Laboratory Experiments: Professionals versus Students*

Guillaume R. Fréchette
New York University

1 Introduction

Most economic theories make general claims. That is, economic models rarely start with assumptions such as: the firms are picnic tables producers in Korea... Hence, whenever one engages in applied work testing a model, there will be issues of generalizability of the results to other data sets, environments, subject population, etc. Such concerns extend to experimental data sets since, in particular, subjects tend to be undergraduate students while much of economic activity outside the laboratory is done by professionals. This paper establishes what is known about the impact of the specific subject sample typically used in experimental economics by reviewing prior studies that have used both the standard subject pool as well as an unusual pool of professionals. Hence, two questions immediately come to mind: 1) why make comparisons across subject pools? and 2) what is the usual subject pool and what are professionals?

Which are some of the reasons why a researcher might want to study a non-standard subject pool?

1. Substantive questions: some fact or phenomenon about a particular group is observed outside of the laboratory, and the lab is used to identify the cause(s) of that phenomenon. An example of this is the work of Niederle and coauthors looking at gender and

*I am grateful to Chloe Tergiman for re-creating the tables and Emanuel Vespa for helping with research. I also wish to thank David Cooper for comments and all those who have sent me papers that I had missed in earlier drafts. I gratefully acknowledge the support from of NSF via grants SES-0519045, SES-0721111, and SES-0924780 as well as support from the Center for Experimental Social Science (CESS), the C.V. Star Center.

how it interacts with competition and other factors relevant for salaries and careers of females versus males (Gneezy, Niederle, and Rustichini 2003).

2. Comparison of parameter estimates across subject pools. For example: Are professional investors more risk averse than the usual experimental subjects (Zhang 2008)? How does risk aversion differs in the standard subject pool and in a representative sample of the population (Andersen, et al 2009)?

3. Would you reach a similar conclusion about behavior in a game or about the predictions of a theory if you used the usual experimental subjects versus if you used subjects who are career professionals at the task the game being tested is supposed to represent? An example of this is the work of Kagel and coauthors looking at how professionals from the construction industry bid in common value auctions (as opposed to the usual experimental subjects). In particular do professionals also fall prey to the winner's curse (Dyer, Kagel, and Levin 1989)?

4. Using animals to do experiments that cannot be done with human subjects. An example of this is the work of Battalio, Kagel, and coauthors using pigeons and rats (Kagel, Battalio, and Green 1995).

5. There are other cases that do not fit neatly in the above categories, here are two examples. Comparing children to adults, see the work of Harbaugh and coauthors (Harbaugh, Krause, and Berry 2001). Difference in behavior across subgroups in an experiment where this is observed ex-post but is not a purpose of the original study: for example Casari, Ham, and Kagel observed differences in bidding behavior between males and females in an auction experiment (Casari, Ham, and Kagel 2007).

Readers interested in issues of subject pool differences in general are referred to Ball and Cech (1996), who have an excellent survey of these issues. My paper is concerned with the third of these questions, would one reach similar conclusions using professionals as opposed to the standard subject pool.[1] However, the papers included in the review might have been motivated by questions of the type 2 highlighted above.

[1] Ball and Cech (1996) also review studies comparing students and professionals. However, my review excludes some of the papers they include because they do not fit the guidelines I impose, but more importantly I also cover papers they did not.

The two categories of subjects, students and professionals, are rather vague concepts. First, who are the typical subjects? The typical subject pool for economic experiments is undergraduate students. In many cases mostly economics and business undergraduate students (in some cases also graduate students). This is, mostly, a result of the recruitment technology, which used to involve going to large undergraduate classes. Nowadays, with email recruiting, targeting a larger population is feasible, however it is still the case that the subject pool consists almost exclusively of undergraduate students. At NYU, for instance, we recruit from all fields, but only undergraduate students. Other labs (such as Pittsburgh) allow non-students, but that remains a minority of the subject pool. Recruiting from undergraduate students means that most experiments are composed of a non-representative sample of the population at large in a few dimensions: gender, race, education, (family) wealth, age, etc. Professionals are loosely defined as people working in an industry where the game under study is thought to be relevant. This definition leaves room for interpretation. For instance, one paper which I hesitated to include is the one by Cadsby and Maynes (1998a) where the game is a public good provision game and the professionals are nurses. Although it is sensible to think that nurses sometimes face situations that have a public good provision structure, it is unclear what makes nurses professionals at public good provision; they are paid for the work they provide and their wage might internalize all the positive externalities of their services.

The papers included in the review had to meet the following criteria. They had to include both a sample of typical experimental subjects[2] and a sample of subjects who are professionals at the task with which the experiment is involved.[3] They also needed to follow what I will loosely define as standard laboratory procedures. That is, subjects are

[2]Studies comparing professionals to students were discarded if their student group was mostly composed of graduate students. Anderson and Sunder (1995) used MBA students. Hong and Plott (1982) used graduate students in engineering, business, and law.

[3]Riker (1979) was excluded even though it has two groups, one with more experience than the other, because it is unclear who composes that second group. All that is known is that they are male evening students in the University of Rochester graduate "college of business ranging in age from 25 to 55 and in occupation from management trainee and mailman to department head at the largest local industrial plant." (p. 58)

Banks, Camerer, and Porter (1994) studies refinements in signaling games and compares students to members of the technical staff at the Jet Propulsion Laboratory. The employees of the Jet Propulsion Laboratory must have a high level of technical and mathematical expertise, but since those attributes do not seem to make them professionals in signaling, this study was not included either.

recruited to participate in an experiment and thus they know that they are participating in an experiment.[4] Furthermore, subjects are given instructions written using neutral language.[5] Finally, the papers included do not merely compare these two groups, they do so in an environment aimed at testing certain theoretical predictions.[6] Although such narrow focus excludes some interesting studies, it allows to focus on a very precise question: do professionals behave differently than the standard experimental subject pool in a typical experimental setting as it pertains to evaluating a certain theory? If they do, it suggests that experimental results are not robust, and to the extent that we care about the behavior of professionals more than that of other groups, this could be a serious problem.[7] If they don't, it suggests that the subject pool per se is not a threat to the external validity of the standard experiment. This is very different, however, from saying that ecological validity is not a concern. These issues will be clarified after concepts relating to the validity of knowledge claims are defined more clearly.

The papers are loosely organized in four thematic groups: other regarding preferences, market experiments, information signals, and a miscellaneous group. The grouping is not important to the analysis, it is only meant to make the review easier to read. Within each section the papers are arranged chronologically. Whenever it was possible, a short description of the typical results for a certain type of experiments is given. Prior to reviewing

[4] Some studies were excluded on the basis that there were not enough details on experimental procedures or because the treatments with professionals had variations with respect to the treatments with students. These include Mestelman and Feeny (1988), Smith, Shuchaneck, and Williams (1988), and King, Smith, Williams, and Van Boening (1992).

The only variations in treatment between students and professionals that were allowed were those having to do with incentives and the number of rounds per sessions.

[5] However I have included a couple of experiments with incentives that are not typical by experimental economics standards (the Burns (1985) study as well as the DeJong, Forsythe, and Uecker (1988)). Those study are included mainly because they are some of the oldest. However, the unusual incentives should be kept in mind when considering the results.

The requirement of neutral language eliminated studies such as Alatas et al (2006).

[6] Based on this last criterion, Montmarquette et al (2004) was not included although it meets all of the other criterions.

[7] It is not always obvious that one cares only about the behavior of professionals. For one thing, there are situations for which there are no obvious professionals (e.g. voluntary contributions to charity). Second, sometimes it is not the only groups of interest (e.g. investment behavior by small individual investors). Third, there are situations that people face only once in their life, and thus they are by definition inexperienced when they play the relevant game (e.g. medical students going through the residency match).

the papers, I will define some terms and mention issues of interpretation of the results.

1.1 Some Methodological Concepts Defined

Since economists only very rarely discuss methodology, I will define a few concepts that will help clarify the relevance of the question asked in this paper.[8] Campbell (1957) first introduced the concepts of internal and external validity, which were later on expended to four aspects of the validity of an inference (Cook and Campbell, 1979): statistical conclusion validity, internal validity, construct validity, and external validity. To start, it is important to clarify that internal validity has made its way into the experimental economists lexicon with a different meaning. Most (experimental) economists use internal validity to mean the extent to which the environment that generated the data corresponds to the model being tested. This, as will become clear, is much closer to construct validity. For simplicity, in what follows, we will refer to x and y where a claim is made that x causes y.

