The problem of social conventions*

Game-theoretic view

Maxim Gladyshev

Higher School of Economics

magladyshev@gmail.com

NATIONAL RESEARCH UNIVERSITY

Abstract

The paper presents an attempt to analyze the problem of the emergence and stability of conventions which was first raised by Lewis as a result of interaction and communication of a group of rational agents. It presents formalization of such interactions in both cooperative and non-cooperative contexts with the game-theoretic approach. Fundamental properties and characteristics of the phenomenon of conventions are singled out, pragmatic phenomena of hint, politeness and trust are considered.

Introduction

Modern discussion about the nature, definition and properties of conventions begins with the book by D. Lewis "Convention"(1969). The merit of Lewis is that he asserts the independence of conventions from any agreements or arrangements and points to a different nature of this phenomenon. The conventions act as some stable patterns of the actions of intelligent agents and solve the problem of coordination. D. Lewis experienced the influence of game theory, in particular, T. Schelling. So, game theoretic approach was deemed relevant by Lewis. The main achievement of Lewis is the definition of conventions he formulated:

A regularity R in the behaviour of members of a population P when they are agents in a recurrent situation S is a convention if and only if it is true that, and it is common knowledge in P that, in almost any instance of S among members of P:
1) almost everyone conforms to R;
2) almost everyone expects everyone else to conform to R;
3) almost everyone has approximately the same preferences regarding all possible combinations of actions;
4) almost everyone prefers that any one more conform to R', on condition that almost everyone conform to R;
5) almost everyone would prefer that any one more conform to R', on condition that almost everyone conform to R',
where R is some possible regularity in the behaviour of members of P in S, such that almost no one in almost any instance of S among members of P could conform both to R' and to R. [Lewis, p. 78]

It's difficult to underestimate Lewis's contribution to the development of the theory of conventions, but it can be noted that his definition has a number of weaknesses. Moreover, Game Theory has undergone a huge number of changes and additions in recent years, new concepts and approaches have emerged. All of this requires a rethinking and application of new tools for analyzing the phenomenon of the emergence and stability of conventions.

Main Objectives

1. Analyze the nature of conventions by the example of such pragmatic phenomena as politeness and hint.
2. Assess the explanatory potential of contemporary Game Theory, in particular the concept of correlated equilibrium, for conventions.
3. Clarify and generalize the Lewis definition.

Materials and Methods

The methodology of this research is reduced to the contemporary Game Theory and Game-Theoretic pragmatics. The definitions of basic standard concepts are below.

1. *The normal-form game is a structure* $G = (N, \{S_i\}_{i\in N}), \{u_i\}_{i\in N})$ *, where*
- N *– set of players (agents)*
- S_i *– set of strategies of* i*'s player*
- $S = \underset{i\in N}{\times} S_i$ *– set of all pure strategies profiles*
- $u_i : S_1 \times ... \times S_n \mapsto \mathbb{R}$ *– payoff function*

2. *Strategy* s_i *is the best response for* s_{-i} *iff* $u_i(s_i, s_{-i}) = \max\limits_{s'_i \in S_i} u_i(s'_i, s_{-i})$.

Set of best responses of i*'s player for stratagies profile* s_{-i} *will be denoted as* $BR_i(s_{-i})$.
$BR_i(s_{-i}) = \{s' \in S_i \mid \forall s''_i \in S_i : u_i(s'_i, s_{-i}) \geq u_i(s''_i, s_{-i})\} = arg\max\limits_{s' \in S_i} u_i(s'_i, s_{-i})$.

Strategies profile $s = (s_i, s_{-i})$ *is* ***Nash equilibrium*** *iff* $\forall i \in N : \; s_i \in BR_i(s_{-i})$.

Strategies profile $s = (s_i, s_{-i})$ *is* ***Nash equilibrium*** *iff* $\forall i \in N \; \forall s'_i \in S_i : u_i(s_i, s_{-i}) \geq u_i(s'_i, s_{-i})$. *(alternative definition).*

For further analysis of pragmatics of bribes we need to use the concept of signaling game. Signaling game is a structure $G = (\{S, R\}, W, Pr, F, A, S_0, R_0, U_S, U_R)$, where:
- $\{S, R\}$– set of players,
  where S – Sender of signal (Speaker), R – Receiver of signal (Listener);
- W – set of possible worlds;
- Pr – the probability distribution defined on set W;
- F – set of Sender's messages;
- A – set of Receiver's actions;
- $S_0 : W \to F$ – set of Sender's strategies;
- $R_0 : F \to A$ – set of Receiver's strategies;
- $U_S : W \times F \times A \to \mathbb{R}$ – utility function for Sender;
- $U_R : W \times F \times A \to \mathbb{R}$ – utility function for Receiver.

Signaling game is a game for two players in which Nature makes the first action choosing the situation in which the speaker appears to be according to distribution of probabilities known to all players (Nature chooses the Sender type). After that, the Speaker selects the message, and the Listener selects some action, knowing the choice of the Speaker, but not knowing the choice of Nature. Probability distribution Pr is known to all players. In other words, the signal game is a game in extensive form, in which the second player has incomplete information.

Results

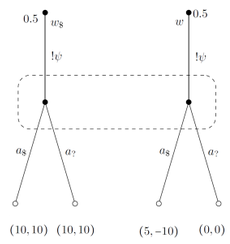
	Fair	Bribe-taker
"I'll give you 1000 rubles, and you let me go "	Arrest	Discharge
"Perhaps, we will find a solution?"	Bewilderment	Discharge

Consider an example when a policeman stops a driver for speeding. The driver does not know if this policeman takes bribes. He must choose one of two communication strategies: ϕ' : "I'll give you 1000 rubles, and you let me go." ϕ : "Perhaps, we will find a solution?".

