

Assessment of Vojtech Bartos' PhD Dissertation: "Essays in Behavioral and Development Economics."

By Peter Martinsson

July 13, 2016

Main assessment

My assessment of Bartos' PhD Thesis is based on the requirements in the "CERGE-EI Handbook for Graduate Students". This dissertation is composed of a short introduction followed by three chapters. The first two chapters may be regarded as a manuscript for a journal article submission while the last chapter is forthcoming in American Economic Review. Chapter 1 is solo-authored and chapter 2 and 3 are co-authored, where only the latter has, to my understanding, a co-author from the Dissertation Committee Chair. Thus, the dissertation fulfills the requirements of consisting of three chapters, at most one of the chapters is co-authored with someone from the student's Dissertation Committee Chair and at least one paper is solo-authored. To the best of my knowledge it is independent work by Bartos that has not previously been used to obtain PhD or any other academic degree elsewhere. Below I will turn to a detailed assessment of each chapter. It follows as a conclusion, that the dissertation is an original and substantive contribution to knowledge and is of a quality that warrants publication in scholarly journals, which is already confirmed by one chapter being forthcoming in American Economic Review. It should also be noted that conducting experiments in rural Afghanistan is impressive. My evaluation of the thesis is acceptable, using the scale provided in the "CERGE-EI Handbook for Graduate Students" (acceptable, unacceptable, or acceptable with revisions). In my view Bartos' thesis very well satisfies formal and content requirements for a PhD thesis in economics and I would recommend it for a defense.

Details on the Three Chapters

1. "Seasonal Scarcity and Sharing Norms"

This chapter experimentally investigates if and in that case how scarcity affects people's willingness both to share and to enforce sharing from others. A lab-in-the-field experiment is conducted using a subject pool of Afghan subsistence farmers during a lean and a post-harvest season to test for the effect of seasonal scarcity. By using both a dictator game and a third party punishment game, the effect of enforcement of sharing norms is obtained. The main results show that giving in the dictator game is stable over time and that the enforcement of sharing norms is weaker during the lean season.

Main comments

Overall it is a nice chapter and the experiment is well designed, executed and analyzed. My main comments relate to additional information that I think should be incorporated and some robustness checks. An experiment has its strength of controlling for causality, but the drawback might be correlation with real life behavior. Although an effect is found in the experiment but does it exist in real life, i.e., to what extent can we generalize the findings? There are two issues to consider: (i) counter punishment and (ii) covariant risk. We know from previous work, especially work by Nikiforakis, that the threat of counter-punishment has a negative effect on the willingness of third parties to punish. In farming communities with covariate risk, i.e., people are more or less equally hit during adverse weather; the need will differ since not all households are equally bad off. Thus, those who potentially could share and those who are in need might differ between seasons. I think the discussion of the chapter should reflect on this issue and the discussion related to policy makers to be strengthened. How would for example provision of formal insurance change the situation, and would it be feasible in rural Afghanistan?

The villages are perfectly correlated with religious denomination (Sunni and Shia Muslims) and therefore religious denomination is not investigated. The village effect reported in Table A1 seems to be small if looking for stars indicating significant variables. This is however an effect of chosen reference group. If for example Koche Aghaz would have been chosen then more significant effects would have been reported. There are pretty large differences between

villages that should be discussed and also if there is a correlation between religious denomination and villages. Is the prediction that Sunni and Shia should be similar or different in their behavior? An important issue related to the denomination is the denomination of the helpers. Were the helpers the same in all sessions and were they Shia or Sunni Muslims? In relation to the experimental set-up, potential experimental demand effect should be discussed, especially when people outside the community conduct the experiment and potentially from a different religious denomination. This brings in another issue namely post-experimental effect. This could both mean that the sender in the dictator game will share with a third person after the game who did not participate as well as the potential perceived threat, or actual threat, after the experiment when people would like to find out who gave so little etc. I know these are notoriously difficult questions, and I don't ask for an answer, but more a reflection, and what was done to reduce these effects.

