Will You Accept Without Knowing What? A Thuringian Newspaper Experiment of the Yes-No Game

Werner Güth* Oliver Kirchkamp†

17 June 2010

Abstract

Many economic experiments are run in the laboratory with students as participants. In this paper we use a newspaper experiment to learn more about external validity of lab research. Our workhorse is the Yes-No game. Unlike in ultimatum games responders of the Yes-No games do not know the proposal when deciding between whether to accept it or not. We use two different amounts that can be shared (100€ and 1000€). Unlike in other experiments with the ultimatum game we find a (small) effect of the size of the stakes. In line with findings for the ultimatum game, we find more generosity among women, older participants, and participants who submit their decision via mail rather than via Internet. By comparing our results with other studies (using executives or students), we demonstrate, at least for this type of game, the external validity of lab research.
Keywords: newspaper experiment; external validity; Yes-No game
JEL-Code: C91, C93

1 Introduction

A major aim of this study is to learn more about external validity of laboratory experiments. Many laboratory experiments are done with students.

* Max Planck Institute of Economics, Kahlaische Strasse 10, D-07745 Jena, Germany.
†Friedrich-Schiller-University Jena, School of Economics and Business Administration, Carl-Zeiss-Str. 3, 07743 Jena, Germany. Phone: +49-3641-943241, Fax: +49-3641-943242, oliver@kirchkamp.de

Members of this subset of the population have a similar age and a similar level of education. Hence, traditional laboratory experiments do not allow us to measure how age and education affect behaviour. They also leave open the question how far results from the lab can be generalised.

Experimentalists can increase the variance of socio-demographic characteristics in the subject pool in several ways. In particular the ultimatum game and the trust game have been studied with a more heterogeneous population.

Roth et al. (1991) study the ultimatum game with 79 students from different nationalities and find clear differences in behaviour between these groups. Murnighan and Saxon (1998) look at the behaviour of 331 children and find that generosity in the ultimatum game decreases with age. In a similar study with 310 children Harbaugh et al. (2003) find that, once on controls for size, generosity increases with age. Güth et al. (2003) run a newspaper experiments with the ultimatum game. With 1035 participants they find that the medium of participation, internet or email, has an effect on generosity. Güth et al. (2007) look at a three-person ultimatum game. With 5132 participants they find that fairness and rejection rates increase with age. Köhler et al. (2007) play an ultimatum game with a heterogeneous sample of 334 German adults. In their sample generosity increases with age and income. Bellemare et al. (2008) integrate experiments into an existing survey, the Dutch CentER panel. Their 1214 participants play either the the ultimatum or the dictator game. One finding is that generosity increases with age.

The trust game has been studied by Fehr and List (2004) who compare the behaviour of 126 students with that of 76 CEOs. CEOs turn out to be more trusting, more trustworthy, and punish less. Fehr et al. (2003) report data from a trust game with a randomly selected sample of 429 German households. Bellemare and Kröger (2007) compare behaviour in the trust game played by 100 students and 499 households of the CentER panel. They find a hump-shaped relation between age and trust, and a u-shaped relation between age and trustworthiness. Bornhorst et al. (2010) play a trust game with 110 Ph.D. students of different nationalities and find significant differences in trust and trustworthiness between different regions of origin. Sutter and Kocher (2007) study a trust game played by 662 participants from dif-

ferent age groups. They find a hump-shaped relation between age and trust and increasing trustworthiness with age.

Other games that have been studied with heterogeneous groups of participants include the beauty-contest game of Bosch-Domenech et al. (2002) or the prisoners' dilemma in the TV show "Friend or Foe" studied by List (2006).

All these studies show that non-students outside the lab and students within the lab behave differently. Pull (1999) and Selten (2000) claim that student participants in a lab environment react more clearly to subtle strategic details than non-students outside the lab. It is, hence, essential to compare behaviour of student participants with a more heterogeneous population. This is what we want to do in this paper.

Other limitations of laboratory experiments are due to the controlled or artificial situation in the lab (see Levitt and List, 2007). We concede that these limitations exist, but in this paper we do not deal with these limitations. We will use a very simple and abstract game in this paper. We concentrate on analysing the effect of stakes and the effects of the population.

Our workhorse in this paper is the Yes-No game. Quite often in life one has to decide whether to accept a proposal or not without knowing what exactly is offered. Examples are so-called experience goods whose quality is not known to customers or partnership proposals without knowing how reliable the partner(s) will be, as often in the case of employment, joint ventures, or spouse relationships. Studying situations where one has to accept or reject without knowing what exactly has been offered is therefore of enormous importance. In our experiment we assume that proposers suggest how to share a given positive monetary amount and responders decide whether to accept or not without knowing the proposal how to share.

To simplify the implementation in the newspaper experiment we use the strategy vector method and we announce that only 40 participants will actually be paid according to their choices. Each participant does not only decide for both pie sizes (100€ and 1000€) but also as a proposer and as a responder. Furthermore, we elicit first order expectations about the typical

behavior in the other role. Participants in the newspaper experiment either cut out the newspaper part, filled it out, and mailed it or used an internet platform which had the same structure and provided the same information as the page in the newspaper.

