University of Warsaw
Faculty of Economic Sciences

Working Papers

No. 13/2011 (53)

Michał Krawczyk

To answer or not to answer? A field test of loss aversion

Warsaw 2011

University of Warsaw
Faculty of Economic Sciences

To answer or not to answer?
A field test of loss aversion

Michał Krawczyk
University of Warsaw
Faculty of Economic Sciences
e-mail: mkrawczyk@wne.uw.edu.pl

Abstract

This study is a field experiment on loss aversion. The framing of scoring rules was differentiated in two exams at the University of Warsaw, with only half the students facing explicit penalty points in the case of giving an incorrect answer. Loss aversion predicts that less risk will be taken (less questions will be answered) when losses are possible but in fact, no treatment effect was observed.

Keywords:
loss aversion, framing, field experiments, gender differences

JEL:
C93, D81

1 Introduction

Loss aversion is believed to be one of the most robust phenomena of decision making under risk. Consequently, it represents a "core idea" (Kahneman 2003) of the leading alternative theoretical framework, the Cumulative Prospect Theory or CPT (Tversky and Kahneman 1992). Indeed, it has been observed in numerous laboratory experiments.[1] While this is very useful evidence, loss aversion is essentially a matter of framing rather than any well-defined preference, so the issue of external validity may be of special importance here. Indeed, it may well be that in an unfamiliar, artificial situation created in the laboratory, the experimenter is able to control subject's perception and thus induce differentiated behavior. In the field, however, experienced decision makers remaining in their natural environment may be less likely to divert from whatever course of action represent their genuine best interest. Even if many puzzling empirical findings, e.g. instances of labor supply decreasing in response to a temporary hike in the wage rate, may be understood in terms of loss aversion, this and most of other cases involve fairly complex phenomena and alternative explanations are typically possible.

In view of this, carefully designed natural field experiments seem to be an attractive way to identify and possibly measure loss aversion. The current study represents a new attempt in this respect. I make use of a unique opportunity to induce perception of a possible loss in a simple situation of decision making under uncertainty. The findings can be summarized succinctly: no evidence of loss aversion is observed and it seems to be difficult to ascribe this null result to any of the "usual suspects" – sample size, stakes, confusion, boredom, ceiling effect etc. In view of the predominantly positive findings in previous studies, it is an intriguing result that calls for future research. Together with some field experiments and other studies that fail to find the hypothesized effect, it may suggest that loss aversion is not as widespread as it is commonly believed.

The rest of the paper is structured as follows: section 2 reviews the relevant literature, section 3 describes the details of the design, section 4 – predictions, section 5 shows the results while section 6 concludes.

2 Literature review

Most direct evidence on loss aversion comes from a number of laboratory experiments such as, for example, (Thaler, Tversky, Kahneman, and Schwartz 1997) and more recently (Novemsky and Kahneman 2005, **?**). Some studies,

[1] See, however, (Ert and Erev 2010).

including (De Martino, Kumaran, Seymour, and Dolan 2006) and (Tom, Fox, Trepel, and Poldrack 2007) are beginning to investigate neural underpinning of loss aversion. There are also several papers reporting observations from the field that might result from loss aversion, typically combined with other factors. A prominent example is the equity premium puzzle (Mehra and Prescott 1985) associated with the "myopic loss aversion" by Benartzi and Thaler (1995). Hardie, Johnson, and Fader (1993) and others using scanner data find that decreases on dimensions such as price and quality have greater impact on the sales of consumer goods then identical increases. In another often-cited study, Camerer, Babcock, Loewenstein, and Thaler (1997) report that NY cab drivers tend to drive less on good days – seemingly exhibiting daily earnings targeting. Generally, the field phenomena such as those discussed above are complex and cannot be taken as a proof of the hypothesis.

Trying to combine the merits of lab experiments and field studies, *field experiments* may provide the most valuable evidence. Unfortunately, these are scarce to date. Interestingly, the findings seem to be predominantly negative; Andersen, Harrison, Lau, and Rutström (2006) do not find any evidence of loss aversion in their study of a high-stake TV show participants. The same can be said of Blavatskyy and Pogrebna (2006). Gächter, Orzen, Renner, and Starmer (2009) find that framing of the difference between early and late conference registration fee (as a discount or as a penalty) has some impact on early registration rate but only in junior economists. As for this study, it has to be emphasized that whether one registers early or not is not a clear case of decision making under uncertainty (although not being sure whether one would eventually attend the conference is surely one of the important factors).

