ON THE EXPERIMENTAL PRACTICES OF ECONOMISTS AND PSYCHOLOGISTS
I: Lab environs, the „real world“, and the importance of context (Lecture 5, 2009_02_11)
[Outlook:
II: Financial Financial and/or social incentives Incentives (L6)
III: Deception, conversational inferences and rational judgement (L7)
IV: ... Context continued: Abstract instructions and related conundrums ... (L8)}

Please recall my remarks in Lecture 1 about the nature of the lecture notes.

Review last lecture:

Wilson, Experimental Economics and Antitrust: What Can We Learn from Laboratory Markets?

„Experimental markets with as few as three sellers-buyers regularly converge to the competitive outcome.“ (fn 5, p. 58)

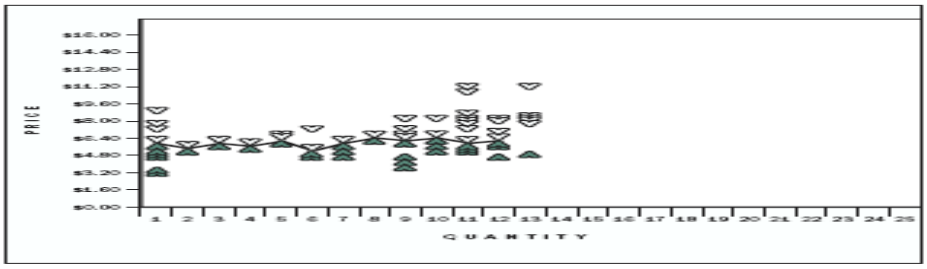


Figure 1

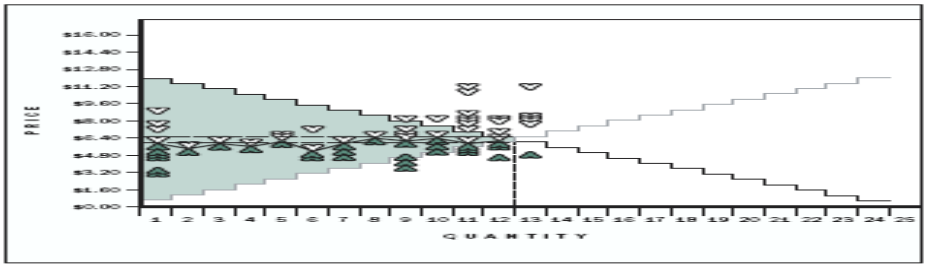


Figure 2

Smith, Induced Value Theory (1976) – a precursor to Smith (AER 1982)

„The laboratory becomes a place where real people earn real money [„are suitably motivated“, p. 274] for making real decisions about abstract claims that are just as „real“ as a share of General Motors.“ (275)

Control – „the essence of the experimental method“ (275); a specific aspect of control is the possibility of replication across experimental studies. Specifically what we need is a way to make sure that individual values (that constitute demand and supply) do not fluctuate in an unspecified way. For experimental exchange studies, this is accomplished through a reward structure that induces prescribed monetary value on actions, the concept of induced valuation.

Qualifications:

1. Subjective costs? (Effects of boredom? Subjective costs of decision making?)
2. Game values? (Utility of taking part in an experiment)

Do these qualifications matter?

- evidence presented on pages 277 - 8:
  - „imagine that trade and profits and commissions are real“: fewer trades, first price convergence, then price divergence (starting in period 4) attributed to gaming boredom that follows an initial (pleasant) experience of learning.
  - more complex tasks (higher subjective decision costs) induces volume below competitive prediction
  - random reward structure retards the market's convergence

TABLE 1—MEAN CONTRACT PRICE BY TRADING PERIOD

Experiment	1	2	3	4	5	6	7
Excess Supply	5	5	5	5	8	8	8
Reward Condition	Complete	Complete	Random	Complete	Complete	Complete	Complete
Information Condition	Incomplete	Incomplete	Incomplete	Complete	Incomplete	Incomplete	Complete
Trading Period 1	3.48	3.67	3.60	3.51	3.26	3.49	3.56
Trading Period 2	3.29	3.26	3.44	3.40	3.15	3.28	3.25
Trading Period 3	3.19	3.12	3.31	3.34	3.11	3.13	3.20
Trading Period 4	3.14	3.10	3.24	3.37	3.10	3.12	3.17

3. „Individuals may not be autonomous own-reward maximizers. Interpersonal utility criteria may qualify the theory of induced valuation. ... with information about one other's payoffs, the way is open for ‚equity' considerations to modify self-interest choices.“ (278)

   „The tendency of prices to be higher under complete information is contrary to the view of those who have argued that ‚perfect' information is essential for establishing competitive prices. The results are consistent with the game-theoretic proposition that more information increases the prospect of collusion (), and with the results of Fouraker and Siegel () in which the tendency of the competitve equilbrium to prevail under duopoly bargaining is reduced under complete information.“ (279) [This result is also very interesting in light of recent debates about social preferences!]

Smith, Microeconomic Systems as an Experimental Science (1982)

Perhaps the most important general feature of the experimental results summarized in all the above *DA* propositions is the support they provide for what might be termed the Hayek hypothesis: *Markets economize on information in the sense that strict privacy together with the public messages of the market are sufficient to produce efficient C.E. outcomes*. This statement is offered as an interpretation in hypothesis form of what Hayek meant in emphasizing that "the most significant fact about this (price) system is the economy of knowledge with which it operates, or how little the individual participants need to know in order to be able to take the right action..."(Hayek, 1945, p. 35).

„At the heart of economics is a scientific mystery: How is it that the pricing system accomplishes the world's work without anyone being in charge? ... The pricing system – How is order produced from freedom of choice? – is a scientific mystery as deep, fundamental and inspiring as that of the expanding universe. It would appear that after 200 years, we know and understand very little." (952)

- Sufficient conditions:
    - Precept 1: Nonsatiation
    - Precept 2: Saliency
    - Precept 3: Dominance (recall little discourse about Harrison's payoff dominance critique)
    - Precept 4: Privacy
    - Precept 5: Parallellism

f. *Precept 5: Parallelism.* Nonsatiation and saliency are sufficient conditions for the existence of an experimental microeconomy, that is, motivated individuals acting within the framework of an institution, but they are not sufficient for a *controlled* microeconomic experiment. For this we also must have dominance and privacy, since individuals may experience important subjective costs (or values) in transacting, and may bring invidious, egalitarian, or altruistic cannons of taste to the laboratory from every day social economy. Precepts 1–4 permit us to study laboratory microeconomic environments in which real economic agents exchange real messages through real property right institutions that yield outcomes redeemable in real money.

Insofar as we are only interested in testing hypotheses derived from theories, we are done, that is, Precepts 1–4 are sufficient to provide rigorous controlled tests of our ability as economists to model elementary behavior. Microeconomic theory abstracts from a rich variety of human activities which are postulated not to be of relevance to human economic behavior. The experimental laboratory, precisely because it uses reward-motivated individuals drawn from the population of economic agents in the socioeconomic system, consists of a far richer and more complex set of circumstances than is parameterized in our theories. Since the abstractions of the laboratory are orders of magnitude smaller than those of economic theory, there can be no question that the laboratory provides ample possibilities for falsifying any theory we might wish to test.

A sufficient condition for this transferability of results can be summarized as a final precept (compare my 1980 article).

