Lab Labour:
What Can Labour Market Research Learn from the Lab?

CLSRN 2011 Summer School Lecture notes

Peter Kuhn

Aside from some minor updates, these notes are based on:

There has been an explosion of papers (we review close to 400) conducting laboratory experiments on topics of interest to labor economists in the past 15 years.

Laboratory Experiments have been used to address:

A. The Supply of Effort; Principal-Agent Interactions
B. Bargaining, Strikes and Arbitration
C. Search
D. Labor Market Equilibrium, including minimum wage effects
E. Gender and Racial Differences and Discrimination
In this review, we ask:

What have we learned from these papers?

What can labor economists learn from lab experiments more generally?

The plan:

I. A sample ‘lab labor’ paper.

II. Advantages and disadvantages of lab experiments.

III. Designing and evaluating lab experiments.
IV. Lab Labor Results: What have we learned?

A. The Supply of Effort; Principal-Agent Interactions

B. Bargaining, Strikes and Arbitration

C. Search

D. Labor Market Equilibrium, including minimum wage effects

E. Gender and Racial Differences and Discrimination

V. An Assessment, and Where to go next?
I. “Lab Labor” In Action: An Example


Agents are endowed with 120 lab dollars in each round.

They choose effort, \( x \) (which can be any integer between 0 and 120), which costs them \( x \), but yields a benefit of \( 2x \) to the principal.

Principals (who move first) only make one decision: whether to impose a minimum effort level, \( x_{\text{min}} \) on the agent.

This restricts the agent’s choice set to \( \{x_{\text{min}}, \ldots, 120\} \). \( x_{\text{min}} \) was set by the experimenters to be either 5, 10, or 20.

All interactions are anonymous and one-shot.
**Predictions:**

1. If the subjects behave like ‘rational economic man’, all agents will choose the minimum effort the principal permits, and all principals will decide to restrict their agent’s choice set.

2. Consider an agent who thinks that the ‘fair’ level of effort to provide is \( x^* > x_{\text{min}} \). Suppose also that the agent’s perception of what is fair is unaffected by actions taken by the principal that have no material consequences for the agent. [fairness independent of others’ intentions]. Then any such agent’s choices should be unaffected by whether or not the principal imposes the requirement that \( x > x_{\text{min}} \).

Put a different way, (2) says that the empirical distribution of effort levels to the right of \( x_{\text{min}} \) should be unaffected by the principal’s decision to impose the minimum effort requirement.

Is it? **Figure 1 from Falk-Kosfeld:**
Figure 1. Cumulative Distribution of Agents’ Choices in Treatment C5 (Panel A), C10 (Panel B), and C20 (Panel C)

Notes: The figure shows all observations $x \leq 50$. In each treatment, there were a few $x$-choices above 50. These observations are summarized as $x > 50$. 
Decidedly not:

While the effort levels of a smaller number of ‘opportunistic’ agents were mechanically increased by the minimum requirement,

-many of the remaining agents (who are not directly affected) reduce their effort when firms impose the minimum.

In most treatments, the latter reductions in effort were so substantial that principals who ‘controlled’ earned lower payoffs than those who did not.
**Why did this happen?**

-when the minimum effort requirement was imposed by the experimenter, ‘high-effort’ subjects did not respond by reducing effort.

-in a follow-up survey, the authors asked agents “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.

-in a clever variation, 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:

(a) When 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.

(b) When the principal is given a small endowment (in contrast to zero in Falk-Kosfeld), agents do not seem to resent effort requirements that simply prevent agents from ‘pilfering’ this endowment.
II. When to do a Lab Experiment?

A. Some Advantages

1. Lab Experiments are well suited to testing explicit predictions of simple game theoretic models, under ideal conditions.

   - No other technique allows such precise control of possible confounding factors (e.g. reputation effects, nonrandom treatment assignment, information structure).

   - No other technique allows one to test actual numerical predictions of the theory so easily.

2. Lab Experiments are highly replicable. When repeated with a comparable subject pool, the same experimental protocol generally yields the same results.
3. Lab Experiments are **quick and cheap**. It is usually fast and inexpensive to assess alternative explanations of a main result by running a modified version of one’s design.

They are also a good way to pre-test designs for more expensive and cumbersome field experiments.

(so, to some extent, lab experiments are the labor economist’s equivalent of animal trials in medical research).

4. Lab experiments often provide unique opportunities to study **behavior that is illegal or hidden** in the field: sabotage, discrimination, spite.
5. Game-theoretic models (e.g. of bargaining or signalling) often have crisp predictions about agents’ beliefs: it is relatively easy to study these in the lab.

6. Some theoretical models have crisp predictions about the form of optimal strategies (e.g. the reservation wage property); it is typically much easier to elicit strategies directly in the lab than anywhere else.

