2): 353-394. (March)

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

Fehr, Ernst, Oliver D. Hart and Christian Zehnder (2008), "Contracts as Reference Points: Experimental Evidence", Working Paper No. 14501 (National Bureau of Economic Research). (November)

Fehr, Ernst, Georg Kirchsteiger and Arno Riedl (1993), "Does Fairness Prevent Market Clearing?: An Experimental Investigation", *Quarterly Journal of Economics* 108(2): 437-459. (May)

Fehr, Ernst, Georg Kirchsteiger and Arno Riedl (1996), "Involuntary Unemployment and Non-compensating Wage Differentials In An Experimental Labour Market", *Economic Journal* 106(434): 106-121. (January)

Fehr, Ernst, Georg Kirchsteiger and Arno Riedl (1998), "Gift Exchange and Reciprocity in Competitive Experimental Markets", *European Economic Review* 42(1): 1-34. (January)

Fehr, Ernst, Alexander Klein and Klaus M. Schmidt (2007a), "Fariness and Contract Design", *Econometrica* 75(1): 121-154. (January)

Fehr, Ernst, Alexander Klein and Klaus M. Schmidt (2007b) "Adding a Stick to the Carrot? The Interaction of Bonuses and Fines" *American Economic Review* papers and proceedings, May 2007, vol. 97(2),177-181

Fehr, Ernst and Armin Falk (1999), "Wage Rigidity in a Competitive Incomplete Contract Market", *Journal of Political Economy* 107(1): 106-134. (February)

Fehr, Ernst, Armin Falk, and Christian Zehnder (2006), "The Behavioral Effects of Minimum Wage Laws", *Quarterly Journal of Economics* 121(4): 1347: 1381. (November)

Fehr, Ernst, Urs Fischbacher, Bernhard von Rosenbladt, Jürgen Schupp, and Gert Wagner (2002), "A Nation-wide Laboratory – Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Surveys", Schmollers Jahrbuch 122: 519-542.

Fehr, Ernst and Simon Gächter (2000), "Fairness and Retaliation: The Economics of Reciprocity", *Journal of Economic Perspectives* 14(3): 159-181. (Summer)

Fehr, E. and K. M. Schmidt (1999), "A Theory of Fairness, Competition and Co-operation" *Quarterly Journal of Economics* 114: 817-868. (August)

Fehr, Ernst, Erich Kirchler, Andreas Weichbold, and Simon Gächter (1998), "When Social Norms Overpower Competition: Gift Exchange in Experimental Labor Markets", *Journal of Labor Economics* 16(2): 324-351. (April)

Fershtman, C. and U. Gneezy (2001), "Discrimination in a Segmented Society: An Experimental Appraoch", *Quarterly Journal of Economics* 116(1): 351-377. (February)

Fershtman, C., U. Gneezy and F. Verboven (2005), "Discrimination and Nepotism: The Efficiency of the Anonymity Rule", *Journal of Legal Studies* 34(2): 371-394. (June)

Fershtman, Chaim, Uri Gneezy, and John List (2009), "Inequity, Selfishness and Social Norms", mimeo.

Ferster, C. B. and B. F. Skinner. *Schedules of Reinforcement*. New York: Appleton-Century-Crofts, 1957.

Forsythe, R., Kennan, J. and Sopher, B., 1991a. An experimental analysis of strikes in bargaining games with one-sided private information. *American Economic Review* **81** 1, pp. 253–278.

Forsythe, Robert; Kennan, John; Sopher, Barry. "Dividing a Shrinking Pie: An Experimental Study of Strikes in Bargaining Games with Complete Information." *Research in experimental economics*. Volume 4. Isaac, R. Marc, ed., Greenwich, Conn. and London: JAI Press. p 223-68. 1991b.

Frechette, Guillaume; Kagel, John H; Morelli, Massimo. Behavioral Identification in Coalitional Bargaining: An Experimental Analysis of Demand Bargaining and Alternating Offers. *Econometrica*. Vol. 73 (6). p 1893-1937. November 2005.

Frey, B. "Does monitoring increase work effort? The rivalry between trust and loyalty" *Economic Inquiry* 31 (1993): 663-670.

Frey, Bruno S., and Felix Oberholzer-Gee. "The Cost of Price Incentives: An Empirical Analysis of Crowding-Out" *American Economic Review* 87(4) (Sept. 1997): 746-755.

Frohlich, Norman, John Godard, Joe A. Oppenheimer and Frederick A. Starke (1998), "Employee versus Conventionally-owned and Controlled Firms: An Experimental Analysis", *Managerial and Decision Economics* 19(4-5): 311-326. (June-August)

Fryer, R., J. Goeree, and C. Holt (2005), "Experience-Based Discrimination: Classroom Games," *Journal of Economic Education*, 36(2), 160-170 (Spring)

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg (2006): "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model," *American Economic Review*, 96, 1043-1068.

Gabuthy, Yannick, Nicolas Jacquemet and Nadège Marchand (2008), "Does Resorting to Online Dispute Resolution Promote Agreements? Experimental Evidence", *European Economic Review* 52(2): 259-282.

Gächter, Simon, Daniele Nosenzo, and Martin Sefton (2008), "The Impact of Social Comparisons on Reciprocity", CeDEx Discussion Paper No. 2008-09. (August)

Gibbons, R. and M. Waldman. "A Theory of Wage and Promotion Dynamics Inside Firms" *Quarterly Journal of Economics* 114(4) (November 1999): 1321-58.

Gill, David and Victoria L. Prowse. "A Structural Analysis of Disappointment Aversion in a Real Effort Competition" IZA working paper no. 4356, October 2009.