Statistical conclusion validity has to do with whether the x and y covary, and how strong is that relation. Any factors that increase type I or type II error rates for instance have a negative impact on statistical conclusion validity. Misspecification in the estimation of an equation would be an example of a source of this problem.

Internal validity is about the correctness of the interpretation of variation in x *causing* variation in y, i.e. is the attribution of causality warranted. An example of a violation of internal validity would be a case where a within subject design is used, the order in which the treatment is applied is always the same, and an order effect is confused with a treatment effect.

Construct validity is the ability of the specifics of the experiment to represent the concepts they are intended to capture. For instance, an experiment by Kagel, Battalio and Green (1995) investigated the impact of wealth on time discounting in rats by having animals with relatively higher and relatively lower weight (weight was controlled by the amount of food the animals were given access to) participate in choices with different rewards and delays. Here the ability to interpret the results as saying that wealthier animals have high or lower discount factor than poorer ones depends on the validity of representing the construct of wealth by weight and access to food. The use of the term internal validity by economists often corresponds to construct validity as defined here. Experiments that test models are

[8]Most of the substance here is from Shadish, Cook, and Campbell (2002).

less subject to issues of construct validity insofar as the theorist is often the one choosing the constructs and defining the environment. The experimenter working with a model often simply operationalizes the constructs as defined in the model. That doesn't mean that there are no issues of construct validity as, for instance, the way the constructs are explained to the subjects matters for construct validity, and is an issue independent of whether a model is being tested or not. However, when a theorist writes a model where a firm is modeled as an agent and an experimenter has a subject take decisions to represent that agent, the issue of construct validity seems more acute for the theorist modeling a firm as a single agent, than for the experimenter using one subject to take the decision of that agent.

Finally, external validity is the ability of the causal relation from x to y to generalize over subjects and environments. External validity does not have to be about generalizing from the subjects or environment in the experiment to subjects and environment outside the laboratory. It can also be about variations in subjects and environments within the experiment (for instance, does the result apply to both men and women in the experiment?).

One concept that economists sometimes blend in with external validity is ecological validity. The two are very different however. Ecological validity is not of the same nature as the four concepts defined previously, it is rather a method where the experiment and its participants are made as similar as possible to what is found in the setting of interest. Hence, an experimenter asking a question relevant to stock trading can design a computer interface exactly like the one at the New York Stock Exchange and recruit professional traders from New York which would make for an experiment with high ecological validity. However, suppose that the causal relationship established between x and y in that experiment did not generalize to non-American traders or to the case when the computer interface from other stock exchanges around the world is used, then that knowledge claim would have little external validity.

Although this paper speaks to questions of ecological validity, my interest is with respect to the question of external validity.

1.2 Some Caveats and Notes

If a model is rejected using undergraduate students while it finds support using professionals because professionals are better at the task at hand, it clearly undermines the external validity of the result, but does it imply that the practice of using undergraduate students is misguided? Models usually do not start with "an agent, who is a professional at doing..." The model specifies the environment and the incentives and that is what an experiment, whether it uses undergraduate students or not, recreates. Thus, a failure of the predictions of the model is a failure whether the subjects were undergraduate students or professionals. However, it is true that, to the extent that the model is used to think about the world, and that outside the laboratory it is often professionals who evolve in the environment of interest, then the failure of the model with undergraduate students might be of little interest.

On the other hand, if the predictions of a model are not rejected using undergraduate students but they are using professionals,[9] then, in terms of testing the model, one could argue that the students were a better population than the professionals. All that is indicated by the failure with professionals is that some element that matters in the environment where those professionals evolve is lacking in the model being tested.

One point worth noting, since it influences my take on the papers reviewed, is that I am more interested in the comparative statistics and qualitative nature of the results than in tests of point predictions when it comes to testing models. To illustrate what I have in mind, consider the following hypothetical example. Suppose that in an ultimatum game experiment (a proposer offers a split of a pie to a responder who can accept, in which case each party receives the proposed split, or reject, in which case they receive nothing) with students the results are the following. The average offer is 0.45 of the pie, and 95% of offers at or above 0.4 are accepted while 85% of offers below 0.4 are rejected. Furthermore, when the game is instead a dictator game (a proposer offers a split of a pie to a responder and each party receives the proposed split) then offers decrease to about 0.15 of the pie. To me the interesting results are: the fact that the average offer in the ultimatum game gives a large fraction of the pie to the responder, the fact that non zero offers are rejected, and the fact that offers are substantially reduced in the dictator game. Suppose that a similar experiment with professional negotiators finds that the average offer is 0.4 of the pie, and 95% of offers at or above 0.35 are accepted while 80% of offers below 0.35 are rejected; and that when the game is instead a dictator game then offers decrease to about 0.1 of the pie. I would conclude from those results, that the findings are robust, even if all the numbers

[9] This could be the result, for instance, of professionals using, in the way they think about what to do, elements which matter in their professional environment even though they are not in the particular environment of the experiment.

mentioned above are statistically different for the students and the professionals. Hence, in my review, I sometimes classify the results as the same for the two groups even if there were quantitative differences. To be considered different, the two groups should produce results which lead to a different interpretation of behavior with respect to the model's prediction.

One much earlier paper about issues relating to subject pools is John Kagel's 1987 chapter in "Laboratory Experiments in Economics: Six Points of View" entitled "Economics According to the Rats (and Pigeons too): What Have We Learned and What Can We Hope to Learn?" The introduction to that paper presents a case for why experiments using animals may be informative for economics. I highly recommend reading that paper to people interested in the issues discussed in this review (even if they are not interested in animal experiments). For instance, reading the Kagel piece, one can easily replace "animal" with "undergraduate student" and "human" with "professional", and many of the arguments will still resonate.

Finally, let me point out that some of the papers covered here did not set out to compare students and professionals, while others are simply interested in different aspects of the results than I am, and thus my description of the results does not always parallel the focus of the papers in question.

2 Review

2.1 Other Regarding Preferences

2.1.1 Bargaining Behavior: A comparison between Mature Industrial Personnel and College Students

Siegel and Harnett (1964) study male undergraduates at The Pennsylvania State University (42 subjects) and compare their behavior to that of General Electric employees working in the Industrial Sales Operation division (32 subjects) in a bargaining game that they refer to as a price leadership bilateral monopoly.[10]

The seller (the price leader) picks a price and the buyer then selects a quantity to be traded. A given price-quantity pair implied certain profits to the buyer and seller, such that for a given quantity, a higher price generates more profits for the seller and less for the buyer,

[10] A more detailed analysis of the students experiment is available in Fouraker, Siegel, and Harnett (1962).

and for a given price, a higher quantity first increases profits and then decreases profits for both buyers and sellers. There were no communications besides price and quantity. Prices could vary between 1 and 16, and quantities between 0 and 18 such that the (inefficient) equilibrium involves a price of 9 and a quantity of 10. However, there exists a price-quantity pair (4, 15) which is efficient and involves equal profits for both players (higher for both than in equilibrium). There are two treatments: complete information, where both sides know the profits of each other; and incomplete information, where subjects are only informed of their own profits.

Although the particulars of this game make it different than most bargaining experiments, it seems fair to say that the central tension is the same as in most bilateral bargaining experiment where one side moves first and the equilibrium predicts that an advantage to that player, such as in the ultimatum game.[11] In those games, it is frequent to observe divisions of the pie that are much closer to being equal than what is predicted by the subgame perfect equilibrium (Roth 1995). Although these authors precede the current wave of research on other regarding preferences, their intuition was much in line with it, namely that with incomplete information the equilibrium would emerge (since subjects have no way to know it is not equal nor efficient) while with complete information, results would be closer to equal-split outcome.

Incentives were such that at the end of the experiment, professionals would make almost $49 if they played the equilibrium, and a bit more than $54 if they played the equal-split outcome. Professionals played 3 practice rounds, and 11 rounds for money. The authors separately analyze the first 10 of these rounds as regular rounds, and the 11th as different because it was announced to be the final round. This is relevant in this case as subjects were in fixed pairs throughout the experiment and, thus, aspects of repeated play could have mattered. Students were paid such that if the played the equilibrium they would have made about $5.06 on average, while average profits under the equal split outcome would have been $5.68. They played 3 practice rounds, 20 rounds with the same payoffs, with the 20th being announced as the final round, plus a 21st round with triple the payoffs used in the previous 20 rounds. For both professionals and students they used a between subjects

[11] In the ultimatum game, a proposer offers a division of an amount of money and the responder can accept, in which case both players receive the proposed split, or reject, in which case they both receive nothing.

design. The authors offer no discussion of how the incentives compared to each group's typical earnings.