7. In addition to estimating the effects of a randomly- and exogenously-allocated treatment, lab experiments can offer nice, controlled ways to study agents’ self-selection into treatment.
8. Simple ‘Lab’ experiments, combined with field studies and social surveys, can help us understand the **sources of heterogeneity** in individual behavior in survey or field data:

- Mexican Family Life Survey (Eckel et al., 2006)
- GSOEP (Fehr et al. 2003, Dohmen et al. 2005)
- Trucker trainee data (Anderson et al. 2011))

9. The lab is also the only way (so far) to **collect physiological data**, or to **administer physiological interventions**:

- heart rate variability (Falk et al. 2011), skin conductance responses (Coricelli et al. 2010), fMRI (Dohmen et al. 2011),
- oxytocin (Kosfeld et al. 2005), testosterone (Eisenegger et al 2010)
10. Lab experiments can provide an inductive basis for building new theories in situations where:

- theory yields multiple equilibria (e.g. infinitely repeated games and the folk theorem)

- no formal models exist (e.g. ‘extinction’ of behavior after fixed versus random reinforcement)

- existing theory is decisively contradicted by the evidence (e.g. the dictator game).
B. Objections to Lab Experiments

By far, the most common concern about lab experiments is about **external validity**: do lab results apply to the real world? This is an important question.

There are two main responses:

1. ‘Standard’ economic theory should apply everywhere. [If not, who decides where it is expected to apply, and where not?]

2. Many of the specific concerns about what is ‘special’ about the lab environment can in fact be addressed:
a) experience: Perhaps subjects behave differently once they have acquired experience with a particular task or game.

One can give subjects more experience, and vary the level of feedback they receive.

One can conduct (both lab and field) experiments on subjects with real-world experience.

Also: note that many important ‘real world’ decisions do not allow for much ‘learning by doing’: retirement saving, seatbelt use, health insurance plan selection.
b) selection: The people who do a given job in the real world (e.g. managers, lawyers, fund managers) may be highly selected on characteristics like inequity aversion, risk aversion, tastes for competition, spite.

One can conduct experiments on these specific groups. Lab experiments have been done 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)
- employees at large French firms (Charness and Villeval 2009)
- MBAs (Camerer and Lovallo 1989), etc.

One can measure these personal characteristics and study the selection process in the lab. (e.g. Gneezy et al., 2003, Della Vigna et al.).
c) *stakes*: experiments have been done with very high stakes (e.g., Fehr and Touareva 1995, Slonim and Roth 1998). Higher stakes do not necessarily lead to smaller concerns with ‘fairness’ or to fewer mistakes: Ariely, Gneezy, Loewenstein, and Mazar (2008).

Also: many day-to-day decisions of considerable consequence and interest to economists are for small stakes.

d) *scrutiny*: it is fairly straightforward to design experiments that limit the experimenter’s knowledge of the subjects’ choices.

Also: lots of real-world decisions are highly ‘public’ and subject to scrutiny by peers and superiors.
e) ‘artificiality’, e.g. ‘real effort’ versus ‘chosen effort’. Again, up to a point, this can be addressed. Real effort ‘tasks’ that have been used in the lab include:

- 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 et al. Winden 2001)
- simple clerical tasks (Falk and Ichino 2006, Carpenter et al., forthcoming)
- cracking walnuts (Fahr and Irlenbusch 2002)
- data entry (Kube, Maréchal, and Puppe (2008)

f) duration: (‘hot’ versus ‘cold’ decisionmaking). To some extent, this can be manipulated in the lab as well, though there are clear limits to how long a lab session can last.
In sum, is a field experiment always better than a lab experiment?

No. A hammer is not ‘better’ than a screwdriver: 

It depends what you want to do!

- If the goal is to estimate the magnitude of a response to a particular policy change (e.g. output response to the piece rate, unemployment duration to a re-employment bonus) for a particular pool of agents, the ideal place is surely the field.

- If the goal is to test a simple, explicit theoretical model of human behavior of potential relevance to a wider population, the lab may be your best choice.
Finally, note that the **boundary between lab and field experiments** is not clear: How to classify internet experiments (Cooper on entrepreneurs; Barankay on MTurk)?

So, just use the design elements that best suit your goals.
III: Designing and Evaluating Laboratory Experiments

BEFORE YOU START:

1) Begin with a well-specified theory
   and

2) Be prepared to have it soundly rejected.
A. General Design Questions

1. Parsimony versus Richness.

How closely should you try to match the field environment? While a close correspondence is nice, remember that:

- one advantage of the lab is that you can eliminate confounding factors that make field data difficult to interpret. Make use of this advantage!

- too much complexity taxes subjects’ understanding.

