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

Michał Krawczyk, dr, Wydział Nauk Ekonomicznych Uniwersytetu Warszawskiego

Keywords: loss aversion, framing, field experiments, gender differences
JEL classification: 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 [Harrison, List, and Towe, 2007]. 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. It is also crucially important, to the extent that prospect theory is put forward as a general theory of decision making under risk and uncertainty, to establish that findings converge across domains, e.g. for non-monetary rewards. 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,

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

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 and perhaps strongest evidence in favour of 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] [Abdellaoui, Bleichrodt, and Parashiv, 2007], often using sophisticated ways to elicit LA with little parametric assumptions. Some studies, 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]. Market studies using scanner data, e.g. [Hardie, Johnson, and Fader, 1993], 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 rather mixed. Haigh and List [2005] confirm myopic loss aversion in an artificial field experiment with professional traders. However, 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] find 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 endowment effect [Kahneman, Knetsch, and Thaler [1990], often considered as a manifestation of loss aversion in risk-free environment, has been shown to subside with experience in the field [List, 2003].

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[2] in the literature is [Alejos, Paz, Matías, and Javier [2005]. The important differences between the present study and the approach of Alejos et al. are following: Firstly, I explain model the situation in terms of CPT. Secondly, I have a sample which is more than twice as large. Thirdly and most importantly, Alejos, Paz Matías, and Javier2005strategic run a within-subject design, in which they divide their sample in three groups, that would 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 design, refrain from informing the subjects beforehand that rules will be differentiated thus preventing experimenter demand effect and between-subject 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 independence of observations.

3. Design

The study was performed during an exam in Financial System of the Economy at the Faculty of Management of the University of Warsaw. The exam consisted of some open-ended questions and ten yes/no questions. Unbek-

[2] Except for the twin project described in [Krawczyk, 2011].

nownst to the students, exam instructions came in two sorts, as far as the number of points per closed-ended question was concerned: the Mixed Treatment (MT) and the Gain Treatment (GT), see Table 1.

Table 1.

Points awarded for closed-end questions

Treatment	Correct	Missing	Incorrect
T	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.

Before the exam students had not been explicitly told the form of the exam (although they could expect penalty points because these had been used for the FSE exams in previous years). Upon arrival, they were directed to their exam rooms, whereby treatment assignment was determined—all students in the same room were to obtain the same treatment. Following the requirements put forward by the faculty member running the exam, this could not be fully randomized because students following different modes of study (full-time, part-time extramural studies) were expected to be in the different rooms. Still, within each mode of study the treatment assignment was randomized (though not necessarily with identical chances for either treatment). There was thus was no selection on unobservables.

In total nearly 400 subjects took part. The data collected involved the gender of the student, the mode of study, 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. The students were never told they had participated in a study.

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 an uncertain 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, 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 $w^+(p) = p^\gamma \Big/ \left(p^\gamma + (1-p)^\gamma\right)^{1/\gamma}$ 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 $w^-(p) = p^\delta \Big/ \left(p^\delta + (1-p)^\delta\right)^{1/\delta}$ 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.

There are two important caveats to this prediction.

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, the academic program used for the experiment is 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 leniencies 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. Again, 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. On top of that, students in the two treatments could have systematically different expectations regarding the thresholds, which again could affect their willingness to take risk. For example, a poor student expecting a high passing threshold could think he had to take chances. However, in a related project [Krawczyk, 2011] I do elicit predictions of own exam score. The main findings are that the coefficient of correlation between actual and predicted score is only.477 and that there is no tendency to adjust risk posture to the predicted position (e.g. just below or just above the passing threshold). This suggests that students do think in terms of points rather than grades only and that even if treatment manipulation affected beliefs regarding thresholds, this was unlikely to strongly affect risk posture.

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 it is based on assessment on own ability. We therefore predict less unanswered questions in male subjects.

5. Results

Table 2. shows summary statistics.

Table 2.

Summary statistics

# obs.	% in MT	% male	correct	missing	incorrect
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.

Overall, 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 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 extramural and part-time students more often to the GT.

Table 3.

Number of missing answers depending on treatment and mode of study (# of observations in parentheses)

treatment	full-time	part-time	extramural
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 part-time students left more questions unanswered in the GT, the opposite was true for the extramurals. These differences appear to stem from the number of observations being limited in some cells; none of them is significant.