From ultimatum games (see Camerer, 2003, for a survey of ultimatum experiments), Yes-No games differ since responders in ultimatum games know what they accept or reject. Unlike to dictator experiments (e.g. Forsythe et al., 1993), the responder in Yes-No games still has full veto power in the sense that without his consent the pie of 100€ or 1000€ is lost.

Compared to former Yes-No experiments (Gehrig et al., 2007), our study differs in particular with respect to the group of participants: Gehrig et al. (2007) study the behaviour of 112 students in the lab[1] while we look at 871 readers of a Thuringian newspaper, the OTZ, and 128 executives of a large business company.[2] Participants could voluntarily reveal their age, gender and their profession.

Of course, we can not directly compare the results of the lab experiment of (Gehrig et al., 2007) with the results of our newspaper experiment since the two differ in several dimensions: the group of participants (students versus newspaper readers), the elicitation method (sequential versus strategy vector method), and the payment scheme (only 40 participants were selected in the newspaper study for payment). Some variables we can control, e.g. the age, gender and profession of the newspaper readers. But behavior may also be influenced by the way choices are elicited or by the payment scheme. In this paper we do not want to discuss these different methods, instead we want to describe properties of a heterogeneous population. We show results from the lab experiment next to the results from the newspaper experiment mainly as an illustration. Our comparisons and our tests will focus on the heterogeneity *within* the population of newspaper readers.

[1] We are grateful to Gehrig et al. (2007) for providing the raw data of their experiment. Here we refer only to what Gehrig et al. (2007) call their "first experiment series". The games in their "second experiment series" were embedded in a bidding mechanism which can not easily be compared to the game we study here.

[2] As usual the specifics of the company will not be revealed by us.

In section 2, we introduce the design of the newspaper experiment which essentially coincides with that of the experiment with business executives. Section 3 discusses some hypotheses which, in section 4, are tested with the help of the rather large data set (involving 871 participants from the newspaper experiment and 128 business executives, altogether 999 participants). Section 5 concludes.

2 Experiment

On Saturday, 6 September 2008, and on Saturday, 13 September, the Ostthüringer Zeitung (Gera, Germany) published the instructions to an experiment in their weekend supplement. A translation of the instructions can be found on our webpage `http://www.kirchkamp.de/ja-nein/`. These instructions also contained a link to a web page with essentially the same instructions and the same format. Furthermore, on Tuesday, 30 September, the newspaper published a note with the link to the web page. Readers of the newspaper knew that they could participate in the experiment either by mail or through the internet. They were also told that we would select 40 participants who would actually play the game.

The game can be described as follows:

- First, the 40 participants are randomly grouped into 20 pairs of two players.
- One of these two players will be the proposer in the Yes-No game (X-player), the other the responder (Y-player).
- A random draw decides for each pair the amount that is to be divided. For 10 pairs the amount is 100€, for the other 10 pairs the amount is 1000€.
- The X-player chooses a division. To simplify the evaluation of the questionnaires we only allow 10 divisions between 5€ and 95€ when 100€ could be divided. Similarly, 10 divisions between 50€ and 950€ were

Table 1 Participants of the newspaper experiment

platform	num. of participants	known sex [%]	females [%]	known age [%]	mean age [years]	median age [years]	known profession [%]	blue collar [%]
mail	303	100.0	57.1	96.0	49.8	50.0	65.0	27.4
internet	568	100.0	44.5	82.2	39.0	40.0	61.4	13.5
all	871	100.0	48.9	87.0	43.2	43.0	62.7	18.5

allowed when 1000€ could be divided (see also the translation of the instructions on our webpage `http://www.kirchkamp.de/ja-nein/`).

- Simultaneously, the Y-player chooses "yes" or "no".
- In case of "yes", the amount is divided according to the proposal of the X-player. In case of "no" both players receive zero.

We used the strategy vector method, i.e. all participants submitted strategies and expectations for both amounts (100€ and 1000 €) and for both positions in the game (X and Y).

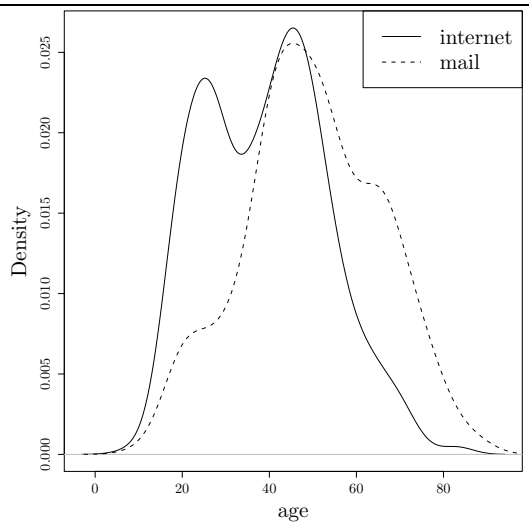
Table 1 shows characteristics of participants from the newspapers experiment. Some participants did not reveal their age or their profession. The table, hence, shows proportions only for those participants who revealed these properties. Figure 1 shows the estimated density of age in our sample.

Essentially the same method was also used in June 2008 to elicit choices of 128 business executives. For this subset of the data we have no information about age and sex.

3 Hypotheses

A materially opportunistic responder should accept the unknown but necessarily positive offer. Anticipating such opportunism, an equally opportunistic proposer should offer the lowest possible amount. We, however, do expect only few participants to behave in line with such common opportunism.