When in doubt, we believe it’s wise to err on the side of simplicity, while ensuring that the experiment captures all the aspects of reality that are relevant to your theory.
2. How much to explain to subjects?

Should you provide examples? How many? A practice period? Even coaching?

Current practice is to:

- provide (a balanced set of) examples.

- in many cases, allow a practice session or two

- administer a ‘comprehension questionnaire’ to ascertain that participants understand the structure of the interaction and the consequences of various choices.
3. Framed or unframed?

-in some experiments (e.g., Charness and Rabin 2002), the researchers are careful to choose completely neutral terms, to limit the role of any preconceptions/rules of time.

-it is increasingly common, however, to ‘frame’ experiments in a context that is familiar to the subjects. For example, the experiment might explicitly label the subjects’ roles as ‘firms’ or ‘workers’ and to label the choices as ‘wages’ or ‘effort’, even though (strictly speaking) the experiment is simply a game with no actual work performed.

-context helps with understanding, and may improve external validity with respect to the context that is supplied, but it may limit the generality of the findings.

And remember that framing matters: Liberman, Samuels, and Ross (2004) implement identical prisoner’s dilemma games, labelled either as “The Wall St. Game”, or “The Community Game”. The rate of cooperation was less than 30% in The Wall St. Game, but was over 70% in The Community Game.
4. Repeated versus One-shot Interactions?

It depends what you are interested in:

(a) one-shot interactions (Theory tends to have the sharpest predictions here, but how common is this?):

- Either run the experiment once only (expensive; subjects may not understand the game yet), or

- run repeated periods with random rematching (“strangers” design). Most people do this.

(b) repeated interactions with a clear end date: (Theory usually predicts ‘unravelling’): Run a “partners design”.

(c) repeated interactions over an indefinite horizon: This can be done with a random probability of the game ending at any time. Dan Houser’s recent paper has an economical way of doing this. (Because of the folk theorem, precise theoretical predictions are often unavailable here.)
5. One Role per Subject, or More?

Can the role assigned to subjects change from period to period? It’s not clear which approach best facilitates learning, but role change permits the experimenter to compare an individual’s behavior across the various roles.

6. Within-Subjects or Between-Subjects Design?

- *between-subjects design* exposes each subject to one treatment only

- with random assignment of treatments, this is the cleanest design, but it can be expensive. Also problematic if subjects require practice to understand the experiment.
-within-subjects design exposes an individual subject to multiple treatments

-in an experimental context, adding person fixed effects provides efficiency gains only

-potentially problematic if framing and sequencing matter, e.g. if a subject’s behavior under one condition is influenced by his prior exposure to other conditions.

(one compromise: do within-subjects, but verify using first-period, cross-subject comparisons).

Charness, Gneezy and Kuhn (2011) review the literature on within- versus between designs.
7. Calibration: How to pick Parameter Values (and Functional Forms)?

- this is as much art as science.
- look at the literature
- sometimes trial and error is required
- the key is to parameterize the baseline situation so that there is ample ‘room’ for the treatment to move the outcome in both directions.

8. How Much, and How to Pay?

- Incentives should be large enough to induce thoughtful and motivated behavior.

- Pay each period separately, or for one (or more) randomly-chosen period? The latter is getting more common:
  - avoids wealth effects, bankruptcy effects
  - mitigates boredom in later rounds
9. Play the Game Through, or use the Strategy Method?

In games where subjects move in turn, it’s traditional (and natural) to inform the responder of the first mover’s actual choice.

But, especially when some first-round choices are rare (but matter a lot to the outcome), the ‘strategy method’ (Selten 1967) is much more economical. Here, the responder states a response to every possible action of the ‘sender’.

But, how ‘real’ is this? Brandts and Charness (2009) examines many comparisons of results, with generally no qualitative difference.

10. Use Students, or “Real People”?

Of course, if you are interested in a specific real-world group, use them!

If you are interested in the broader population, students span a pretty broad spectrum (relative, for example, to Safelite employees). Students are also relatively poor and smart: a good combination for the experimenter, and are readily available. On the other hand, they have relatively little experience as employees, and essentially none as employers.
B. Design Questions in principal-agent/effort experiments

1. Is there a labor market (and if so, how to model it)?

   - Standard, and most common approach is fixed outside options (often zero) if firms and workers choose not to interact.

   - But some of the earliest (e.g. Fehr, Kirchsteiger, and Riedl (FKR) 1993) and some of the best (e.g Brown, Fehr and Falk (BFF) 2004) papers explicitly implement markets for labor contracts. FKR 1993 showed that workers’ fairness concerns can cause these markets to fail to clear. [Charness, Kuhn and Villeval (CKV, 2009) have shown causation can run the other way: introducing ex post labor markets can eliminate the well-known ratchet effect in the repeated principal-agent problem.]