The results for both groups of subjects (confining attention to the rounds played for money) are that prices start close to the equilibrium, but move downwards over rounds in the complete information treatment, towards the equal-split price. After about half of the regular rounds (which differs across groups), prices have settled at the equilibrium in the incomplete information case and at the equal-split price in the complete information case. Besides the speed of adjustment, the only main difference is that in the final round, professionals move slightly towards the equilibrium price in the complete information case (about a quarter of the way on average), something which doesn't happen with students.

Overall, results from professionals are in line with results from students, namely they display a tendency toward outcomes that equalize payoff when they have the information allowing them to do so, and toward equilibrium otherwise. However, in the final round, professionals did show a tendency for strategic behavior that the students did not display. Unfortunately, the changes in protocol between the two studies make it difficult to draw strong conclusions one way or the other.

2.1.2 Choosing Between A Socially Efficient And Free-Riding Equilibrium: Nurses versus Economics And Business Students

Cadsby and Maynes (1998a) study undergraduate and graduate economics and business students at York University and the University of Guelph (6 sessions of 10 subjects labelled experiments E1-E6) and compared their behavior to that of registered nurses at four hospitals and one college (6 sessions of 10 subjects labelled experiments N1-N6) in a threshold public goods game.

At the beginning of each period, each participant receives 10 tokens. They can contribute any portion to a group fund. If 25 or more tokens are put in the group fund, each player receives 5 additional tokens. Payoffs are $10 - contribution$ if the threshold is not met, and $15 - contribution$ if the threshold is met. This is repeated 25 times.

There are two pure strategy Nash equilibria: either each player contributes 0, or the group contributes just enough to meet the threshold (which is Pareto optimal). There are an infinite amount of mixed strategy equilibria.

It is not clear what is the standard result in this environment. Croson and Marks (2000)

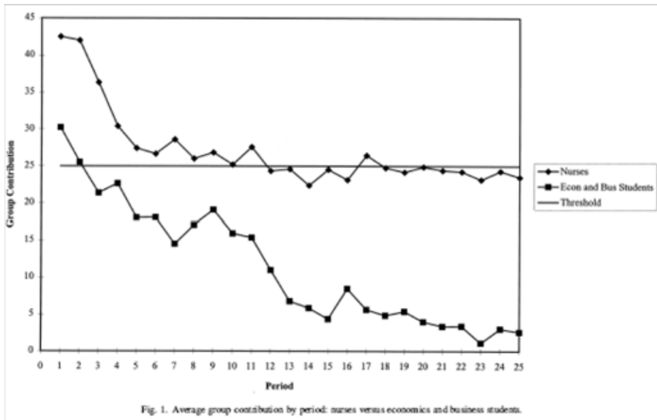

Figure 1: Nurses and Students in Cadsby and Maynes 1998

report, in their meta analysis, three other studies with a similar design (threshold public goods game with no refund, no rebate, no communication, no heterogeneity in endowments or valuation, and a step return of 2 – the ratio of the aggregate group payoff of the public good to the total contribution threshold), the differences are the number of players, the endowments, and the subject pools (Canada or the Unites States). In the paper by Cadsby and Maynes (1998b), contributions go toward 0, whereas in the papers by Croson and Marks (1998 and 2001) contributions go to the efficient level. According to the meta study, this divergence in results would result from the different endowment since the impact of the number of players is supposed to be in the opposite direction (the country where the study was conducted is not considered as a possible factor).

The incentives were the same for both types of subjects: 12 cents per token. Thus, if they were paid the average of the two equilibria times 25, they would make, \$31.56 which was about the opportunity cost for nurses for 1 hour and a half.

The results are that nurses started well above the threshold and finished close to the threshold. On the other hand, students started close to the threshold and finished far below the threshold. The average contributions are plotted in figure 1. In the last 5 periods, the authors report that neither student nor nurses are closer to a NE equilibrium, that is the distance between their contributions and the closest NE is not statistically different.

Thus, it seems that nurses contributed more. However, the frequency with which they provide the public good, albeit being higher than for students, is not statistically significant.

This raise a question, would these differences persists if the experiment had lasted longer? In the face of the threshold not being met, one can wonder how much longer nurses would have continued contributing.

Another question left open is how would a nurse in a group of students have behaved, and how would a student in a group of nurses have behaved? Is the difference driven by perceptions (or expectations) of others or by intrinsic differences. The downward trend seems to indicate that it is about perceptions (which they continually adjust).

Nonetheless, this does suggest that subject pools may matter for inferences about coordination situations. On the other hand, neither group is closer to the theory's prediction than the other.

2.1.3 The Hidden Costs And Returns Of Incentives – Trust And Trustworthiness Among CEOs

Fehr and List (2004) study students from the University of Costa Rica (126 subjects) and CEOs from the coffee mill sector (76 subjects) in a trust game.

Two versions of the game are studied. In both there is a principal and an agent and they are paired anonymously. Both receive 10 experimental currency units (ECUs). In the trust treatment, the principal transfers $x \in \{0, 1, 2, ..., 10\}$ and announces a desired back transfer $\hat{y}$. Then, the agent receives $3x$ and returns $y \in \{0, 1, 2, ..., 3x\}$. In the trust with punishment (TWP) treatment, in addition to x and $\hat{y}$, the principal chooses if he wants to impose a fine of 4 if $y < \hat{y}$. The agent is told x, $\hat{y}$ and f before choosing y.

In both treatments $x = y = 0$ is an equilibrium. In the TWP treatment there are other equilibria such that $f = 4$ and $y = 3$ or 4.

The incentives were 1 ECU converted to \$0.20 for students. Average earnings were \$5.65 for 45-60 minutes. The typical earnings are \$2 per hour. Professionals faced a conversion rate of 1 ECU equal \$2.00. This resulted in average earnings of \$65 for 45-60 minutes.

The typical results in trust games are that senders send money and responders send back money. However, it is often the case that on average responders send slightly less then senders have sent on average (Berg, Dickaut, and McCabe 1995).

The results that apply to both groups are the following. Principals choose $x > 0$ on average (in both treatments). Principals choose $3x > \hat{y} > 0$ on average (in both treatments). Agents send money back $3x > \hat{y} > y > 0$ on average (in both treatments). The effect of the treatment manipulation is negligible. More principals opt to use $f > 0$ than not. Payback is higher if $f = 0$. The differences between students and professionals are that CEO principals transfer more money than students. Also, CEO agents send back more money than students. Finally, CEO principals use the $f > 0$ less than student principals.

In conclusion, the main result seems to be that the insights one gains from the trust game experiments (and TWP) using students are supported by experiments with CEOs.

2.1.4 Do Social Preferences Increase Productivity? Field experimental evidence from fishermen in Toyama Bay

Carpenter and Seki (2005) compare university students from Japan (26 students - 2 sessions) to Japanese shrimp fishermen (27 subjects - 2 sessions) in voluntary contribution mechanisms experiments. In their dayly fishing activities some of the fishermen are organized in a group that shares both income and operating expenses (referred to as poolers) and a group that doesn't (nonpoolers).

The experiment has two parts for a total of 10 rounds. Throughout they are divided in groups of 4 (since the group size were not multiples of 4, some subjects were "active" in more than 1 group). In the first part of five rounds subjects do a standard voluntary contribution mechanism game. That is, each subject is endowed with 500 yen. They can keep the money for themselves or contribute to a public good. Everything contributed to the public good is doubled and divided equally by the members of the group. This was followed by five rounds of the social disapproval protocol. The social disapproval protocol is a modification of the standard VCM in which after subjects have decided on their contributions, they are told each member's contributions and asked if they want to send an unhappy face to the rest if the group at a cost of 10 yen to themselves. That signal is displayed to the group at the beginning of the following round.

The subgame perfect equilibrium of both games (with standard preferences) implies no contributions. In the case of the social disapproval protocol it implies no signals.

The typical result in experiments on the VCM games can be summarized as follows: positive contributions to start, but contributions decrease to relatively low levels over repetitions of the game, and settle slightly above 0 (Ledyard 1995).

The experiments typically lasted less than one hour. Average earnings were \$73.19.

Combining all 10 periods, poolers contribute slightly more than nonpoolers, and together they cooperate significantly more than students. However, in the standard VCM poolers and nonpoolers are not statistically different, while post disapproval rates are significantly different. Social disapproval rates are lowest for poolers and highest for students, but of course this is partly driven by the different contribution levels. Going from the standard VCM to the social disapproval protocol, contributions increase for poolers while they decrease for both nonpoolers and students.

Thus, contribution levels vary across subject pools, including across professional subject pools. The direction of change across treatments is the same for students and nonpoolers but is different for poolers.

