to Victorina

DRAFT

Introduction

These lectures do not follow classic game theory. The concepts of ‘strategy’, ‘utility’, ‘game matrix’, ‘guaranteed payoff’, ‘equilibrium point’ are not used there. The theory of reflexive games is designed for solving problems different from those solved by the classic game theory. Its goal is to predict an individual choice made by a subject in a group and indicate the possibility of controlling this choice. We call this control *reflexive control*. The term ‘subject’ refers to single individuals or to various types of organizations: political parties, military units, states, and even civilizations. Relations between the interests of a group and those of individual subjects are regulated by the rule called *a principle of prohibition for selfishness*: a subject in a group while pursuing one’s own goal must not cause damage to the group as a whole. This principle is as important in the reflexive game theory as the principle of guaranteed payoffs in the classic game theory.

The principle of prohibition for selfishness makes the actions unacceptable if they are advantageous for a subject but harmful for a group to which the subject belongs. It does not forbid, however, antisocial actions if a subject does not pursue his own interests. Therefore, the subject’s unselfishness justifies his actions that harm a group or a society. Note that the indication of the actions preferable by a subject and those preferable by a group are not assigned in advance, but is generated by the formalism of the model.

The essential difference from the classic game theory is that in the reflexive game theory, the special assumptions are made concerning mental mechanism generating choice. We assume that the subject possesses a partially ordered set of self-images, that is, the subject has several images of the self each of which may have images of the self etc. A hierarchy of images is depicted by a special formula called *diagonal form*. Let us emphasize that the hierarchy of images is not arbitrary. It is always finite and predetermined by the graph of relationships between the subjects. This is an important moment. People have known for many centuries that a human mental domain can be described with chains of the type “he knows that he knows that he knows . . . ” However, these chains have not been used in the models of human cognition because there was no rule to stop. There is such a rule in the reflexive game theory and that makes the theory efficient.

A diagonal form defines a mental procedure of choice and the mathematical function describing the choice, at the same time. Thus, having written a diagonal form according to empirical data we automatically obtain a subject’s function of choice. One of the values in the diagonal form is interpreted as the subject’s intention. Then we assume that the subject is goal-oriented, which means that he has only such intentions that can be turned into reality, so, intentions are not assigned in advance. The goal-oriented subject corresponds to an equation whose solution is interpreted as an alternative the subject chooses. A case when the equation does not have solutions is interpreted as the subject's inability to make decision under given circumstances. The equation may have several solutions, then each of them is considered to be a potential subject's choice. Finally, it is possible that any alternative can be the equation solution, we assume then that the subject has freedom of choice, i.e., a group does not impose limitations on his decisions.

In this work, we give examples of applying the reflexive game theory in the areas of personal relations, social life, politics, international relationships, military decision making, and law.

Is it possible to conduct objective testing of this theory? It is unlikely that the principle of falsification offered by Karl Popper for the natural sciences can be used here. General social-psychological theories such as the classical game theory cannot be rejected as a result of small number of failures, because the area of their applications is defined vaguely. The idea formulated by Cambell (1997) will better fit here: in science, the principle of natural selection is realized, and only those theories survive that are interesting for researchers. The theory of reflexive game as well as the classical game theory cannot be rejected due to a small number of wrong predictions. Its fate depends on specialists: will they use it and for how long.

The first step toward construction of the reflexive game theory has been done more than forty years ago. The recursive chains "I know that he knows that I know . . . " have been put at the base of the model of the subject and the term "reflexive games" has been introduced (Lefebvre, 1965, 1966, 1967). Then a special formal apparatus for modeling human choice was developed (Lefebvre, 1982), which later allowed using the theory of reflexive games for analysis of concrete situations (Lefebvre, 2003, 2007; ?Lefebvre, 2001).

Important contributions to the theoretical understanding of reflexion were done by numerous scientists. Tatiana Taran (1998, 2001) constructed a multivalued Boolean model of choosing social norms. Vladimir Krylov (2000) studied problems related to axiomatic of reflexive models. Yuli Schreider (1999) considered continuously-valued logics as languages of reeflexion. Pavel Baranov and Vladimir Lepsky developed a formal model of the subject with reflexion and inner value (see Lefebvre, Baranov, Lepsky, 1969). Anatoly Trudolubov (1972) created a reflexive game model on dependency nets. Tim Kaiser and Stefan Schmidt (2008) found relation between the reflexive game theory and the theory of functors and categories. There were also two attempts to combine reflexive games with the classic game theory. Novikov and Chkhartashvili (2003) included reflexive games into formalism of the classic game theory; Lefebvre (2001, ?2003) included the classic game theory into the reflexive game theory. Future study will show if such connections can be productive.

The book has two appendices. The proofs of some mathematical statements used in the work are given in the Appendix A. Appendix B contains problems and exercises relevant to the material described.

I am thankful to Viacheslav Filimonov, Tim Kaiser, Jonathan Farley and Stefan Schmidt for their advices and corrections. Invaluable help was given to me by Victorina Lefebvre, with whom I discussed my plans and main ideas of the book. In addition, she technically prepared these lectures for publication and made many significant remarks in the course of work. Without her participation this book would never be completed.