   - FKR did a one-sided oral auction where firms propose wages and any worker can accept. BFF allow for ‘private offers’. Both have an excess supply of workers. CKV take a simpler approach and consider both excess workers and excess firms. Another question is whether a firm can employ more than one worker.
2. ‘Real’ or ‘Stated’ Effort?

(“Stated” effort is just a number that implies an (increasing, convex) cost to the worker.)

-obvious advantage of real effort is realism.

-disadvantage is that it’s messy: subjects will differ in ability (or effort disutility); these differences are:
   -initially unobserved by the experimenter
   -correlated with other sources of individual heterogeneity (e.g. risk aversion, competitiveness, reciprocity).

-useful feature of stated effort is that we know the agent’s disutility of effort function and can manipulate it experimentally, independently of the above confounding factors.
3. Do the subjects work for each other, or for the experimenter?

If you are only interested in the agents’ behavior, one approach is to have the agents work for you (the experimenter). Advantages are

- cost

-subjects may view the experimenter more like a ‘real’ employer (as opposed to their schoolmate).

-college students’ behavior in the role of principals is not very informative about real firms’ behavior.

If you also want to study principals’ behavior (or labor market equilibrium), you must have some subjects ‘work for’ others. Advantages:

-If you are interested in social preferences towards persons other than the experimenter, this may be more accurate.

-Also, if you use ‘real’ managers, you can study their behavior.
4. Other design decisions:

- the reward scheme
- can workers self-select among reward schemes?
- does the experimenter induce reference points?
- can principals coerce agents?
- how many agents per principal?
- can the agents observe each other’s performance?

These, and many more, constitute the research agenda in empirical personnel economics, both in the lab and elsewhere.
3. Reading papers involving laboratory experiments: some pointers

-understand the design: insiders often read the instructions first.

-understand the flow of information: what the participants knew and when they knew it. Did the participants understand the game/task?

-how well does the experiment test the theory, or match the relevant field environment?

-is the framing likely to bias behavior?

-what’s an observation? statistically, how do the authors treat correlations within subjects, or within sessions?
  -one extreme view is that one session = one observation
  -another approach is to average each participant’s choices in a session and treat this as one observation
  -otherwise, need to handle within-subject and within-session correlations in some way.
-how **parametric** are the statistical tests? (‘workhorse’ tests are Wilcoxon rank-sum test, exact binominal tests).

-are the variables of interest **randomly assigned**? (Sometimes they can’t be, even in a lab experiment: e.g. gender, race, or the wage offered by subject A to subject B.)
IV. Lab Labor Results: What have we learned?

Labor economics topics that have been addressed using laboratory experiments include:

A. Bargaining, Strikes and Arbitration
B. Search
C. Labor Market Equilibrium, including minimum wage effects
D. Gender and Racial Differences and Discrimination
E. The Supply of Effort; Principal-Agent Interactions
A. Bargaining, Strikes and Arbitration

-lab experiments on bargaining go back at least to Deutsch and Krauss (1960) in psychology and Siegel and Fouraker (1960) in economics.

Deutsch and Krauss’s apparatus for bargaining experiments:
the other can go forward, or both can back up, or both
can sit there head-on without moving.

There is another way for each S to reach the desti-
nation on the map, labeled the "alternate route." The
two players' paths do not cross on this route, but the
alternate is 56% longer than the main route. Ss were
told that they could expect to lose at least $.10 each
time they used the alternate route.

At either end of the one-lane section there is a gate
that is under the control of the player to whose start-
ing point it is closest. By closing the gate, one player
can prevent the other from traveling over that section
of the main route. The use of the gate provides the
threat potential in this game. In the bilateral threat
potential condition (Two Gates) both players had gates
under their control. In a second condition of unilateral
-acceleration in economics with arrival of formal non-cooperative bargaining theory, both under full and asymmetric information. For early reviews, including experiments and enthusiasm for labor applications, see Kennan (1987) Kennan and Wilson (1993), Roth (1995).