2.2 Market Experiments

The two experiments under this heading are different in their focus. The first one is a variation on the standard oral double auction (ODA). The other could have been categorized differently, it studies a principal-agent problem within a sealed bid market structure (sellers make offers). However, they have been grouped together due to the extreme power of market environments at bringing about the "correct" outcomes, in the case of the ODA even in the face of extremely limited agents (Dhananjay and Sunder 1993). One could say that under most conditions, ODA quickly results in equilibrium outcomes (Holt 1995). However, "posted-offer markets tend to competitive predictions more slowly than comparable DA and they tend to converge to the competitive price prediction from above (if at all), implying a transfer of surplus from consumers to producers in the adjustment process." (Davis and Holt p. 217.)

2.2.1 Experience And Decision Making: A Comparison Of Students And Businessmen In A Simulated Progressive Auction

Burns (1985) compares the behavior of second year microeconomics undergraduates (who were part of an economic course) and experienced wool buyers (senior buyers with an average 35 years of experience) in a progressive oral auction with homogeneous commodities.

There are two groups of 9 student subjects and one group of 9 professional subjects (4 are officers of the Australian Wool Corporation and 5 represent the leading wool brokerage houses.) The progressive oral auctions with homogeneous commodities is close to the market structure faced by wool buyers.

In each market every buyer has a demand for 2 units, the second having lower returns than the first. There is a supply of 12 units (unknown to the bidders). Every bid must be an improvement bid. When no bid has been placed for 5 seconds, the item is declared sold. There are penalties for untraded units, which is a way to stimulate trade and is argued to represent the fact that the requirement to meet the full demand quota is very serious in the wool market.

15 auctions are conducted in total (5 per week). Conditions are constant within week but demand changed across weeks. Traders are informed of the direction but not the magnitude of the change.

The standard result for oral double auction experiments is that price and quantity traded converge rapidly to equilibrium predictions (Holt 1985). Hence comparative static predictions of the impact of demand and supply change on the direction of change of the quantity and price variables is as predicted.

Table 1 is taken from the paper and indicates the values for each bidder and the predicted market prices. The key prediction of interest is that under perfect competition, the market clearing prices would be: 96.5, 90.5, and 98.5 in weeks 1, 2 and 3 respectively.

The incentives for the students were very unusual and rather vague. The experiment was part of a course exercise for which an essay worth 10% of the students final assessment must be written. The students do not know the subject of the essay, but they are advised that "only by striving to maximize their profits would they gain the understanding necessary to successfully complete the assignment." The incentives for the professionals are also very unusual. It was announced to the wool buyers that "the "best" trader would be revealed at the end of the session."

The results can be summarized as follows. Wool buyers bid up to their marginal values on the first lot, then marginal value plus penalty on the second. Students behave similar to wool buyer on day 1 of week 1, and then the curve flattens out in subsequent days (more contracts at or close to the market equilibrium prediction). In each successive week the period of adjustment to profit maximizing level of prices grows shorter and they slowly incorporate the penalty in their bids over time. As a result, students make much more money then wool buyers. From one week to another, both students and wool buyers change in behavior is such that price adjusts in the right direction.

Discussions and interviews with the wool buyers suggest that their behavior was driven

	Week 1			Week 2			Week 3		
	Lot 1	Lot 2	Av.	Lot 1	Lot 2	Av.	Lot 1	Lot 2	Av.
Buyer	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
1	115	96	105.5	96	93	9405	114	97	105.5
2	113	95	104	95	92	93.5	111	96	103.5
3	112	95	103.5	94	91	92.5	109	95	102
4	109	94	102.5	109	90	99.5	106	95	100.5
5	107	93	100	107	89	98	104	101	102.5
6	104	93	98.5	106	89	97.5	103	100	101.5
7	102	99	100.5	103	88	95.5	102	99	100.5
8	101	98	99.5	101	87	94	117	98	107.5
9	101	97	99	98	87	92.5	115	97	106
Expected average market price (= equil. value) if marginal values are used.			96.5			90.5			98.5
Expected average market price (= equil. value of column 3) if average values are used.			99.75			93.75			101.75
Expected average market price (= average of 2nd to 13th valuations) if full-value bidding is used.			103.25			97.25			105.25
Throughout this paper, Australian dollars are used as the monetary unit.									

Table 1: Buyers Valuations'

by the behavior they have learned in the market they know. More specifically, one aspect of their professional experience that seems to conflict with the specific of the experiment is the fact that the good they deal with in their everyday experience is homogeneous. That, in particular, means that these traders are not accustomed to noticing within day price variations as these can represent different quality wool. Prices at which wool trades is noted by a junior and traders receive changed instructions from the head office across days; those changes reflect price trends. As a result, despite the fact that each auction featured a sharp decline in prices in the course of the session, 7 of the 9 professional buyers reported not noticing that pattern. Another practice that stems from their experience is to bid on units for which they are not interested in to "keep others honest."

A key question left open is whether the professionals would have generated higher earnings if they had been paid, i.e. would they have relied less on the reflexes they have developed in the usual environment and paid more attention to the specific details of the new environment they were confronted with.

2.3 A Note On the Use Of Businessmen As Subjects In Sealed Offer Markets

DeJong, Forsythe, and Uecker (1988) study the behavior of students from the College of Business at the University of Iowa (1 session of 7 subjects) and of members of the Professional Accounting Council of the Department of Accounting at the University of Iowa (1 session with 5 partners in public accounting or auditing and two corporate financial officers) in an elementary principle-agent problem.

Buyers and sellers are provided with an initial endowment of $1 for the buyers and $6 for the sellers. Each seller submits a sealed bid offer to buyers that specifies a quality of service and a price. Buyers decide which offer to accept. Then a random draw is performed to determine if the buyer incurs a loss or not given the quality of service he has bought. The parameters were such that the payoffs are as presented in Table 2.

The due care standard is set as 3. That is, the seller is responsible for a loss if he provided a level of quality lower than 3.

In a perfectly competitive market the predicted outcome is service of quality 3 provided at a price of $0.26.

Students were paid the sum of their earnings, which ranged between $10 and $25. Pro-

Level of service quality x	Prob. of loss P(x)	Expected loss[a] P(x)l	Change in prob. of loss ΔP(x)	Change in expected loss -ΔP(x)l	Cost of service C(x)	Change in cost ΔC(x)	Seller expected cost[b]	Expected social cost[c]
No purchase	0.90	72c			0c		0c	72c
1[d]	0.60	48c	-0.30	24c	5c	5c	53c	53c
3[e]	0.20	16c	-0.40	32c	26c	21c	26c[e]	42c[e]
5	0.05	4c	-0.15	12c	53c	27c	53c	57c

a. Loss(l)=80c throughout.

b. Seller's expected cost when the seller bears liability for the loss is predicted to be P(x)l+C(x).

Seller's expected cost when the buyer bears liability for the loss is predicted to be C(x).

Under the assumption o frisk neutrality, the fee, r, in a competitive market is predicted to equal the expected cost.

c. Expected social cost if P(x)l+C(x).

d. Seller bears the loss under the negligence liability rule.

e. Predicted equilibrium price and quality of the service. This quality of service is also the due care standard and level of service that minimizes expected social costs.

Table 2: Parameter values for markets

fessionals faced a different payment structure. First, each professional subject was given an envelope with the average earnings of the student who had the same role as they have. They played for points, at the end, if they had made more points per period then the student they were paired with, they received a University of Iowa Pewter souvenir, otherwise they received nothing.

The results indicate that the prices for each quality of service are not statistically different for students and professionals. As for the quality of service, students met or exceeded the due care standard more often than professionals, however that difference is not statistically significant. The same conclusions can be reached looking at average expected profits, namely they are not statistically different for professionals and students (only one of the 18 comparison they make is statistically different). Similarly, out of the 18 comparisons the authors conduct relating to allocative efficiency, only 2 are statistically different.

Hence, behavior of students and businessman is very similar in this experiment, but just as with the previous study it is impossible to tell what, if any, role the unusual (and different for the two groups) incentives played.

2.4 Information Signals

This section includes two signaling experiments and one experiment about information cascades. All these models rely on Bayesian updating, something which in general humans have shown not to be extremely good at. (Camerer 1995) However, models of signaling games have found support when subjects are given ample experience (Cooper, Garvin, and Kagel 1997) and work on information cascade also finds that many of the model's predictions are borne out in the data (Anderson 1997).