Figure 1 Age of participants in the newspaper experiment



Whether "stakes" matter is often explored by using the same stakes in rich and poor countries, i.e., stake variation relies on large discrepancies of living conditions (see, e.g. Cameron, 1999). The possible disadvantage of confounding "stake" and culture is avoided by stake variation in our within-subjects design: the very same participants decide for a small (100€) and a much larger (1000€) pie. But which stake effects do we expect?

Hypothesis S: "Stakes"

1. (stake independent offers) In view of the stake independence observed for ultimatum games (Hoffman et al., 1996) we expect the relative shares, offered by proposers for both pie sizes, to be similar.

2. (stake dependent rejections) Although game theoretically (assuming non monetary payoff maximisation) responders should accept, some responder participants will use the small pie (100€) to "teach fairness to proposers", by rejecting in case of small pie while accepting in case of the large one (1000€).

3. (low rejection rates) In line with Gehrig et al. (2007), we expect rejection rates close to zero. Rejection rates should be clearly lower than in standard ultimatum games.

Although Gehrig et al. (2007) use a different elicitation method, we hope to confirm the following hypothesis:

Hypothesis ROB: "Robustness"

At least for the large pie (1000€) the results do not differ much between newspaper participants, business executives in our study and student participants in Gehrig et al. (2007). With respect to socio-demographic variables, we will test

Hypothesis SD: "Socio-Demographics"

1. As in Eckel and Grossman (1998) and Eckel and Grossman (2001) we expect women to be more generous than men.
2. As in Harbaugh et al. (2003), Güth et al. (2007), Köhler et al. (2007), Bellemare et al. (2008) we expect that older participants offer more as proposers and reject less as responders.
3. In line with Güth et al. (2003) we expect "more fairness in the mail than in the internet".

Hypothesis E: "Expectations"

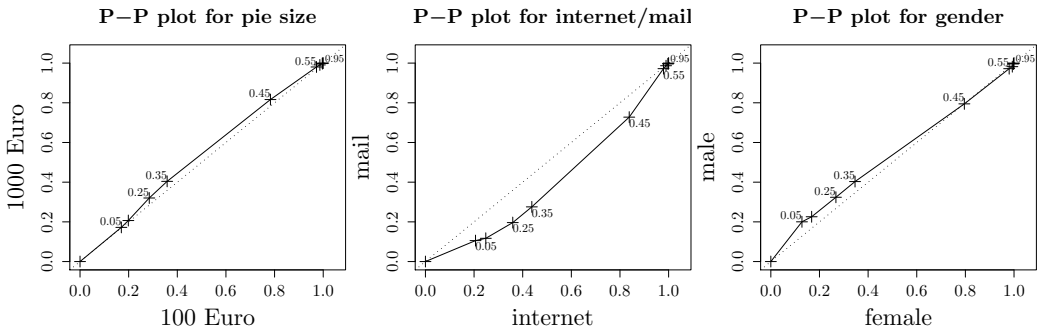
We expect expectation data to be more reactive to individual heterogeneity.

4 Results

We present the results by investigating whether they confirm the hypotheses stated above.[3]

[3]The data in this paper was analysed with R version 2.11.0 (2010-04-22) (R Development Core Team, 2010).

Figure 2 Percentile-percentile plots for offers



Numbers next to the points denote relative offers. Figure 6 in the appendix provides the same data as histograms. The dotted line is the 45°-line.

Hypothesis S-1 (stake independent offers): Since the effect of the pie size is small, histograms are not the most efficient way to compare empirical frequencies. Instead, we use (similar to Forsythe et al., 1993) a percentile-percentile plot.[4] We see that for both amounts, 100€ and 1000€, the majority of players offers less than one half of the pie to the other player. The P-P line is (slightly) above the 45°-line, i.e. smaller offers are (slightly) more frequent for larger amounts. For an amount of 100€ the average offered share is 0.375 of the entire amount, for an amount of 1000€ the average offered share is with 0.361 slightly smaller. This difference is small, but significant[5] which contradicts hypothesis S-1.

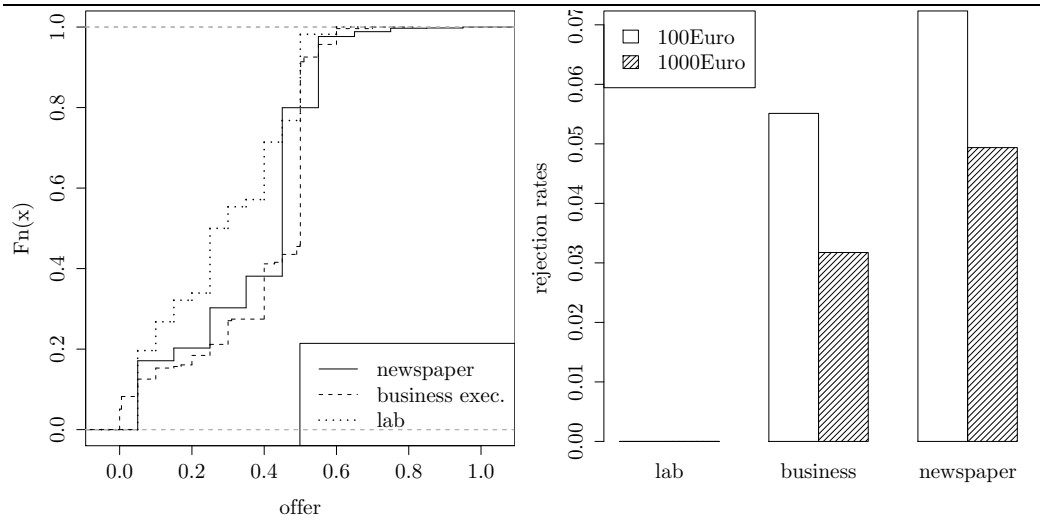
S-2 (stake dependent rejections) When the amount is 100€ then 7.2% of all participants reject, whereas for an amount of 1000€ only 4.9% of all participants reject. The difference is significant.[6] This is in line with hypothesis S-2.