-major anomalies in the early experiments sparked the laboratory study of extremely simple bargaining games, such as the ultimatum and dictator games, to isolate their causes (e.g. Charness and Rabin 2002). This literature includes lab experiments on bargaining and:

- 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)
- more than two players (Fréchette, Kagel and Morelle 2005; Charness, Corominas-Bosch and Frechette 2007).
Bargaining Experiments with *Labor Applications*:

*a) Union Bargaining, Strikes, and Bargaining Structure:*
Sopher (1990); Forsythe, Kennan and Sopher (1991, 1991b)


*b) Arbitration and Dispute Resolution*

-testing (and “pre-testing”) arbitration mechanisms in the lab has a long history in psychology (e.g. Johnson and Tullar 1972)

-earliest economic analysis is Farber and Bazerman (1986): their subjects were professional labor arbitrators
Arbitration Systems Studied in the lab include:


- Tri-offer arbitration, Combined arbitration, Double-offer arbitration, Amended-final-offer arbitration, and Automated negotiation. (See Kuhn 2009 for a review).

c) Gender wage gaps

Gender differences in bargaining have been studied in the lab and may be relevant to the gender wage gap (e.g. Eckel and Grossman 1997, Babcock and Laschever 2003).
In sum:

- experimental study of bargaining has a long history in psychology and economics

- has rejected some early, theoretical models of bargaining and spawned a large theoretical and empirical literature on behavior in very simple bargaining games.

- has yielded some interesting labor applications, including one labor economics success story:

**The laboratory study of dispute resolution** has yielded insights not only of interest to *Econometrica* readers, but also to policymakers who are faced with the real problem of designing mechanisms that work well in the real world, e.g. in public-sector labor markets where strikes are prohibited by law.
B. Search

Labor economists have conducted well-known field experiments loosely based on the job search paradigm; see Meyer (1995, 1996) for excellent reviews.

We review only search experiments based very tightly on the theory (as reviewed by Lippman and McCall 1987a,b).

Earliest lab experiments on search are Schotter and Braunstein (1981, 1982).

Other lab experiments on search include:

- Hey (1982) (included “protocol analysis” subjects report their thinking as they play)
  
  - Sonnemans (1998) examines subjects’ use of an information board; also implements strategy method.

  - Cox and Oaxaca (1989, 1992) looked at finite horizon
- Harrison and Morgan (1990) allow for variable search intensity

- Cox and Oaxaca (2000) considers search from unknown distribution

- Carbone and Hey (2004) look at search and consumption

- Gabaix, Laibson, Moloche, and Weinberg (2006) look at limited cognition

- Schunk (2009); Schunk and Winter (2009) look at reference point updating

-Most results are quite consistent with basic theory, except:

- reservation wages decline

- subjects stop searching too soon

- subjects adopt strategies that depend on factors (number of searches, total earnings to date) other than a reservation wage.
Most recently:

Brown, Flinn and Schotter (forthcoming) argue that sharply declining reservation wages and ‘early stopping’ are an artifact of experimental design:

Their results suggest subjects have sharply rising time costs, perhaps related to students’ short-term time commitments outside the economics laboratory.

**In sum**, while search in the lab is “not too suboptimal”, even under ideal laboratory conditions subjects systematically choose search strategies that are suboptimal *in a particular way*.

This raises interesting hypotheses for research using field data, and for structural models of job search.
C. Labor Market Equilibrium

a) The “Market Design” Literature (originated with Plott; reviewed by Roth 2008), has studied the following markets, among others:

- airport landing rights (Grether, Isaac and Plott 1981; Rassenti, Smith and Bulfin 1982)
- 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)

Lab experiments are used in these studies for a variety of purposes, including pre-testing proposed institutions.
Labor markets studied in the ‘market design’ literature include


Clinical psychologists: Roth and Xing (1997)

Law clerks: Haruvy, Roth, and Unver (2006); Avery, Jolls, Posner and Roth (2007).

Also of interest to labor economists are market-design studies of:

- college admissions (Roth and Sotomayor 1989)
- school choice mechanisms Chen and Sonmez (2006)
- admissions to New York City high schools (Abdulkadiroglu, Pathak and Roth 2009).

Of the above, Nalbantian and Schotter (1995); Kagel and Roth (2000); McKinney, Niederle and Roth (2005), Chen and Sonmez (2006) and Haruvy, Roth and Unver (2006) all use laboratory experiments as at least part of their research design.
**Key questions** studied in labor market design:

- the causes of ‘unraveling’

- how to ensure enough market thickness to allow for efficient matching.

**Uses of lab experiments in market design** analyses:

- test hypotheses about particular market mechanisms
- compare mechanisms’ performance
- pre-test final designs before field implementation.

-isolate the effects of different institutional changes that co-vary in the field (e.g. Niederle and Roth 2009 disentangle exploding offers vs. binding acceptances).

-assess the robustness of market mechanisms to large potential swings in demand and supply (McKinney, Niederle, Muriel and Roth 2005): impractical or unethical in the field

-provide detail on how individual actions change during transitions that is unavailable in field data (Kagel and Roth 2000).
In sum, laboratory study of market design is another (labor) economics success story:

Economic theory, combined with a carefully chosen set of research methods *including* lab experiments, has created institutions that improve the efficiency of economic exchange in the real world.
b) Effects of Contractual Incompleteness (Moral Hazard) on Labor Markets

If effort is noncontractible, markets tend to exhibit rent-sharing, long-term relationships, and involuntary unemployment:

Fehr, Kirchsteiger and Riedl (1993)

-the labor market is a one-sided auction in which firms posted contract offers to all workers in the session.