2.4.1 Gaming Against Managers In Incentive Systems: Experimental Results with Chinese Students and Chinese Managers

Cooper, Kagel, Lo, and Gu (1999) compare college students from China Textile University (10 sessions) to managers (12 sessions). Managers include older managers (high-ranking managers, mid-level managers, and senior foremen) from textile factories in Shanghai (10 sessions) as well as younger managers who are graduate students at China Textile University and had spent at least 5 years in factories before returning to school for an M.B.A.-type degree as well as China Textile University alumni working in the area with at least 2 years working experience (2 sessions). They compared the behavior of these groups in the ratchet effect game.

The game played in the lab is a simplified version that generates a standard signaling game which is thought to represent a typical problem in centrally planned economies, namely the fact that production targets assigned by the central planner to specific firms increase with productivity, thus giving an incentive to firms to misrepresent their true productivity. Prior to the start, firms learn their type (50% probability of each). Firms choose output and at any output payoffs are higher if the central planer chooses "easy" rather than "tough". Central planners see the output level chosen and select "easy" or "tough". The central planner's payoffs are a function of the target and the Firms type. The payoffs are indicated in Table 3.

About half the sessions were conducted in generic terms and half were in context (referring to easy and tough contracts, and high-productivity firms, etc). Each session had 12 to 16 subjects. The game was repeated 36 times with role reversed after every 6 games.

The game has 3 pure strategy sequential equilibria: pooling at output levels 1, 2, or 3.

The incentives were the following. The students had two payment schedules. In the standard pay: pooling at 2 corresponds to an expected payoffs of 30 yuan ($3.75) which is about equivalent to the earnings in a typical U.S. experiment. In the high-pay cases, payments were multiplied by 5, which represents 150 yuan in the pooling equilibrium at 2. As a point of comparison, the monthly wage for an associate professor was 1200 yuan. Managers had the same incentives as the students in the high-pay condition. Note that the vast majority of managers in the experiment earned less than 2000 yuan per month.

The results are the following. First, looking at all treatments as a whole. Firm's choices start clustered around their type's full information output. Central planners give easier contracts to outputs 1-3 than higher. Experience increases the frequency of 1-3 outputs by high firms (strategic play) with play converging to 2. However, a sizable frequency of nonstrategic play by high firms remains even in the last 12 games. Second, we focus on students only and the impact of incentive variation. Increased pay promoted more strategic play by firms initially. However, by the end there is no difference. Increased pay has no impact on central planner's choice. Third, we focus on the impact of the variation in context. There are no effects on students acting as Firms. Mistakes by central planners are reduced for students but only in standard pay. There is an increased level of strategic play for

Firm Payoff					
	Low-Productivity firm Production target		High-productivity firm Production target		
Output	EASY	TOUGH	EASY	TOUGH	Output
1	710	542	1,108	815	1
2	730	561	1,145	852	2
3	697	528	1,230	937	3
4	527	357	1,308	1,015	4
5	273	103	1,328	1,035	5
6	220	48	1,298	1,005	6
7	190	15	1,250	966	7

Planner Payoff		
Production target	Facing low-productivity firm	Facing high-productivity firm
EASY	645	764
TOUGH	528	820

Table 3: Firm Payoff and Planner Payoff

Payoff to S,R		R's choice	
		B1	B2
Low message cost (c=0.5)			
State	White (P=2/3)	2, 3	4, 1
k	Black (P=1/3)	1, 0	7, 1
High message cost (c=1.5)			
State	White (P=2/3)	1.5, 3	3.5, 1
k	Black (P=1/3)	1.5, 0	5.5, 1

Table 4: Firm Payoff and Planner PayoffPayoffs (in Dutch guilders) to sender and receiver in the two treatments depending on the color of the disk (white or black) and the choice of the receiver (B1 or B2)

managers in their role as firms in later cycles of play. There is a strong effect on managers as central planers, promoting higher target rate differentials than in the generic sessions.

To conclude, similar behavior is observed between students and managers. However, context helps managers.

2.4.2 Professionals And Students In A Lobbying Experiment Professional Rules Of Conduct And Subject Surrogacy

Potters and van Winden (2000) compare the behavior of undergraduate students (mostly economics majors) from the University of Amsterdam (12 sessions) and public affairs and public relations officers from both the private and public sector (3 sessions) in a lobbying game.

The game is a signaling game with a sender (S) and a receiver (R). R decides between B1 and B2, the payoffs of S and R depend on the realization of a stochastic variable (k) which can take 2 values (w with $p = \frac{2}{3}$ and b with $p = \frac{1}{3}$). S sees k and decides to send a message or not. Sending a message costs c to S. After R is informed of S's decision, he picks B1 or B2. Two treatments: low message cost (L) and high message cost (H). The specific payoffs are shown in Table 4.

Each session had 10-12 subjects and lasted for 10 rounds of play.

The game has a unique sequential equilibrium that satisfies "all" the refinements. In treatment L: if $k = w$, S sends a message with $p = \frac{1}{4}$, otherwise he sends a message for sure.

If R gets no message, he picks B1, otherwise he picks B2 with $p = \frac{1}{4}$. In treatment H: if $k = w$, S sends a message with $p = \frac{1}{4}$, otherwise he sends a message for sure. If R gets no message, he picks B1, otherwise he picks B2 with $p = \frac{3}{4}$.

The incentives for the students were such that the expected equilibrium earnings were 20 guilders per hour (typical earnings are 15 guilders per hour). The professionals received 4 times as much, hence in equilibrium that represents 80 guilders per hour (approximately their estimated wage per hour).

The results that apply to both subject pools are the following. S engages in costly signaling. R responds in the expected direction. R chooses B2 more after a message in the H treatment than in the L treatment. More messages are sent when $k = w$ in the L treatment than in the H treatment (not predicted by the theory). Two differences emerge between the two groups. First, the frequencies with which professionals send messages following w versus b are closer to the equilibrium predictions than for the students. Second, professionals in the role as S earn slightly more than students (it is not clear from the paper if that is significant).

To conclude, the results indicate only minor substantive differences. The core result seems to be that students, just as professionals, understand the central strategic tension in this game.

2.4.3 Information Cascades: Evidence From A Field Experiment With Financial Market Professionals

Alevy, Haigh, and List (2007) compare students at the University of Maryland (10 sessions with 54 subjects) and market professionals from the floor of the Chicago Board of Trade (10 sessions with 55 subjects) in a game about information cascade formation.

The game is the following. The state of nature (A or B) is selected with the roll of a die, but not revealed. Subjects draw a signal from an urn. Subjects make their choice sequentially with their choice revealed to others. In the symmetric treatment, urn A contains two type-a ball and one type-b ball; urn B contains 2 type-b balls and 1 type-a ball. In the asymmetric treatment four type-a balls are added to both urns. The other treatment variation is that some treatments are framed as gains and others as losses (gain for correct guess or loss for incorrect guess). This is repeated for 15 rounds in groups of 5 or 6.

The theoretical predictions for the state of the world being A are presented in Table 5.

Posterior Probabilities: Symmetric (upper) and Asymmetric (lower) Urns							
a \ b	0	1	2	3	4	5	6
0	0.500	0.330	0.200	0.110	0.060	0.030	0.020
	0.500	*0.330*	*0.200*	*0.110*	*0.060*	*0.030*	*0.020*
1	0.670	0.500	0.330	0.200	0.110	0.060	
	0.545	*0.375*	*0.231*	*0.130*	*0.070*	*0.036*	
2	0.800	0.670	0.500	0.330	0.200		
	0.590	***0.419***	*0.265*	*0.153*	*0.083*		
3	0.890	0.800	0.670	0.500			
	0.633	***0.464***	***0.302***	*0.178*			
4	0.940	0.890	0.800				
	0.675	*0.509*	***0.341***				
5	0.970	0.940					
	0.703	*0.554*					
6	0.980						
	0.749						

Table 5: Entries represent the posterior probabilities for all possible sequence of draws for both symmetric (upper) and asymmetric (*lower*)treatments based on choice histories (a,b). The prior probability of an urn is 0.5 in (0,0). Bold entries for the asymmetric urn are those in which counting and the posterior probability make different predictions about the state.

Students received \$1 per correct guess or -\$1 per incorrect guess. Professionals received \$4 per correct guess or -\$4 per incorrect guess. which made the median payout greater than \$30 for 30 minutes.

For both subject pools, a majority of choices are consistent with Bayesian updating, and a non-negligible fraction of cascades are realized. Earnings and the rate of cascades are not different across subject pools. The differences are the following: professionals are slightly less Bayesian than students; professionals enter in fewer reverse cascades; and, finally, although students' Bayesian behavior is affected by the gain and loss domain the professionals' is not.

Thus, although there are differences in the details, the main insights from the student population extend to the professionals.