> *Parallelism*: Propositions about the behavior of individuals and the performance of institutions that have been tested in laboratory microeconomies apply also to nonlaboratory microeconomies where similar *ceteris paribus* conditions hold.

Plott, Dimensions of parallelism: some policy applications for experimental methods

„ ... 10 instances in which I have been involved personally at some level ...

... the topic is policy research as opposed to basic research. The issues are: What was attempted, what seemed to work and why, what was a flop and why? ... Five different strategies are identifiable ...

7.2. Ex post evaluation of a decision: the flying club

7.3. Demonstration: landing slot allocations

7.4. Shift in the burden of proof

... Some of the demonstration arguments used by Grether, Isaac, and Plott in the Polinomics airport slot study could be counted as a third instance of shift of burden of proof strategies.

7.4.1. Inland waterways barge traffic

The difficult part was to determine an appropriate scale. ...

7.4.2. The Ethyl case

...

A interesting feature of all three attempts to use the shift of burden of proof strategy is that the experimenters were designed to mirror the industry as closely as possible. Relative sizes of buyers and sellers, demand elasticities, numbers of participants, and so on were all similar to those of target industries. ...

7.5. Direct extrapolation: air freight posting

7.6. Potential design: prepolicy research

7.7. Design

Sometimes economic problems require completely new types of organization and decision processes. ...

7.7.1. Slot exchanges

7.7.2. Westchester County airport

7.7.3. Space station

7.8. Closing remarks

... the use of laboratory methods in policy contexts seems to be an art involving a skillful and very subjective choice of experimental conditions. ...

Roth, The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics.

A particular interesting example that Roth explored is motivated by Table 1 (p. 1351): It shows that stable mechanisms seem indeed to be more successful than unstable mechanisms. However, unstable mechanisms are not necessarily doomed: the unstable mechanisms at Cambridge and London Hospital were still in use when the article was published. See the discussion on pages 1350 – 1354 on the (Edinburgh) deferred acceptance (firm-proposing) mechanism [see p. 1349 for description or the worker-proposing analogue] vs the (Newcastle) priority algorithm (which defines a „priority“ for each firm-worker pair as a function of their mutual rankings.)

Here are the relevant Table 1 and Figure 1 (data from Kagel & Roth 2000):

TABLE I

STABLE AND UNSTABLE (CENTRALIZED) MECHANISMS

Market	Stable	Still in use [halted unraveling]
American medical markets		
NRMP	yes	yes (new design in '98)
Medical Specialties	yes	yes (about 30 markets)
British Regional Medical Markets		
Edinburgh ('69)	yes	yes
Cardiff	yes	yes
Birmingham	no	no
Edinburgh ('67)	no	no
Newcastle	no	no
Sheffield	no	no
Cambridge	no	yes
London Hospital	no	yes
Other healthcare markets		
Dental Residencies	yes	yes
Osteopaths (<'94)	no	no
Osteopaths (≥'94)	yes	yes
Pharmacists	yes	yes
Other markets and matching processes		
Canadian Lawyers	yes	yes (except in British Columbia since 1996)
Sororities	yes (at equilibrium)	yes

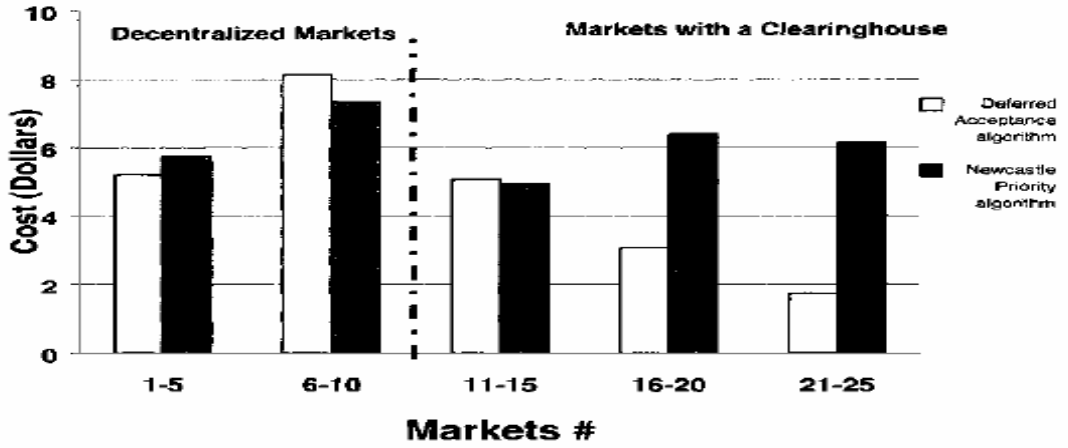


FIGURE 1.— Average costs of early markets, over time. In the first ten markets, #1–10, only the decentralized matching technology is available, and the participants match early, despite the cost. Starting with Market 11, and continuing through market 25, participants who wait until the last market period can use a centralized matching mechanism. In one cell of the experiment this was a stable, deferred acceptance algorithm of the kind used in Edinburgh, and in the other it was the unstable priority algorithm used in Newcastle. With the stable mechanism, costs of going early fell over time, indicating that more matches were being made at the efficient time.

Binmore & Klemperer, The Biggest Auction Ever: The Sale of the British 3G Telecom Licenses.

- Experiences from the 3G mobile-phone licence auction that concluded on 27 April 2000 (and started on 6 March 2000), after being in the works for almost three years (C90).
- Auction took 150 rounds of bidding (C90)
- Auction raised ($34 billion in 2001 exchange rate) or 2.5 % of GNP, or „enough money to build 400 new hospitals“ (C76), very successful compared to other auctions, specifically ... the Dutch [which raised only about a fourth of what it was expected to raise, see p. C93; or the Turkish [which led to a monopoly, see p. C93; or the French ...
- „The exact total raised was English Pound 22,477.4 million (or about English Pound 22,477.3 after deducting the cost of the economics consultants – primarily on programming simulations, running experiments, etc.)“ (C74, fn 1)
- **„The investment bankers advising the government (N.M. Rothschild and Sons Ltd.) were paid a fee that depended on the number of bidders who participated in the auction. By attracting 13 bidders, Rothschilds earned English Pound 4,770,000, or over forty times the total expenditure of economic consultancy.“ (C90, fn 49)**
- Insights:
    - „Beginning the planning so far in advance of the auction () proved a shrewd move by the UK government. It allowed us plenty of time to develop and test our ideas and, just as importantly, it allowed for a sustained marketing campaign without being overtaken in the race to be the first on the European scene (indeed worldwide) with a 3G auction.“ (C90)
    - „A well-designed auction is the method most likely to allocate resources to those who can use them most valuably. Rather than relying on government bureaucrats to assess the merits of competing firms‘ business plans, an auction forces businessmen to put ‚their money where their mouths are‘ when they make their bids. An auction can therefore extract and use information otherwise unavailable to the government. ... the difficulty of specifying and evaluating criteria for a beauty contest [see footnote 11 on Negroponte, „the technology guru who is one of the most prominent advocates of beauty contests“, C76] makes this a time-consuming and opaque process that leads to political and legal controversy, and the perception, if not the reality, of favoritosm and corruption. [„favoring ‚national champions‘“] ... protectionism is unlikley to benefit consumers or taxpayers. ... There are several common objections to auctions. They are said to be unfair to firms, raise consumer prices, and to reduce investment. But all of these complaints are based on misperceptions.“ (C76)

- **„Perhaps the most important lesson of all is not to sell ourselves too cheap. Ideas that seem obvious to the trained eonomist are often quite new to layfolk. Our marginal product in preventing mistakes can therefore sometimes be surprisingly large.“ (C95)**

On other optional readings for today:

THE JOURNAL OF FINANCE • VOL. LVIII, NO. 5 • OCTOBER 2003

Corporate Board Composition, Protocols, and Voting Behavior: Experimental Evidence

ANN B. GILLETTE, THOMAS H. NOE, and MICHAEL J. REBELLO*

ABSTRACT

We examine voting by a board designed to mitigate conflicts of interest between privately informed insiders and owners. Our model demonstrates that, as argued by researchers and the business press, boards with a majority of trustworthy but uninformed "watchdogs" can implement institutionally preferred policies. Our laboratory experiments strongly support this conclusion. Our model also highlights the necessity of penalties on insiders when there is dissension among board members. However, penalties for dissent appeared to have little impact on the experimental outcomes.