Now we have a normal-form game with 4 possible outcomes: Let's try to construct a formal model of this situation using the game-theoretic pragmatics, that is, consider the signaling game G_1. Let:
- $W = \{w_\$, w\}$, $w_\$$ – situation in which a policeman is a bribe taker, w – situation in which a policeman is fair
- $Pr(W) = \left[\frac{w_\$ \mapsto 0.5}{w \mapsto 0.5}\right]$
- $F = !\psi$, where $!\psi$ public announcement of policeman "You have to pay a fine!"
- $A = \{a_?, a_\$\}$, set of answers for the driver, where $a_?$: "Perhaps, we will find a solution?"; $a_\$$: direct offer of bribery
- The only strategy available in this game for the speaker: $S_1 = \left\{s_1 : \left[\frac{w \mapsto !\psi}{w_\$ \mapsto !\psi}\right]\right\}$,
- Listener has two strategies: $R_1 = \left\{s_1 : \left[\frac{!\psi \mapsto a_?}{!\psi \mapsto a_?}\right], r_2 : \left[\frac{!\psi \mapsto a_\$}{!\psi \mapsto a_\$}\right]\right\}$.
- The pay-off function is not symmetric for the Listener and the Speaker, since we are dealing with a non-cooperative context of communication:

$$U_S = \begin{bmatrix} (w, !\psi, a_?) \mapsto 0 \\ (w, !\psi, a_\$) \mapsto 5 \\ (w_\$, !\psi, a_?) \mapsto 10 \\ (w_\$, !\psi, a_\$) \mapsto 10 \end{bmatrix}, \qquad U_R = \begin{bmatrix} (w, !\psi, a_?) \mapsto 0 \\ (w, !\psi, a_\$) \mapsto -10 \\ (w_\$, !\psi, a_?) \mapsto 10 \\ (w_\$, !\psi, a_\$) \mapsto 10 \end{bmatrix}.$$

From these conditions we obtain the following game:



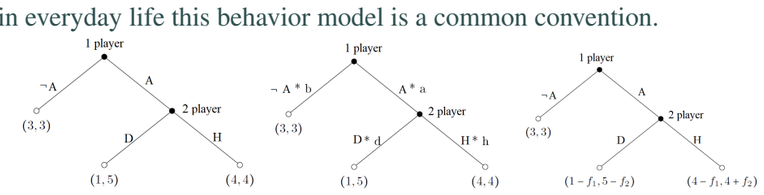
Find the expected utility for all outcomes of the game:

$$EU_R(!\psi, a_?) = 5$$
$$EU_R(!\psi, a_\$) = 0.$$

The solution of the game is the unique Bayes-Nash equilibrium: $\left(s_1 : \left[\frac{w \mapsto !\psi}{w_\$ \mapsto !\psi}\right], r_2 : \left[\frac{!\psi \mapsto a_?}{!\psi \mapsto a_?}\right]\right)$.

Thus, the use of an indirect speech act (hint) is the only rational strategy of the Speaker. From this we can conclude that conventional behavior of the agent is due to a rational desire to maximize utility. This behavior pattern is Nash equilibrium, which explains the stability of hint convention.

Another important phenomenon of social convention is politeness or trust. Formal models for these cases were provided by J. Quinley [3]. Let's consider a Trust Game as a model for decision on a request for help. In this game 1 player decides whether to Ask (A) 2 player about help or not ($\neg A$). 2 player can Help (H) or Deny (D). It's clear that 1 player will not ask for help if the game is repeated only once (a). But we can see that 1 player could get higher utility and 2 player would not deny his request if every player knew that this situation would be repeated in the future (b). In this case (4,4) payoff will become stable outcome of the game (actually, mixed strategy equilibrium is ($\frac{3}{4}$A, $\frac{2}{3}$H). And we see that in everyday life this behavior model is a common convention.



This model also makes it possible to take into account the influence of the social factor as social approval or social pressure (c). It allows to count the fact that politeness is a more costly communicative mechanism. Many excess speech structures are usually explained by the influence of culture or traditions, but conventions cannot be reduced to these factors. And Game Theory provides a better explanation.

Conclusions

Unfortunately, the poster format allows to demonstrate only half of the research. An equally important part is analysis of such concepts as correlated equilibrium and evolutionary stable strategies. Both of them have great potential for formalizing and clarifying the phenomenon of conventions and remain the most promising method for further research. The work demonstrates the effectiveness of Game Theory as the main tool for formalizing and analyzing conventions. It's clear, that the phenomenon of conventions still needs a more precise and complete definition which would take into account all or almost all cases.

References

1. *Benz, A., Ja- ger, G., van Rooij R.* Game Theory and Pragmatics, Basingstoke: Palgrave Macmillan, 2005.
2. *Lewis, D.* Convention: A Philosophical Study, Harvard University Press 1969.
3. *Quinley, J.* Trust Games as a Model for Requests // New Directions in Logic, Language and Computation. Springer, 2012.
4. *Van Rooy, R.* Being polite is a handicap: Towards a game theoretical analysis of polite linguistic behavior. TARK, Vol. 9, 2003.
5. *Shoham, Y., Leyton-Brown, K.* Multiagent Systems, Cambridge: Cambridge Univeristy Press, 2008.

Acknowledgements

I wish to thank Vitaliy Dolgorukov for his supervision on my research project and Elena Dragalina-Chernaya for her contribution to the creation and development of research group and international research laboratory.

*The article was prepared within the framework of the Academic Fund Program at the National Research University Higher School of Economics (HSE) in 2017 – 2018 (grant 17-05-0040) and by the Russian Academic Excellence Project 5-100.