2.5 Other Topics

2.5.1 A Comparison Of Naive And Experience Bidders In Common Value Offer Auctions: A Laboratory Analysis

Dyer, Kagel, and Levin (1989) compare undergraduate students from upper-level economics majors at the University of Houston (3 sessions labelled experiments 1-3) to executives from local construction companies (1 session labelled experiment 4) in a series of sealed-bid first-price common value auctions. The professionals have on average 20 years of experience in the construction industry – all but one have many years experience in bid preparation.

The game is the following. Subjects have the right to supply a single unit of a commodity, and the commodity is awarded to the lowest bidder using a sealed-bid first-price auction. The cost, C, is unknown at the time of the bid. The winner earns $bid - C$ while the others earn 0. C is drawn from a uniform on $[\$50, \$250]$. Each bidder receives a private signal c_i drawn from a uniform on $[C - \varepsilon, C + \varepsilon]$. Distributions for C and c_i, ε and N are common knowledge.

Experiments 1-4 start with 4 active bidders. In experiment 4, after 24 periods 1 professional and 2 "experienced" students subjects are added both with and without an announcement of the highest cost estimate C_H.

Theory predicts that for costs in the interval $[\$50 + \varepsilon < c_i < \$250 - \varepsilon]$, the symmetric risk neutral Nash equilibrium (SRNNE) is

$$b(c_i) = c_i + \varepsilon - Y$$

where $Y = \left[\frac{2\varepsilon}{N+1}\right] \exp\left[-\left(\frac{N}{2\varepsilon}\right)(250 - \varepsilon - c_i)\right]$. Y diminishes rapidly as c_i moves below $\$250 - \varepsilon$. Thus, in equilibrium, bids mark-up signals by close to ε.

The incentives are the same for both types of subjects (although the number of periods differ and the details of the auction are varied for part of experiment 4). The starting balance is \$10. Losses and profits from every auction played are added to the starting balance. Thus under the SRNNE this represented about \$100 in experiment 4 which lasted about 3 hours.

The result that really stands out in the experimental literature on common value auction is the failure to account for the winner's curse and thus the persistent finding that bids are such that loses would result on average. One result which is somewhat of a hallmark of the winner's curse is the fact that many experiments show that losses increase as the number of bidders is increased (Kagel and Levin 2002).

In their experience, the authors observe that for executives: negative or near zero profits dominate, there is little evidence of learning over time, losses increase as N increases, announcing C_H raises the offer price (contrary to the theoretical prediction).

Comparing the executives to the students, they find that both fall prey to the winner's curse. There are no difference (at 10% level) for any of the following: percentage of times the low bid is submitted by the low signal holder, average actual profits, and percentage of times the low bid is such that it would result in losses on average. There are small differences in terms of the impact of changes in ε. There are also differences in the details of the reaction to changes in N as compared to other experiments with students. The impact of C_H is consistent with prior experiments with students.

Thus, executive and students behavior is similar in that both exhibit the winner's curse and are similar on many other relevant dimensions. However, there are some differences in the details. One question this raises is how are these executives successful in their business? Based on discussions with the executives, the authors conclude that the executives have learned a set of situation specific rules of thumb which permit to avoid the winner's curse in the field but could not be applied in the lab. This is explored in more details in Dyer and Kagel (1996).

		Goalkeeper	
		Left	Right
Kicker	Left	0.6, 0.4	0.95, 0.05
	Right	0.9, 0.1	0.7, 0.3

Table 6: Penalty Kick Game

		2			
		Red	Brown	Purple	Green
	Red	0, 1	1, 0	1, 0	0, 1
1	Brown	1, 0	0, 1	1, 0	0, 1
	Purple	1, 0	1, 0	0, 1	0, 1
	Green	0, 1	0, 1	0, 1	1, 0

Table 7: O'Neil Game

2.5.2 Experientia Docet: Professionals Play Minimax in Laboratory Experiments

Palacios-Huerta and Volij (2006) study how non-economics and non-mathematics majoring male Spanish university students (160 students) and male professional soccer players (40 kickers and 40 goalkeepers) behave in two zero-sum games.[12]

The first game is the 2x2 zero-sum game in Table 6.

The numbers used correspond to the probabilities that a penalty kick would be successful. However, no soccer terminology or reference was made in the experiment. Subjects played 15 rounds for practice and 150 rounds for money, at 1 Euro per round. Experiments lasted on average 70 minutes for professionals and 81 minutes for students. The second Game is taken from O'Neil (1987). It corresponds to the zero-sum game in Table 7.

The payoffs are in Euros. This was played 15 rounds for practice and 200 rounds for money. On average, sessions lasted 61 minutes for professionals and 68 minutes for students.

The unique mixed strategy equilibrium of the penalty kick game is for the kicker to play Left with probability 0.3636 while the goalkeeper plays Left with probability 0.4545. The unique mixed strategy equilibrium of O'Neil's game is for both players to play Green with

[12] Although I will simply base my analysis on the authors own account of their data, Wooders (2008) reaches different conclusions from reanalyzing the Palacios-Huerta and Volij (2006) data.

probability 0.4, and to play each of the other choices with probability 0.2.

The results are clearly supportive of the minmax model for the soccer players. Namely, choice frequencies in both games are either not statistically different from that predicted by the model, or so close that the difference is not important. Furthermore, the data supports the implication of the theory that choices are independent draws. On the other hand, the results for students are much further away from the theory. In the penalty kick game, the aggregate frequencies are not too far off, however, there aren't enough differences in the choice frequencies of kickers and goalkeepers. Furthermore, unlike soccer players, students do not generate "random" sequences. Similarly for O'Neal's game, students aggregate behavior is similar to the model's predictions, but not once one looks at individual behavior.

Non-economics and non-mathematics majoring male Spanish university students do not behave as male professional soccer players in the zero-sum games studied.

2.5.3 Option Pricing by Students and Professional Traders: A Behavioural Investigation

Abbink and Rockenback (2006) compare students (mostly from the economics and law departments) from the University of Bonn (108 students in 6 treatments) to 24 employees (2 treatments) from"an influential German Bank in Frankfurt/Main" who were "decision makers" in their "departments of foreign exchange, security, futures, bonds, and money trade" (p. 503) in an option pricing (individual decision) experiment. They also add an additional 2 treatments of students from a different institution: the University of Erfurt, which has mainly social science students in non-technical fields. In particular, option pricing is not part of the curriculum in any courses offered at University of Erfurt, while it is part of the Finance module, a popular one with economics students, at the University of Bonn.

Subjects have 600 units of experimental currency to invest in one of three potential forms, 2 risky investments, X and Y, and a riskless investment which is to keep the currency (at an interest rate of 10%). The returns to the risky investments depend on the state of the world tomorrow: *Red* (with probability q) or *Blue* (with probability $1 - q$). The subjects are first offered to buy or sell up to 9 units of X at price P. Following that, they can buy, sell, or not trade in Y at a price of 100 per unit with the constraints that they can only trade up to a quantity such that they will not be bankrupt tomorrow. The cash flows per unit traded are represented in Table 8.

Cash flows of the investment forms

When selling				When buying			
	Today	*Red*	*Blue*		Today	*Red*	*Blue*
X	$+P$	-62	-2	X	$-P$	$+62$	$+2$
Y	$+100$	-150	-90	Y	-100	$+150$	$+90$

The interest rate for cash is 10%.

Table 8: The Investment Possibilities

Experience Phase	Students		Professionals	
	Treatment, q (%)			
	20	70	20	70
Rounds 1-15	25.17	38.70	27.48	36.49
Rounds 16-30	18.64	35.89	32.08	39.35

Table 9: The Separating Prices Over Time

A rational investor in this environment follows an option pricing strategy with a separating price of 20. That is, the investor buys X when P is less than 20 and sells it when it is above 20. This result is independent of q, the probability of states *Red* and *Blue*.

The experiments used a between subjects design varying q across values 10, 20, 33, 50, 70, and 90 for students; and values 20 and 70 for professionals. Students subjects played 50 rounds, and in each round a different price P between 2 and 62 was randomly selected. The professionals played 30 rounds with prices drawn from the odd numbers between 2 and 62.

The experimental currency was exchanged for money at an unspecified rate. The student sessions lasted between 1 and 2 hours whereas the professional sessions lasted 1 hour. The authors say the following of the exchange rate for professionals: "The exchange rate of the thalers [the experimental currency] earned in the experiment was chosen such that it was comparable to the one used in the corresponding students treatment." (pages 503-504)

Four results stand out. First, students react more to q than professionals, that is, the separating price that best fits the subjects behavior varies more with q. Remember that in theory q should not affect the separating price. Second, the estimated average separating price is closer to the theoretical one for students than for professionals. Third, over time students move closer to the theoretically predicted price, whereas the professionals move in the opposite direction. A subset of the results (for the cases where data are available for students and professionals) are reproduced in Table 9. Fourth, the professional traders exploit arbitrage opportunities less than students and, as a consequence their behavior implies lower expected value exploitation.