SO, HOW TO DO EXPERIMENTS? – certainly
a debate between economists and psychologists (but also between experimental and behavioralist economists)

According to H&O (2001)

Economists

- bring a precisely defined „script“ to the (context-free) experiment for subjects to enact
- often use repeated experimental trials to allows subjects to learn about the task and the environment
- pay subjects on the basis of clearly defined performance criteria (which are informed by models that assume maximization of some function)
- virtually never deceive subjects

Psychologists

- typically don't (provide scripts)
- typically don't (use repeated experimental trials) [Reexamination of Koehler's review of Bayesian reasoning studies]
- usually pay a flat fee or grant course credit [JBDM study]
- do (use deception) [Table 2]

[Note that these statements refer to those areas where economists and psychologists pursue similar topics and questions: behavioral decision making and related areas in social and cognitive psychology such as social cognition, problem solving, and reasoning. These statements do not necessarily apply to research practices in sensation and perception, biological psychology, psycho-physics, neuro-psychology, learning and related fields [behaviorism] (See 1.2.) Note also that our statements are based on empirical (!!!) exercises.]

Ironically, psychologists seem much more laissez-faire than economists in their experimental standards and practices!

Why?

And is it indeed so that the wider range of practices in psychology (and the resultant lack of „procedural regularity“) are responsible for the problems of replicability that psychologists seem to face?

Explanations as to Why (See R8)?

- Historically, experimental economists were the „new kid on the block“ (Lopes, Roth) that had to prove itself to a sceptical crowd. (And sceptical that crowd was; recall the Wallis-Friedman critique, or read Smith Nobel Prize speech to get an idea how sceptical!) One obvious way to go about it was to anticipate, and react to earlier, criticism of their methodology.
- Historically, all the way through 50's, 60's, 70's and 80's experimental economists were working in about half a dozen places (e.g., Arizona, Caltech, Texas A & M, Pittsburgh, Frankfurt, …), facilitating the codification of methodological practices (e.g., Smith 1976, 1982)
- Blaich & Barreto make the interesting suggestion that economists and psychologists – partially in reaction to the different availability of a unifying framework – engage in different experimental practices because psychologists are mostly interested in rejecting null hypotheses while economists are more interested in estimating parameters.

Basic policy prescription: Do-it-both-ways! (Do-it-n-ways!) when in doubt !!!

Enacting a script vs. „ad-libbing“ (See R3, see also 2.)

Definition: A script is a clear and comprehensive set of instructions to participants (and possibly the experimenter) that increases procedural regularity, and hence experimental control and replicability.

- might constrain subjects‘ interpretation of the social situation experiment by focusing participants on the experimenter’s intentions [this aspect interacts with deception]
- might promote subjects‘ active involvement in the experiment by making their choices consequential
- might restrict the sets of perspectives that subjects can take
- makes it possible to study the potentially large influence of seemingly small procedural or wording differences (see Plott & Zeiler’s work later in Lecture 8 and also Hoffman et al. in the same Lecture.)

Interesting problem: Demand characteristics! (We suggest that the benefits from clear and comprehensive instructions typically outweigh the costs of demand effects.)

Repeated trials versus (?) snapshot studies (See R4, see also 3.)

No doubt both are useful!

Why repetition?

- „Practice effects“ (in „games“ against nature as well as interactive decision situations) [the first kind of learning]
- „Learning about others‘ choices“ [the second kind of learning]

[Recall the Plott & Uhl middlemen experiment.]

Interesting issues:

- „Groundhog-day“ replication? [Admittedly, a poor abstraction of real life]
- Learning without feedback? (e.g., Weber 2004)

An Aside on Field Experiments (drawing on Harrison & List JEL 2004)

Field experiments are not just lab experiments
- with less control
- more external validity

Overall taxonomy

- conventional lab experiment (Cherry et al., Plott & Uhl, ...)
  - employs standard subject pool of students, an abstract framing, an imposed set of rules (e.g., trading rules)
- field experiment
- social experiment
  - deliberate government policy with treatments
- natural experiment
  - serendipity („phenomenon of finding valuable things not sought for") observed
- thought experiment
  - „all talk and no action"

Defining field experiments

- „Used attributively to denote an investigation, study, etc., carried out in the natural environment of a given material, language, animal, etc., and not in the laboratory, study or office." (OED 2nd edition, definition for „field")
- Criteria to differentiate the word
  - Subject pool (e.g., non-students)
  - Information that the subjects have (e.g., rules of thumb from the domain we are interested in)
  - Commodity (e.g., non-abstract goods or services)
  - Task (e.g., List's article)
  - Stakes (e.g., more than 2 – 3 times minimum wage)
  - Environment (e.g., not the lab)

Taxonomy of field experiments

- artefactual field experiment
  - a conventional lab experiment with a non-standard subject pool
- framed field experiment
  - adds field context in either the information set, commodity, task, or stakes
- natural field experiment
  - adds an environment in which the subjects naturally undertake these tasks, such that the subjects do not know that they are in an experiment.

Pros and cons of lab experiments

Students as subject pools:

- similar to rat labs that are bred to be comparable in terms of „unobservables“; also relatively smart crowd that climbs up the learning curve fast; convenient to use because „at your fingertips“
  problems: might get too much homogeneity and a very selected crowd („professional“ subjects)

- are likely to be clean slates when it comes to certain tasks and task performances; hence are likely to bring fewer priors to the lab
  problem: they might not have the rules of thumb that we are interested in studying (e.g., Dyer & Kagel, Management Science,1996)

Abstract commodities:

- are not confounded by uncontrollable affections
  problem: that's something we might want to study

Tasks:

- are controllable and tend to be simplistic
  problem: we might want to study how people fare with less simplistic tasks (e.g., how do they fare with complicated investment decisions – what kind of model of the situation do they have! -- and how they try to influence the environment in this context)
  another problem: we might still not be able to get rid of „field heuristics“ from The Game of Life. I.e., subjects might (try to) „pattern recognize“ the experimental task and bring in solution strategies for seemingly related situations (e.g., social dilemma situations)
  yet another problem: subjects might construe („represent“) the task quite different from the experimenter. E.g., -- encouraged by the typical opening language of instructions – they may conceptualize the game as „Us against the Experimenter“. The experimental situation is – after all – itself a social situation.