A priori, order effects cannot be ruled out, i.e., that subjects try to behave consistently in the experiments conducted in the lean season and post-harvest season. Since 84 of the original 291 subjects did not participate in the second experiment, it is possible to test for order effect by comparing the behavior of the 207 who already participated in the first experiment with the 82 new recruited subjects. I am not surprised that attrition happens and this was largely explained by migration. I think the discussion of this issue in 1.5.4 should be done in 1.4.1 since this would make the results discussion more convincing. An important issue, however, is how the replacement was selected. Were they selected in a way that if a person with some specific observable characteristics in village B did not participate instead a "twin" of this person was selected from the same village, or was selection made such that number of subjects was the same in each village? The selection strategy of the new subjects needs to be discussed and motivated. If it is a "twin" matching, then the behavior of these two subjects could be matched as a robustness check. The non-parametric test and regressions in 1.4.1 is based on using the whole sample. In terms of regression, it is an unbalanced panel since for some individuals there are two observations while for others there are one observation (same logic applies to the non-parametric test). If missing observations (attrition) are not random, then the regression results might be biased. At least, this issue should be discussed.

Why do people give in a dictator game? One reason to give is based on the perceived need of the other person, is there any information on this? To use income among farmers in a

developing country is difficult and here cash income during last 30 days is used. Is this cash income to household or income per equivalence scaled household member? I think choice of income variable should be discussed, and also consider that their main source of food are, I presume, from their own supplies and hence is not included in income. The link to social preference models in section 1.5.2. is rather weak. For example, in the Fehr-Schmidt model if we assume that beta is less than alpha (more disutility from earning less than more than others), then an individual would give zero if beta is less than 0.5 and otherwise 5. In other words, this prediction can be easily tested although it is confounded with other explanations. It is discussed in 1.5.1 that the drop in prosocial behavior is not caused by changes in individual preferences but caused by weaker social norm enforcement. I would like a more elaborative discussion on “changes in individual preferences” since this is fundamental for how we model individual behavior. I would think of our preferences to be robust but the changes in our environment causes us to behave differently. The key reference given when describing the dictator game is Kahneman et al. (1986). It should be noted that they only gave the dictator two options either a 18 USD and 2 USD split OR a 10 USD and 10 USD split of the total pie of 20 USD. This is not the case in the experiment in the paper where the dictator can give any integer number between 0 and 10 to the receiver. In the third-party punishment, the punisher does not observe the actual giving but is asked what to do if dictator the money 10-0, 9-1 etc., i.e., using the strategy method. This way of eliciting behavior is sometimes described as “cold” compared to the case where the punisher that the dictator divided for example 8-2, how much would you like to punisher. There is a discussion about “hot and cold”, which should be discussed both in selection of this specific experimental design as well as when interpreting the results (an early reference on this topic is Brandt and Charness (2000) in *Experimental Economics*).

Minor comments

I would consider to rescale some of the variables. For example in Table 1.A.3. for variables age: 0.01** 0.00** -0.00 -0.00** -0.00** -0.00*

(0.00) (0.00) (0.00) (0.00) (0.00) (0.00)

In some cases variable is 0.00 and insignificant while in other case -0.00 and significant at 5% level. It is hard to evaluate the marginal effect in an ordered probit per se since it is, due to non-linearity, evaluated at mean but also to understand the economic significance of a variable that is reported to have a value of 0.00.

Clearly state number of subjects who are assigned the role as PA, PB and PC and this should be shown in a table. In appendix this information should be provided by village.

It would be helpful to show where in the text the tables (and figures) will be located by writing "Table X here", or even better paste the table to its location in the chapter.

Specify what α is in equation (1.1), which for example includes dummy variables for villages.