In the additional treatments conducted with students from the University of Erfurt (using the same protocol and treatment parameters as for professionals) they find behavior closer to that of professionals.

As a whole, it seems that professionals behavior is further from the predictions of the theory than that of students, although students fail more severely in terms of their reaction to the treatment variable q.

2.5.4 Are Experienced Managers Experts at Overcoming Coordination Failure?

Cooper (2006) compares 20 Case Western Reserve University undergraduate students to 19 students of the executive MBA program of that university's Weatherhead School of Management as managers in what he calls a Turnaround game. That is the managers set bonuses in a weak-link game (also known as a minimum game). The difference between the two pools are covered in details in the paper, but some of the key facts about the professionals are that they have at least 10 years of relevant work experience, with at least 5 years involving substantial managerial responsibilities. They come mostly from manufacturing, but there are also many in the healthcare and service sectors and their average income is \$122, 000.

Subjects are in groups of 5, 1 manager and 4 employees. The employees are all undergraduate students whereas the managers vary with treatment (professionals or students). For 10 rounds, employees play the minimum game shown in Table **??**. Managers and employees both observe the minimum of the group choice. In the following 20 rounds, managers select a bonus (B can be any integer between 6 and 15) and can send a message to the employees before the employees select their effort (E). The bonus is costly to the manager but transforms the game of the employees, the payoff functions are

$$\begin{aligned} \text{Manager payoff} &= 100 + \left(\left(60 - 4B \times \min_{i \in \{1,2,3,4\}} (E_i) \right) \right) \\ \text{Employee i payoff} &= 200 - 5E_i + \left(B \times \min_{i \in \{1,2,3,4\}} (E_i) \right) \end{aligned}$$

Table **??** shows the employees stage game for a bonus of 14.

		Minimum Effort by Other Employees				
		0	10	20	30	40
	0	200	200	200	200	200
Effort	10	150	210	210	210	210
By	20	100	160	220	220	220
Employee i	30	50	110	170	230	230
	40	0	60	120	180	240

Table 10: Employee's Payoff When the Bonus is 6

		Minimum Effort by Other Employees				
		0	10	20	30	40
	0	200	200	200	200	200
Effort	10	150	290	290	290	290
By	20	100	240	380	380	380
Employee i	30	50	190	330	470	470
	40	0	140	280	420	560

Table 11: Employee's Payoff When the Bonus is 14

The incentives are the same for both groups of subjects. Payoffs are in ECUs and converted to dollars at a 500:1 ratio. Hence average payoffs were $23.89, including the $10 show-up fee. Furthermore, participants were told that a list of earnings ranked from top to bottom would be sent to them. A few aspects of the design that are atypical are the following. First, the managers were separated from the employees, but in particular the professional managers were participating from home over the internet. Second, the instructions changed slightly over time to deal with the fact that professionals were having difficulty operating the software. Third, the instructions were neutral but used words that might be more evocative than what is typical (which the author refers to as naturalistic) For instance, employees were told they had to choose hours per week of work.

The results can be summarized as follows. By round 10 effort is very low (a typical result in minimum games, see for instance Van Huyck et al 1990), averaging 3.08 and is exactly 0 in 31 of the 39 groups. Bonuses and communication allow both groups two increase profits and efforts by coordinating on higher effort levels. The increase in effort is about 5 fold between round 10 and round 30 and the increase in profit is about 3 fold. Over the last 10 rounds, average minimum efforts are essentially indistinguishable between undergraduates and professionals, and so are bonuses and profits. Even more so if one focuses on groups which started at a minimum effort level of 0 in round 10.

Some differences emerge however. First, professionals manage to increase effort and profits much quicker than undergraduates do. From the graphs in the paper, it looks like it takes about 5 rounds for undergraduates to increase profits close to the levels of professionals. Second, professionals communicate different messages than undergraduates. In particular, professionals are much more likely to request a specific effort level, they are less likely to lay out a long term plan, and they are more likely to offer encouragement. They are also much more likely to refer to trust.

Hence, one would have reached similar conclusions using professionals and undergraduates in this experiment in the sense that both learned to use bonuses and communication to improve profits by increasing efforts.

	Other-Reg.	Market	Signaling	Other	Total
Pros closer				PHV	1
Similar	SH FL (CM)	DFU	CKLG PvW AHL	DKL AR Cooper	9
Pros further	CS	Burns			2
Different	CM				1
Total	4	2	3	4	13

Table 12: Summary of the distance to the theoretical prediction

3 Discussion

Thus, there are 13 papers that allow us to compare students and professionals in a standard laboratory environment (papers are referred to by the initials of its authors). The results are summarized in Table 12. In 9 of those 13, professionals are not closer or further from the theory in a way that would lead us to draw different conclusions. In the remaining 4, only 1 finds behavior on the part of professionals which is substantially closer to the prediction of the theory, and that is the study of Palacios-Huerta and Volij (2006). On the other hand, both Burns (1985) and Carpenter and Seki (2005) find that professionals and students have different behavior, but the difference goes the other way, namely students are closer to the theoretical prediction than professionals. In both of these cases, it would seem that the source of the difference are elements of the professionals work environment or preferences which leads to behavior different from that of the students. As for Cadsby and Maynes (1998a), although neither professionals nor students are closer to the theory, the behavior is so different that it needs to be categorized differently form the other studies where neither professionals nor students are closer to the theory in a way that changes how one thinks of the theory's performance. That is, even though both group are at a similar distance from the theory, one reaches qualitatively very different conclusions by studying these two groups.

One issue that such a review brings to light is the difficulty of having comparable incentives when comparing students and professionals. For instance, should the average payoffs be the average opportunity cost of time of professionals, or that of students, or each group receives different incentives? The most thorough approach in that regard might be that adopted in Cooper, Kagel, Lo, and Gu (1999) where professionals are paid on average their opportunity cost of time, while students are divided in one group that faces those same incentives while another receive payoffs equal to the opportunity cost of time for students.

Another difficulty is that to the extent that differences between professionals and students are observed, although it is tempting to attribute those differences to the experience of professionals, one cannot ignore that the two samples may vary along many other dimensions: gender, race, age, socio-economic background, etc. These factors could cause differences which have nothing to do with the fact that one group is composed of professionals and the other is not. In that respect, the most careful paper is Cooper 2006 in which he tries to disentangle the role of these various factors.

It also raises the question of what is the group of interest? Who are the agents that are supposed to be represented in those models being tested? Are Japanese fishermen organized in cooperative closer to the group of interest than students? What about soccer players? In that last case for instance, although it is true that soccer players have frequent exposures to situations where randomization is key, they also have more brain lesions than people who do not play soccer (Autti, Sipilä, Autti, Salonen 1997)?

In my view, this survey of the studies that allow to compare professionals and students, although it does indicate that there are situations where focussing on students is too narrow, does not give overwhelming evidence that conclusions reached by using the standard experimental subject pool cannot generalize to professionals. Studying professionals can prove very insightful in ways that studying undergraduates is not. For instance, the reaction of chinese managers to context was informative (Cooper, Kagel, Lo, and Gu 1999), and the fact that professional managers use different messages to get employees out of "bad" equilibria faster (Cooper 2006) was also an opportunity to learn something that could not have been learned with students. Nonetheless, overall much of the big picture seems the same whether one looks at professionals or students in laboratory experiments testing economic models.

References

Abbink, K. and Rockenbach, B.: 2006, Option pricing by students and professional traders: A behavioral investigation, *Managerial and Decision Economics* **27**, 497–510.

Alatas, V., Cameron, L., Chaudhuri, A., Erkal, N. and Gangadharan, L.: 2006, Subject pool effects in a corruption experiment: A comparison of indonesian public servants and indonesian students. THE UNIVERSITY OF MELBOURNE RESEARCH PAPER NUMBER 975.
URL: *http://www.economics.unimelb.edu.au/SITE/research/workingpapers/wp06/975.pdf*

Alevy, J. E., Haigh, M. S. and List, J. A.: 2009, Information cascades: Evidence from a field experiment with financial market professionals, *Journal of Finance* **62**(1), 151–80.

Andersen, S., Harrison, G., Lau, M. and Rutstrom, E.: 2009, Preference heterogeneity in experiments: Comparing the field and lab, *Journal of Economic Behavior and Organization* . forthcoming.