Stakes:

- we can usually motivate students reasonably well with 2 – 3 times minimum wage; studying the impact of increased financial incentives (e.g., increased risk aversion) is still relatively inexpensive
  problem: it still might be quite expensive to run certain experiments

Environment:

- is controllable and tends to be simplistic
  problem: we might want to study how people fare in less simplistic environments (e.g., how do they fare in investment environments that are highly uncertain – what kind of model of the environment to they construct! – and how do they try to influence the environment in this context)

Excursion: social experiments

- deliberate government policy with treatments
  - need to randomize properly
  - need to overcome ethical concerns and willingness to implement
  - Deception? Experimenter effects?

Excursion: natural experiments (e.g., twin studies, or game shows)

- serendipity („phenomenon of finding valuable things not sought for") observed
  - Not there when you want them!
  - Exogeneity provided by others (e.g., monozygotic twins)
  - Typically no influence on design and implementation ..
    - but ... controlled replications, large stakes, well-known rules, many years of data, interesting cross-country and time-series variations

Ortmann & Gigerenzer's „Reasoning in Economics and Psychology: Why Social Context Matters" (JITE 1997)

Two models of reasoning in psychology:

- people claimed to make decisions according to the laws of logic and probability theory (e.g., Wason selection task).
- people found (by the school of Kahneman and Tversky 1996) to be quite deficient in following the laws of logic and probability theory. Example:

Figure 1

Letters-and-numbers rule:

If there is an "A" on one side (p), then there is a "2" on the other side (q).

Each of the following cards has a letter on one side and a number on the other. Indicate only the card(s) you definitely need to turn over to see if the rule has been violated.

Can people identify whether a rule has been violated?

What would they have to do? What do people actually do?

- select card with A on one side, or 7 on the other
- overwhelming number of people select the *p-card* or the *p-card* and the *q-card*, with only about 10 % of the people taking the normatively correct *p- and not-q card*

Are there ways around this poor performance?

Figure 4

Day-off rule:

If an employee works on the weekend, then that person gets a day off during the week.

The cards below have information about four employees. Each card represents one person. One side of the card tells whether the person worked on the weekend, and the other side tells whether the person got a day off during the week. Indicate only the card(s) you definitely need to turn over to see if the rule has been violated.

- Putting things in thematic garb seems to help: Gigerenzer and Hug (1992) report that performance goes up to 70 – 90 %.

- Is it the social contracts/social situation that facilitates logical reasoning? (No, say Gigerenzer & Hug 1992)
- Is it pragmatic goals such as cheater detection, and implicit cost-benefit analyses, that help us solve problems of this kind via evolutionary pressures? (Yes, say Gigerenzer & Hug 1992, drawing on Cosmides 1989 who had argued that selective cooperation demands the ability to detect cheaters. Indeed, Gigerenzer & Hug, 1992, tease apart the relative importance of cheating detection ability and social contract. They argue that about half of the content effect is really a context effect in that being party to a social contract affects reasoning. See the detailled discussion on pages 706-7. Their results, however, have been disputed: Evans & Chang, 1998, Cheater Detection and Altruistic Behaviour: An Experimental and Methodological Exploration, Managerial and Decision Economics 19, 467 – 480.)
- Choices in the selection task are a systematic function of perspective (e.g., employer versus employee) and guided by the goal of detecting cases „benefit taken“ and „cost not paid“. Reasoning is functional rather than logical: Reasoning consistent with propositional logic is entailed by some perspectives (e.g.. the employee's) but not by others (e.g., the employer's). But again, see the already mentioned Evans & Chang study.

The two models of reasoning in economics:

- people claimed to make decisions according to the laws of logic, probability (Bayesian updating), dominance etc.
- people found to be quite deficient in following the laws of logic and Bayesian updating and even in the use of dominance arguments.

Example: The Principal Agent Game

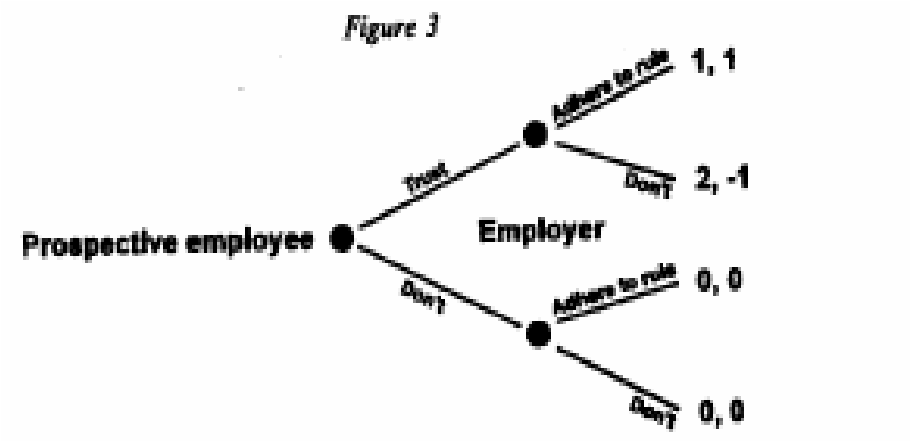
- an intensively researched problem because of the fundamental role it plays in the literature on gift-exchange, trust & reciprocity (e.g., the the sequential prisoner's dilemma game in List 2006 below), and as a building block in the analysis of of markets involving adjustable quality, the internal organization of for-profits and nonprofits, etc. (references in article)

- Here is a simple example of such a game in both the strategic and the extensive form:

Figure 2

		Prospective employee: Trust	Prospective employee: Don't
Employer	Adhere to rule	1, 1	0, 0
	Don't	2, −1	0, 0

Figure 3



- Note that this is the day-off problem in a particular parameterizationan and perspective: the employer faces a moral-hazard problem. (We could also habe constructed a moral-hazard problem in which the employee would have faced temptation (e.g., shirking on effort).)

  - Assume that you are the employee: Will you trust the employer? -Obviously, your answer depends on whether you interact only one time, or – more likely in organizational contexts of the kind described here – repeatedly (and without a clear-cut endpoint). For the one-shot interaction it all depends on whether people get the backward induction right, i.e. whether they (can) reason in dominance relations.
  - As it is, the behavior of subjects does often not conform to the predictions of the theory for one-shot, of finitely repeated, games. The reasons for that are disputed. Is it altruism? Is it trust? Is reciprocity? Can we preclude the possibility that subjects model the situation as if there is some possibility of future interaction? [„'phantom future' explanation"] Do people bring rules-of-thumb into the lab from The Game of Life (which tend to be indefinitely repeated and therefore is likely to have „nicer" equilibria?) Can we preclude the possibility that people do not understand the

situation fully? List's article on the behavioralist meeting the market speaks to some of these issues.

- Some related empirical evidence: In the TV show *Friend Or Foe?* subjects play as teams of 2 to earn money through answering questions. Then they have to decide privately to play Friend or to play Foe with a stake X. The payoff function is

	Friend	Foe
Friend	X/2,X/2	0,X
Foe	X,0	0,0

Clearly, it is a a weakly dominant strategy to play Foe.
In the show, roughly 50% play Foe.
In the lab, roughy 10% play Foe.
What gives? (We'll have more to say more about this in a second.)

In general, context effects remain an understudied area of research in economics.