-effort was either contractible (a worker was forced to provide the effort specified in the contract), or not.

-when effort is not contractible, firm post above-market-clearing wages; these wages elicited more effort than market-clearing wages.

-firms that tried to underbid such wages did poorly due to workers’ effort responses

-involuntary unemployment resulted.

Note the parallels to Shapiro and Stiglitz (1984), but the mechanism is different: worker reciprocity rather than threat of dismissal.
Brown, Falk and Fehr (2004)

-in a session, ten workers interact 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.

-the market for contracts is, 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.

-Now, firms can make both private and public offers: private offers were only conveyed to the worker with whom the firm wanted to trade.
BFF compare two main experimental conditions:

Complete contracts (C): the firm’s required effort level was automatically implemented (by the computer) if a worker accepted a contract.

Incomplete contracts (ICF), this third party enforcement of desired effort levels was absent.

In both treatments, firms only observe the (current and past) effort levels of their ‘own’ workers.
BFF’s Findings:

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
- rent sharing is driven toward the competitive equilibrium (where firms retain all the surplus).

Under incomplete contracts, successful exchange usually proceeds as follows:

- the firm makes a generous contract offer that a worker reciprocates with a high effort level.
- then, the firm repeatedly seeks out the same worker with a private offer.
- rents are \textit{shared} in these long-term bilateral relationships
- these long-term relationships are disciplined by the threat of non-renewal.
- unemployment is \textit{involuntary}

Which of these two scenarios strikes you as more representative of ‘real’ labor markets?
In my view, the BFF view of labor markets (where moral hazard problems inside the firm have a profound effect on how the outside labor market works), may be a more useful way to conceptualize labor markets than either the perfectly competitive one, or one of the many equilibrium search models that are so popular nowadays.
Variations on BFF:

Brown, Falk and Fehr (2008) replace the excess supply of workers by an excess supply of firms.

Perhaps surprisingly, not much changes: 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.

Falk, Huffman and Macleod (2008) introduce two institutions -- dismissal barriers and bonus pay.

EPLs 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.
c) Effects of Adverse Selection on Labor Markets

Experimental studies of signaling and screening models are surprisingly plentiful, dating back at least to Miller and Plott (1985).

Most are cast in an insurance context (especially screening), or product quality choice (signaling). As a result, they are not familiar to most labor economists.

Kübler, Müller, and Normann (2008) conduct signalling and screening experiments framed in a labor market context, and reviews the earlier experimental literature in other contexts.

Cabrales, Charness and Villeval (2006) and Charness, Kuhn and Villeval (2008) also consider hidden information in labor markets in the laboratory.
D. Gender, Race and Discrimination in the Lab

1. Gender differentials.

Babcock and Laschever (2003) suggest that women are more likely to shrink from negotiation.

Charness and Gneezy (2009) find that women are more financially risk averse than men.

Niederle and Yestrumskas (2008) find that men choose harder tasks than women.

Schwieren (2003) finds that women receive lower pay for comparable work in a gift-exchange experiment, regardless of whether men or women are in their role of the firm.

Several papers, starting with Gneezy, Niederle, and Rustichini (2003), show that women perform worse in competitive environments, in a variety of settings and populations, including schoolchildren (Gneezy and Rustichini 2004).

Niederle and Vesterlund (2007) provide evidence that women avoid competitive environments.
Gneezy, Leonard, and List (2009) show this differs between patriarchal (Maasai) and matrilineal societies (Khasi) societies.

Booth and Nolen (2009b) find this varies (among girls) between coed or single-sex schools.

Kuhn and Villeval (in progress) are studying gender gaps in preferences for a cooperative work environment.

For a review of evidence on gender differences (most of it experimentally-based), see Croson and Gneezy (2009).
2. Race and ethnicity

Overall, surprisingly little experimental work by economists:


Fershtman, Gneezy, and Verboven (2005) try to distinguish between discrimination and nepotism using an investment game in Israel and Belgium.

Michele Belot (Oxford) is studying facial recognition and race.

Fryer et al. (2005) study market equilibrium in the lab when ‘green’ and ‘purple’ workers have different investment costs).
Psychologists commonly study race and sex discrimination:

- using ‘mock’ job interviews (often with real recruiters). See for example Cohen and Bunker 1975 re. sex; Dovidio and Gaertner 2000 re. race.

- using the implicit association test (IAT): [https://implicit.harvard.edu/implicit/](https://implicit.harvard.edu/implicit/)

- using on-line computer games, where avatars with different names, genders and skin tones are randomly assigned (Goodwin et al, 2010).