Georg, Sebastian, Sebastian Kube and Ro'i Zultan "Treating Equals Unequally: Incentives in Teams,Workers' Motivation and Production Technology"IZA discussion paper No. 3959, January 2009

Gibbons, Robert. "Piece-Rate Incentive Schemes" *Journal of Labor Economics*, Vol. 5, No. 4, Part 1 (Oct., 1987), pp. 413-429

Gibbons, Robert. "Incentives in Organizations" *Journal of Economic Perspectives* 12(4) (Fall 1988): 115-132/

Gneezy, U. and J.A. List (2006), "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments", *Econometrica* 74(5): 1365-1384. (September)

Gneezy, U. and Aldo Rustichini (2000), "Pay Enough or Don't Pay At All", *Quarterly Journal of Economics* 115(3): 791-810. (August)

Gneezy, Uri, Ernan Haruvy, and Alvin E. Roth, "Bargaining Under a Deadline: Evidence from the Reverse Ultimatum Game," *Games and Economic Behavior*, Special Issue in Honor of Robert W. Rosenthal, 45, 2, November 2003, 347-368.

Gneezy, U. and A. Rustichini (2004), "Gender and Competition at a Young Age", *American Economic Review* 94(2): 377-381. (May)

Gneezy, U., K.L. Leonard and J.A. List (2008), "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society", *Econometrica* (Forthcoming).

Gneezy, U., M. Niederle and A. Rustichini (2003), "Performance in Competitive Environments: Gender Differences", *Quarterly Journal of Economics* 118(3): 1049-1074. (August)

Guth, Werner, Steffen Huck, and Peter Ockenfels [1996], "Two- Level Ultimatum Bargaining with Incomplete Information: An Experimental Study," *Economic Journal*, 106, May, 593-60

Guth, Werner; Levati, M. Vittoria; Maciejovsky, Boris. Deadline Effects in Sequential Bargaining--An Experimental Study.*International Game Theory Review.* Vol. 7 (2). p 117-35. June 2005

Greenberg, Jerald (1990), "Employee Theft as a Reaction to Underpayment Inequity: The Hidden Cost of Pay Cuts", *Journal of Applied Psychology* 75: 561–568.

Grether, David M; Isaac, R. Mark; Plott, Charles R. "The Allocation of Landing Rights by Unanimity among Competitors" *American Economic Review*. Vol. 71 (2). p 166-71. May 1981.

Griffith, Rachel and Andrew Neely. "Performance Pay and Managerial Experience in Multitask Teams: Evidence from within a Firm" *Journal of Labor Economics* 27(1) (2009): 49-82.

Gürtler, Oliver and Christine Harbring (2007), "Feedback in Tournaments under Commitment Problems: Theory and Experimental Evidence", Discussion Paper No. 3111 (Institute for the Study of Labor). (October)

Gu, Wulong and Peter Kuhn. "A Theory of Holdouts in Wage Bargaining", *American Economic Review* 88 (June 1998): 428-449.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze (1982), "An experimental analysis of ultimatum bargaining", *Journal of Economic Behavior and Organization* 3(4): 367-388.

Güth, W., M. Königstein, J. Kovács and E. Zala-Mezõ (2001), "Fairness within Firms: The Case of One Principal and Many Agents," *Schmalenbach Business Review*, **53(2)**: 82-101

Hall, Robert E and Edward P. Lazear. "The Excess Sensitivity of Layoffs and Quits to Demand" *Journal of Labor Economics*. Vol. 2 (2). p 233-57. April 1984.

Hamermesh, Daniel S. (2010), "Beauty Pays", *Princeton University Press* (Forthcoming).

Hamilton, B., J. Nickerson, and H. Owan, "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation". *Journal of Political Economy* 111(3) (June 2003): 465-497.

Hamman, John, George Loewenstein and Roberto A. Weber (2008), "Self-interest Through Agency: An Alternative Rationale for the Principal-agent Relationship", Unpublished Paper (Carnegie Mellon University). (July)

Hannan, R. Lynn (2005), "The Combined Effects of Wages and Firm Profit on Employee Effort", *Accounting Review* 80(1): 167-188.

Hannan, R. Lynn, John H. Kagel and Donald V. Moser (2002), "Partial Gift Exchange in an Experimental Labor Market: Impact of Subject Population Differences, Productivity Differences and Effort Requests on Behavior", *Journal of Labor Economics* 20(4): 923-951. (October)

Harbring, Christine and Bernd Irlenbusch (2003), "An Experimental Study on Tournament Design", *Labour Economics* 10(4): 443-464. (August)

Harbring, Christine and Bernd Irlenbusch (2005), "Incentives in Tournaments with Endogenous Prize Selection", *Journal of Institutional and Theoretical Economics* 161(4): 636-663. (December)

Harbring, Christine and Bernd Irlenbusch. "How Many Winners are Good to Have? On Tournaments with Sabotage" *Journal of Economic Behavior and Organization*. Vol. 65 (3-4). p 682-702. March 2008.

Harbring, Christine and Bernd Irlenbusch "Sabotage in Tournaments: Evidence from a Laboratory Experiment" IZA Discussion Paper No. 4205 June 2009

Harbring, Christine; Bernd Irlenbusch, Matthias Krakel, and Reinhard Selten. "Sabotage in Corporate Contests--An Experimental Analysis" *International Journal of the Economics of Business*. Vol. 14 (3). p 367-92. November 2007.

Hargreaves, Heap S. and Y. Varoufakis (2002), "Some Experimental Evidence of the Evolution of Discrimination, Co-Operation and Perceptions of Fairness", *Economic Journal* 112(481): 679-703. (July)

Harrison, Glenn W., and Peter Morgan. "Search Intensity in Experiments" *Economic Journal*. Vol. 100 (401). p 478-86. June 1990.