- Levitt & List, What Do Laboratory Experiments [Measuring Social Preferences Reveal] (Tell Us) About the Real World?
(27 June 2006 version, 16 pages of references; the JEP version that you have read is much less ambitious: „An earlier version of this paper that discussed the generalizability of a much wider class of experiments circulated under ... „ (p. 171))

„The allure of the laboratory experimental method in economics is that, in principle, it provides *ceteris paribus* observations of motivated individual economic agents, which are otherwise exceptionally difficult to obtain using conventional econometric techniques. Lab experiments provide the investigator with a means to directly influence the set of prices, budget constraints, information sets, and actions available to actors, and thus measure the impact of these factors on behavior within the context of the laboratory.“ (1)

„A critical [cut: maintained] assumption underlying (added: the interpretation of data from many) laboratory experiments is that insights gained in the lab can be extrapolated to the world beyond, a principle we denote as generalizability.“ (pp. 1-2 old version, p. 153 JEP version) [often used synonyms: -> parallelism, -> external validity, -> ecological validity; see fn 2; see also the list of keywords] „For physical laws and processes like gravity, photosynthesis, and mitosis, the evidence supports the idea that what happens in the lab is equally valid in the broader world. ... The basic strategy underlying laboratory experiments in the physical sciences and economics is similar, but the fact that humans are the object of study in the latter raises special questions about the ability to extrapolate experimental findings beyond the lab, questions that do not arise in the physical sciences.“ (pp. 153/4)

„We argue (added: , based on decades of research in psychology and recent findings in experimental economics,) that [cut: human decisions are] (added: behavior in the lab is) influenced not just by [cut: simple] monetary calculations, but also by at least [cut: three] (added: five) other [cut: considerations] (added: factors):

1) (added: the presence of moral and ethical considerations;)
2) The nature *and* extent [cut: to which one’s actions will be scrutinized by others] (added: of scrutiny of one’s actions by others;)
3) The [cut: particular] context and process in which [cut: a] (added: the) decision is embedded;
4) self-selection of the individuals making the decisions; and
5) the stakes of the game.“ (p. 3 old version, p. 154 JEP version)

New: Summary of Experimental Games Used to Measure Social Preferences – See Table 1 on p. 155.

Their model

$U_i(a,v,n,s) = M_i(a,v,n,s) + W_i(a,v)$
Where the first and second RHS term denote the utility from moral and wealth arguments, respectively, a ε (0,1), denotes an action, v what is at stake („The higher the stakes or monetary value of the game, which we denotev, the greater is the decision's inpact on W_i"), n norms, and s the moral costs of being watched („For instance, in a dictator game, keeping a greater share for oneself increases an individual's wealth, but doing so may cause the agent moral disutility."). Note that actions and what's at stake both affect the utility from moral and wealth arguments.

„ ... predictions. First, when the wealth maximizing action has a moral cost associated with it, the agent will (weakly) deviate from that action towards one that imposes a lower moral cost. Second, the greater is the social norm against the wealth maximizing choice or the degree of scrutiny, the larger the deviation from that choice. In both cases, we envision the agent trading off morality and wealth until an equilibrium is reached. Third, to the extent that individuals follow different moral codes (that is, $M_i \neq M_j$ for individuals i and j), they will generally make different choices when faced with the same decision problem. Fourth, in situations where there is no moral component (e.g., the choice between investing in a stock or bond index), the model reverts back to a standard wealth maximization problem. Imposing a number of further restrictions on the utility function beyond those noted above yields crisper insights. Typically, we expect that as the stakes of the game rise, wealth concerns will increase in importance relative to issues of fairness [niceness is price-sensitive, AO] ... While it remains an open empirical issue, we would also expect that the cross-partial derivatives between v,n, and s in the morality function are potentially important. ... It is also worth recognizing that the relevant social norms and the amount of scrutiny are not necessarily exogenously determined, but rather, subject to influence by those who will be affected by the choices an agent makes. ... [panhandlers][churches]„ (pp. 8 - 9)

„Implications of the model for experiments designed to measure physical constants such as social preferences" (pp. 9 – 29) [Physical constants?]

- Tipping incurs a cost but also benefit. „This is true if one is eating alone, but probably even more so when there is a higher degree of scrutiny (e.g., you are with your clients, on a first date, or when another diner is closely observing your actions)." (p. 9)
- Key question is to what extent lab and real-world setting may differ along the key dimensions of the model (v,n,s, and the type of individuals engaged in the activity.) „Our primary interest, however, is developing the implications of the model for the generalizability of lab experiments to naturally occuring contexts. ... If the lab diverges from the environment of interest [real world?] on certain dimensions, the model provides a framework for predicting in what direction behavior in the lab will deviate from that outside of the lab." (p. 10)
- This concern applies to all experimental results but „is likely to be greatest for those games in which there is the potential for a strong moral component to behavior and for those experiments that use the lab for quantitative purposes, such as to measure deep structural parameters (or „physical constants"). One of the most influential areas

of research in experimental economics in recent years precisely fits this bill: games that provide insights into social preferences. ... Findings from gift exchange games, for example, have been interpreted as providing strong evidence that many agents behave in a reciprocal manner even when the behavior is costly and yields netiher present nor future material rewards. Further, the social preference results have bene widely applied outside the laboratory, based on the assumption that the experimental findings are equally prescriptive of the world at large ... [references].“ (pp. 10 – 11)

LL then review the empirical evidence regarding each of the possible complications highlighted in the model to extrapolating the experimental findings outside the lab

- Scrutiny (pp. 11 – 18)
  Lab effects and Non-anonymity effects. Pages 13 – 14 provides summary of List (2006), here listed as List (2005a). Curiously, the study by Cherry et al. (2002) is not mentioned.
  [In the JEP version:
  - Scrutiny That Is Unparallled in the Field (Orne 1962; Schultz 1969; Pierce 1908; List 2006; etc.)
  - Anonymity in the Lab and the Field (Hoffman et al. 1994; 1996; Davis & Holt 1993; Andreoni & Bernheim 2006; List et al. 2004; Haley & Fessler 2005 – pair-of-eyes study; Allen 1965 – reduced privacy in conformity study; Harris & Menger 1989 – more-hand-washing when observed study)
  - the study by Cherry et al. (2002) still not mentioned]
- Context matters (pp. 18 – 26 -> Ortmann & Gigerenzer 1997)
  „As various authors have pointed out (), however, activities in the spirit of public good provision, dictator, ultimatum game, trust, gift exchange games, and other social dilemmas, are typically not one-time encounters in the outside world, but rather repeated games. O&H (2000) go further and speculate that, there is a good chance that many of the experimental results for the classes of games discussed here [social dilemmas] are artifacts of experimental design.“ (p. 22)
  [In the JEP version:
  - Context Matters and Is Not Completely Controlled By the Experimenter; we’ll take about a couple of other examples in L8; „While researchers might hope that experimental subjects will make clear strategic adjustments from repeated contexts to one-shot games, the empirical evidence is mixed. ... “ (p. 163)]
- Stakes (pp. 25 – 26) – poor discussion, also in the JEP version – the Parco, Rapoport and Stein (2002) study is explicity motivated by Hertwig & Ortmann (2001)
- Self-selection into the subject pool (pp. 26 – 29) – poor discussion, also in the Jep version – there is simply no clear-cut sampling strategy defined. For a better approach, see Hertwig & Ortmann (2001)

- Artificial Restrictions on Choice Sets and Time Horizons (new section)
  „In naturally occuring environments, the choice set often is almost limitless and institutions arise endogenously.“ (p. 166)
  ... Gneezy & List 2006 library study

„Implications of our model for experiments in which the moral and wealth-maximizing actions are not competing objectives.“ (pp. 29 – 40)

I.e. experiments exploring general economic theory, Bayesian updating, risk and uncertainty, psychological phenomena such as loss aversion, hyperbolioc discounting, impersonal auctions, market experiments where the demand and cost functions are unknown.