- using confederates to study subjects’ responses to actual and reported racist behavior/comments in the lab (Kawakami, et al. 2009)

- examining effects of interracial interaction on mental ‘resource depletion’ (Richeson and Trawalter 2005)

- using dot-probe detection paradigm to study attention to black versus white faces (Richeson and Trawalter 2008)
3. Beauty

Mobius and Rosenblat (2006) investigate the “beauty premium” in experimental labor market in which firms choose wages for workers in a real-effort task.

Wilson and Eckel (2006) also investigate the beauty premium using the investment game.

Andreoni and Petrie (2008) study the beauty premium in a public-goods game.
4. Policy

A number of experiments model affirmative action as a ‘handicap’ imposed on abler participants in a tournament setting. Essentially all have found that, for the right parameter values (e.g. a large-enough ability gap), AA is either costless or profit-enhancing:

Schotter and Weigelt (1992) (framed as effort supply)

Corns and Schotter (1999): similar result for price-preference auctions.

Niederle, Segal, and Vesterlund (2008) ask what happens when women are guaranteed equal representation among winners.

Balafoutas and Sutter (2009) compare AA to other policy interventions that promote women in competitive environments.


Psychologists have done ‘scenario’-based experiments on affirmative action (e.g. Levi and Fried 2008).
E. Principal-Agent Interactions and the Supply of Effort

Plan:

a) Implementing “Classic” P/A Models in the Lab: Results and Anomalies

b) Understanding the Anomalies: *Towards a Behavioral Personnel Economics*
a) “Traditional” P/A Models that have been implemented in the lab include:

1. The Basic Principal-Agent Problem (one agent, one period, one task)
2. Multiple Agents: Tournaments
3. Multiple Agents: Teams
4. Multiple Tasks
5. Multi-period Principal-Agent Interactions
A Sample of Findings (evidence of ‘social preferences’ in blue):

1. The Basic Principal-Agent Problem

a) For the most part, higher piece rates raise effort (Swenson 1988, Sillamaa 1999a,b, Dickinson 1999). Agent self-selection reinforces these effects (Lazear 2000; Dohmen and Falk 2006, Cadsby et al. 2007).

b) Zero wages can yield higher effort than low wages. Gneezy and Rustichini (2000); very high stakes can reduce performance.

d) Effort also responds to the intercept of the compensation schedule, at least when generosity is seen as intentional (Fehr, Kirchsteiger and Riedl 1993; Charness JLE 2004). [gift exchange game]

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).
f) **Costs of Control**: Restrictions on agents’ choice sets can reduce the efforts of agents on whom they are not binding (Falk and Kosfeld 2006); not when the restrictions are ‘legitimate’ (Schnedler and Vadovic 2007).

g) A principal’s *decision to use* piece rates can reduce agents’ efforts; this may be because it changes agents’ interpretation of the implicit contract (Irlenbusch and Sliwka 2005).

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

i) **Cheap talk**: Unenforceable promises by principals to reward ‘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).

k) **Peer effects**: Even when there is no strategic independence between workers, workers’ efforts sometimes depend on their co-workers’ efforts (Falk and Ichino 2006). *Co-workers’* wages, however, do not appear to affect effort in gift-exchange games (Charness and Kuhn 2007).
2. Tournaments in the Lab

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

b) handicaps improve the performance of tournaments between unequal agents, though some of this is because in the absence of handicaps the actions of less-able agents often diverge from Nash behavior (sometimes working too hard, sometimes dropping out completely) (Schotter and Weigelt 1992).

c) excessive entry into tournaments is not always a problem (Rapaport 1995), but may be for certain subgroups, such as MBAs (Camerer and Lovallo 1989) and men (Niederle and Vesterlund 2007). 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 though (Irlenbusch et al. 2007).

f) collusion is rare in anonymous tournaments with more than two contestants (Harbring and Irlenbusch 2008).
3. Team Incentives in the Lab (closely related to VCM literature).

a) free riding can be significant, especially with experience, larger teams, and in the absence of communication and/or repeated interaction. **Altruistic punishment** can dramatically improve team performance.

b) the forcing contracts suggested by Holmstrom (1982) don’t work, due to co-ordination problems (Nalbantian and Schotter 1997).

c) adding competition between teams can be more effective than any of the above strategies (Nalbantian and Schotter 1997).

d) when there is complementarity between the efforts of team members, co-ordination failures can cause severe efficiency losses. (Van Huyck, Battalio and Beil 1990). Adding cheap talk can generate dramatic improvements (Brandts and Cooper 2007).
4. The Multi-Task Principal-Agent Problem

a) as predicted (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) cheap talk ‘bonuses’ for ‘satisfactory’ overall performance by the agent work remarkably well (Fehr and Schmidt 2004).

5. Multi-Period Principal-Agent Interaction

a) ‘Ratchet effects’ can be generated in the lab (Cooper, Kagel and Lo 1999); labor market competition eliminates them (Charness, Kuhn and Villeval 2008).

b) Signal-jamming ‘career concerns’ equilibria can be generated in the lab (Irlenbusch and Sliwka 2006; Koch, Morgenstern and Raub 2009), but do not respond as predicted to making ability publicly observable. (Irlenbusch and Sliwka 2006).
b) Understanding the ‘anomalies’ in Principal/Agent Experiments: Towards ‘Behavioral Personnel Economics’

Two main types of ‘social preference’ models:

a) **Distributional** (inequity or difference aversion) (Bolton and Ockenfels 2000; Fehr and Schmidt 1999): agents willing to sacrifice money to equalize payoffs.

b) **Reciprocity** (Rabin 1993). “kindness” = payoff options made available to the other player by one’s choices. Agents are willing to sacrifice money to reward kindness, punish unkindness. Key: *Intentions matter.*

Various versions and hybrids exist, including models where social efficiency matters (Charness and Rabin 2002), ‘guilt aversion’ (Charness and Dufwenberg 2006) and ‘lying aversion’ (not yet formalized).

Distinguishing (a) and (b) can be tricky; Charness (2004) did so. Negative reciprocity is much stronger than positive.

The gift exchange game (Fehr, Kirchsteiger and Riedl 1993) provides highly robust evidence of gift exchange in a ‘labor’ context in the lab.
A gap in the literature:

There has been surprisingly little work on extending the above social preference models to multi-worker firms: it seems unlikely the ‘distributional’ models will be helpful here.

Only a few experimental studies of gift exchange with multiple workers: Charness and Kuhn 2007, Maximiano et al., 2007).
The $64,000 question: How important are gift exchange and reciprocity in real labor markets?

a) ‘Real effort’ experiments:

Gneezy and List (2006) show very low persistence of gift exchange in real-effort field experiments; but their results pertain to positive reciprocity only.

Kube, Marechal and Puppe (2006) find stronger and more persistent effects of negative reciprocity.

Together, these two papers suggest that workers’ effort choices may be highly subject to framing effects. Repeatedly convincing workers they have received a ‘gift’ may be challenging indeed.
b) Evidence of gift exchange in real workplaces:


Krueger and Mas (2004) and Mas (2006, 2008) show retribution with real firms and workers.

Bewley (1999) presents evidence that managers expect such retribution.

Bandiera et al. illustrate several ways in which social preferences affect the response to incentives among fruit pickers.

Very recently, Cappelli has shown large, durable effort responses to ‘gifts’ of stock options.

But note: in the field it is much harder to rule out instrumental reciprocity.
Overall, robust evidence of social preferences in real workplaces remains relatively scarce, and much remains to be learned about which social preferences matter when in real firms.

A widespread conjecture is that factors like competition, experience and stakes probably drive out many of the laboratory’s ‘behavioral’ anomalies in the real world.

Perhaps, but it could go either way:


- Not all important ‘real’ world decisions allow for a lot of experience/feedback.

- In the lab, Fehr, Kirchsteiger and Riedl 1993 showed that social preferences cause markets to fail to clear, rather than markets crowding out social preferences.
V. Summing Up

1. Lab experiments are *one* useful component of a labor economist’s toolkit. They are especially well suited to

   - testing specific *theoretical models*

   - studying basic, *general* principles of social and strategic behavior.

They are not so well suited to measuring the *magnitude* of responses to specific policy changes in a specific population.
2. Lab experiments have already made important contributions to our understanding of labor markets. ‘Success stories’ include:

- market design for professional workers, school admissions
- mechanism design for arbitration of disputes,
- new insights into the sources of gender differentials in the labor market

and perhaps most important:

The discovery of ‘social preferences’:

Early lab experiments on bargaining and on principal-agent interactions identified robust departures from rational, selfish behavior. “Social preference” models have been developed to explain these.
What is needed next?

-Social preference models that are better suited to real, multi-worker firms.

-Theory and more evidence on *which* social preferences matter *when*. Evidence on this will come from lab experiments, field experiments, econometric studies (‘insider’ and otherwise), and calibration.
Epilogue: A “Personnel Economics” Retrospective

**Personnel economics in 1994:**

“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.

“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. For instance… 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: Internal Labor Markets: Too Many Theories, Too Few Facts” 1995).
**Personnel economics in 2009**: Too many facts?

Perhaps, but:

- some of the models of 1994 perform well in the lab
- others do not.

This is real progress.
New references (not in Handbook):