The same is generally true of studies such of persuasiveness of marketing messages, reviewed by O Keefe and Jensen (2006). For example, Ganzach and Karsahi (1995). These authors found that credit card holders' reaction to a letter informing of gains from the card usage was weaker than to a letter describing equivalent losses from not using it.

The main methodological problem with field experiments is that it is difficult to identify situations in which a sizable population is involved in simple decisions under risk with non-trivial stakes and potential losses that can be actively manipulated by the researcher, observed and registered without the participants' knowledge (and thus consent). I have fortunately come across such an opportunity. The only similar attempt that I was able to find in the literature is (Alejos, Paz, Matías, and Javier 2005). The important diifferences between the present study and the approach of Alejos et al. are following: Firstly, I expclitly model the situation in terms of prospect theory. Secondly, I have a sample which is three times larger. Thirdly and most im-

portantly, Alejos, Paz, Matías, and Javier (2005) run a within-subject design, in which they divide their sample in three groups, that will all be subject to three different scoring rules at three different points in time (thus the groups would only differ in terms of the order of being exposed to these rules). These rules are announced at the beginning of the course. I, on the other hand, run a one-time between-subject desing, refrain from informing the subjects beforehand that rules will be differentiated thus preventing communication regarding "optimal" risk-taking strategies.

The latter point may be crucial, given that Alejos, Paz, Matías, and Javier (2005) observe rather perplexing results. While comparison of their "normalized reward" and "normalized penalty" conditions, which roughly correspond to my treatments, yields no difference, the "penalty" condition results in *more* risk taking than normalized reward in one case and *less* risk taking in another. Furthermore, despite random treatment assignment, they observe robust differences between groups' behavior, even when controlling for a number of covariates. The natural explanation is that intense within-group communications took place, destroying independance of observations.

3 Design

The study was performed during two exams at the University of Warsaw - a mid-term exam in Microeconomics at the Faculty of Economic Sciences (henceforth: Micro) and a final exam in Financial System of the Economy at the Faculty of Management (FSE). The Micro exam involved some open-ended questions and 10 multiple choice questions (with one correct and three incorrect answers each). Unbeknownst to the students, exam instructions came in two sorts, as far as the number of points per multiple choice question was concerned: the Mixed Treatment (MT) and the Gain Treatment (GT), see the first rows of Table 1.

Table 1: *Points awarded for closed-end questions*

exam	treatment	correct	missing	incorrect
micro	MT	2	0	-1
	GT	3	1	0
FSE	MT	1	0	-1
	GT	2	1	0

First note that an incorrect answer would bring a penalty point in the MT

only (so giving an answer yields a mixed prospect, hence the label). Second, whatever answer was given to any question, a student would be awarded one point more in the GT than in the MT. As a result, any combination of answers would automatically bring 10 points more in the control than in the loss condition. Accordingly, all grade thresholds were 10 points higher in the former than in the latter.

What regards information provided to the students, before the exam they had only been told that leaving a question unanswered would yield more points than giving an incorrect answer. Assignment to exam rooms and hence, to the two treatments was random. At the start of the exam the students were given the accurate description of the scoring rules (including grade thresholds) relevant for their treatment only (the translation of the exam copies is available from the author).

The FSE exam proceeded in a similar fashion, with the following modifications (some of which resulted from the fact that the author was not in charge for the exam and thus had only limited Gain Treatment over the procedure). First, students had not been explicitly told the form of the exam in advance but they could expect penalty points because these had been used for the FSE exams in previous years. Second, grade thresholds were not specified. Third, while assignment to treatments was, again, random, it was not orthogonal to the mode of study (full-time, part-time or evening studies) which tends to correlate i.a. with age, social background and skills; within each mode of study the treatment assignment was randomized (though, obviously, not necessarily with identical chances for either treatment). There was thus was no selection on unobservables. Last but not least, the 10 closed questions were of the yes/no type rather than a multiple choice one; the points awarded are given in the last two rows of Table 1.