Potential distortions that remain are the impact of stakes and non-random selection of participants into the experiment. Four extensions of model that highlight additional issues:

(1) experience (pp. 33 – 36)
(2) endogenous emergence of real-world institutions that try „to exploit behavioral effects of naive players“ (pp. 36 – 38)
(3) more learning opportunities in real-world situations (pp. 38 – 39)
(4) group differences (pp. 39 – 40)

Plus, stakes and cognitive costs (pp. 30 – 31), non-random selection of participants (pp. 31 – 32)

„The Virtues of Lab Experiments“ (pp. 40 – 42)

- Design and implement experiments so as to minimize biases discussed in Levitt & List.
- Focus on qualitative rather than quantitative insights.
- A combination of lab and field experiments „is likley to provide more insight than either in isolation.“

- List, The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects (JPE 114.1, 1 - 37)

What is the basic research issue? Whether social preferences, as documented in the „gift-exchange“ literature, are robust to transfer from the field to the lab. (See fn 1, p. 2: „I explore social preferences under this broad definition and am not interested in pinpointing whether the behavior consistent with social preferences is due to altruism, reciprocity, fairness, inequality aversion, or another motive. Yet within the gift exchange literature, reciprocity motives have been highlighted; thus I shall continue this spirit in the discussion below.“

What is the typical laboratory gift-exchange scenario?
And what is a typical parameterization? (pp. 7 – 9)

- Buyer (Employer!) determines two integer values p and q
  (the first offered, the second requested)
- Seller (Employee!) decides whether to provide the requested quality (effort!)
  - the cost of quality is increasing monotonically with prodcut quality
  - product quality choice is revealed to buyer partner (and the monitor!)

Individual p and q choices combine to determine monetary payoffs for the pair according to the following payoff functions:

$$\text{seller payoff: } \Pi_s = p - c(q),$$

$$\text{buyer payoff: } \Pi_b = (v - p)q, v = \$80, p \in [\$5, \$80], q \in [0.1, 1]. \quad (1)$$

These parameter values yield a standard prediction under the assumption of common knowledge, self-interest theory, and appropriate backward induction. Since product quality is costly, sellers will choose the minimum level ($q_{\min} = 0.1$). A buyer's best response is to choose $p_{\min}$, which is $p = \$5$. Thus the subgame perfect equilibrium outcome is $q^* = 0.1$ and $p^* = \$5$, with associated profits of $\Pi_s = \$5$ and $\Pi_b = \$7.50$, much less than more efficient profit levels (i.e., $p = \$30$ and $q = 0.5$ yields $\Pi_s = \$24$ and $\Pi_b = \$25$).

Previous experimental efforts have found that typically $q > q^*$ and $p > p^*$ and that $\partial q/\partial p > 0$ in a reduced-form regression model, leading

What is the purpose of the Lab-R treatment and what happens as List goes from Lab-R to Lab-RF? That's the key calibration step!

In column 2 of table 1, Treatment Lab-RF (RF denotes replication with field values) simply manipulates the environment in Treatment Lab-R by setting[7]

$$\text{seller payoff: } \Pi_s = p - c(q),$$

$$\text{buyer payoff: } \Pi_b = v(q) - p,\ p \in [\$5, \$80],\ q \in [1, 5]. \qquad (2)$$

For $c(q)$ values, I use $c(q) = \$4, \$5, \$8, \$15$, and $\$50$ for $q = 1, 2, 3, 4$, and 5; for $v(q)$ values, I use $\$6, \$8, \$15, \$30$, and $\$80$ for $q = 1, 2, 3, 4$, and 5 (PSA 6, 7, 8, 9, and 10).[8] While these chosen values are admittedly only a rough estimate of the gains to trade available in this market, use of these parameters provides the necessary tension between the dominant strategy and the joint-profit maximization actions. Under this design, the Nash purely selfish prediction is $p^* = \$5$; and for sellers to send minimal card quality, $q^* = 1$. These actions result in $\Pi_s = \$1$ and $\Pi_b = \$1$. The efficient quality level is $q = 5$, which ensures a joint surplus of $30. Note that there could be losses of up to $74 (the buyer sends $80 and receives the lowest-quality Frank Thomas card); as in the other induced value laboratory treatments herein—Treatments Lab and Lab-Context—after these treatments were carried out, I had subjects participate in other unrelated experiments that did not involve interaction to ensure that they would leave with positive cash balances.

What is the purpose of the Lab-RF1 treatment? (To make sure that one-shot game lead to the same outcome as the five rounds from which one is randomly paid off.)

In Harrison's and List's terminology, what kind of experiments are these? (Artefactual? Framed? Natural?) What is a Leaf Frank Thomas card anyway? And why does it have any value whatsoever?

What does Lab-Context add? „Context that resembles the subjects' naturally occuring environment"

In row 2 of table 1, Treatment Lab-Context adds context to Treatment Lab-RF1. In this case, rather than buyers and sellers transacting with abstract commodities, Treatment Lab-Context adds context that resembles the subjects' naturally occurring environment. For example, buyers make an offer to a seller to buy one 1990 Leaf Frank Thomas card, and the buyer requests a certain PSA grade. As in Treatment Lab-RF1, sellers have five PSA grades available (PSA 6, 7, 8, 9, and 10) and subsequently choose the quality of the Thomas baseball card to give the buyer if they accept the buyer's offer.[9] Treatment Lab-Context includes 32 buyers and 32 sellers.

[9] PSA grades 6–10 were chosen because little trading of Thomas cards below PSA 6 is carried out in the actual market. Note that in this treatment I am not actually having agents transact with real commodities; rather subjects are told to act as though they are using graded Thomas cards.

And what do Lab-Market ($20) and Lab-Market ($65) add?
And how exactly were they set up? Compute how much List spend on these treatments alone! (pp. 10 – 11)

> Completing the laboratory treatments is Treatment Lab-Market, more specifically, Treatment Lab-Market($20) and Treatment Lab-Market ($65). Treatment Lab-Market is the laboratory market parallel to Treatment Lab-Context: buying agents approach dealers in the experimental market to purchase 1990 Leaf Thomas baseball cards in face-to-face transactions. Each participant's experience in Treatment Lab-Market followed four steps: (1) consideration of the invitation to participate in an experiment, (2) a session to learn the market rules, (3) actual market participation, and (4) conclusion of the experiment and exit interview.

Step 1: Potential consumer subjects were (that approached the monitor's dealer table) were asked ... Potential dealer subjects were identified by List as fulfilling certain conditions (whether the dealer had a fair number of ungraded 1990 LT cards) ...
Step 2: Market rules were explained privately to buyers and sellers
(What's the potential problem with that? See Ortmann 2005)

Step 3: Buying agents were instructed to approach dealers and buy (first round grade 9 cards LFT cards, second round grade 10 LFT cards), no haggling.

List spend 30 ($20 + $65) = $2,550 on these „two" treatments.

Step 4: Exit interviews, professional grading of cards.

What do the two Treatment Floor Markets add to the two Treatment lab markets? (pp. 12 – 13) Explain exactly the differences between these two sets of treatments?

Again four steps.