Harrison, Glenn W., John A. List and Charles Towe (2007), "Naturally Occuring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion", *Econometrica* 75(2): 433-458. (March)

Haruvy, Eman, Alvin E. Roth and M. Utku Unver (2006), "The Dynamic of Law Clerk Matching: An Experimentl and Computational Investigation of Proposals for Reform of the Market", *Journal of Economics Dynamics and Control* 30(3): 457-486. (March)

Hashimoto, Masanori and Ben T. Yu. "Specific Capital, Employment Contracts, and Wage Rigidity" *Bell Journal of Economics*. Vol. 11 (2). p 536-49. Autumn 1980.

Hayes, Beth. "Unions and Strikes with Asymmetric Information" *Journal of Labor Economics*. Vol. 2 (1). p 57-83. January 1984.

Healy, Paul J. (2007), "Group Reputations, Stereotypes and Cooperation in a Repeated Labor Market", *American Economic Review* 97(5): 1751-1773. (December)

Hennig-Schmidt, Heike, Betting Rockenbach and Abdolkarim Sadrieh (2009), "In Search of Workers' Real Effort Reciprocity: A Field and a Laboratory Experiment", *Journal of the European Economic Association* (Forthcoming).

Hey, John D. "Search for Rules for Search" *Journal of Economic Behavior and Organization*. Vol. 3 (1). p 65-81. March 1982.

Hicks, J. R. *The Theory of Wages*. London: Macmillan, 1932, second edition 1963.

Hoffman, Elizabeth; Kevin McCabe and Vernon Smith "Preferences and Property Rights in Ultimatum and Dictator Games" Chapter 47 in Plott, Charles R. and Vernon L. Smith, editors. *Handbook of Experimental Economics Results, vol. 1.* Amsterdam, North Holland (2008). pp. 417-422

Holm, Hakan, Engseld, Peter. Choosing Bargaining Partners--An Experimental Study on the Impact of Information about Income, Status and Gender. *Experimental Economics.* Vol. 8 (3). p 183-216. September 2005

Holm, Håkan, and Paul Nystedt (2005), "Intra-generational trust – a semi-experimental study of trust among different generations", *Journal of Economic Behavior and Organization*, 58(3): 403-419.

Holmstrom, B. "Moral Hazard in Teams", Bell Journal of Economics, Autumn 1982.

Holmstrom, B. "Managerial Incentive Problems: A Dynamic Perspective" *Review of Economic Studies* 66 (1999): 169-182. Originally published in *Essays in Honor of Lars Wahlbeck*. Helsinki, Finland: 1982.

Holmstrom, B. and P. Milgrom, "Multi-Task Principal-Agent Problems: Incentive Contracts, Asset Ownership, and Job Design", *Journal of Law, Economics and Organization* 7 (Special issue, 1991): 24-52.

Holt, Charles A. (1995), "Industrial Organization: A Survey of Laboratory Research", in The Handbook of Experimental Economics, Alvin E. Roth and John H. Kagel, eds., 349-443.

Hvide, H. "Tournament Rewards and Risk-Taking". *Journal of Labor Economics* 20(4) (October 2002); 877-898.

Irlenbusch, Bernd and Dirk Sliwka (2005a), "Transparency and Reciprocal Behavior in Employment Relations", *Journal Economic Behavior and Organization* 56(3): 383-403. (March)

Irlenbusch, Bernd and Dirk Sliwka (2005b), "Incentives, Decision Frames and Crowding Out: An Experimental Investigation", Discussion Paper No. 1758 (Institute for the Study of Labor). (September)

Irlenbusch, Bernd and Dirk Sliwka (2006), "Career Concerns in a Simple Experimental Labour Market", *European Economic Review* 50(1): 147-170. (January)

Irlenbusch, Bernd and Dirk Sliwka. "Career Concerns in a Simple Experimental Labour Market" *European Economic Review*, 2006, 50 (1), 147-170

Irlenbusch, Bernd and Gabriele K. Ruchala (2008), "Relative Rewards within Team-based Compensation", *Labour Economics* 15(2): 141-167. (April)

Isaac, R. M. and J. M. Walker. "Group size effects in public goods provision: The Voluntary Contribution Mechanism" *Quarterly Journal of Economics* 103 (1988): 179-200.

Ivanova-Stentzel, Radosveta and Dorothea Kubler (2005), "Courtesy and Idleness: Gender Differences in Team Work and Team Competition", Discussion Paper No. 1768 (Institue for the Study of Labor). (September)

Johnson, D. F., and W. Tullar 1972 "Style of third party interven-tion, face-saving, and bargain-ing behavior." *Journal of Experimental Social Psychology***,** 8: 319-330.

Jones, Stephen R. G. (1992) "Was There a Hawthorne Effect?" *American Journal of Sociology* 98(3) (November): 451-468.

Kagel, John H. "Economics According to the Rats (and Pigeons Too): What have we learned and what can we hope to learn? In Alvin E. Roth, ed. *Laboratory Experimentation in Economics: Six Points of View* Cambridge: Cambridge University Press, 1987.

Kagel, J. H., R. C. Battalio and L. Green. *Economic Choice Theory: An experimental Analysis of Animal Behavior* Cambridge: Cambridge University Press, 1995.

Kagel, John H. and A.E. Roth (2000), "The Dynamics of Reorganization in Matching Markets: A Laboratory Experiment Motivated by a Natural Experiment", *Quarterly Journal of Economics* 115(1): 201-235. (February)

Kahan, J.P., A. Rapoport, and L. V. Jones. "Decision making in a sequential search task". *Perception and Psychophysics* 2 (1967): 374-376.