Both exams took place in the winter of 2009/2010. The Micro exam involved nearly 100 students while FSE - about 400. The data collected involved the gender of the student, the mode of study (which actually did not vary among the Micro students), the number of points earned for the open questions, the number of correct and incorrect answers and hence also and most importantly, the number of missing answers.

4 Predictions

Choosing the best answer is obviously a task involving subject-specific skill and knowledge. However, whether to actually choose it or rather leave the question unanswered is a decision under uncertainty.

Crucially, for any level of certainty about one's ability to pick the right

answer, answering the question is a risky option, while giving no answer guarantees a fixed payoff. Prospect theory proposes that a mixed prospect (one involving possible losses) resulting from answering the question in the loss treatment is generally relatively unattractive.

More precisely, if there is little curvature of the value function $v(\cdot)$ for small stakes, as CPT predicts, we can assume that $v(2) = 2, v(1) = 1, v(0) = 0, v(-1) = -\lambda$, whereby $\lambda > 1$ is a loss aversion parameter. Let us denote by p the probability of getting the answer right that makes an individual indifferent between guessing and skipping in the Gain Treatment. Under standard prospect-theoretic notation of $w^+(\cdot)$ as probability weighting function for gains and $w^-(\cdot)$–for losses, for FSE we have

$$w^+(p)v(2) = v(1)$$

thus

$$w^+(p) = 1/2$$

Assuming as in (Tversky and Kahneman 1992) that the probability weighting function for gains is given as $w^+(p) = p^\gamma/(p^\gamma + (1-p)^\gamma)^{(1/\gamma)}$ and for median participant $\gamma = 0.61$ we obtain $p = 0.645$. We ask now whether a subject would answer or skip a question with this level of certainty if she was in the Mixed Treatment? Skipping a question under MT yields zero, while answering it gives

$$w^+(p)v(1) + w^-(1-p)v(-1)$$

It is easy to verify that if the probability weighting function for losses is $w^-(p) = p^\delta/(p^\delta + (1-p)^\delta)^{(1/\delta)}$ with $\delta = 0.69$ for a typical participant, then moderate loss aversion of 1.4 would suffice to make this expression negative. Thus she would abstain from taking the risk if she was in the MT and her perceived probability of finding the correct answer was p. In other words, higher level of certainty is required under MT, so a prediction of less questions answered follows.

Similarly for the Micro exam we obtain $w^+(p) = 1/3$, which gives p close to .3 for a broad range of γ. Then the weight of the negative payoff in the MT for $\delta = .69$ is close to .6, such that, again, very moderate loss aversion parameter of 1.1 would be sufficient to prevent risk taking under MT.[2] We thus submit that CPT predicts less risk taking (i.e. more skipping of questions) in the mixed treatment.

There are two important caveats to this prediction.

[2]The details of the calculation and sensitivity analysis are available from the author upon request.

The first concerns broad vs. narrow framing: the reasoning presented above implicitly assumed that students considered each question separately, in isolation from the rest of the exam. However, students could combine many similar decisions ("should I answer this question or not") into a single meta-prospect. Then, loss aversion (if any) could stop playing a role, e.g. with 10 questions, each with an independent probability of a correct answer of, say, .8, an aggregate loss is relatively unlikely. However, there are reasons to believe that most subjects did not engage in such an aggregation. First, combining several gambles would turn a series of relatively straightforward decision problems (a choice between a sure thing and a two-outcome gamble) into a much more complex one. Second, for most subjects there was probably only a relatively narrow range of certainty for which the decision whether or not to answer the question was a true dilemma. A typical subject could be reasonably certain that she knew the answer to, say, eight questions (and thus surely wanted to answer them), had absolutely no clue about one question (and thus was not willing to gamble) and only had a real decision problem for the one remaining question. This conjecture is corroborated by the fact that, as the data shows, almost half of subjects eventually answered all the questions or all but one question – perhaps for most of the other questions they were reasonably sure about the correct answer (and indeed got most of them right in the end).