Step 1: Potential consumer subjects were (that approached the monitor's dealer table) were asked ... Potential dealer subjects were identified by List as fulfilling certain conditions (whether the dealer had a fair number of ungraded 1990 LFT cards) BUT ... (they were not told they would be in the experiment)

Step 2: Market rules were explained privately to buyers, no haggling. „Throughout the experiment the sports card dealers were not aware that an experiment was occurring." (12)

Step 3: Buying agents were instructed to approach dealers and buy (first round grade 9 cards LFT cards, second round grade 10 LFT cards).

Step 4: Exit interviews, professional grading of cards.

[Design issues discussed on page 13?]

What are the final three treatments and why did List do them? (pp. 13 – 14)

He tried to tease out whether his dealers had reputation effects!

So, what's the overall design? (Discuss Table 1!)

TABLE 1
EXPERIMENTAL DESIGN

	(1)	(2)	(3)
Treatment Lab	Treatment Lab-R: Replicate lab studies $N = 25$	Treatment Lab-RF: Extend to field values $N = 25$	Treatment Lab-RF1: Extend to one-shot environment $N = 27$
Treatment Lab-Context	Treatment Lab-Context: Adds market context $N = 32$	Treatment Lab-Market($20): Adds market interaction $N = 30$	Treatment Lab-Market($65): Adds market interaction $N = 30$
Treatment Floor (Cards)	Treatment Floor-$20: Naturally occurring sports card market $N = 50$	Treatment Floor-$65: Naturally occurring sports card market $N = 50$	
Treatment Floor (Tickets)	Treatment Floor-NoGrading: Naturally occurring ticket market before grading was available $N = 60$	Treatment Floor-AnnounceGrading: Naturally occurring ticket market after grading announcement $N = 54$	Treatment Floor-Grading: Naturally occurring ticket market when grading service was available $N = 36$

NOTE.—Each cell represents one (or two, in the case of Treatment Floor [Tickets]) unique treatment. For example, Treatment Lab-R in row 1, col. 1, denotes that 25 dealer and 25 nondealer observations were gathered to replicate the laboratory gift exchange studies in the literature.

And what are the predictions of Self-Interest Model vs. Social Preference Model? (p. 16)

TABLE 2
PREDICTIONS: SELF-INTERESTED MODEL VS. SOCIAL PREFERENCE MODEL

	FLOOR TREATMENTS		
		Tickets	
LAB TREATMENTS	Sports Cards	Floor-NoGrading	Floor-Announce Grading and Floor-Grading
Self-Interested Model			
Without experimenter effects: no relationship between price and quality With experimenter effects: positive relationship between price and quality due to, e.g., reputation effects	Local dealers: positive relationship between price and quality due to, e.g., reputation effects Nonlocal dealers: no relationship between price and quality	Local dealers: no relationship between price and quality Nonlocal dealers: no relationship between price and quality	Local dealers: positive relationship between price and quality due to, e.g., reputation effects Nonlocal dealers: no relationship between price and quality
Social Preference Model			
Without experimenter effects: positive relationship between price and quality due to social preferences With experimenter effects: positive relationship between price and quality due to, e.g., reputation effects and social preferences	Local dealers: positive relationship between price and quality due to, e.g., reputation effects and social preferences Nonlocal dealers: positive relationship between price and quality due to social preferences	Local dealers: positive relationship between price and quality due to social preferences Nonlocal dealers: positive relationship between price and quality due to social preferences	Local dealers: positive relationship between price and quality due to, e.g., reputation effects and social preferences Nonlocal dealers: positive relationship between price and quality due to social preferences

NOTE.—Each column represents predictions of the self-interested model vs. the social preference model across the three major experimental types. In a split of the dealer types, a dealer is labeled as a nonlocal if he or she is unlikely to be concerned with reputation effects; e.g., if he or she rarely attends sports card shows in the area (fewer than three times in a typical year), does not plan to attend more frequently than this in the future, does not own a sports card shop, and does not have an Internet sports card business. All other dealers are labeled as locals.

Results? (p. 17)

TABLE 3
RESULTS SUMMARY

	$\bar{p}$ (1)	$\bar{q}$ (2)	q_r (3)
	Treatment Lab		
Treatment Lab-R	28.4 (16.1)	3.5 (2.0)	6.1 (2.1)
Treatment Lab-RF	22.6 (20.7)	2.3 (1.4)	4.1 (.9)
Treatment Lab-RF1	24.8 (22.1)	2.5 (1.7)	4.0 (1.3)
	Treatment Lab-Context		
Treatment Lab-Context	19.5 (19.6)	2.3 (1.5)	4.2 (1.1)
Treatment Lab-Market($20)	$20	3.1 (.9)	4
Treatment Lab-Market($65)	$65	4.1 (.6)	5
	Treatment Floor (Cards)		
Treatment Floor-$20	$20	2.1 (.9)	4
Treatment Floor-$65	$65	3.2 (1.0)	5
	Treatment Floor (Tickets): Treatment I		
Treatment Floor-NoGrading	$10	2.7 (.6)	4
Treatment Floor-AnnounceGrading	$10	2.9 (.6)	4
Treatment Floor-Grading	$10	3.1 (.8)	4
	Treatment Floor (Tickets): Treatment II		
Treatment Floor-NoGrading	$30	2.7 (.7)	5
Treatment Floor-AnnounceGrading	$30	3.4 (.8)	5
Treatment Floor-Grading	$30	3.6 (1.1)	5

NOTE.—Summary statistics from one (or two in the case of Treatment Floor [Tickets]) unique treatment. $\bar{p}$ is the average price, $\bar{q}$ is the average quality, and q_r is the average requested quality. Treatment I-R data are scaled to range from 1 to 10, and PSA 6, 7, 8, 9, and 10 are denoted as quality levels 1, 2, 3, 4, and 5 in the table. Standard deviations are in parentheses.

Note that while the data suggest a lot of gift exchange as suggested by List's definition of gift exchange, there is a lot of cheating going on leading very likely to highly asymmetric payoffs. (We don't know for sure, and would have to compute based on average price, quality, and required quality what the earnings were.) Andrea suggested that the word gift exchange seems not particularly appropriate; the market is one of lesser cheaters. Point well taken.

RESULT 1. Behavior of sports card enthusiasts in laboratory games is in line with the gift exchange literature using student subjects, and the results extend well to one-shot environments.

RESULT 2. Adding natural context influences behavior, but gift exchange remains alive and well.

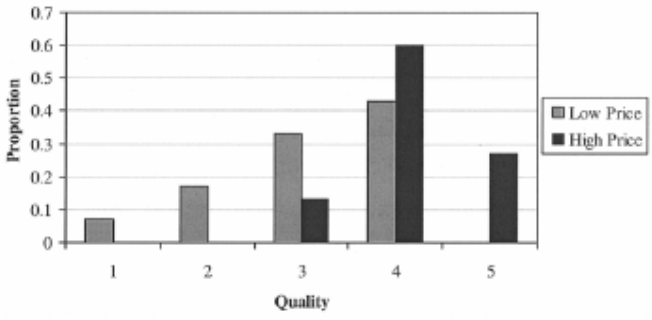


FIG. 1.—Treatment Lab-Market

Results 1 and 2 „a nice validity check on the extant gift exchange literature.“ (p. 21) Explain! (See Figure 1.) Why does this evidence not show unequivocally that subjects exhibit social preferences? What is an alternative explanation? How does List relax „the controls of the experiment?“ (p. 22) What does he find? (See Figure 2)

RESULT 3. When third-party verification is available, behavior in naturally occurring transactions is consonant with both gift exchange and a concern for reputation.