2. "Contract Enforcement and Trustworthiness Across Ethnic Groups: Experimental Evidence from Northern Afghanistan"

The paper uses a la-in-the-field experiment in Afghanistan to investigate how a formal institution affects trust and trustworthiness based on a design developed by Fehr and Rockenbach (2003). The specific focus is how people from same and different groups (in this case Sunni and Shia Muslims) respond to the formal institutions in an environment which mostly lack formal institutions. The key finding is that formal sanctioning does not have an effect on money sent back when the pair comes from same ethnic group but it does when from different ethnic groups.

Main comments

The paper is nicely written and the experiment is carefully designed including testing for ordering effect. The analyses are thorough done and implications are discussed. My main comments relate to choice of theoretical model and better motivation of the groups included in the experiment to create an out-group. The theoretical model chosen to describe trustee's behavior is based on Bowles and Polania-Reyes (2012). Model choice is to some extent a matter of taste but the choice for this model needs to be better motivated. To some extent the decision by trustee does not differ much from the sender in a dictator game, just that the endowment for the trustee is based on the investor's decision. In chapter 1, it was argued that Fehr and Schmidt model of inequality aversion and Falk and Fischbacher's model of reciprocity can explain behavior (these models are briefly mentioned in 2.4.3, but I think these models and behavior predicted by these models need to be better integrated in the chapter). I understand that a model is needed to be chosen, but at least the model choice should be discussed and motivated.

The paper investigates differences in behavior towards in-group and out-group represented by Sunni and Shia Muslims in Afghanistan. I am surprised that the investors state that others can be trusted more in outgroup than in in-group as shown in Table 2.1. (is this true if separated by denomination?). As a reader with little or no knowledge about similarities and differences between these groups, what would I expect? Do readers generally know the differences between Sunni and Shia? Why do the people from these groups consider each other as out-groups, and do they do it to the same extent? Why is it of interest to investigate in-group and out-group between these groups? How do these groups interact in their daily life – market places etc. although they live in different villages? Have there been any conflicts between the groups recently? These are important pieces of information for the reader to put the key finding that formal sanctioning does not have an effect on money sent back when the pair comes from same ethnic group but it does when from different ethnic groups in perspective.

Minor comments

In general, it is difficult to read the effect from regression tables since there are many dummy variables and interaction effects. However, the tests if sanctioning matters for in-group and out-group respectively are clear. In the same panel of the tables, I would also like to see if there are differences between outgroups and in-groups given sanctioning and no sanctioning, respectively. Moreover, a figure presenting the data would facilitate the understanding of data, for example showing transfer and required back transfer etc. There might be several observation for a specific combination, and this could be shown in the figure using a “ball” where size is related to number of observations.

How is order effect consider in the regression model in 2.6 (is it in E?), and in the analyses? How was income collected and is the income used equivalence scaled?

Tests in Table 2.1 should be non-parametric instead of t-tests. I would like Table 2.1 to be shown by religious denomination.

There are a number of trust experiments investigating in-group and out-group and I think these studies could be mentioned (e.g., Falk and Zehnder, 2013 J of Pub Ec (different parts of Zurich); and Johansson-Stenman et al, 2009 *Economica* (Muslims and Hindus in Bangladesh)).

When reporting significance level, is it true that it is $p=0.00$, or is it $p<0.01$ or $p<0.001$?

Were the helpers the same in all sessions and were they Shia or Sunni Muslims?

Motivate and discuss the pros and cons of using the strategy method.

3. "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition"

This chapter reports a field experiment on information acquisition in the labor and housing market in two countries. In the labor market experiment the focus is on employers' effort to inspect resumes of applicants for a job interview while in the housing market experiment on landlords' review of resumes for apartment viewing. By signaling if the applicant belongs to a minority or majority group in the country, the paper investigates if discrimination occurs already at the stage of information acquisition. The finding from labor market is that employers' are less likely to read resumes from applicants belonging to minority while in the housing market landlords are more likely to read resumes from applicants belong to minority. It is a very nice and well written paper, which has already been accepted paper for American Economic Review. Congratulations! Thus due to copyright infringement issues it cannot be changed as discussed in the "CERGE-EI Handbook for Graduate Students", and therefore I do not have any comments except that I enjoyed reading the paper.