Anderson, L. R. and Holt, C. A.: 1997, Information cascades in the laboratory, *The American Economic Review* **87**(5), 847–862.

Anderson, M. J. and Sunder, S.: 1995, Professional traders as intuitive bayesians, *Organizational Behavior and Human Decision Processes* **64**(2), 185–202.

Autti, T., Sipilä, L., Autti, H. and Salonen, O.: 1988, Brain lesions in players of contact sports, *The Lancet* **349**(19), 1144.

Ball, S. B. and Cech, P.-A.: 1996, *Research in Experimental Economics*, Vol. 6, JAI Press, Greenwich, CT, chapter Subject pool choice and treatment effects in economic laboratory research, pp. 239–292.

Banks, J., Camerer, C. and Porter, D.: 1994, An experimental analysis of nash refinements in signaling games, *Games and Economic Behavior* **6**, 1–31.

Berg, J., Dickaut, J. and McCabe, K.: 1995, Trust, reciprocity and social history, *Games and Economic Behavior* **10**, 122–142.

Burns, P.: 1985, *Research in Experimental Economics*, Vol. 3, JAI Press, Greenwich, CT, chapter Experience and Decision Making: A Comparison of Students and Businessmen in a Simulated Progressive Auction, pp. 139–153.

Cadsby, C. B. and Maynes, E.: 1998a, Choosing between a socially efficient and free-riding equilibirum: Nurses versus economics and business students, *Journal of Economic Behavior and Organization* **37**, 183–192.

Cadsby, C. B. and Maynes, E.: 1998b, Gender and free riding in a threshold public goods game: Experimental evidence, *Journal of Economic Behavior and Organization* **34**, 603–620.

Camerer, C. F.: 1995. "Individual Decision Making" in *Handbook of Experimental Economics*, ed. by John H. Kagel and Alvin E. Roth. Princeton: Princeton University Press, 587-703.

Campbell, D. T.: 1957, Factors relevant to the validity of experiments in social settings, *Psychological Bulletin* **54**(4), 297–312.

Carpenter, J. and Seki, E.: 2005, Do social preferences increase productivity? field experimental evidence from fishermen in toyoma bay. IZA DP No. 1697.

Casari, M., Ham, J. C. and Kagel, J. H.: 2007, Selection bias, demographic effects and ability effects in common value auction experiments, *The American Economic Review* . forthcoming.

Cook, T. D. and Campbell, D. T.: 1979, *Quasi-experimentation: Design & analysis issues for field settings*, Rand McNally & Co, US.

Cooper, D.: 2006, Are experienced managers experts at overcoming coordination failure?, *The B.E. Journal of Advances in Economic Analysis & Policy* **6**(2), 1479–1479.

Cooper, D. J., Garvin, S. and Kagel, J. H.: 1997, Signalling and adaptive learning in an entry limit pricing game, *RAND Journal of Economics* **28**, 662–683.

Cooper, D. J., Kagel, J. H., Lo, W. and Gu, Q. L.: 1999, Gaming against managers in incentive systems: Experimental results with chinese students and chinese managers, *The American Economic Review* **89**, 781–804.

Croson, R. and Marks, M.: 2001, The effect of recommended contributions in the voluntary provision of public goods, *Economic Inquiry* **39**(2), 238–249.

Croson, R. T. A. and Marks, M. B.: 2000, Step returns in threshold public goods: A meta- and experimental analysis, *Experimental Economics* **2**, 239–259.

Davis, D. D. and Holt, C. A.: 1993, *Experimental Economics*, Princeton University Press, Princeton, NJ.

Dejong, D. V., Forsythe, R. and Uecker, W. C.: 1988, A note on the use of businessmen as subjects in sealed offer markets, *Journal of Economic Behavior and Organization* **9**, 87–100.

Dyer, D. and Kagel, J. H.: 1996, Bidding in common value auctions: How the commercial construction industry corrects for the winner's curse, *Management Science* **42**, 1463–1475.

Dyer, D., Kagel, J. H. and Levin, D.: 1989, A comparison of naive and experienced bidders in common value offer auctions: A laboratory analysis, *The Economic Journal* **99**(394), 108–115.

Fehr, E. and List, J. A.: 2004, The hidden costs and returns of incentives – trust and trustworthiness among ceos, *Journal of the European Economic Association* **2**(5), 743–771.

Fouraker, L. E., Siegel, S. and Harnett, D. L.: 1962, An experimental disposition of alternative bilateral monopoly models under conditions of price leadership, *Operations Research* **10**(1), 41–50.

Gneezy, U., Niederle, M. and Rustichini, A.: 2003, Performance in competitive environments: Gender differences, *Quarterly Journal of Economics* **CXVIII**, 1049–1074.

Gode, D. K. and Sunder, S.: 1993, Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality, *Journal of Political Economy* **101**(1), 119–137.

Harbaugh, W. T., Krause, K. and Berry, T. R.: 2001, Garp for kids: On the development of rational choice behavior, *American Economic Review* pp. 1539–1545.

Holt, C. A.: 1995. "Industrial Organization: A Survey of Laboratory Research" in *Handbook of Experimental Economics*, ed. by John H. Kagel and Alvin E. Roth. Princeton: Princeton University Press, 349-444.

Hong, J. T. and Plott, C. R.: 1982, Rate filing policies for inland water transportation: An experimental approach, *The Bell Journal of Economics* **13**, 1–19.

John H. Kagel, Ray C. Battalio, R. and Green, L.: 1995, *Economic choice theory: An experimental analysis of animal behavior*, Cambridge University Press.

Kagel, J. H.: 1987, Economics according to the rats (and pigeons too); what have we learned and what can we hope to learn?, *Laboratory Experimentation in Economics: Six Points of View* pp. 155–192.

Kagel, J. H. and Levin, D.: 2002, *Common value auctions and the winner's curse*, Princeton University Press.

King, R. R., Smith, V. L., Williams, A. W. and Boening, M. V.: 1993, *Evolutionary Dynamics and Nonlinear Economics – A Transdisciplinary Dialogue*, Oxford Press, Oxford, England, chapter The Robustness of Bubbles and Crashes in Experimental Stock Markets, pp. 183–200.

Ledyard, J. O.: 1995. "Public Goods" in *Handbook of Experimental Economics*, ed. by John H. Kagel and Alvin E. Roth. Princeton: Princeton University Press, 111-194.

Marks, M. and Croson, R.: 1998, Alternative rebate rules in the provision of a threshold public good: An experimental investigation, *Journal of Public Economics* **67**, 195–220.

Mestelman, S. and Feeny, D.: 1988, Does ideology matter?: Anecdotal experimental evidence on the voluntary provision of public goods, *Public Choice* **57**, 281–286.

Montmarquette, C., Rullière, J.-L., Villeval, M.-C. and Zeiliger, R.: 2004, Redesigning teams and incentives in a merger: An experiment with managers and students, *Management Science* pp. 1379–1389.

Palacios-Huerta, I. and Volij, O.: 2006, Experientia docet: Professionals play minimax in laboratory experiments. Working Paper.
URL: *http://www.najecon.org/naj/cache/122247000000001050.pdf*

Potamites, E. and Zhang, B.: 2008, Measuring ambiguity attitudes:a field experiment among small scale stock holders in china. Working Paper.

Riker, W. H.: 1967. "Experimental Verification of Two Theories about n-Person Games" in *Mathematical Applications in Political Science III*, ed. by Joseph L. Bernd. The University Press of Virginia Charlottesville, 52-66.

Roth, A. E.: 1995. "Bargaining Experiments" in *Handbook of Experimental Economics*, ed. by John H. Kagel and Alvin E. Roth. Princeton: Princeton University Press, 253-348.

Shadish, W. R., Cook, T. D. and Campbell, D. T.: 2002, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Hougton Mifflin.

Siegel, S. and Harnett, D. L.: 1964, Bargaining behavior: A comparison between mature industrial personnel and college students, *Operations Research* **12**(2), 334–343.

Smith, V. L., Suchanek, G. L. and Williams, A. W.: 1988, Bubbles, crashes, and endogenous expectations in experimental spot asset markets, *Econometrica* **56**, 1119–1151.

Van Huyck, J., Battalio, R. C. and Beil, R.: 1990, Tacit coordination games, strategic uncertainty, and coordination failure, *American Economic Review* pp. 234–248.

Wooders, J.: 2008, Does experience teach? professionals and minimax play in the lab. Working Paper.
URL: *http://www.u.arizona.edu/~jwooders/Reexamination 2006 2009 202008.pdf*