[4]We show the same data as histograms in figure 6 in the appendix. Forsythe et al. (1993) call the plot in figure 2 a quantile-quantile plot. Since the axes show percentiles we call it percentile-percentile here. Each dot corresponds to the empirical percentiles of one of the 10 possible offers. Two identical distributions would be represented by a 45°-line.

[5]An exact (Streitberg and Röhmel) paired Wilcoxon test yields a p-value of 0.00001, a paired t-test yields a p-value of 0.00003.

[6]A one-sided Fisher's exact test for independence yields a p-value of 0.0282.

Figure 3 Results from different experiments

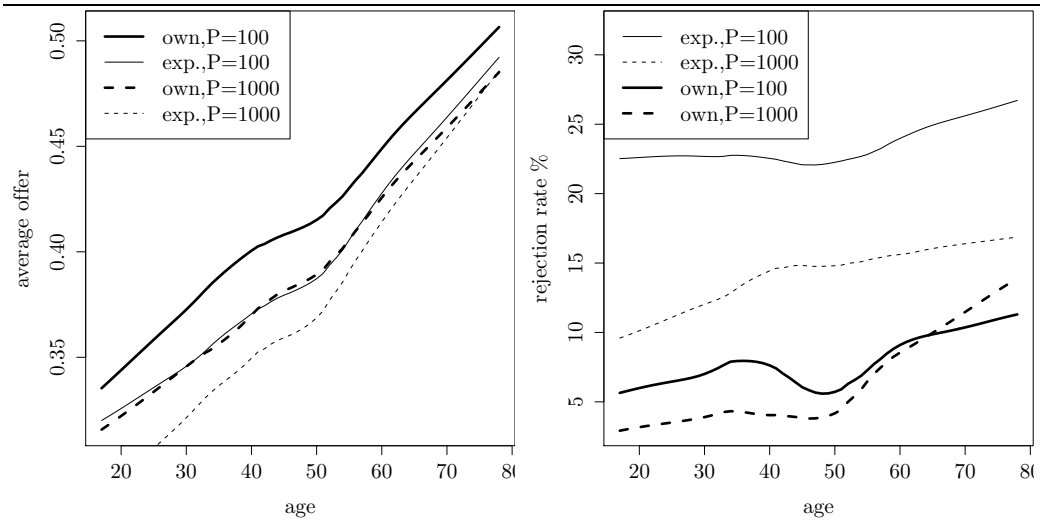


S-3 (low rejection rates) In our experiment we find rejection rates of 7.2% and 4.9% for amounts of 100€ and 1000€, respectively. Hence, the rejection rates we find for the Yes-No game are much larger than those in the Yes-No game of Gehrig et al. (2007) and pretty much in the same range as rejection rates in laboratory ultimatum games.[7] This is not in line with hypothesis S-3.

ROB (robustness) The left graph in figure 3 shows the empirical cumulative distribution of offers for different experiments. The solid line shows offers in a newspaper experiment, the dashed line shows results from a very similar experiment with business executives (conducted in June 2008), the dotted line shows offers from the lab experiment by Gehrig et al. (2007). One main observation from the lab is confirmed by newspaper readers and business executives: most offers are larger than the game theoretically predicted one (the smallest positive offer). However, we also see that relative offers are clearly smaller for the student population in the lab than for readers of the

[7] E.g., the rejection rate is 9.52% for players who play the ultimatum game for the first time in Güth et al. (1982), it is 2.81% in the ultimatum game of Gehrig et al. (2007), it is 5.97% in the ultimatum game with payment of Forsythe et al. (1993), and (depending on the size of the pie) between 13% and 18.5% in Hoffman et al. (1996).

Figure 4 Average offers and rejection rates



Graphs show own choices (own) and expected choices (exp.) of the other player. The lines are lowess-splines based on R's `plsmo` function.

newspaper or business executives.[8]

The graph on the right side in figure 3 shows the empirical relative frequency of rejections. We see that the pattern is similar in the newspaper experiment and with business executives: with larger stakes participants are more cautious. Business executives are generally more cautious anyway. We also see that rejection rates for the Yes-No game in the lab are (for the smaller pie) significantly smaller than those for business executives,[9] they are also smaller than those for newspaper readers.[10] We do not find significant differences in rejection rates between business executives and newspaper readers.[11]

SD (offers) The left part of figure 4 shows how average offers depend on the age group. We see that for both amounts, 100€ and 1000€, and also for actual decisions as well as for expected decisions, the offer increases with age. This is in line with what Harbaugh et al. (2003), Güth et al. (2007),

[8]Exact Wilcoxon test: $p < 0.0001$.

[9]Fisher's exact test $p = 0.0155$ and $p = 0.124$ for 100€ and 1000€ respectively.

[10]$p = 0.00069$ and $p = 0.0111$ for 100€ and 1000€ respectively.

[11]$p = 0.579$ and $p = 0.502$ for 100€ and 1000€ respectively.

Table 2 Random effects estimation of equation 1

	β	σ	t	p value	95% conf	interval
1	0.326	0.0183	17.8	0.0000	0.29	0.362
1000€	-0.0158	0.00775	-2.04	0.0419	-0.031	-0.000583
age	0.00209	0.000312	6.67	0.0000	0.00147	0.0027
internet	-0.0469	0.00993	-4.73	0.0000	-0.0664	-0.0274
♂	-0.0205	0.00913	-2.25	0.0248	-0.0385	-0.00261

Confidence intervals and p-values are based on a parametric bootstrap with 1000 replications. The estimations in tables 2– 5 do not include data on business executives, since there we neither know age nor gender.

Köhler et al. (2007), Bellemare et al. (2008) find for ultimatum games.

More formally, we estimate the following random effects model:

$$\frac{\text{offer}}{\text{amount}} = \beta_1 \cdot d_1 + \beta_{1000€} \cdot d_{1000€} + \beta_{\text{age}} \cdot \text{age} + \beta_{\text{internet}} \cdot d_{\text{internet}} + \beta_{♂} \cdot d_{♂} + \nu_i + \epsilon_{ik} \quad (1)$$

where $d_{1000€}$ is a dummy which is one if the amount is 1000€ and zero otherwise, d_{internet} is a dummy that is one for participants who submitted their strategy through the internet and zero otherwise, and $d_{♂}$ is a dummy that is one for male participants and zero otherwise. ν_i is a random effect for each participant and ϵ_{ik} is a random effect for the individual decision. Results are shown in table 2.[12] Alternative models where age enters as a polynomial of second or higher degree do not lead to a significant change in the estimation results. Also, when we add a dummy for white collar workers to equation (1) or to the following equations (2) and (4) we do not find a significant effect nor a substantial change in the estimated coefficients.

Estimation results in table 2 confirm what we see in figure 4: offers increase significantly with age. We also confirm what we see in figure 2: Male participants offer significantly less than female participants. This is in line with Eckel and Grossman (1998) and Eckel and Grossman (2001) who find

[12]All the estimations shown in tables 2–5 exclude the data from business executives, since there we neither know age nor gender.

men to be less generous in dictator experiments and ultimatum games.[13]

The effect of the medium of participation, internet or mail (see also the diagram in the center of figure 2), is highly significant. Even when we control for age, offers on the internet are significantly smaller and closer to the game theoretic solution behavior based on material opportunism. This finding is in line with Güth et al. (2003) who also observe "more fairness in the mail than in the internet".

Increasing the pie size by a factor of ten decreases the relative offer by -1.58%. This is only a small amount and here (where, due to missing answers about gender and age, we can only use 753 observations) only weakly significant. Note, that with all 999 observations we found a significant effect in our discussion of hypothesis S-2 (see also the left diagram in figure 2).

We find it interesting to note that, once we control for age, the offers in our newspaper experiment do not differ too much from the average offer in the lab (29%). The prediction we obtain for newspaper readers from equation (1) is with 29.4% ($\mathrm{CI}_{95} = [27.7, 31.1]$) only slightly larger.[14]

SD (rejection behaviour) The right part of figure 4 shows the relation between rejection rates and age. To measure this effect we estimate the rejection probability as a logistic function of age and other explanatory variables. Since we see in the left part of figure 4 that older people have more pessimistic expectations than younger people we include expectations as an explanatory variable in the following random effects model:

$$P(\text{reject}) = \mathcal{L}\Big(\beta_1 \cdot d_1 + \beta_{1000€} \cdot d_{1000€} + \beta_{o^E} \cdot o^E + \beta_{\text{age}} \cdot \text{age} + \beta_{\text{internet}} \cdot d_{\text{internet}} + \beta_{♂} \cdot d_{♂} + \nu_i\Big) \quad (2)$$

$\mathcal{L}$ is the standard logistic function and o^E is the expected relative offer. Estimation results can be found in table 3. There are fewer significant effects

[13] More basically, one can control for the idiosyncratic testosterone level of male participants (see Burnham, 2007).

[14] We assume age to be 20, gender to be half male and half female, large stakes, and that computerised experiments in the lab are similar to using the internet in the field.

Table 3 Random effects estimation of equation 2

	β	σ	z	p value	95% conf	interval
1	-8.09	2.54	-3.19	0.0014	-13.1	-3.11
1000€	-1.03	0.368	-2.79	0.0053	-1.75	-0.305
o^E	-1.12	1.85	-0.606	0.5443	-4.74	2.5
age	0.0327	0.0406	0.806	0.4201	-0.0469	0.112
internet	-0.351	1.39	-0.252	0.8009	-3.08	2.38
♂	0.264	1.3	0.202	0.8396	-2.29	2.82

than in equation (1). This is not surprising, since most offers are accepted anyway (we have 1413 accepted and only 93 rejected offers). The small number of rejection decisions does not yield the variance needed for highly significant results. The only significant factor is the amount: Participants are significantly less likely to reject a share of a large (1000€) pie than a share of a small (100€) pie.

What looks like increasing stubbornness of the elderly in figure 4 turns out to be insignificant in the estimation.

Eckel and Grossman (2001) find fewer rejections by women in ultimatum games. In our study we find with 0.264 a positive coefficient for $d_{♂}$, i.e. males are, indeed, more likely to reject, but this effect is not significant.

E (expectations) Similar to equation (1) we explain expected relative offers o^E:

$$o^E = \beta_1 \cdot d_1 + \beta_{1000€} \cdot d_{1000€} + \beta_{\text{age}} \cdot \text{age} + \beta_{\text{internet}} \cdot d_{\text{internet}} + \beta_{♂} \cdot d_{♂} + \nu_i + \epsilon_{ik} \quad (3)$$

Results are shown in table 4. We see that coefficients are very similar to those estimated in equation 2). Also the standard deviation of residuals is very similar,[15] i.e. unlike expected in hypothesis E offers are neither more reactive to individual heterogeneity in a dimension that we observe (age, gender,

[15]The standard deviation of residuals is 0.0536 for equation (1) and 0.0584 for equation (3).

Table 4 Random effects estimation of equation 3 for expected offers

	β	σ	t	p value	95% conf	interval
1	0.33	0.0183	18	0.0000	0.294	0.365
1000€	-0.0151	0.00805	-1.88	0.0605	-0.031	0.000671
age	0.00157	0.00032	4.91	0.0000	0.000942	0.0022
internet	-0.0464	0.0101	-4.58	0.0000	-0.0663	-0.0265
♂	0.0048	0.00935	0.513	0.6079	-0.0136	0.0232

Confidence intervals and p-values are based on a parametric bootstrap with 1000 replications.

Table 5 Random effects estimation of equation 2 for expected rejection rates

	β	σ	z	p value	95% conf	interval
1	-1.33	0.247	-5.39	0.0000	-1.82	-0.848
1000€	-0.368	0.127	-2.89	0.0038	-0.618	-0.119
age	0.0056	0.00435	1.29	0.1979	-0.00292	0.0141
internet	0.0357	0.139	0.257	0.7975	-0.237	0.308
♂	-0.163	0.129	-1.26	0.2071	-0.415	0.09

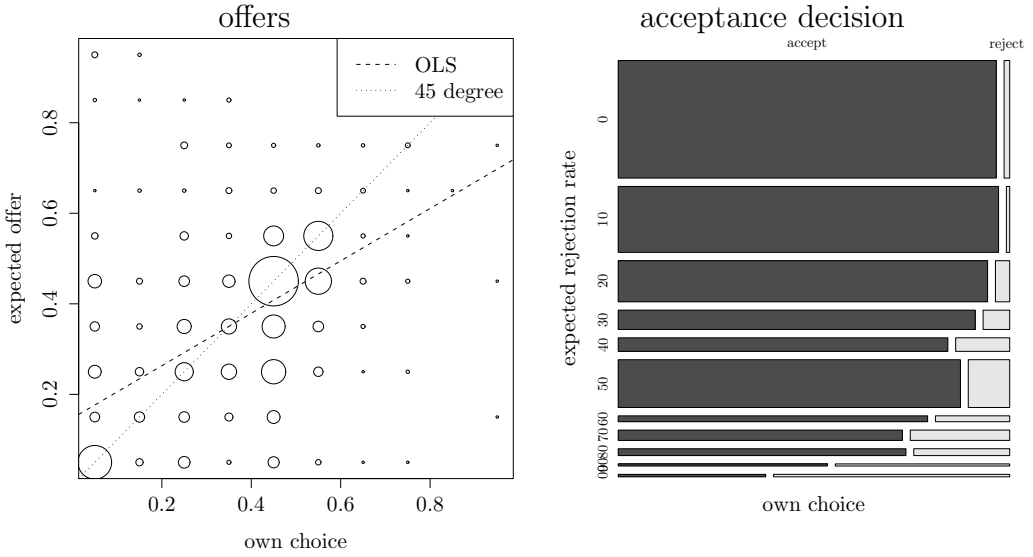
medium of participation) nor in a dimension we do not observe (measured as residuals). Only the (weakly significant) gender effect that we found for actual offers disappears for expected offers. Men expect the same offers as women.

As in equation (2) we explain expected rejection rates (elicited as "expected average frequencies" in the experiment) with a logistic regression with random effects. Since the rejection decision of another person can not depend on the own expectation equation 4 does not contain the expected offer o^E.

$$P(\text{reject}) = \mathcal{L}\Big(\beta_1 \cdot d_1 + \beta_{1000€} \cdot d_{1000€} + \beta_{\text{age}} \cdot \text{age} + \beta_{\text{internet}} \cdot d_{\text{internet}} + \beta_{♂} \cdot d_{♂} + \nu_i\Big) \quad (4)$$

We show results in table 5. As in the comparison of equation (1) and (3), also estimation results for equations (2) and (4) are not too different from each other. Correctly, participants expect smaller rejection rates when stakes are larger.

Figure 5 Own choices and expectations in the newspaper experiment



The area of the circles (in the left diagram) and the area of the rectangles (in the right diagram) is proportional to the number of observations. The dashed line shows an OLS regression of expectations on offers.

Let us next have a look at consistency of expectations. Do participants who make generous offers also expect those offers? And are participants who expect frequent rejections more likely to reject themselves? The answer to both questions is "yes". The left graph in Figure 5 shows a bubbleplot of expectations over offers. We clearly see that on the individual level offers and expectations are correlated. Participants who make small offers expect others to make small offers, too. Participants who are generous expect other to be generous as well. The right graph in figure 5 shows a mosaicplot of actual and expected rejection decisions. Again, we find that expectations are in line with choices. Participants who expect a rejection rate of 0 do not reject themselves. There are not many participants who expect a high rejection rate but those who do will reject rather frequently.

5 Conclusion

Lab research is often questioned by arguing that the stakes are minor, student participants are not representative, and experimental games are far too

abstract. All three issues concern the external validity of typical lab research in experimental economics.

In this paper we do not discuss the third issue, whether experimental games are abstract. We concentrate on the first two issues, i.e. whether student participants are representative and whether the size of the pie matters. All what we can say to defend our choice of game, the Yes-No-game, is that it is simple enough to be understood by reasonably educated newspaper readers and executives and that it captures some important aspect of life, namely the need to accept or reject some deal whose profitability has already been determined or manipulated but is not known to the responder.

With respect to the first issue, size of stakes, we have explored stake dependence by quite high pie sizes of 100€ and 1000€. Of course, the random selection of only 40 participants questions the stake size. Still, there is little evidence for random payment effects (see, for instance Cubitt et al., 1998). In contrast to Hoffman et al. (1996), who find no significant effect of the stake size for the ultimatum game, we find for the Yes-No game a significant effect of stakes on the aggregate level. Even in the estimation of equation (1), where we include only a subset of the data, the effect remains weakly significant. Of course, the effect is small and is only significant due to the large number (999) of our participants. Still, this raises the possibility that also for other games the impact of stakes could be close to zero, but still different from zero. With a limited number of participants in the lab small effects might be harder to detect than with a large number of newspaper readers.

With respect to the second point, representativeness of the student population, we agree that students belong to a rather narrow age bracket. We have found three important socio-demographic variables: age, gender, and the medium of participation (mail vs. internet). Here our observations confirm most of our expectations. In line with Eckel and Grossman (1998, 2001) women are more generous. Similar to (Güth et al., 2003) we find "fairness in the mail and more material opportunism in the internet". Consistent with Harbaugh et al. (2003), Güth et al. (2007), Köhler et al. (2007), Bellemare et al. (2008) we find older people to be more generous.

In particular we find that—once we control for age—our results from the field are surprisingly consistent with behaviour in the lab. Newspaper readers with the age of laboratory participants behave as if they were laboratory participants. Older participants are more generous and reject more frequently. The newspaper study can, thus, be seen as a natural extension of the laboratory. It allows us to easily access a more heterogeneous subject pool.

Ideally, we should know for each laboratory experiment how sensitive it is to changes in parameters. Here we have studied a single game and we have mostly confirmed our expectations that where, however, based only on a very small number of related studies. This does not mean that we can stop here and continue to solely experiment in the lab. There are many games which have so far only been researched in the lab. We have to be careful when extrapolating from these laboratory results to the field. We have seen in this paper that newspaper experiments can help us to learn more about external validity since the group of participants is more heterogeneous than the group of participants in the laboratory. The group of participants in a newspaper experiments is also typically larger than the group of participants in the laboratory. This allows us to detect small, but perhaps essential, effects which are easy to overlook in the lab.

References

Charles Bellemare and Sabine Kröger. On representative social capital. *European Economic Review*, 51(1):183 – 202, 2007. ISSN 0014-2921. doi: DOI:10.1016/j.euroecorev.2006.03.006. URL ↦.

Charles Bellemare, Sabine Kröger, and Arthur van Soest. Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica*, 76(4):815–839, 07 2008. URL ↦.

Fabian Bornhorst, Andrea Ichino, Oliver Kirchkamp, and Eyal Winter Karl H. Schlag. Similarities and differences when building trust: the role of cultures. *Experimental Economics*, 2010. forthcoming.

Antoni Bosch-Domenech, Jose G. Montalvo, Rosemarie Nagel, and Albert Satorra. One, two, (three), infinity,... and lab beauty-contest experiments. *The American Economic Review*, 92(5):1687–1701, 2002.

Terence C. Burnham. High-testosterone men reject low ultimatum game offers. *Proceedings of the Royal Society B*, 274:2327–2330, July 2007.

Colin F. Camerer. *Behavioral game theory*. Princeton University Press, New York, 2003.

Lisa A Cameron. Raising the stakes in the ultimatum game: Experimental evidence from indonesia. *Economic Inquiry*, 37(1):47–59, January 1999. URL ↦.

Robin P. Cubitt, Chris Starmer, and Robert Sugden. On the validity of the random lottery incentive system. *Experimental Economics*, 1(2):115–131, September 1998.

Catherine C. Eckel and Philip J. Grossman. Are women less selfish than men?: Evidence from dictator experiments. *The Economic Journal*, 108: 726–735, 1998.