Kahneman, Daniel, Jack Knetsch, and Richard Thaler (1986), "Fairness and the Assumptions of Economics", *Journal of Business* 59:S285-S300.

Kanemoto, Yoshitsugu and W. Bentley MacLeod (1992), "The Ratchet Effect and the Market for Secondhand Workers" *Journal of Labor Economics* 10(1) (January): 85-98.

Karlan, Dean (2005), "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions, *American Economic Review* 95(5), 1688-1699. (December)

Kelsey, J. E. and Allison, J. Nov. 1976. "Fixed ratio level pressing by UVH rats: Work vs. accessibility of sucrose reward." *Physiology and Behavior 17(5):749-754.*

Kennan, John. (1987) "The economics of strikes", in Richard Layard and Orley Ashenfelter, eds., *Handbook of Labor Economics*, volume 2. Amsterdam: North-Holland, 1987.

Kennan, John and Robert Wilson. "Bargaining with Private Information" *Journal of Economic Literature*, Vol. 31, No. 1 (Mar., 1993), pp. 45-104

Kidd, Michael P., Paul S. Carlin and Jonathan Pot (2008), "Experimenting with Affirmative Action: The Coate and Loury Model", *Economic Record* 84(266): 322-337. (September)

Kiyotaki, Nobuhiro; Wright, Randall. "A Search-Theoretic Approach to Monetary Economics" *American Economic Review*. Vol. 83 (1). p 63-77. March 1993.

Knez, Marc and Duncan Simester. "Firm-Wide Incentives and Mutual Monitoring at Continental Airlines" *Journal of Labor Economics*. Vol. 19 (4). p 743-72. October 2001.

Koch, Alexander, Albrecht Morgenstern and Philippe Raab (2009), "Career Concerns Incentives: An Experimental Test", Working Paper No. 2009-1 (University of Aarhaus). (January)

Kosfeld, Michael and Ferdinand A. von Siemens (2009), "Worker Self Selection and the Profits from Cooperation", *Journal of the European Economic Association* 7(2-3): 573-582. (April)

Koszegi, B. & Rabin, M. (2006), `A model of reference-dependent preferences', *Quarterly Journal of Economics* 121(4), 1133-1165.

Krakel, Matthias, Petra Nieken and Judith Przemeck (2008), "Risk Taking In Winner-take-all Competition", Discussion Paper No. 233 (University of Munich).

Kritikos, Alexander S. (2006), "The Impact of Compulsory Arbitration on Bargaining Behavior: An Experimental Study", *Economics of Governance* 7(3): 293-315.

Kritikos, Alexander S. and Jonathan H.W. Tan (2009), "Indenture As a Self-enforced Contract Device: An Experimental Test", *Southern Economic Journal* 75(3): 857-872. (January)

Krueger, Alan and Alexandre Mas (2004), "Strikes, Scabs and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires", *Journal of Political Economy*112(2): 253:289. (April)

Kube, Sebastian, Clemens Puppe, and Michel Maréchal. "The currency of reciprocity". Unpublished manuscript, University of St. Galllen, 2006a.

Kube, Sebastian, Clemens Puppe, and Michel Maréchal. "Putting reciprocity to work- positive versus negative responses in the field" Unpublished manuscript, University of St. Galllen, 2006b.

Kübler, Dorothea, Wieland Müller and Hans-Theo Normann (2008), "Job Market Signalling and Screening: An Experimental Comparison", *Games and Economic Behavior* 64(1): 219-236. (September)

Kuhn, Michael A. "To Settle or Not To Settle: A Review of the Literature on Arbitration in the Laboratory", unpublished manuscript, University of California, San Diego, August 2009.

Kuhn, Peter. "Unions and the Economy: What we know; what we should know", Canadian Journal of Economics 31(5) (November 1998): 1033-1056.

Kuhn, Peter and Wulong Gu. 25. "Learning in Sequential Wage Negotiations: Theory and Evidence" *Journal of Labor Economics* 17 (January 1999): 109-140.

Lazear, Edward P., 1989. Pay Equality and Industrial Politics. *Journal of Political Economy*, 97(3), 561-580.

Lazear, E. P. *Personnel Economics for Managers*. New York: Wiley, 1998.

Lazear, E. P. " Performance Pay and Productivity". *American Economic Review* 90(3) (December 2000): 1346-1361.

Lazear, E. and S. Rosen, "Rank-Order Tournaments as Optimum Labor Contracts", *Journal of Political Economy*, Oct. 1981.

Lazear, Edward P., Ulrike Malmendier and Roberto A. Weber (2006), "Sorting in Experiments with Application to Social Preferences", Working Paper No. W12041 (National Bureau of Economic Research). (February)

Ledyard, J. O. "Public goods: A survey of experimental research", in J. Kagel and A. Roth, eds., *Handbook of Experimental Economics* Princeton: Princeton University Press, 1995.

Ledyard, John; David Porter, and Randii Wessen. "A Market-Based Mechanism for Allocating Space Shuttle Secondary Payload Priority" *Experimental Economics*. Vol. 2 (3). p 173-95. March 2000.

Levitt, Steven D., John A. List, and David H. Reiley, Jr. "What Happens in the Field Stays in the Field: Exploring Whether Professionals Play Minimax in Laboratory Experiments" NBER working paper no 15609 (December 2009).

Lippman, S. and McCall, J., "The Economics of Job Search: A Survey - Part I." *Economic Inquiry,*June 1976, 14,155-90.

Lippman, S. and McCall, J., "The Economics of Job Search: A Survey - Part II" *Economic Inquiry*, September 1976,14,347-89.