RESULT 4. When third-party verification is possible, local dealer behavior in naturally occurring transactions is consonant with both gift exchange and a concern for reputation, whereas nonlocal dealers' behavior is in line with self-interest theory.

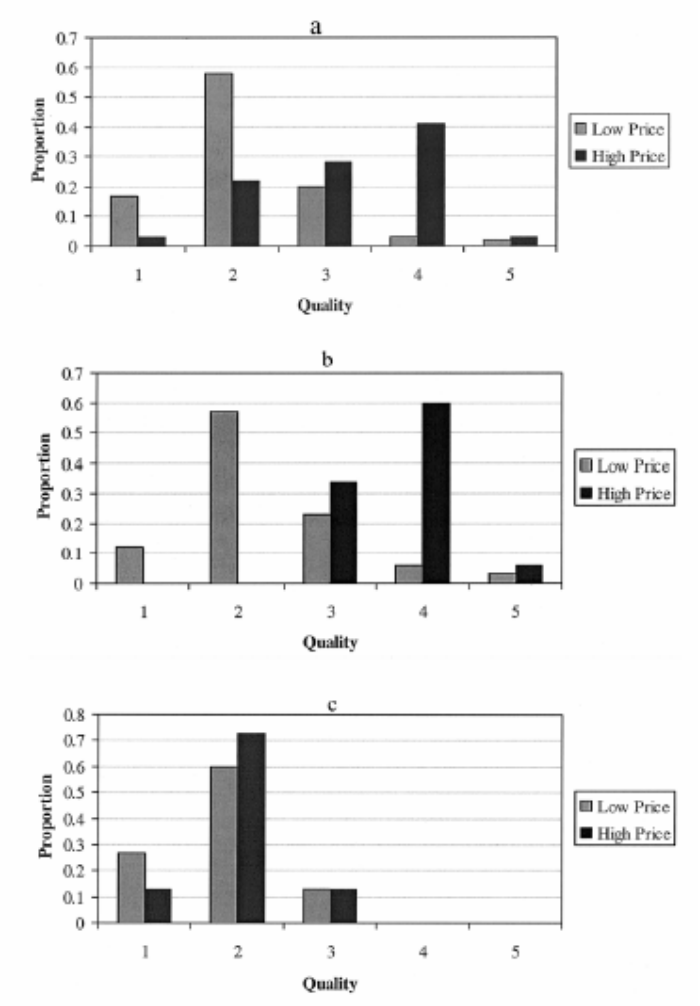


FIG. 2.—Treatment Floor: *a*, sports cards; *b*, local dealers; *c*, nonlocal dealers

It looks as if „some of the dealers in the sample may have had an economic reason to uphold their reputations, whereas others may not have had similar incentives.“ (p. 24, see Figure 2, panels b,c; see also Figures 3 and 4, and Table 5)

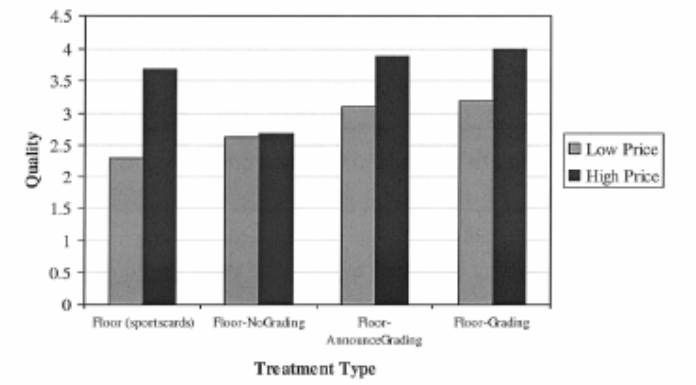


FIG. 3.—Price/quality relationship for local dealers

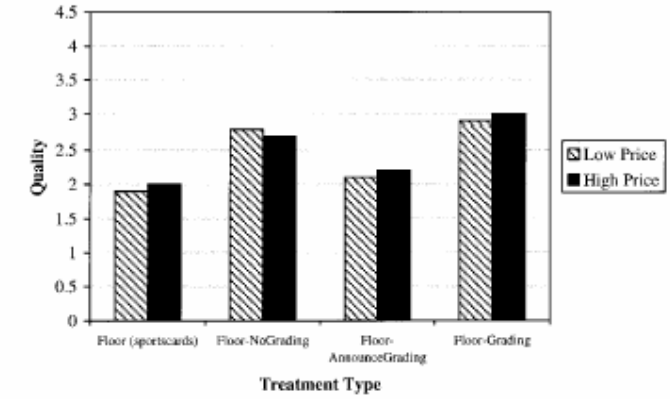


FIG. 4.—Price/quality relationship for nonlocal dealers

That leaves the „natural question“ whether local dealer behavior is driven primarily by reputation effects or social preferences.“ (27) How does List try to address this question? (See Result 5 and the Treatment Floor-NoGrading results in figures 3 and 4).

RESULT 5. When third-party verification is not available, supply-side behavior in naturally occurring transactions is consonant with purely selfish money-maximizing theory, suggesting that reputational considerations, rather than social preferences, are driving the earlier results.

What does „mendacious“ mean? See www.m-w.com: „given to or characterized by deception or falsehood or divergence from absolute truth“

Whose mendacious claims does List talk about? And what does he find? (pp. 28 – 32, Results 6 and 7)

RESULT 6. When third-party verification is possible, local dealers provide fewer claims of quality than nonlocal dealers in the field and, conditional on claiming quality, shirk less frequently.

As in the spirit of the inquiry into result 4, one can question whether the increased quality promises and deliveries from local dealers are due purely to reputational concerns or have an element of social preferences. Examining data collected in Treatment Lab-Market and Treatment Floor (Ticket Stubs) lends insights into this issue and leads to the final result.

RESULT 7. In the laboratory, or when third-party verification is not possible, local and nonlocal dealers make similar claims of quality, and conditional on claiming quality, shirk to the same extent.

„Control is the hallmark of good experimental practice, whether it be undertaken by economists or psychologists.“ (Harrison & Rutstroem 2001, p. 413)
So is replicability (see H&O 2001 on that issue on, for example, p. 386: „This practice [of providing scripts] is also likely to enhance the replicability of experimental results.“)

Enumerate all those instances where the reader can not verify/falsify details of the implementation of List's study, or where it is possible that experimenter/demands effects (e.g., Ortmann 2005 – See Levitt & List 2006) have affected his results.

- his evasive identification of the locations (Necessary?)
- his apparently unscripted identification of sellers
- his instructions to buyers and sellers (Was there a script? Could it be that we see a lot of HansTheHorse effects here, see Ortmann 2005 for details)

Other issues and observations:

Jesse: How exactly were the buying agents paid? If they were paid ex ante, why did they not take the money and ran? Doesn't there behavior contradict the basic message of List's article? Were they reputationally constrained?

Andrea: The result had to be expected because these kind of markets are really markets of sharks. [In other words, this market is the result of some form of selection bias.]

???: Did not allowing to haggle, bring out the worst in sellers?

???: What kind of information flows were there. Can we believe that these guys did not talk about the experiment in progress? (Very unlikely!)