Analysis also showed a strong gender effect-on average males tried to answer almost one question more, a highly significant difference (M-W p =.001). This was not associated with any superior knowledge—scores obtained (for either part of the exam) were nearly identical across genders. 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.

6. Conclusion

The present study was planned as a straightforward field tests of loss aversion in decision-making in risk and uncertainty. With some 400 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. In other words, my findings suggest that the domain on which this particular component of the CPT remains valid is somewhat limited.

Factors such as subjects' experience could play a role and call for additional observations but the similarity of the results obtained 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

Abdellaoui, M. and Bleichrodt, H. and Paraschiv, C., 2007, "Loss aversion under prospect theory: A parameter-free measurement," *Management Science* Vol. 53, No. 10, p. 1659–1674.

Alejos, E. and Paz, M. and Matías, G. and Javier, F., 2005, "On the strategic equivalence of multiple-choice test scoring rules: evidence from a field experiment," *DFAE-II WP Series* No. 20, p. 1.

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

Benartzi, S. and Thaler, R.H., 1995, "Myopic loss aversion and the equity premium puzzle," *The Quarterly Journal of Economics* Vol. 110, No. 1, p. 73–92.

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

Byrnes, J.P. and Miller, D.C. and Schafer, W.D., 1999, "Gender differences in risk taking: A meta-analysis," *Psychological Bulletin* Vol. 125, No. 3, p. 367.

Camerer, C. and Babcock, L. and Loewenstein, G. and Thaler, R., 1997, "Labor Supply of New York City Cabdrivers: One Day At A Time*," *The Quarterly Journal of Economics* Vol. 112, No. 2, p. 407–441.

De Martino, B. and Kumaran, D. and Seymour, B. and Dolan, R.J., 2006, "Frames, biases, and rational decision-making in the human brain," *Science* Vol. 313, No. 5787, p. 684.

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

Gächter, S. and Orzen, H. and Renner, E. and Starmer, C., 2009, "Are experimental economists prone to framing effects? A natural field experiment," *Journal of Economic Behavior and Organization* Vol. 70, No. 3, p. 443–446.

Ganzach, Y. and Karsahi, N., 1995, "Message framing and buying behavior: A field experiment," *Journal of Business Research* Vol. 32, No. 1, p. 11–17.

Haigh, M.S. and List, J.A., 2005, "Do professional traders exhibit myopic loss aversion? An experimental analysis," *The Journal of Finance*, Vol. 60, No. 1, p. 523–534.

Hardie, B.G.S. and Johnson, E.J. and Fader, P.S., 1993, Modeling loss aversion and reference dependence effects on brand choice, *Marketing Science*, p. 378–394.

Harrison, G.W. and List, J.A. and Towe, C., 2007, "Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion," *Econometrica* Vol. 75, No. 2, p. 433–458.

Kahneman, D., 2003, "Maps of Bounded Rationality: Psychology for Behavioral Economics†," *American Economic Review* Vol. 93, No. 5, p. 1449–1475.

Kahneman, D. and Knetsch, J.L. and Thaler, R.H., 1990, "Experimental tests of the endowment effect and the Coase theorem," *Journal of Political Economy* Vol.?, No.?, p. 1325–1348.

Krawczyk, M., 2011, "Framing in the field: a simple experiment on the reflection effect," Discussion paper.

List, J., 2003, "Does Market Experience Eliminate Market Anomalies?," *The Quarterly Journal of Economics* Vol. 118, No. 1, p. 41–71.

Mehra, R. and Prescott, E., 1985, "The equity premium: A puzzle," *Journal of Monetary Economics* Vol. 15, No. 2, p. 145–61

O'Keefe, D.J. and Jensen, J.D., 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* Vol. 30, p. 1.

Thaler, R.H. and Tversky, A. and Kahneman, D. and Schwartz, A., 1997, "The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test," *The Quarterly Journal of Economics* Vol. 112, No. 2, p. 647–661.

Tom, S.M. and Fox, C.R. and Trepel, C. and Poldrack, R.A., 2007, "The neural basis of loss aversion in decision-making under risk," *Science* Vol. 315, No. 5811, p. 515.

Tversky, A. and Kahneman, D., 1992, "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty* Vol. 5, No. 4, p. 297–323.

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

This study is a field experiment on loss aversion. The framing of scoring rules was differentiated in an exam 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 classification: C93, D81.