Loewenstein, George, Max Bazerman and Leigh Thompson, (1989), "Social Utility and Decision Making in Interpersonal Contexts," *Journal of Personality and Social Psychology* 57: 426-441.

MacLeod, W. Bentley and James M. Malcomson. "Investments, Holdup, and the Form of Market Contracts" *American Economic Review*. Vol. 83 (4). p 811-37. September 1993.

Malcomson, James M. "Contracts, Hold-up, and Labor Markets", Journal of Economic Literature 35 (December 1997): 1917-1957.

Mas, Alexandre. "Pay, Reference Points, and Police Performance" *Quarterly Journal of Economics*. Vol. 121 (3). p 783-821. August 2006.

Mas, Alexandre. "Labour Unrest and the Quality of Production: Evidence from the Construction Equipment Resale Market." *Review of Economic Studies*. Vol. 75 (1). p 229-58. January 2008.

Mas, A. and E. Moretti. "Peers at Work" *American Economic Review* 99(1) (March 2009): 112-143.

Mathewson, Stanley B. (1931), *Restriction of Output among Unorganized Workers*, New York, Viking Press.

Maximiano, Sandra., Randolph Sloof, and Joep Sonnemans (2006), "Gift Exchange in a Multi-worker Firm," *Economic Journal* 117 (522) (July 2007): pp. 1025-1050.

Maximiano, Sandra, Randolph Sloof and Joep Sonnemans (2006), "Gift Exchange and the Separation of Ownership and Control", Discussion Paper No. 2006-037/1 (Tinbergen Institute). (April)

McKinney, C. Nicholas, Niederle, Muriel and Alvin E. Roth (2005), "The Collapse of a Medical Labor Clearinghouse (and Why Such Failures Are Rare)", *American Economic Review* 95(3): 878-889. (June)

McMillan, John. (1994): "Selling Spectrum Rights," *Journal of Economic Perspectives*, 8, 145-162.

Meidinger, Jean-Louis Rullière and Marie-Claire Villeval (2003), "Does Team-based Compensation Give Rise to Problems When Agents Vary in Their Ability?", *Experimental Economics* 6(3): 253-272. (November)

Meyer, Bruce D. "Lessons from the U.S. Unemployment Insurance Experiments". *Journal of Economic Literature*. Vol. 33 (1). p 91-131. March 1995.

Meyer, Bruce D. "What Have We Learned from the Illinois Reemployment Bonus Experiment?" Journal of Labor Economics. Vol. 14 (1). p 26-51. January 1996.

Miller, R. and Plott, C. "Product quality signalling in experimental markets". *Econometrica* 53 (1985): 837-872.

Mobius, Markus M. and Tanya S. Rosenblatt (2006), "Why Beatuy Matters", *American Economic Review* 96(1): 222-235. (March)

Moffitt, Robert, and Walter Nicholson. "The Effect of Unemployment Insurance on Unemployment: The Case of Federal Supplemental Benefits" *Review of Economics and Statistics*. Vol. 64 (1). p 1-11. February 1982.

Mongell, Susan; Roth, Alvin E. "Sorority Rush as a Two-Sided Matching Mechanism" *American Economic Review*. Vol. 81 (3). p 441-64. June 1991.

Mortensen, Dale, and Christopher Pissaridies. "Job Creation and Job Destruction in the Theory of Unemployment." *Review of Economic Studies*. July 1994; 61(3): 397-415

Munier, Bertrand; Zaharia, Costin. High Stakes and Acceptance Behavior in Ultimatum Bargaining: A Contribution from an International Experiment. Theory and Decision. Vol. 53 (3). p 187-207. November 2002.

Murnighan, J. Keith. "Fairness in Ultimatum Bargaining", chapter 50 in Plott, Charles R. and Vernon L. Smith, editors. *Handbook of Experimental Economics Results*, vol. 1. Amsterdam, North Holland (2008) pp. 436-453,

Nalbantian, Haig R. and Andrew Schotter. "Matching and Efficiency in the Baseball Free-Agent System: An Experimental Examination" *Journal of Labor Economics*, Vol. 13, No. 1 (Jan., 1995), pp. 1-31

Nalbantian, Haig R. and Andrew Schotter (1995), "Matching and Efficiency in the Baseball Free-Agent System", *Journal of Labor Economics* 13(1): 1-31. (January)

Nalbantian, Haig R. and Andrew Schotter (1997), "Productivity under Group Incentives: And Experimental Study", *American Economic Review* 87(3): 314-341. (June)

Nash, John [1950], "The Bargaining Problem," *Econometrica*, 18, 155- 162.

Nash, John [1953], "Two-Person Cooperative Games," *Econometrica*, 21, 128-40.

Neale, Margaret A. and Max H. Bazerman (1983), "The Role of Perspective-taking Ability in Negotiating and under Different Forms of Arbitration", *Industrial & Labor Relations Review* 36(3): 378-388.

Niederle, Muriel and Lise Vesterlund (2007), "Do Women Shy Away from Competition? Do Men Compete Too Much?", *Quarterly Journal of Economics* 122(3): 1067- 1101. (August)

Niederle, Muriel and Alexandra H. Yestrumskas (2008), "Gender Differences in Seeking Challenges: The Role of Institutions", Working Paper No. 13922 (National Bureaum of Economic Research).

Niederle, Muriel; Roth, Alvin E. "Unraveling Reduces Mobility in a Labor Market: Gastroenterology with and without a Centralized Match" *Journal of Political Economy.* Vol. 111 (6). p 1342-52. December 2003.

Niederle, Muriel; Roth, Alvin E. "Market Culture: How Rules Governing Exploding Offers Affect Market Performance" *American Economic Journal: Microeconomics*. Vol. 1 (2). p 199-219. August 2009.