Catherine C. Eckel and Philip J. Grossman. Chivalry and solidarity in ultimatum games. *Economic Inquiry*, 39(2):171–188, 2001.

Ernst Fehr and John A. List. The hidden costs and returns of incentives — trust and trustworthiness among ceos. *Journal of the European Economic Association*, pages 743–771, 2004.

Ernst Fehr, Urs Fischbacher, Bernhard von Rosenbladt, Juergen Schupp, and Gert G. Wagner. A nation-wide laboratory. examining trust and trustworthiness by integrating behavioral experiments into representative survey. CESifo Working Paper Series CESifo Working Paper No., CESifo GmbH, 2003. URL ↦.

R. Forsythe, J. L. Horowitz, N. E. Savin, and M. Sefton. The statistical analysis of experiments with simple bargaining games. *Games and Economic Behavior*, 1993. Forthcoming.

Thomas Gehrig, Werner Güth, Vittoria Levati, René Levínský, Axel Ockenfels, Tobias Uske, and Torsten Weiland. Buying a pig in a poke: An experimental study of unconditional veto power. *Journal of Economic Psychology*, 28(6):692 – 703, 2007. ISSN 0167-4870. doi: DOI: 10.1016/j.joep.2007.06.005.

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

Werner Güth, Carsten Schmidt, and Matthias Sutter. Fairness in the mail and opportunism in the internet: A newspaper experiment on ultimatum bargaining. *German Economic Review*, 4(2):243–265, 2003.

Werner Güth, Carsten Schmidt, and Matthias Sutter. Bargaining outside the lab - a newspaper experiment of a three-person ultimatum game. *The Economic Journal*, 117:449–469, 2007.

William T. Harbaugh, Kate Krause, and Jr. Steven G. Liday. Bargaining by children. Working paper, Economics Department, University of Oregon, 2003.

Elizabeth Hoffman, Kevin A. McCabe, and Vernon L. Smith. On expectations and the monetary stakes in ultimatum games. *International Journal of Game Theory*, 25:289–301, 1996.

Christoph Köhler, Wiebke Kuklys, Olaf Struck, and Sven Fischer. Ultimatum behavior, individual characteristics, and moral attitudes on the labor market. Arbeitspapier 15, SFB 580, University of Jena, 11 2007.

Steven D. Levitt and John A. List. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2):153–174, Spring 2007. URL ↦.

John A. List. Friend or foe? a natural experiment of the prisoner's dilemma. *The Review of Economics and Statistics*, pages 463–471, 2006.

J.Keith Murnighan and Michael Scott Saxon. Ultimatum bargaining by children and adults. *Journal of Economic Psychology*, 19:415–445, 1998.

Kerstin Pull. What is the fair wage?-a model of as-if-co-operation. Discussion paper 58, Quint-Essenzen, 1999.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. URL ↦. ISBN 3-900051-07-0.

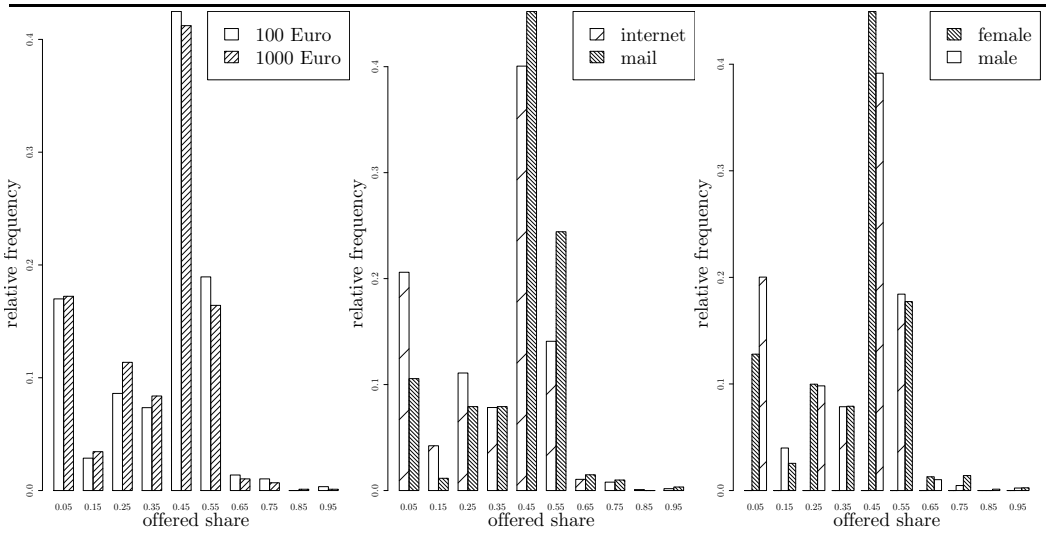
Alvin E. Roth, Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir. Bargaining and market power in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *American Economic Review*, 81:1068–1095, 1991.

Reinhard Selten. Thünen-Vorlesung: Eingeschränkte Rationalität und ökonomische Motivation. *Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 9 (special issue):129–157, 2000.

Matthias Sutter and Martin G. Kocher. Trust and trustworthiness across different age groups. *Games and Economic Behavior*, 59:364–382, 2007.

Appendix

Figure 6 Histograms for offers



These graphs show the same empirical frequencies as figure 2.