The second important consideration is the fact that students may not be interested in maximizing the number of points *per se* but rather in obtaining a possibly high grade (or simply maximizing the probability of passing). This could obviously affect the risk posture. In a way, the same is true of monetary rewards: the value of money is in the goods and services that it can buy. It seems that in our case points could play the role of a "prime reinforcer" like we generally believe money does in standard lab experiments. First, academic programs used for the experiment are among the best in the country and students are typically highly motivated. Clearly, scoring 17 points out of 40 is less unpleasant and humiliating than ending up with just 5, although both would result in the same grade. In this sense, the grade is not all that matters. Second, losing by a small margin only, students may generally expect more leniency on the part of the lecturer (e.g. his or her willingness to grant an additional chance to pass). Third, students did not know how well they did with other questions and thus could not possibly guess whether they are, say, below the threshold (and need to take a chance) or not.[3] Again,

[3]I do elicit predictions of own exam score in a related project. The coefficient of correlation between actual and predicted score is only .477 and there is no tendency to adjust risk posture to the predicted position (e.g. just below or just above the passing

this consideration most probably has led them to think mostly in terms of points. Even if all of the above did not apply, the loss framing could carry over from points to grades.

To the extent therefore that students in different treatments had systematically different expectations regarding how their scores may compare to grade thresholds, treatment effects could result from rational calculation. Obviously, this does not apply to the Micro exam, where students were explicitly told the thresholds, it could however to the FSE.

As for other effects, a great many studies (see Byrnes, Miller, and Schafer (1999)) show that males are generally more willing to take risk than females, especially when its based on assessment on own ability. We therefore predict less unanswered questions in male subjects.

5 Results

Table 2 shows summary statistics for both samples.

Table 2: *Summary statistics for both samples*

sample	# obs.	% in MT	% male	correct	missing	incorrect
micro	96	55	46	4.10	2.09	3.81
FSE	397	48	58	5.14	1.85	3.01

It is reassuring to note that the average number of missing answers was non-trivial; in fact, about two-thirds of subjects omitted at least one question; this leaves sufficient space for the treatment effect (if any) to show.

Because of the numerous differences between the samples (outlined in section 3), we report the results separately for the two, starting with the Micro exam.

The simplest test of the treatment effect is the comparison of number of missing observations in the two treatments. While the number of missing observations is slightly higher for the Mixed Treatment (2.21, as compared to 1.95 for GT), the difference was not significant ($p = .16$ in the M-W-W test and $p = .19$ in the Fisher exact test).

Males tried to answer more questions than females (2.48 vs 1.63, $p = .02$ in M-W-W). This was not associated with any superior knowledge – scores obtained (for either part of the exam) were nearly identical across genders.

threshold).

This confirms our auxiliary hypothesis of males being more self-assure and/or more risk-seeking. There was no significant interaction between gender and treatment.

Results for the FSE sample were remarkably similar. While the mean is slightly higher in the GT (1.98 vs 1.71), the difference is far from significant ($p = .53$ in the Mann-Whitney-Wilcoxon test or $p = .99$ in the Fisher exact test). However, as mentioned before, assignment to the treatments was not orthogonal to the mode of study. Specifically, full-time students were mostly assigned to the MT while part-time and evenings students more often to the GT.

Table 3: *FSE: number of missing answers depending on treatment and mode of study (# of obs. in parentheses)*

treatment	full-time	evening	part-time
MT	1.65	1.44	2.59
	(142)	(31)	(17)
GT	1.86	2.31	1.55
	(57)	(94)	(56)

Table 3 shows that, first, differences between the numbers of missing observations between modes of study were limited (though *scores* were in fact highly divergent, with full-time students performing much better than the other two categories). Second, there was no clear pattern of treatment effect – while evening students left more questions unanswered in the GT, the opposite was true for the part-times. These differences appear to stem from the number of observations being limited in some cells; none of them is significant.

Further analysis showed that gender effect was strong and highly significant – on average males tried to answer almost one question more (M-W-W $p = .001$).