Niederle, Muriel; Roth, Alvin E; Unver, M. Utku. "Unraveling Results from Comparable Demand and Supply: An Experimental Investigation" National Bureau of Economic Research, Inc, NBER Working Papers: 15006. 2009.

Niederle, Muriel, Carmit Segal and Lise Vesterlund (2008), "How Costly is Diversity? Affirmative Action In Light of Gender Differences in Competitiveness", Working Paper No. W13923 (National Bureau of Economic Research). (April)

Nieken, Petra and Dirk Sliwka (2009), "Risk-taking Tournaments: Theory and Theory and Experimental Evidence", *Journal of Economic Psychology* (Forthcoming).

Notz, William W. and Frederick A. Starke (1978), "Final-offer Versus Conventional Arbitration as Means of Conflict Management", *Administrative Science Quarterly* 23(2): 189-203.

Ochs, Jack and Alvin E. Roth (1989), "An Experimental Study of Sequential Bargaining", *American Economic Review* 79(3): 355–384. (June)

Ockenfels, Axel; Selten, Reinhard. An Experiment on the Hypothesis of Involuntary Truth-Signalling in Bargaining. *Games and Economic Behavior*. Vol. 33 (1). p 90-116. October 2000

Oettinger, Gerald S. "An Empirical Analysis of the Daily Labor Supply of Stadium Vendors". *Journal of Political Economy*. Vol. 107 (2). p 360-92. April 1999.

Offerman, Theo (2002), "Hurting hurts more than helping helps", *European Economic Review* 46(8): 1423-1437. (September)

O'Keefe, Viscusi and Zeckhauser, "Economic Contests: Comparative Reward Schemes", *Journal of Labor Economics,* January 1984

Oosterbeek, Hessel, Randolph Sloof and Joep Sonnemans (2006), "Rent-seeking versus Productive Activites in a Multi-task Experiment", Discussion Paper No. 06-083/1 (Tinbergen Institute). (September)

Oosterbeek, Hessel, Randolph Sloof and Joep Sonnemans. "Who Should Invest in Specific Training?" *Journal of Population Economics*. April 2007a; 20(2): 329-57.

Oosterbeek, Hessel, Randolph Sloof and Joep Sonnemans. "Promotion Rules and Skill Acquisition: An Experimental Study" *Economica*. Vol. 74 (294). p 259-97. May 2007b

Oosterbeek, Hessel, Randolph Sloof and Joep Sonnemans. "Does Making Specific Investments Unobservable Boost Investment Incentives?" *Journal of Economics and Management Strategy*. Vol. 16 (4). p 911-42. Winter 2007c.

Orrison, Alannah, Andrew Schotter and Keith Weigelt (2004), "Multiperson Tournaments: An Experimental Examination", *Management Science* 50(2): 268-279. (February)

Ortona, Guido, Steffane Ottone, Ferruccio Ponzano and Francesco Scacciati (2008), "Labour Supply in Presence of Taxation Financing Public Services: An Experimentl Approach", Journal of Economic Psychology 29(5): 619-631. (November)

Osborne, Martin and Ariel Rubinstein. *Bargaining and markets*. Boston: Academic Press, 1990

Owens, M. and J. Kagel (2009), "Minimum Wage Restrictions and Employee Effort in Incomplete Labor Markets: An Experimental Investigation," *Journal of Economic Behavior and Organization*, forthcoming.

Oxoby, Robert J. and Colette Friedrich (2008), "Incentive Design and Trust: Comparing the Effects of Tournament and Team-based Incentives on Trust", Working Paper No. 3434 (Institute for the Study of Labor). (March)

Paarsch, Harry J. and Bruce S. Shearer (2007), "Do Women React Differently to Incentives? Evidence from Experimental Data and Payroll Records", *European Economic Review* 51(7): 1682-1707. (October) Update available online, 21 November 2008.

Pecorino, Paul and Mark Van Boening (2001), "Bargaining and Information: An Empirical Analysis of a Multistage Arbitration Game", *Journal of Labor Economics* 19(4): 922-948.

Pereira, Paulo T., Nuno Silva and João Andrade e Silva (2006), "Positive and Negative Reciprocity in the Labor Market", *Journal of Economic Behavior and Organization* 59(3): 406-422. (March)

Plott, Charles R. "Dimensions of Parallelism: Some Policy Applications of Experimental Methods" in Alvin E. Roth, ed., *Laboratory Experimentation in Economics: Six points of view.* Cambridge: Cambridge University Press, 1987.

Plott, Charles R. and Vernon L. Smith, editors. *Handbook of Experimental Economics Results, vol. 1.* Amsterdam, North Holland (2008). pp. 417-422

Prendergast, Canice. "The Role of Promotion in Inducing Specific Human Capital Acquisition" *Quarterly Journal of Economics*. Vol. 108 (2). p 523-34. May 1993.

Rabin, Matthew, (1993), "Incorporating Fairness into Game Theory and Economics," *American Economic Review* 83: 1281-1302. (December)

Rapaport, A. "Individual Strategies in a Market-Entry Game" *Group Decision and Negotiation* 4 (1995):117-133.

Rapoport, A. and A. Tversky. "Cost and accessibility of offers as determinants of optimal stopping". *Pyschonomic Science* 4 (1966): 145-146.

Rapoport, A. and A. Tversky. “Choice behavior in an optimal stopping task” *Organizational Behavior and Human Performance* 5 (1970): 105-120.

Rassenti, S. J; Smith, V. L; Bulfin, R. L. “A Combinational Auction Mechanism for Airport Time Slot Allocation” *Bell Journal of Economics*. Vol. 13 (2). p 402-17. Autumn 1982.

Riedl, Arno and Frans van Winden (2007), “An Experimental Investigation of Wage Taxation and Unemployment In Closed and Open Economies”, *European Economic Review* 51(4): 871-900. (May)

Rigdon, Mary L. (2002), “Efficiency Wages in an Experimental Labor Market”, *Proceedings of the National Academy of Sciences* 99(20): 13348-13351. (October)

Roth, Alvin E. “The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory” *Journal of Political Economy*. Vol. 92 (6). p 991-1016. December 1984.

Roth, Alvin E. “On the Allocation of Residents to Rural Hospitals: A General Property of Two-Sided Matching Markets” *Econometrica*. Vol. 54 (2). p 425-27. March 1986.

Roth, Alvin E. “Bargaining Experiments” chapter 5 in John H. Kagel and Alvin E. Roth, editors, The *Handbook of Experimental Economics*, Volume 1. Princeton University Press, 1995, pp. 253-348.

Roth, Alvin E. “The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics”. *Econometrica.* Vol. 70 (4). p 1341-78. July 2002.

Roth, Alvin E. 2008. "What Have We Learned from Market Design?," *Economic Journal*, vol. 118(527), pages 285-310, 03.

Roth, Alvin E; Peranson, Elliott. “The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design” *American Economic Review.* Vol. 89 (4). p 748-80. September 1999.

Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir (1991) “Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study,” *American Economic Review* 81(5): 1068-95. (December)

Roth, Alvin E; Sonmez, Tayfun; Unver, M. Utku. “Efficient Kidney Exchange: Coincidence of Wants in Markets with Compatibility-Based Preferences” *American Economic Review*. Vol. 97 (3). p 828-51. June 2007.

Roth, Alvin E; Sotomayor, Marilda. “The College Admissions Problem Revisited” *Econometrica*. Vol. 57 (3). p 559-70. May 1989.

Roth, Alvin E; Xing, Xiaolin. “Turnaround Time and Bottlenecks in Market Clearing: Decentralized Matching in the Market for Clinical Psychologists” *Journal of Political Economy.* Vol. 105 (2). p 284-329. April 1997.

Rubinstein, Ariel [1982], "Perfect Equilibrium in a Bargaining Model," *Econometrica*, vol. 50, pp97-109.

Rydval, Ondrej and Andreas Ortmann (2004), “How Financial Incentives and Cognitive Abilities Affect Task Performance in Laboratory Settings: An Illustration”, *Economics-Letters* 85(3): 315-320. (December)

Sarver, Todd (2008), “Anticipating Regret: Why Fewer Options May Be Better”, *Econometrica* 76(2): 263-305. (March)

Shiller, R. J. (1993): *Macro Markets: Creating Institutions for Managing Society's Largest Economic Risks*, Clarendon Lectures in Economics. Oxford: Oxford University Press.

Schminki, Marshall, Russell Cropanzano, and Deborah Rupp (2002), "Organizational structure and fairness perceptions: the moderating effects of organizational level", *Organizational Behavior and Human Decision Processes* 89: 881-905.

Schnedler, Wendelin and Radovan Vadovic (2007), "Legitimacy of Control", Discussion Paper No. 3013 (Institute for the Study of Labor). (August)

Schotter, Andrew, and Yale M. Braunstein. "Economic Search: An Experimental Study", *Economic Inquiry*. Vol. 19 (1). p 1-25. January 1981

Schotter, Andrew, and Yale M. Braunstein. "Labor Market Search: An Experimental Study. *Economic Inquiry*. Vol. 20 (1). p 133-44. January 1982

Schotter, Andrew and Keith Weigelt (1992), "Asymmetric Tournaments, Equal Opportuniuty Laws and Affiemative Action: Some Experimental Results", *Quarterly Journal of Economics* 107(2): 511-539. (May)

Schotter, Andrew; Zheng, Wei; Snyder, Blaine. "Bargaining through Agents: An Experimental Study of Delegation and Commitment". *Games and Economic Behavior*. Vol. 30 (2). p 248-92. February 2000.

Schram, Arthur, Jordi Brandts and Klarita Gërxhani (2007), "Information Networks and Worker Recruitment" Unpublished Paper (University of Amsterdam). (June)

Schunk, Daniel. "Behavioral Heterogeneity in Dynamic Search Situations: Theory and Experimental Evidence" *Journal of Economic Dynamics and Control*. Vol. 33 (9). p 1719-38. September 2009.

Schunk, Daniel; Winter, Joachim. "The Relationship between Risk Attitudes and Heuristics in Search Tasks: A Laboratory Experiment." *Journal of Economic Behavior and Organization.* Vol. 71 (2). p 347-60. August 2009.

Shapiro, Carl, and J. E. Stiglitz. "Equilibrium Unemployment as a Worker Discipline Device". *American Economic Review* 74 (1984): 433-444.

Shearer, Bruce. "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment" *Review of Economic Studies.* Vol. 71 (2). p 513-34. April 2004.

Siegel, Sidney and Lawrence E. Fouraker [1960], *Bargaining and Group Decision Making: Experiments in Bilateral Monopoly*, New York, McGraw-Hill.

Sillamaa, M.A. (1999a), "How Work Effort Responds to Wage Taxation: A Non-linear versus a Linear Tax Experiment", *Journal of Economic Behavior and Organization* 39(2): 219-233. (June)

Sillamaa, M.A. (1999b), "How Work Effort Responds to Wage Taxation: An Experimental Test of a Zero Top Marginal Tax Rate", *Journal of Public Economics* 73(1): 125-134. (July)

Skarlicki, Daniel P. and Robert Folger (1997), "Retaliation in the workplace: The roles of distributive, procedural, and interactional justice", *Journal of Applied Psychology* 82: 434–443.

Slade, Margaret E. "Multitask Agency and Contract Choice: An Empirical Exploration" *International Economic Review* 37(2) (May 1996): 465-486.

Sloof, Randolph C. and Mirjam van Praag (2008a), "The Effect of Noise in a Performance Measure on Work Motivation", Discussion Paper No. 08-074/1 (Tinbergen Institute). (August)

Sloof, Randolph C. and Mirjam van Praag (2008b), "Performance Measurement, Expectancy and Agency Theory: An Experimental Study", *Journal of Economic Behavior and Organization* 67(3): 794-809. (September)

Slonim, R. and A.E. Roth "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic," *Econometrica*, 66,3, May 1998, 569-596.

Sonnemans, Joep. "Strategies of Search". *Journal of Economic Behavior and Organization*. Vol. 35 (3). p 309-32. April 1998.

Sonnemans, Joep. "Decisions and Strategies in a Sequential Search Experiment" *Journal of Economic Psychology*. Vol. 21 (1). p 91-102. February 2000.

Sopher, Barry. "Bargaining and the Joint-Cost Theory of Strikes: An Experimental Study."*Journal of Labor Economics*. Vol. 8 (1). p 48-74. Part 1 January 1990.

Sprinkle, Geoffrey B. (2000), "The Effect of Incentive Contract on Learning and Performance", *Accounting Review* 75(3): 299-326. (July)

Starke, Frederick A. and William W. Notz (1981), "Pre- and Post-intervention Effects of Conventional Versus Final Offer Arbitration", *Academy of Management Journal* 24(4): 832-850.

Sterbenz, Frederic P, Phillips, Owen R. "Bargaining Experiments with Deadlines and Random Delays"*Economic Inquiry*. Vol. 39 (4). p 616-26. October 2001.

Stevens, Carl M. (1966), "Is Compulsory Arbitration Compatible with Bargaining?", *Industrial Relations* 5: 38-52.

Sutter, Matthias and Martin Kocher (2007), "Trust and Trustworthiness across Different Age Groups", *Games and Economic Behavior* 59: 364-382.

Sutter, Matthias and Christina Strassmair (2009), "Communication, Cooperation and Collusion in Team Tournaments: An Experimental Study", *Games and Economic Behavior* 66(1): 506-525. (May)

Svejnar, J. and M. Singer (1994): "Using Vouchers to Privatize an Economy: The Czech and Slovak Case," *Economics of Transition*, 2, 43-69

Swenson, C. W. "Taxpayer behavior in response to taxation: An experimental analysis". *Journal of Accounting and Public Policy* 7 (1988): 1-28.

Tournadre, Fabienne; Villeval, Marie-Claire. "Learning from Strikes" *Labour Economics.* Vol. 11 (2). p 243-64. April 2004.

Van Boening, Mark, Tanja F. Blackstone, Michael McKee and Elisabet Rutström (2006), "Benefit Packages and Individual Behavior: Choices over Discrete Goods with Multiple Attributes", *Managerial and Decision Economics* 27(6): 511-526. (September)

Van der Heijden, E.C.M., J.H.M. Nielsen, J.M. Potters and H.A.A. Verbon (2001), "Simple and Complex Gift Exchange in the Laboratory", *Economic Inquiry* 39(2): 280-297. (April)

Van Dijk , Frans, Joep Sonnemans and Frans Van Winden (2001), "Incentive Systems in a Real Effort Experiment", *European Economic Review* 45(2): 187-214. (February)

Van Huyck, John B., Raymond C. Battalio, and Richard O. Beil. "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure" *American Economic Review* Vol. 80 (1). p 234-48. March 1990.

Vanberg, Christopher (2008), "Why Do People Keep Their Promises? An Experimental Test of Two Explanations", *Econometrica* 76(6): 1467-1480. (November)

Vandegrift, Donald and Paul Brown (2003), "Task Difficulty, Incentive Effects and the Selection of High-variance Strategies: An Experimental Examination of Tournament Behavior", *Labour Economics* 10(4): 481-497. (August)

Vandegrift, Donald, Abdullah Yavas and Paul M. Brown (2007), "Incentive Effects and Overcrowding in Tournaments: An Experimental Analysis", *Experimental Economics* 10(4): 345-368. (December)

Von Siemens, F. (2004), "Social Preferences, Sorting, and Employment Contracts", Chapter 2 of Ph. D. dissertation, University of Munich.

Wakai, Katsutoshi (2008), "A Model of Utility Smoothing", *Econometrica* 76(1): 137- 153. (January)

Weber, Roberto A. "Managing Growth to Achieve Efficient Coordination in Large Groups" *American Economic Review*. Vol. 96 (1). p 114-26. March 2006.

Williamson, Oliver. *The Economic Institutions of Capitalism.* New York: The Free Press, 1985.

Wilson, Rick K. and Catherine C. Eckel (2006), "Judging a Book by Its Cover: Beauty and Expectations in the Trust Game", *Political Research Quarterly* 59(2): 189- 202. (June)

Wilson, R. B. (2002): "Architecture of Power Markets," *Econometrica*, 70, 1299-1340.

Winter, Eyal (2004). "Incentives and Discrimination", *American Economic Review* 94 (3): 764-773.

Zeng, Dao-Zhi (2003), "An Amendment to Final-bid Arbitration", *Mathematical Social Sciences* 6(1): 9-19.

Zeng, Dao-Zhi, Shinya Nakamura and Toshihide Ibaraki (1996), "Double-offer Arbitration", *Mathematical Social Sciences* 31(3): 147-170.