6 Conclusion

The current study was planned as a straightforward field tests of loss aversion in decision making in risk and uncertainty. With some 500 subjects motivated to make best choices possible, hardly any trace of loss aversion was detected. This null effect cannot be ascribed to the specificity of the sample in the sense that laboratory studies finding loss aversion used university students

as well. Nor was it likely that subjects did not notice or understand the scoring rules. Apparent lack of loss aversion could result from the fact that the underlying reward medium involved academic success rather than money; but to the extent that loss aversion goes away with experience, the all too familiar monetary domain is the one where we would least expect it to show. Besides, previous experiments on endowment effect seem to suggest that loss aversion works just fine for coffee mugs, chocolate bars etc. where no money is involved at all. Combining my results with those of some other field experiments mentioned before, loss aversion may not be as ubiquitous in the field as it would seem given the bulk of evidence from the laboratory.

Factors such as subjects' experience could play a role and call for additional observations but the similarity of the results obtained for the two samples seem to suggest that lack of loss aversion is somewhat robust in this context. Another interesting factor is the impact of feedback. Studies on myopic loss aversion show that reluctance to take risks involving possible losses might be strengthened if information about winning or losing is instantly provided. While this poses a technical difficulty within the framework proposed in this study, the impact of feedback frequency will be investigated in a follow-up project.

References

Alejos, E., M. Paz, G. Matías, and F. Javier (2005): "On the strategic equivalence of multiple-choice test scoring rules: evidence from a field experiment," *DFAE-II WP Series*, (20), 1.

Andersen, S., G. Harrison, M. Lau, and E. Rutström (2006): "Dynamic choice behavior in a natural experiment," *University of Central Florida Department of Economics Working Paper*, pp. 06–10.

Benartzi, S., and R. Thaler (1995): "Myopic loss aversion and the equity premium puzzle," *The Quarterly Journal of Economics*, 110(1), 73–92.

Blavatskyy, P., and G. Pogrebna (2006): "Loss Aversion? Not with Half-a-Million on the Table!," *IEW-Working Papers*.

Byrnes, J., D. Miller, and W. Schafer (1999): "Gender differences in risk taking: A meta-analysis.," *Psychological bulletin*, 125(3), 367.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler (1997): "Labor Supply of New York City Cabdrivers: One Day At A Time*," *The Quarterly Journal of Economics*, 112(2), 407–441.

De Martino, B., D. Kumaran, B. Seymour, and R. Dolan (2006): "Frames, biases, and rational decision-making in the human brain," *Science*, 313(5787), 684.

Ert, E., and I. Erev (2010): "On the descriptive value of loss aversion in decisions under risk," *Harvard Business School Working Papers*.

Gächter, S., H. Orzen, E. Renner, and C. Starmer (2009): "Are experimental economists prone to framing effects? A natural field experiment," *Journal of Economic Behavior and Organization*, 70(3), 443–446.

Ganzach, Y., and N. Karsahi (1995): "Message framing and buying behavior: A field experiment," *Journal of Business Research*, 32(1), 11–17.

Hardie, B., E. Johnson, and P. Fader (1993): "Modeling loss aversion and reference dependence effects on brand choice," *Marketing Science*, pp. 378–394.

Kahneman, D. (2003): "Maps of Bounded Rationality: Psychology for Behavioral Economics†," *American economic review*, 93(5), 1449–1475.

Mehra, R., and E. Prescott (1985): "The equity premium: A puzzle," *Journal of Monetary Economics*, 15(2), 145–61.

Novemsky, N., and D. Kahneman (2005): "The boundaries of loss aversion," *Journal of Marketing Research*, 42(2), 119–128.

O Keefe, D., and J. Jensen (2006): "The advantages of compliance or the disadvantages of noncompliance? A meta-analytic review of the relative persuasive effectiveness of gain-framed and loss-framed messages," *Communication yearbook*, 30, 1.

Thaler, R., A. Tversky, D. Kahneman, and A. Schwartz (1997): "The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test*," *The Quarterly Journal of Economics*, 112(2), 647–661.

Tom, S., C. Fox, C. Trepel, and R. Poldrack (2007): "The neural basis of loss aversion in decision-making under risk," *Science*, 315(5811), 515.

Tversky, A., and D. Kahneman (1992): "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and uncertainty*, 5(4), 297–323.

Faculty of Economic Sciences
University of Warsaw
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl